Road Infrastructure Safety Management Evaluation Tools (RISMET)

Accident Prediction Models for Rural Junctions on Four European Countries

Deliverable Nr 6.1

December 2011

Project Coordinator: SWOV

Institute for Road Safety Research

Project Partner 2: TUD

Technische Universität Dresden

Project Partner 3: LNEC

National Laboratory for Civil Engineering

Project Partner 4: TØI - Transportøkonomisk institutt Stiftelsen Norsk senter for samterdselsforskning

Project Partner 5: TRL –

Transport Research Laboratory

Project Partner 6: KfV - Kuratorium für Verkehrsicherheit

This project was initiated by ERA-NET ROAD.

Project Nr. 823137

Project acronym: RISMET

Project title: Road Infrastructure Safety Management Evaluation Tools

RISMET

Deliverable Nr 6.1 – Accident Prediction Models for Rural Junctions on Four European Countries

Due date of deliverable: 31.03.2011

Actual submission date: 30.12.2011

Start date of project: 01.09.2009

Planned end date of project: 31.08.2011

Revised end date of project: 30.12.2011

Author(s) this deliverable:

Sofia de Azeredo Lopes, LNEC, Portugal

João Lourenço Cardoso, LNEC, Portugal

Version: final draft

Executive summary

The "ROAD INFRASTRUCTURE SAFETY MANAGEMENT EVALUATION TOOLS (RISMET)" project targets objective A (Development of evaluation tools) of the Joint Call for Proposals for Safety at the Heart of Road Design ("The Call"). This project aims at developing suitable road safety engineering evaluation tools that will support the aims of the Call as described in the Guide for Applicants (GfA) and furthermore those of the Directive for Road Infrastructure Safety Management (2008). These evaluation tools allow the easy identification of both unsafe (from accidents or related indicators) and potentially unsafe (from design and other criteria) locations in a road network. With such evaluation tools estimates of potential benefits at the local and the network level can be calculated and potential effects on aspects such as driver behaviour can be estimated. Such tools empower road authorities to improve their decision making and to implement (ameliorative) measures to improve the road safety situation on the roads.

Since evaluation tools rely on good quality data, RISMET aims at reviewing available data sources for effective road infrastructure safety management in EU-countries, linked to a quick scan and assessment of current practices. Furthermore, RISMET aims at exploiting results related to the development and use of Accident Prediction Models (APMs) in road safety management.

The present deliverable provides APMs for data collected at junctions from the rural road networks of Austria, Norway, Portugal and Holland. For the first three countries it was possible to obtain accident prediction models for each country individually. For Holland, however, and due to restrictions on the dimension of the data set, it was only possible to analyse these data together with the other countries data, i.e. analysing aggregated data sets. The data consists, per junction, of injury accident counts, type of junction, traffic control, speed limit and annual average daily traffics entering from the major and the minor road. The regression models had the injury accident frequencies as the dependent variable and the remaining variables as explanatory and were fitted using Bayesian statistical techniques with vague or non-informative prior and hyper-prior distributions. These models consisted on the Poisson regression model, hierarchical Poisson-Gamma and Poisson Log-Normal hierarchical regression model. The Poisson regression model was found to be not appropriate to model the junction data in any of the data sets due to not being able to capture variations and attributes of the data, namely the over-dispersion. The Poisson-Gamma and the Poisson Log-Normal models obtained similar results and in general performed equally well. It was found that accidents occurring at junctions in all countries depend on the junction's entering traffic volume as well as the other explanatory variables considered. This report provides descriptions of the several data sets, equations for the expected injury accident frequencies, per year, on rural road network junctions for Austria, Norway and Portugal and for the conjoint set of the combined data (including Dutch data) as well as posterior means of the expected number of accidents for minimum, mean, median and maximum profiles obtained by the explanatory variables and measurements of model fit together with the major results obtained.

Table of content

Executive summary 3
Table of content 4
List of Tables 7
List of Figures 11
1 Introduction 23
1.1 Background and objectives 24
1.2 Structure of the report 24
2 Methodological Approach 25
2.1 The Poisson Regression Model 25
2.2 Mixture Models 25
2.2.1 Poisson-Gamma Models 26
2.2.2 Poisson Log-Normal Model 26
2.3 Convergence Assessment 27
2.3.1 Gelman-Rubin Diagnostics 28
2.4 Model Assessment 29
2.4.1 Deviance Information Criterion and Effective Model Dimension 29
2.4.2 Posterior Predictive Checking 30
3 Modelling Norwegian injury accidents 30
3.1 Norwegian Junction Data 31
3.2 The Poisson Regression Model 35
3.2.1 Model Checking 39
3.3 Poisson-Gamma hierarchical regression model 41
3.3.1 Model Checking 44
3.4 Poisson Log-Normal Regression Model 49
3.4.1 Model Checking 52
3.5 Discussion 56
4 Modelling Austrian injury accidents 57
4.1 Austrian Junction Data 58
4.2 The Poisson Regression Model 62
4.2.1 Model Checking 65
4.3 The Poisson-Gamma hierarchical regression model 69

4.3.1 Model Checking 72
4.4 The Poisson Log-Normal regression model....76
4.4.1 Model Checking 78
4.5 Discussion....83
5 Modelling Portuguese injury accidents 84
5.1 Portuguese Junction Data 85
5.2 The Poisson regression model 88
5.2.1 Model Checking 90
5.3 The Poisson-Gamma hierarchical regression model 94
5.3.1 Model Checking 96
5.4 The Poisson Log-Normal regression model....100
5.4.1 Model Checking 102
5.5 Discussion....106
6 Modelling Austrian, Norwegian and Portuguese injury accidents....107
6.1 Aggregated Junction Data....108
6.2 The Poisson regression model....109
6.2.1 Model Checking 111
6.3 The Poisson-Gamma hierarchical regression model 114
6.3.1 Model Checking 115
6.4 The Poisson Log-Normal regression model....119
6.4.1 Model Checking 120
6.5 Discussion....124
7 Modelling Austrian, Norwegian and Portuguese injury accidents on non-roundabout junctions....124
7.1 Aggregated Junction Data (excluding roundabouts) 125
7.2 The Poisson-Gamma hierarchical regression model 127
7.2.1 Model Checking 129
7.3 The Poisson Log-Normal hierarchical regression model....133
7.3.1 Model Checking 135
7.4 Discussion....139
8 Modelling Austrian, Dutch and Portuguese injury accidents on roundabout Junctions.140
8.1 Austrian, Dutch and Portuguese Roundabout Data 140
8.2 The Poisson-Gamma hierarchical regression model 143
8.2.1 Model Checking 145
8.3 The Poisson Log-Normal hierarchical regression model....148
8.3.1 Model Checking 149

8.4 Discussion....153
9 Conclusions....153
9.1 Model and Model Development....154
9.2 Primary Conclusions per Country....155
9.3 General....155
Sources....157

List of Tables

Table 1 Summary statistics for the variables registered on Norwegian junctions from 1997 to 2002. 35

Table 2 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson regression model was fitted to the Norwegian accident data. 37

Table 3 Expected numbers of accidents for Norwegian junctions, for a one year period, obtained by a Poisson regression model, for a baseline/reference of Number_of_Legs='3' and Speed_Limit='60'. 38

Table 4 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fitted to the Norwegian accident data using 3 leg junctions with a 60km/h speed limit as baseline. 42

Table 5 Expected number of accidents per year for Norwegian junctions, obtained by a Poisson-Gamma regression model using 3 leg junctions with a 60km/h speed limit as baseline. 44

Table 6 Posterior means and corresponding (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the Norwegian accident data. 48

Table 7 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the Norwegian accident data using a three leg junction with a 60km/h speed limit as baseline. 50

Table 8 Expected number of accidents per year for Norwegian junctions obtained by a Poisson Log-Normal regression model using a three leg junction with a 60km/h speed limit as baseline. 51

Table 9 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the Norwegian accident data. 56

Table 10 Comparison of DIC and related statistics for the three models fitted to the Norwegian junction data. 57

Table 11 Summary statistics for the variables registered on Austrian junctions from 2007 to 2010. 62

Table 12 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson regression model was fitted to the Austrian accident data using a '*Y*' type junction with a '*yield'* traffic control as baseline. 64

Table 13 Expected number of accidents per year for Austrian junctions obtained by a Poisson regression model using a '*Y*' type junction with a '*yield'* traffic control as baseline. 65

Table 14 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model

was fit to the Austrian accident data using a '*Y*' type junction with a '*yield'* traffic control as baseline. 70

Table 15 Expected number of accidents per year for Austrian junctions obtained by a Poisson-Gamma regression model using a '*Y*' type junction with a '*yield'* traffic control as baseline. 71

Table 16 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the Austrian accident data. 75

Table 17 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the Austrian accident data using a '*Y*' type junction with a '*yield'* traffic control as baseline. 76

Table 18 Expected number of accidents per year for Austrian junctions obtained by a Poisson Log-Normal regression model using a '*Y*' type junction with a '*yield'* traffic control as baseline. 78

Table 19 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the Austrian accident data. 83

Table 20 Comparison of DIC and related statistics for the three models fitted to the Austrian junction data. 83

Table 21 Summary statistics for the variables registered on Portuguese junctions from 2003 to 2007. 88

Table 22 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson regression model was fit to the Portuguese accident data using a 3-leg of 'intersection' type junction as baseline...89

Table 23 Expected number of accidents per year for Portuguese junctions obtained by a Poisson regression model using a 3-leg 'intersection' type junction as baseline. 90

Table 24 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fit to the Portuguese accident data using a 3-leg 'intersection' type junction as baseline. 95

Table 25 Expected number of accidents per year for Portuguese junctions obtained by a Poisson-Gamma regression model using a 3-leg 'intersection' type junction as baseline. 96

Table 26 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the Portuguese accident data. 100

Table 27 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the Portuguese accident data using a 3-leg of 'intersection' type junction as baseline. 101

Table 28 Expected number of accidents per year for Portuguese junctions obtained by a Poisson Log-Normal regression model using a 3-leg of 'intersection' type junction as baseline. 102

Table 29 Posterior means (standard deviations) of expected number of accidents for

minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the Portuguese accident data. 106

Table 30 Comparison of DIC and related statistics for the three models fitted to the Portuguese junction data. 106

Table 31 Summary statistics for the variables registered on the aggregated junctions. ...109

Table 32 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson regression model was fit to the aggregated data set. 110

Table 33 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fit to the aggregated data set. 114

Table 34 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the aggregated accident data. 119

Table 35 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fitted to the aggregated data set. 120

Table 36 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the aggregated accident data. 124

Table 37 Comparison of DIC and related statistics for the three models fitted to the aggregated junction data. 124

Table 38 Summary statistics for the variables registered on the aggregated junctions excluding roundabouts. 127

Table 39 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fit to the aggregated accident data (excluding roundabouts) using 3-leg Norwegian junctions as baseline. 128

Table 40 Expected number of accidents per year for the aggregated junction data set (omitting roundabouts) obtained by a Poisson-Gamma regression model using 3-leg Norwegian junctions as baseline. 129

Table 41 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the aggregated accident data (omitting roundabouts). 133

Table 42 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the aggregated accident data (omitting roundabouts) using 3-leg Norwegian junctions as baseline. 134

Table 43 Expected number of accidents per year for the aggregated junction data set (omitting roundabouts) obtained by a Poisson Log-Normal regression model using 3-leg Norwegian junctions as baseline. 135

Table 44 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the aggregated accident data (omitting roundabouts). 139

Table 45 Comparison of DIC and related statistics for the three models fitted to the

aggregated junction data (excluding roundabouts). 140

Table 46 Summary statistics for the variables registered on the aggregated roundabout junctions from Austria Holland and Portugal. 143

Table 47 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fit to the aggregated roundabout accident data Portuguese junctions as baseline. 144

Table 48 Expected number of accidents per year obtained by a Poisson-Gamma regression model, for the aggregated roundabout junction data set Portuguese junctions as baseline. 145

Table 49 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the aggregated roundabout accident data. 147

Table 50 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the aggregated roundabout accident data. 148

Table 51 Expected number of accidents per year obtained by a Poisson Log-Normal regression model, for the aggregated roundabout junction data set. 149

Table 52 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the aggregated roundabout accident data. 153

Table 53 Comparison of DIC and related statistics for the three models fitted to the aggregated roundabout junction data. 153

List of Figures

Figure 1 Example of Gelman-Rubin statistics of a simulated parameter (beta0) showing that the ratio of between to within variability (Rhat) is close to one (red line) and therefore assuming that the corresponding model has converged. 28

Figure 2 Plots of Accidents, Killed, Critical, Serious and Slight, per junction, from upper left to right, respectively, registered from 1997 to 2002 in Norwegian rural road network junctions. 32

Figure 3 Bar plot giving the frequency of the number of accidents registered from 1997 to 2002 on Norwegian junctions on the rural road network. 32

Figure 4 Bar plots giving the frequencies of the number of fatalities and injured victims registered from 1997 to 2002 on Norwegian junctions on the rural road network. 33

Figure 5 The number of accidents per junction in the Norwegian data set against *AADTmaj* and *AADTmin*, and corresponding polynomial fits, on the left and right panels, respectively. 34

Figure 6 Bar plots giving the frequencies of the number of junctions per categories of variables *Number_of_Legs* and *Speed_Limit*, registered from 1997 to 2002 at Norwegian junctions on the rural road network. 34

Figure 7 Box plots of the number of accidents in the Norwegian junctions by group for *Number_of_Legs* and *Speed_Limit*, on the left and right panels, respectively. 35

Figure 8 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson regression model fitted to the Norwegian junction accident data. 36

Figure 9 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson regression model was fitted to the Norwegian accident data. 37

Figure 10 Histogram of the observed number of accidents in Norwegian junctions (left upper corner in grey) and 19 histograms of replicated data sets (“Acc.rep”) of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model. 39

Figure 11 Dot plot of the observed number of accidents in Norwegian junctions (left upper corner in grey) and 19 dot plots from replicated data sets (“Acc.rep”) of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model. 40

Figure 12 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 40

Figure 13 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distributions of the same measure. The *p* is the estimated probability that the measure obtained by the posterior predictive distribution is greater than the one obtained by the observed data. 41

Figure 14 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson-Gamma regression model fitted to the Norwegian junction accident data. 42

Figure 15 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson-Gamma regression model was fitted to the Norwegian data set. 43

Figure 16 Histogram of the observed number of accidents in Norwegian junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model. 45

Figure 17 Dot plot of the observed number of accidents in Norwegian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model. 46

Figure 18 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 46

Figure 19 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 47

Figure 20 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson Log-Normal regression model fit to the Norwegian accident data. 49

Figure 21 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson Log-Normal regression model was fit to the Norwegian data. .51

Figure 22 Histogram of the observed number of accidents in Norwegian junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 52

Figure 23 Dot plot of the observed number of accidents in Norwegian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 53

Figure 24 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1287 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fit to the Norwegian data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 54

Figure 25 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the

posterior predictive distribution from the Poisson Log-Normal regression model fit to the Norwegian data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 55

Figure 26 Values of the posterior means of the expected number of accidents for Norwegian junctions classified per speed limit and number of legs (as under column *Mean* in Table 9). 57

Figure 27 Plots of *Accidents*, *Killed*, and *Serious*, per junction, from upper left to right, respectively, registered from 2007 to 2010 in the Austrian rural road network junctions. 59

Figure 28 Bar plots giving the frequencies of the total number of accidents, fatalities and seriously injured victims registered from 2007 to 2010 on Austrian junctions from the rural road network. 60

Figure 29 The number of accidents per junction in the Austrian data set against *AADTmaj* and *AADTmin*, and corresponding polynomial fit for the *AADTmaj*, on the left and right panels, respectively. 61

Figure 30 Bar plots giving the frequencies of the number of junctions per categories of variables *Junction_Type* and *Traffic_Control*, registered from 2007 to 2010 on Austrian junctions from the rural road network.....61

Figure 31 Box plots of the total number of accidents registered in 2007 -2010 at Austrian junctions by group for *Junction_Type* and *Traffic_Control*, on the left and right panels, respectively. 62

Figure 32 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson regression model fitted to the Austrian junction accident data.....63

Figure 33 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson regression model was fit to the Austrian data. 64

Figure 34 Histogram of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 histograms of replicated data sets (“Acc.rep”) of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.....66

Figure 35 Dot plot of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets (“Acc.rep”) of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.....67

Figure 36 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson regression model fit to the Austrian data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 68

Figure 37 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson regression model fit to the Austrian data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 69

Figure 38 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson-Gamma regression model fitted to the Austrian junction accident data. 69

Figure 39 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson-Gamma regression model was fit to the Austrian data. 71

Figure 40 Histogram of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Gamma regression model. 72

Figure 41 Dot plot of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model. 73

Figure 42 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fit to the Austrian data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 74

Figure 43 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fit to the Austrian data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 74

Figure 44 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson Log-Normal regression model fitted to the Austrian junction accident data. 76

Figure 45 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson Log-Normal regression model was fit to the Austrian data. 77

Figure 46 Histogram of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 79

Figure 47 Dot plot of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 80

Figure 48 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fit to the Austrian data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 81

Figure 49 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fit to the

Austrian data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 82

Figure 50 Values of the posterior means of the expected number of accidents for Austrian junctions classified per junction type and traffic control (as under column *Mean* in Table 19). 84

Figure 51 Plots of the total number of *Accidents*, *Killed_Acc*, and *KSI_Acc*, per junction, from upper left to right, respectively, registered from 2003 to 2007 in Portuguese rural road network junctions. 86

Figure 52 Bar plots of the frequencies of *Accidents*, number of accidents involving fatalities and number of accidents involving killed and seriously injured victims registered from 2003 to 2007 on Portuguese junctions from the rural road network. 86

Figure 53 The number of accidents per junction in the Portuguese data set against *AADTmaj* and *AADTmin*, and corresponding polynomial fits, on the left and right panels, respectively. 87

Figure 54 Bar plots giving the frequencies of the number of junctions per categories of variables *Junction_Type* and *Number_of_Legs*, registered from 2003 to 2007 on Portuguese junctions from the rural road network.....87

Figure 55 Box plots of the number of accidents in the Portuguese junctions by group for *Junction_Type* and *Number_of_Legs*, on the left and right panels, respectively. 88

Figure 56 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson regression model fitted to the Portuguese junction accident data.....89

Figure 57 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson regression model was fit to the Portuguese data. 90

Figure 58 Histogram of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 histograms of replicated data sets (“Acc.rep”) of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model. 91

Figure 59 Dot plot of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 dot plots from replicated data sets (“Acc.rep”) of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model. 92

Figure 60 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson regression model fitted to the Portuguese data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 93

Figure 61 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson regression model fitted to the Portuguese data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 94

Figure 62 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for

the beta coefficient parameters from the Poisson-Gamma regression model fitted to the Portuguese junction accident data. 95

Figure 63 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson-Gamma regression model was fit to the Portuguese data. 96

Figure 64 Histogram of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model. 97

Figure 65 Dot plot of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model. 98

Figure 66 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fit to the Portuguese data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 99

Figure 67| Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fit to the Portuguese data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 99

Figure 68 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson Log-Normal regression model fitted to the Portuguese junction accident data. 101

Figure 69 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson Log-Normal regression model was fit to the Portuguese data. 102

Figure 70 Histogram of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 103

Figure 71 Dot plot of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 104

Figure 72 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fit to the Portuguese data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 105

Figure 73 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fit to the

Portuguese data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 105

Figure 74 Values of the posterior means of the expected number of accidents for Portuguese junctions classified per junction type and number of legs (as under column *Mean* in Table 26). 107

Figure 75 Plot of *Accidents*, per junction, registered in the aggregated rural road network junctions (on the left) and the histogram giving the frequency of the number of accidents (on the right). 108

Figure 76 The number of accidents on the aggregated data set against *AADTmaj* and *AADTmin*, on the left and right panels, respectively, and corresponding polynomial fits. 109

Figure 77 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters obtained after the Poisson regression model was fit to the aggregated data set. 110

Figure 78 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson regression model was fit to the aggregated junction data. .110

Figure 79 Histogram of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 histograms of replicated data set ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model. 111

Figure 80 Dot plot of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model. 112

Figure 81 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson regression model fitted to the aggregated data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 113

Figure 82 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson regression model fitted to the aggregated data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 113

Figure 83 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters obtained after the Poisson-Gamma hierarchical regression model was fit to the aggregated data set. 114

Figure 84 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson-Gamma hierarchical regression model was fit to the aggregated junction data. 115

Figure 85 Histogram of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 histograms of replicated data set ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma hierarchical regression model. 116

Figure 86 Dot plot of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma hierarchical regression model. 117

Figure 87 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fitted to the aggregated data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 118

Figure 88 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fit to the aggregated data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 118

Figure 89 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters obtained after the Poisson Log-Normal hierarchical regression model was fitted to the aggregated data set. 119

Figure 90 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson Log-Normal hierarchical regression model was fitted to the aggregated junction data. 120

Figure 91 Histogram of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 histograms of replicated data set ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal hierarchical regression model. 121

Figure 92 Dot plot of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 122

Figure 93 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fitted to the aggregated data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 123

Figure 94 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fitted to the aggregated data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 123

Figure 95 Plot of *Accidents*, per junctions, registered in the aggregated rural road network junctions excluding the roundabouts (on the left) and the histogram of the frequency of Accidents (on the right).....125

Figure 96 The number of accidents per junction in the aggregated data set, excluding the roundabouts, against *AADTmaj* and *AADTmin*, and corresponding polynomial fits, on the

left and right panels, respectively. 126

Figure 97 Bar plot giving the number of junctions with 3 and 4 legs (on the left panel) and box plot of the number of accidents per number of legs (right panel) for the aggregated data excluding the roundabouts....126

Figure 98 Box plot of the number of accidents in the aggregated set excluding the roundabouts by *Country*. 127

Figure 99 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters, from the Poisson-Gamma regression model fitted to the aggregated junction data excluding the roundabouts....128

Figure 100 Posterior densities of the beta parameter estimates obtained after a Poisson-Gamma regression model was fitted to the aggregated data set excluding roundabouts. 129

Figure 101 Histogram of the observed number of accidents in the aggregated set, excluding roundabouts (left upper corner) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model. 130

Figure 102 Dot plot of the observed number of accidents in the aggregated set, excluding roundabouts (left upper corner) and 19 dot plots of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model. 131

Figure 103 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fitted to the aggregated data excluding roundabouts. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data....132

Figure 104 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fitted to the aggregated data omitting the roundabouts, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 132

Figure 105 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters, from the Poisson Log-Normal regression model fitted to the aggregated junction data excluding the roundabouts. 134

Figure 106 Posterior densities of the beta parameter estimates obtained after a Poisson Log-Normal regression model was fitted to the aggregated data set, excluding roundabouts. 135

Figure 107 Histogram of the observed number of accidents in the aggregated set, excluding roundabouts (left upper corner) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 136

Figure 108 Dot plot of the observed number of accidents in the aggregated set, excluding roundabouts (left upper corner) and 19 dot plots of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 137

Figure 109 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fit to the aggregated data excluding roundabouts. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.....138

Figure 110| Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fitted to the aggregated data omitting the roundabouts, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 138

Figure 111 Values of the posterior means of the expected number of accidents for the aggregated data set excluding the roundabouts, classified per country and number of legs (as under column *Mean* in Table 44).....140

Figure 112 Plot of *Accidents*, per junctions, registered in the Austrian, Dutch and Portuguese roundabout rural road network junction data (on the left) and the histogram of the frequency of *Accidents* (on the right). 141

Figure 113 The number of accidents at roundabouts for Austrian, Dutch and Portuguese data setsagainst *AADTmaj* and *AADTmin*, and corresponding polynomial fits, on the left and right panels, respectively. 142

Figure 114 Box plot of the number of accidents in the aggregated set including only roundabouts (Austria, Holland and Portugal) by *Country*. 142

Figure 115 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters, from the Poisson-Gamma regression model fitted to the aggregated junction roundabout data. 143

Figure 116 Posterior densities of the beta parameter estimates obtained after a Poisson-Gamma regression model was fitted to the roundabout aggregated data set. 144

Figure 117 Histogram of the observed number of accidents in the aggregated roundabout data set (left upper corner) and 19 histograms of replicated data sets (“Acc.rep”) of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model. 145

Figure 118 Dot plot of the observed number of accidents in the aggregated set, including roundabouts only (left upper corner) and 19 dot plots of replicated data sets (“Acc.rep”) of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.....146

Figure 119 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fit to the aggregated roundabout data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 146

Figure 120 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fit to the

aggregated roundabout data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 147

Figure 121 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters, from the Poisson Log-Normal regression model fitted to the aggregated junction roundabout data. 148

Figure 122 Posterior densities of the beta parameter estimates obtained after a Poisson Log-Normal regression model was fitted to the roundabout aggregated data set. 149

Figure 123 Histogram of the observed number of accidents in the aggregated roundabout data set (left upper corner) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 150

Figure 124 Dot plot of the observed number of accidents in the aggregated set, including roundabouts only (left upper corner) and 19 dot plots of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model. 151

Figure 125 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fitted to the aggregated roundabout data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data. 151

Figure 126 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fitted to the aggregated roundabout data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data. 152

1 Introduction

"ERA-NET ROAD – Coordination and Implementation of Road Research in Europe" was a Coordination Action funded by the 6th Framework Programme of the EC. The partners in ERA-NET ROAD (ENR) were the United Kingdom, Finland, Netherlands, Sweden, Germany, Norway, Austria, Slovenia, Belgium, Hungary and Ireland (www.road-era.net). Within the framework of ENR this joint research programme **Safety at the Heart of Road Design** was initiated in 2008. As part of this joint programme, the project Road Infrastructure Safety Management Evaluation Tools (RISMET) was initiated to provide road authorities with a set of easy to use state of the art guidelines for applying road safety engineering management tools.

This report describes the study carried out by LNEC (National Laboratory of Civil Engineering) as partner of the RISMET EU project in work package WP4 (Development of evaluation tools for the future). It consists of a statistical analysis of road accident data occurring at junctions of the rural road networks of four European countries; Austria, Holland, Norway and Portugal; as well as the analysis of aggregated data sets formed by joining together the data sets from each country. In the RISMET Description of Work (DoW) document (see Schermers and Elvik, 2009) it is stated that reports on six countries will be conducted. However, due to the fact that not all countries possessed complete sets of data, the project steering committee decided to integrate instead the results corresponding to countries where complete data sets were available.

The statistical analyses described in this deliverable made use of data sets from each country and by employing a Bayesian statistical approach, namely hierarchical Bayesian regression models, developed country specific and aggregated accident prediction models for junctions.

Several models have been proposed in traffic safety literature for analysing accident data. These models range from the standard Negative Binomial (see Lord 2006, Hauer 2002, Zhang *et al.* 2007 and Park and Lord 2008) to more complex models such as hierarchical Poisson-Gamma and Poisson Log-Normal (Lord and Miranda Moreno 2008) and more recently the Conway-Maxwell-Poisson models as described in Park and Lord (2009). However, the hierarchical Poisson-Gamma remains one of the most popular.

In the analysis performed and described in this report three regression models were fitted to the data sets; these models included the Poisson regression model, the Poisson-Gamma hierarchical model and the Poisson Log-Normal hierarchical regression model. In a hierarchical Bayesian analysis, the parameters of the prior distributions depend in turn on additional parameters with their own priors who are also referred as hyper-priors, see for e.g. Carlin and Louis (2000), Gelman *et al.* (2004) and Congdon (2010). When the hyper-prior densities are chosen to be "vague" or "non-informative" they guarantee to play a minimal role in the posterior distribution. The rationale for using non-informative prior distributions is often said to "let the data speak for themselves", so that inferences are unaffected by information external to the current data (Gelman *et al.* 2004).

The regression model form in this study was chosen according to studies by, amongst others, Lord (2006), Miaou and Lord (2003) and Eenink *et al.* (2007) and is the functional form most favoured by transportation safety modellers for modelling crash data at intersections (Lord, 2006). Non-informative or vague prior and hyper-prior distributions were used in all models due to lack of previous information and knowledge about the models parameters.

The analyses were performed using the following statistical software: R: *A Language and Environment for Statistical Computing*, developed and maintained by the R Development Core Team (2011) and a software developed for Bayesian analysis based on a programming language: WinBUGS (see Spiegelhalter *at al.* 2003 and Lunn *et al.* 2000). WinBUGS is *S*

based (*S* is a statistical programming language developed at the Bell Laboratories) and offers the basis for sophisticated programming and data manipulation with a distinctive Bayesian functionality. WinBUGS selects appropriate Markov chain Monte Carlo (MCMC) updating schemes via an built-in expert system which may be criticised as being something of a blackbox (Congdon, 2006). This latter software is used to generate random samples from the model parameters posterior distribution of a Bayesian model. The analysis also made use of the R package R2WinBUGS, a package for running WinBUGS from R developed by Sturtz *et al.* (2005).

The parameter estimates of the regression models were obtained by posterior inference resulting from MCMC simulation methods such as Gibbs sampling and Metropolis Hastings algorithm (see Gelman *et al.* 2004 and Ntzoufras, 2009 for more details) which are implemented in the WinBUGS software.

The original data set of Dutch junctions consisted of around 500 junctions of which only some 10 per cent had traffic volume data on all approaches. Due to the limited sample statistical analysis for the purpose of developing accident prediction models (see Elvik, 2010 for more details) could not be performed. However, it was decided to utilise the certain intersection data from the Netherlands in the aggregated set formed by European roundabouts.

In all the models considered, the dependent variable was taken to be the number of injury accidents (i.e. accidents with at least one injured victim) registered per junction. The explanatory variables were the continuous values of the logarithms of the major and minor annual average daily traffic volumes (AADT) and the categorical variables were the type of junction, number of approaches (also referred to as legs in the report), traffic control and speed limit of approaching roads. Three accident prediction models were developed for data in each of the following countries: Norway, Austria and Portugal. The aggregated data of these three countries were analysed taking only the major and minor AADT values. Further, an aggregated set formed by non-roundabout junctions from Austria, Norway and Portugal were analysed with the following explanatory variables: number of legs, AADT volumes and an indicator categorical variable denoting the country. Another set analysed comprised the roundabout junctions from Austria, the Netherlands and Portugal (the junctions collected in Norway did not comprise roundabouts).

1.1 Background and objectives

Safety Performance Fucntions (SPF) establish the relations between the accident frequency of a roadway element (during a fixed period of time) and a selection of its characteristics, namely those related to its geometry, hierarchy in the road network and traffic. Usually, the average annual daily traffic is amongst the most important explanatory variables. SPF are very important for measuring safety through accident and injury frequencies, as they may be used to estimate the long term expected number of accidents at a specific road element (for example, a curve or an intersection), through the empirical Bayes method (Hauer, 1997). The expected number of accidents may be used to compare the safety performance of a roadway location with the expected safety performance of comparable sites; to identify deviant sites, with extremely high expected number of accidents, for safety intervention; to evaluate the safety effect of safety interventions; and to forecast safety performance developments of alternative planning scenarios and preliminary design schemes.

1.2 Structure of the report

Chapter 2 of this report gives a description of some of the theory behind the various models considered and corresponding measurements of assessment. Chapter 3 describes the analysis performed with the Norwegian junction data set, the three models fitted and the overall conclusions. Chapters 4 and 5 describe the analysis performed with the Austrian and Portuguese data sets and present the main conclusions. Chapter 6 presents three regression

models developed using the aggregated data set comprising Austrian, Norwegian and Portuguese junctions and applying the logarithms of AADT major and minor as sole explanatory variables. Chapter 7 describes the analysis performed on an aggregated data set consisting of Norwegian, Austrian and Portuguese non-roundabout junctions. Chapter 8 is concerned with the analysis of Austrian, Dutch and Portuguese roundabout junctions.

The main results and conclusions obtained are presented in Chapter 9.

2 Methodological Approach

This Chapter provides an overview of the three Bayesian regression models that were considered appropriate (Lord and Miranda Moreno 2008; Elvik, 2011) to fit to the data registered at junctions from the rural road networks of three European countries. The chapter provides the theory which is adopted in the following chapters and in which the models are fitted to the country data. For completeness the chapter provides insight into how the model fit and convergence can be assessed.

2.1 The Poisson Regression Model

The benchmark model for (accident) count data is the Poisson distribution (Cameron and Trivedi, 1998). The approach taken to the analysis of count data, especially the choice of the regression framework depends on how the counts are assumed to arise. According to Cameron and Trivedi (1998) they can arise from a direct observation of a point process, examples of which include the number of road accidents. The standard model for count data is, consequently, the Poisson regression model. This regression model is derived from the Poisson distribution by allowing the intensity parameter μ to depend on covariates (regressors). If the dependence is parametrically exact and involves exogenous covariates and no other source of stochastic variation, then one obtains the standard Poisson regression. A usual application of Poisson regression is to data consisting of n independent observations, the i^{th} of which is (y_i,x_i). The scalar dependent variable, y_i, is the number of occurrences of the event of interest, and x_i is the vector of linearly independent regressors that are thought to determine y_i. A regression model based on this distribution follows by conditioning the distribution of y_i on a k-dimensional vector of covariates, $x_i^{'} = [x_{1i},...,x_{ki}]$, and parameters β, through a continuous function $\mu(x_i,\beta)$, such that $E[y_i \mid x_i] = \mu(x_i,\beta)$.

That is, y_i given x_i is Poisson distributed with density:

$$f(y_i \mid x_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \quad y_i = 0,1,2,... \tag{2.1}$$

In the log-linear version of the model the mean parameter is parameterised as:

$$\mu_i = \exp(x_i^{'}\beta), \tag{2.2}$$

to ensure that $\mu > 0$. Equations (2.1) and (2.2) jointly define the Poisson (log-linear) regression model (Cameron and Trivedi, 1998).

2.2 Mixture Models

The Poisson model assumes that the variances of the observations are known functions of the mean parameters. In practice, data of this type may be more dispersed than the Poisson density assumes. According to Congdon (2005) such overdispersion of excess heterogeneity may reflect, amongst others, a few extreme observations, variation between units of widely different exposures or, alternatively, it may be due to unobserved variations between

subjects (or frailties) ϕ_i that are not represented by the observed covariates. Without correction for extra variation the precision of the β parameters will be overstated: their credible intervals will be too narrow.

The Poisson density may need to be modified when the observed variance exceeds the form assumed under the density, and this involves a mixture distribution on the Poisson mean.

Mixture generalizations can be seen as providing greater robustness in inferences (Gelman *et al*., 2004) and as providing a density that is compatible with the data. Another motivation for mixture models is to pool information over units when event counts for each unit may vary considerably; by modelling the rates for individual units in terms of an overall hyperdensity, shrinkage estimates may be obtained for each unit that smoothes towards the average (Congdon, 2005).

2.2.1 Poisson-Gamma Models

The unobserved heterogeneity in a regression analysis of count data, especially if, as according to Congdon (2006), overdispersion is attributable to variations in proneness between individuals or to unknown predictors, can be represented by a multiplicative frailty (ϕ_i) with log-link such that:

$$y_i \sim Poisson(\nu_i)$$
$$\nu_i = \mu_i \phi_i = \exp(\beta X_i)\phi_i \quad (2.3)$$

Assuming a Gamma frailty model:

$$\phi_i \sim Gamma(\delta, \gamma), \quad (2.4)$$

conjugate[1] to the Poisson density, then:

$$P(y_i \mid X_i, \phi_i)P(\phi_i) = \left\{\exp(-\phi_i \mu_i)[\phi_i \mu_i]^{y_i} / y_i!\right\}$$
$$= \left\{\delta^{\gamma} \phi_i^{\delta-1} \exp(-\gamma\phi_i) / \Gamma(\delta)\right\}$$

Integrating out ϕ_i, as in:

$$P(y_i \mid X_i) = \int P(y_i \mid X_i, \phi_i)P(\phi_i)d\phi_i$$

Leads to a marginal negative binomial density for y_i.

The identifiability constraint $\delta = \gamma$ is frequently used. With this constraint $Var(\phi_i) = 1/\delta$ and the Negative Binomial $NB(\mu_i, \delta)$ has the form:

$$P(y_i \mid X_i) = \Gamma(\delta + y_i)\{\Gamma(\delta)\Gamma(y_i + 1)\}\left(\frac{\delta}{\delta + \mu_i}\right)^{\delta}\left(\frac{\mu_i}{\mu_i + \delta}\right)^{y_i}$$

With estimation by repeated sampling it is straightforward to analyse count data using either the Negative Binomial likelihood or the mixed Poisson-Gamma likelihood, with the latter approach having the benefit of providing observation specific frailties (Fahrmeir and Osana, 2006).

There are several alternative Gamma mixtures and generalisations of the Negative Binomial that can be employed. They are described in Congdon (2005, 2006 and 2010).

2.2.2 Poisson Log-Normal Model

[1] For definition of conjugate densities see Gelman *et al*., (2004).

According to Congdon (2005) the main alternative to a model including a conjugate frailty is an additive random error in $g(\mu_i)$ so that:

$$y_i \sim Poisson(\mu_i)$$
$$g(\mu_i) = \beta_0 + \beta X_i + \varepsilon_i \quad (2.5)$$

Where ε_i may follow a parametric density such as the Normal. If,

$$\varepsilon_i \sim N(0, \alpha) \quad (2.6)$$

then to a close approximation, $Var(y_i \mid X) = \mu_i + \alpha\mu_i^2$.

The parameter α can follow a *a priori* non-informative Gamma distribution.

Equations (2.5) and (2.6) jointly define the Poisson log-Normal regression model.

2.3 Convergence Assessment

Bayesian inference has become closely linked to sampling-based estimation methods (Congdon, 2006). Both focus on the entire density of a parameter or function's of parameters. Iterative Monte Carlo methods involve repeated sampling that converges to sampling from the posterior distribution. Such sampling provides estimates of density characteristics (moments, quantiles), or of probabilities relating to the parameters (Smith and Gelfand, 1992).

Monte Carlo methods encompass several algorithms that employ simulation to solve multiple statistical and mathematical problems; one of these algorithms is Markov chains, for more details see Gelman *et al*. (2004), Congdon (2003, 2005, 2006 and 2010) and Ntzoufras (2009). Monte Carlo methods are used for obtaining the sought after *a posteriori* values, namely the probability of the occurrence of a certain event. In order to obtain reliable results it is necessary that the algorithms, in particular the Markov chain algorithm, converge to the respective equilibrium distribution, i.e. its target. If convergence occurs then the obtained sample (simulated sample) comes from the distribution that is sought.

Verifying the convergence of the MCMC algorithm consists in verifying the convergence of the model's parameters, or a set of parameters, if the model has many parameters, estimated by the algorithm. Once the algorithm has converged, the samples from the conditional distributions will be used to summarise the posterior distribution of the parameters of interest. According to the online manual of WinBUGS (see Spiegelhalter *et al*. (2003)) checking the convergence requires a lot of care, being very difficult to state that the chain (simulation) converges; it is only possible to diagnose when it clearly does not converge. Nevertheless, monitoring the convergence of the algorithm is essential for producing results from the posterior distribution of interest. There are many ways to monitor convergence. The simplest way is to monitor the Monte Carlo (MC) error, which measures the variability of each estimate due to simulation; this error should be low in order to calculate the parameter of interest with increased precision. It is sometimes suggested that the MC error should be less than 5% of the posterior standard deviation of a parameter (Congdon, 2010). Monitoring autocorrelations and the plots of iterations versus the generated values can also be very useful.

According to Ntzoufras (2009) all convergence diagnostics work like "alarms" that sound when they detect an unexpected anomaly in the MCMC output. Each diagnostic test is constructed to detect different problems and hence, in most cases, all diagnostic must be applied to ensure that convergence has been reached. Nevertheless, the diagnostics based on the Gelman-Rubin statistics are considered the more formal ones and have consequently the most reliable results.

2.3.1 Gelman-Rubin Diagnostics

Gelman *et al*. (2004) argue that the way to best identify a non-convergence is to simulate several multiple chain sequences of Markov chains with differing starting values. The aim is to verify whether the chains have a similar behaviour. When several chains are simulated in parallel, each one starting from different initial values, it is possible to calculate the Gelman-Rubin convergence diagnostic. The Gelman-Rubin diagnostic consists in obtaining, and later comparing, the between-sample and the within-sample variability (i.e. inter-sample and intra-sample variability).

WinBUGS calculates the Gelman-Rubin statistic by removing *n* samples from *m* parameters θ and calculates the following statistics.

Within-sample variance, *W*:

$$W = \frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{i=1}^{n} (\theta_j^i - \bar{\theta}_j)^2 \tag{2.7}$$

Between-sample variance, *B*:

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\theta_j - \bar{\theta})^2 \tag{2.8}$$

Estimated variance, $V_{hat}(\theta)$:

$$Vhat(\theta) = (1 - \frac{1}{n})W + \frac{1}{n}B \tag{2.9}$$

The Gelman-Rubin statistic is given by *Rhat*:

$$Rhat = \frac{Vhat(\theta)}{W} \tag{2.10}$$

Before attaining the convergence, *W* underestimates the marginal posterior variance in θ since the target distribution has not been completely explored. On the other hand, $V_{hat}(\theta)$, overestimates the variance in θ because the initial values are overdispersed relative to the target distribution. Once convergence is attained, *W* and $V_{hat}(\theta)$ should be equivalent since the inter and intra-sampling variability should coincide. Consequently, *Rhat* should be approximately equal to 1. For more details concerning these statistics, see Ntzoufras (2009) and Gelman *at al*. (2004).

Figure 1 shows an example of a graph obtained by the *W* (blue line), *Vhat* (green line) and *Rhat* (red line) statistics showing no evidence to doubt non-convergence of the three Markov chains.

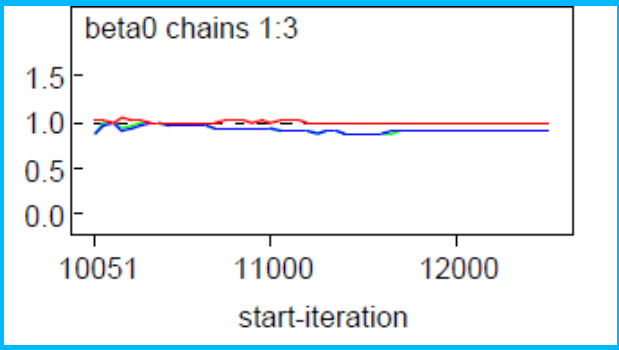

Figure 1 Example of Gelman-Rubin statistics of a simulated parameter (beta0) showing that the ratio of between to within variability (Rhat) is close to one (red line) and therefore assuming that the corresponding model has converged.

For more details concerning these statistics, see Ntzoufras (2009) and Gelman *at al*. (2004).

2.4 Model Assessment

Model assessment involves both the choice between competing models in terms of best fit and checks to ensure model adequacy. Therefore, even if one model has superior fit, it still needs to be checked whether predictions from the model satisfactorily reproduce the observed data. According to Congdon (2010) there are three main strategies to assess model fit and carry out model checks. They include the so-called formal approach, approaches based on posterior analysis of the deviance, and predictive methods based on samples of replicate data. The analysis performed in this report has considered the formal approach and the posterior analysis of deviance.

2.4.1 Deviance Information Criterion and Effective Model Dimension

Spiegelhalter *et al*. (2002) provide a penalized fit criterion called the deviance information criterion (DIC), it is used as a measure of model comparison and adequacy. It can be applicable to comparing non-nested models and also to models including random effects where the true model dimension is another unknown. The DIC is based on the posterior distribution of the deviance statistic:

$$D(\theta \mid y) = -2\log[p(y \mid \theta)] + 2\log[h(y)]$$

Where $p(y \mid \theta)$ is the likelihood of data *y* given the parameters θ, and h(y) is a standardizing function of the data only and so does not affect model choice (see Congdon, 2010).

When the deviance is monitored as an extra in a MCMC run, with *R* iterations, it will produce samples $\{D^{(1)},\ldots,D^{(R)}\}$. The overall fit of a model is measured by the posterior expected deviance obtained by averaging over the posterior density of the parameters:

$$\bar{D} = E_{\theta|y}[D], \tag{2.11}$$

While the effective model dimension, d_e, is estimated as:

$$d_e = E_{\theta|y}[D] - D\left(E_{\theta|y}[\theta]\right) = \bar{D} - D\left(\bar{\theta}\right) \tag{2.12}$$

Basically, the effective model dimension is the expected deviance minus the deviance at the posterior means of the parameters.

The DIC is obtained as the expected deviance plus the effective model dimension:

$$DIC = \bar{D} + d_e = D\left(\bar{\theta}\right) + 2d_e \tag{2.13}$$

Therefore, using DIC, models with lower values of $\bar{D}$ will be favoured, combined with small values of d_e, which indicate a relatively parsimonious model as stated by Congdon (2010).

The examination of the DIC values can also be used to variable selection (see Ntzoufras, 2009) by choosing the model with the lowest DIC value.

According to Carlin and Louis (2009), just like other penalised likelihood criteria, DIC is not intended for identification of the “correct” model, but rather merely as a method of comparing a collection of alternative formulations (all of which may be incorrect). The same authors also state that the values of DIC have no intrinsic meaning; only differences in DIC across models are meaningful, with differences of 3 or 5 normally being thought of as the smallest that are interesting.

2.4.2 Posterior Predictive Checking

Gelman *et al*. (2004) proposed a diagnostic procedure known as posterior checking that makes use of predictive replicates y_{new} or y_{rep}. The idea behind this is the following: if the model fits, then replicated data generated under the model should look similar to the observed data. Or, as Gelman *et al*. (2004) also state: the observed data should look plausible under the posterior predictive distribution.

One basic technique for checking the fit of a model to data is to draw simulated values from the posterior predictive distribution of replicated data and compare these samples to the observed data. Any systematic differences between the samples and the data indicate potential failings of the model.

Various forms of checking function may be calculated for both new data and actual observations to assess whether the model satisfactorily reproduces certain important aspects of the actual data. For example, according to Congdon (2005) if count data are overdispersed, then the model should reproduce such features in the replicates (y_{new} or y_{rep}) which are sampled from the model.

Suppose $T(y;\theta)$ is the observed criterion (e.g. a ratio of observed variance to mean). Let the same criterion based on replicated data be denoted by $T(y_{rep};\theta)$. A reference distribution is obtained from the joint distribution of y_{rep} and θ:

$$P(y_{rep},\theta) = P(y_{rep} \mid \theta)P(\theta \mid y)$$

And the actual value obtained by sampling set against this distribution.

In practice, at each iteration t the criteria $T(y_{rep}^{(t)},\theta^{(t)})$ and $T(y,\theta^{(t)})$ are obtained and the proportion of iterations where $T(y_{rep}^{(t)},\theta^{(t)})$ exceeds the other, is also obtained, namely:

$$\hat{p}_T = \Pr(T(y_{rep},\theta) > T(y,\theta) \mid y) \tag{2.14}$$

This quantity is estimated as:

$$\hat{p}_T = \sum_{t=1}^{Z} 1(T(y_{rep}^{(t)},\theta^{(t)}) > T(y,\theta^{(t)})) > T(y,\theta^{(t)}))/Z \tag{2.15}$$

Where Z is the total number of iterations considered.

According to Congdon (2005) values of $\hat{p}_T$ near 0 or 1 (below 0.1 or above 0.9) indicate discrepancy between the observations and the model. Values relatively close to 0.5 mean that the actual data and the predicted (i.e. replicated) data sampled from the model are closely comparable in terms of the feature that the checking function summarises.

Another model checking procedure based on replicated data is suggested by Gelfand (1996) and involves checking for all sample cases i=1,…,n whether observed y are within 95% intervals of y_{new}.

3 Modelling Norwegian injury accidents

The present Chapter describes the development and the assessment of three accident prediction models, obtained with the employment of statistical Bayesian techniques, for injury accidents occurring at junctions from the Norwegian national road network.

The first section gives a brief description of the data. The following sections detail the model fitting and checking for each regression model described in sections 2.1, 2.2.1 and 2.2.2 of Chapter 2.

Three regression models were fitted to the data; they include the Poisson (section 3.2), Poisson-Gamma (section 3.3) and Poisson Log-Normal models (section 3.4). The results

presented concern a period of annual frequencies even though the data was collected over a six years period of time.

3.1 Norwegian Junction Data

The data analysed was generously made available by Professor Stein Johannessen of the Norwegian University of Science and Technology (NTNU) in Trondheim via Professor Rune Elvik of the Institute of Transport Economics of the Norwegian Centre for Transport Research (TØI). The data set consists of measurements registered at 732 junctions on Norwegian national roads located in the counties of Østfold, Akerhus, Hedmark and Oppland. Traffic at all junctions was controlled by yield signs on the minor approaches and no junctions were roundabout controlled. The data was collected over a six year period from 1997 to 2002.

The variables registered included, per junction:

- ***Junction_number***: The number of the junction (from 1 to 732 in Index of junctions)
- ***Number_of_Legs***: a binary (categorical) variable indicating whether the junction was formed by “3” or “4” legs;
- ***Speed_Limit***: a categorical variable referring to the speed limit allowed on the vicinity of the junction (it takes the values “40”, “50”, “60”, “70”, “80” and “90” representing km per hour) (Note: speed limit is used as a variable describing a certain design standard/philosophy and NOT driving behaviour) ;
- ***AADTmaj***: annual average daily traffic (AADT) volume on the major approaches;
- ***AADTmin***: annual average daily traffic volume on the minor approaches;
- ***Accidents***: the number of injury accidents occurring within 50 metres of the junction;
- ***Killed***: the number of fatalities;
- ***Critical***: the number of critically injured victims (life threatening injuries or accidents associated with permanent impairment);
- ***Serious***: the number of seriously injured victims (requiring in-hospital treatment);
- ***Slight***: the number of slightly injured victims;

Before proceeding to modelling, it is useful to give some graphical descriptions of the data which will also help in investigating the relationship between the number of accidents (*Accidents*) and the eligible explanatory variables (*AADTmaj*, *AADTmin*, *Legs* and *Speed_Limit*).

The dot plots depicted in Figure 2 consist of plots representing the number of accidents (*Accidents*), number of fatalities (*Killed*), number of critically injured (*Critical*), number of seriously injured (*Serious*) and number of slightly injured (*Slight*) per junction registered over the six year period. It can be observed that the highest number of accidents on a junction is 9. Most junctions have zero fatalities and zero critical injured victims. There are few junctions where up to 19 slightly injured victims were registered.

Another way of presenting the same data consists of the use of bar plots which can be observed in Figure 3 and Figure 4, for variable *Accidents* and for the various types of injured victims, respectively.

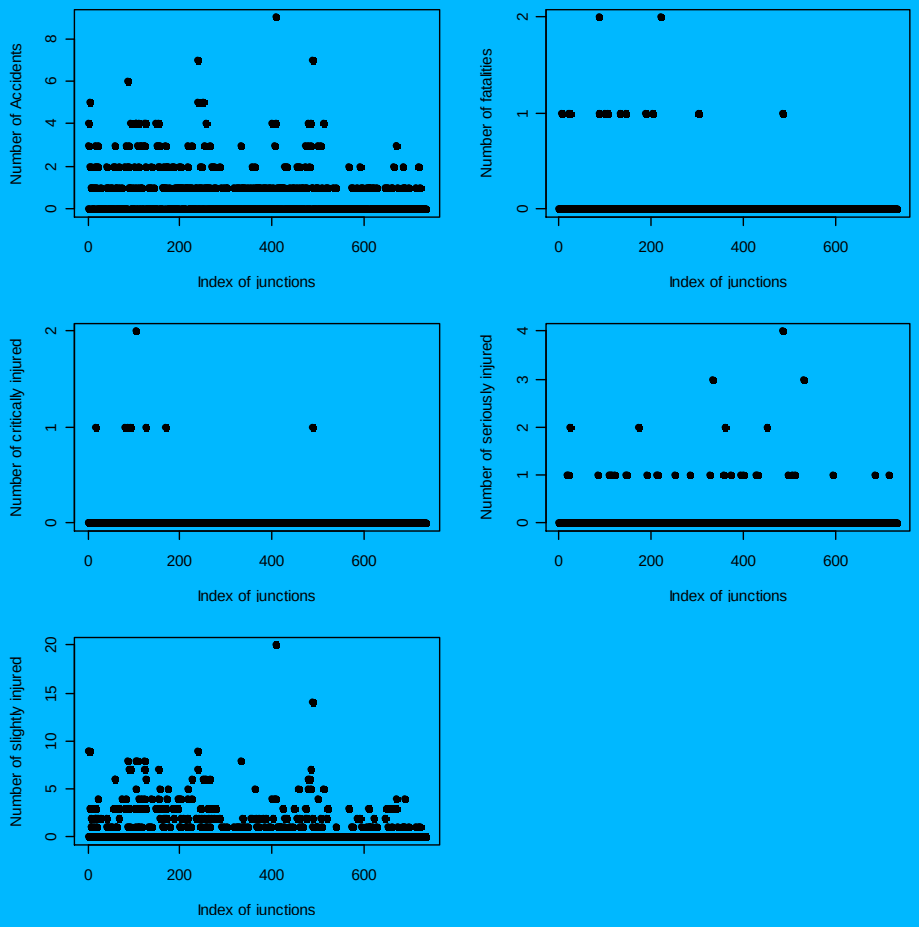

Figure 2 Plots of Accidents, Killed, Critical, Serious and Slight, per junction, from upper left to right, respectively, registered from 1997 to 2002 in Norwegian rural road network junctions.

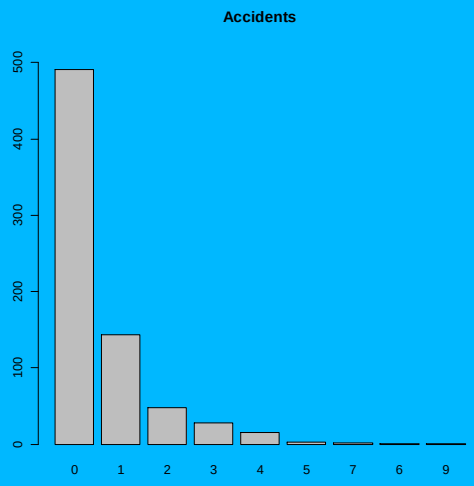

Figure 3 Bar plot giving the frequency of the number of accidents registered from 1997 to 2002 on Norwegian junctions on the rural road network.

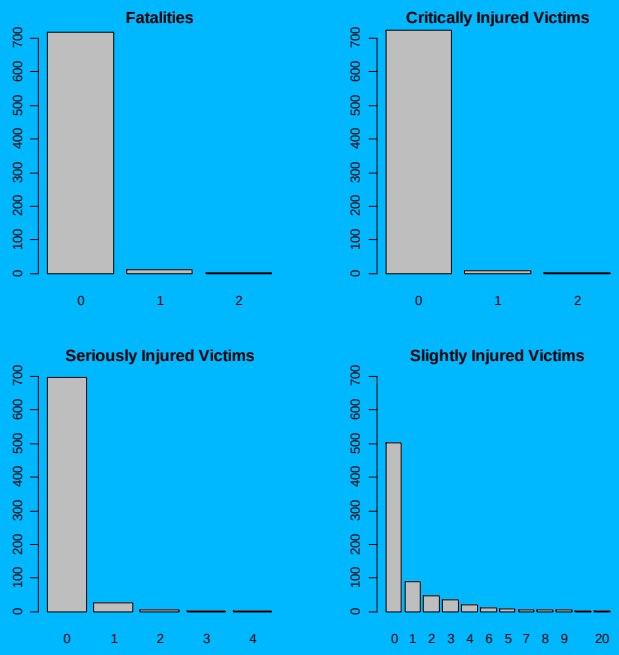

Figure 4 Bar plots giving the frequencies of the number of fatalities and injured victims registered from 1997 to 2002 on Norwegian junctions on the rural road network.

The two graphs depicted in Figure 5 show the *AADTmaj* and *AADTmin* values plotted against the number of accidents (*Accidents*) per junction, on the left and right panels, respectively. A fitted smooth regression curve, obtained with a regression function (loess) developed by Venables and Ripley (2002) is also included in the plots. The numbers on the plots next to some of the points represent the junction's indexes for their better identification. The inclusion of the smooth regression curve served the purpose of providing an overall idea of how the number of accidents varies with the various AADT values. Some junctions posses high values of either *AADTmaj* or *AADTmin*, which certainly influenced the smooth regression equation obtained. Nevertheless these values are not considered to be disproportionate and therefore it was decided not to remove the corresponding junctions from the data set.

Most accidents at Norwegian rural junctions occur at values of *AADTmaj* varying between 133 and 12000 vehicles per day (Figure 5 and Table 1) and *AADTmin* values of between 7 and 3000 vehilces per day. The number of accidents seem to stabilise for values of *AADTmin* greater than 2000 (right panel on Figure 5), however, the smooth regression line on the left panel decreases due to an occurrence of one accident at junction indexed as 43 which has an *AADTmaj* value greater than 30000.

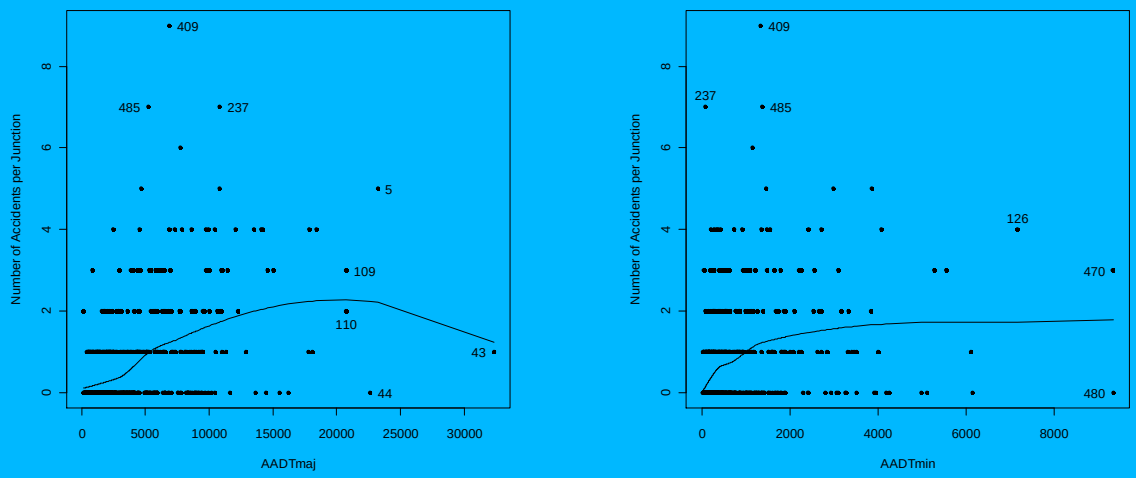

Figure 5 The number of accidents per junction in the Norwegian data set against *AADTmaj* and *AADTmin*, and corresponding polynomial fits, on the left and right panels, respectively.

The bar plots in Figure 6 show the number of junctions with 3 and 4 legs (left panel) and the number of junctions grouped by speed limit (right panel).

The left panel in Figure 7 (box plot) indicates that the distribution of the two groups of variable *Legs* differ as far as the degree of skewness (to the right) is concerned. The median of *Accidents* also differs according with the number of legs, with junctions with 4 legs having higher median (of accident counts) than junctions with 3 legs (with median equal to zero). Consequently, the distribution of the number of accidents seems to differ whether the junction is formed by 3 or 4 legs.

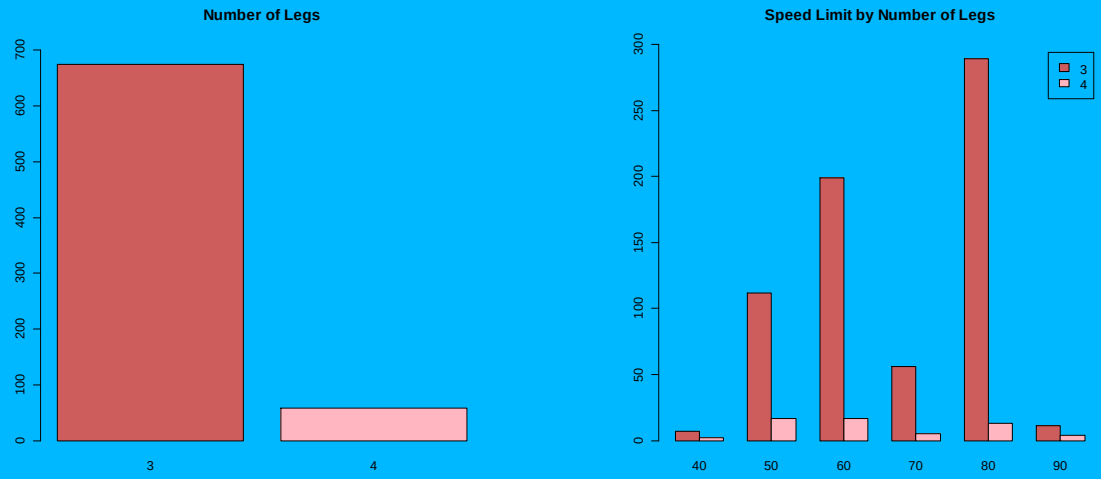

Figure 6 Bar plots giving the frequencies of the number of junctions per categories of variables *Number_of_Legs* and *Speed_Limit*, registered from 1997 to 2002 at Norwegian junctions on the rural road network.

The plot on the right panel in Figure 7 shows the boxplots of the number of accidents per speed limit. It can be observed that the medians for the number of accidents are similar for junctions with various speed limits apart from those with speed limit of 70km/h which have a higher value for the median of accidents when compared with the others.

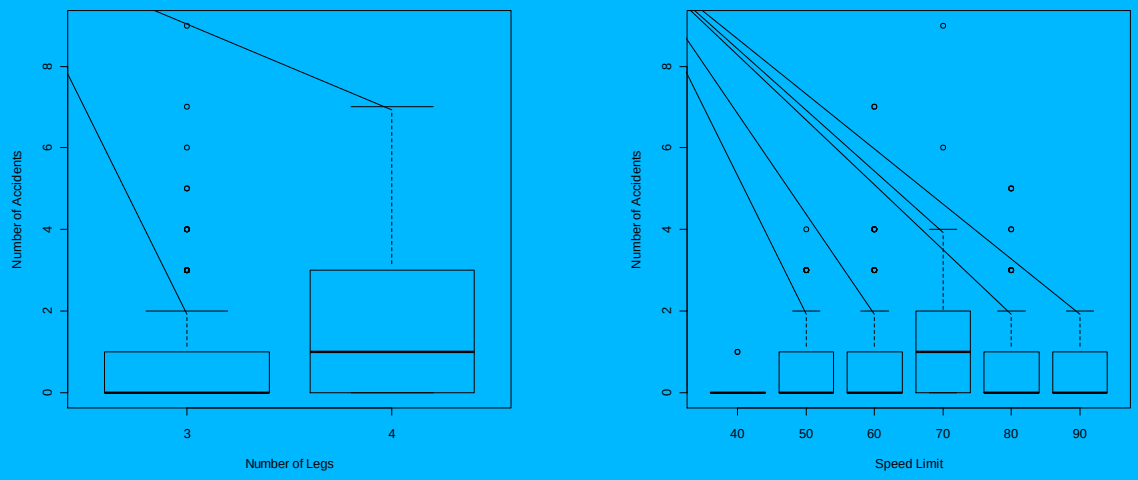

Figure 7 Box plots of the number of accidents in the Norwegian junctions by group for *Number_of_Legs* and *Speed_Limit*, on the left and right panels, respectively.

Table 1 contains descriptive statistics for all the variables registered from 1997 to 2002 at Norwegian junctions on rural roads.

Table 1 Summary statistics for the variables registered on Norwegian junctions from 1997 to 2002.

Variables	minimum	mean	standard deviation	median	maximum
AADTmaj	133	3615.0	3600.1	2266	32311
AADTmin	7	646.7	1011.9	275	9332
Number_of_Legs	3	-	-	-	4
Speed_Limit	40	-	-	-	90
Accidents	0	0.583	1.090	0	9
Killed	0	0.022	0.164	0	2
Critical	0	0.012	0.122	0	2
Serious	0	0.858	0.319	0	4
Slight	0	0.893	1.848	0	20

3.2 The Poisson Regression Model

The Poisson regression model as given by Equations 2.1 and 2.2 was fit to the data where the number of accidents (*Accidents*) was considered to be the dependent variable and the logarithms of *AADTmaj* and *AADTmin*, as well as the categorical variables *Legs* and *Speed_Limit*, were taken as explanatory, as shown on Equation 3.1 (as in Lord 2006 and Elvik 2010) and corresponding to Equation 2.2.

$$ln(\hat{\mu}_i) = \beta_0 + \beta_1 ln(AADTmaj_i) + \beta_2 ln(AADTmin_i) + \beta_3 Number\_of\_Legs_i + \beta_4 Speed\_Limit_i \quad (3.1)$$

The parameter $\hat{\mu}_i$ gives the expected number of accidents for a period of one year. The β parameters were assigned Normal *a priori* distributions with mean zero and precision

(inverse of the variance) equal to 0.0001.

The following baseline, or reference, categories were used for the categorical variables:

Number_of_Legs = 3;

Speed_Limit = 60km/h.

The MCMC algorithm comprised three chains and was run for 35000 iterations with 33000 iterations considered as burn-in with a thinning rate equal to 7. The final sample had dimension 1002. In the burn-in period a chosen number of iterations are eliminated from the sample in order to avoid the influence of the initial values, the thinning rate equal to 7 means that the first generated values in every batch of 7 iterations was kept.

From the observation of the graphs obtained from the Gelman-Rubin statistics (see Chapter 2) for the estimates of the β parameters depicted in Figure 8 it can be seen that the *Rhat* statistic (red line in the plot and given by Equation 2.10) converges to 1 and that both *W* (blue line) and *Vhat* (green line) stabilise as the number of iterations increase. These indicate that there seems to be convergence of the iterative simulation.

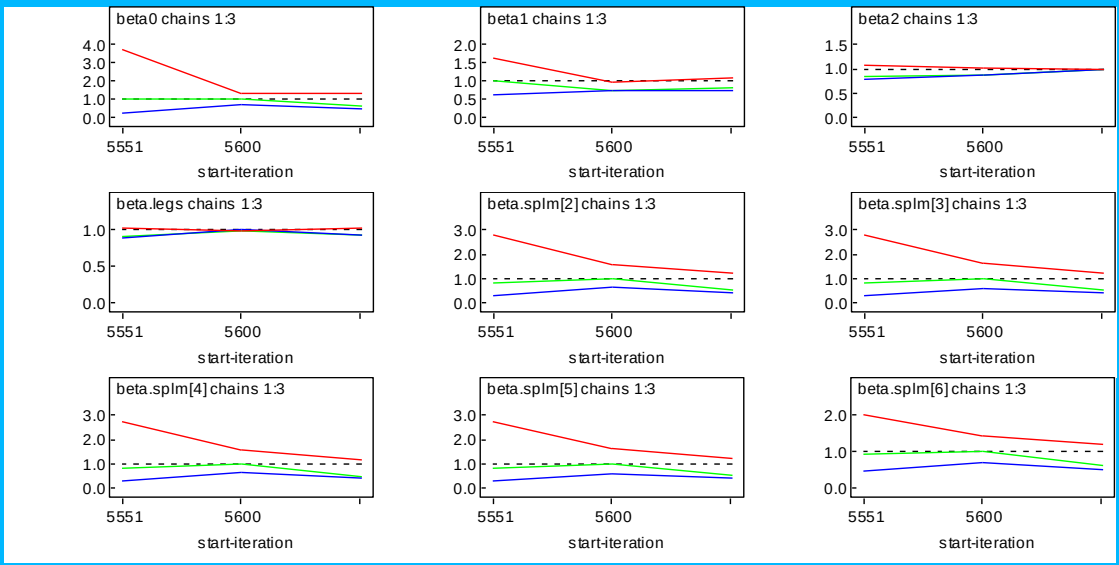

Figure 8 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson regression model fitted to the Norwegian junction accident data.

The point estimates for the model thus obtained are given in Table 2. These point estimates are provided by the posterior means obtained for the distributions of each unknown parameter. Table 2 shows also the standard deviations, Monte Carlo errors and 95% credible intervals for the estimates of the estimates of the β parameters shown in Equation 3.1.

Table 2 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson regression model was fitted to the Norwegian accident data.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-10.860	1.308	1.765E-01	-13.960	-8.915
β_1 (ln (AADTmaj))	0.706	0.072	9.088E-03	0.586	0.862
β_2 (ln(AADTmin))	0.251	0.044	3.825E-03	0.166	0.331
β_3 (Legs='4')	0.776	0.129	5.357E-03	0.508	1.014
β_4 (Speed_Limit='40')	0.608	0.927	1.218E-02	-0.725	2.576
β_4 (Speed_Limit='50')	1.146	0.927	1.228E-02	-0.136	3.119
β_4 (Speed_Limit='70')	1.572	0.929	1.226E-02	0.253	3.520
β_4 (Speed_Limit='80')	1.229	0.944	1.251E-02	-0.085	3.226
β_4 (Speed_Limit='90')	0.863	1.036	1.274E-02	-0.723	3.152

The 95% credible intervals for the estimates of the non-categorical independent covariates do not include zero, indicating that these variables (*ln(AADTmaj)* and *ln(AADTmin)*) have a relevant effect on the prediction of the number of accidents (see Congdon (2005) and Ntzoufras (2009)).

The posterior densities of the parameter estimates are shown in Figure 9. These densities are functions that describe the relative likelihood (in the *y* axis) of the parameter estimates (considered here as a random variable) to occur at the given points of the *x* axis. The observation of Figure 9 indicates that the mean values of all the densities have moved away from zero (the mean value assumed *a priori*), even if the 95% credible interval (see also Table 2) covered zero in some cases.

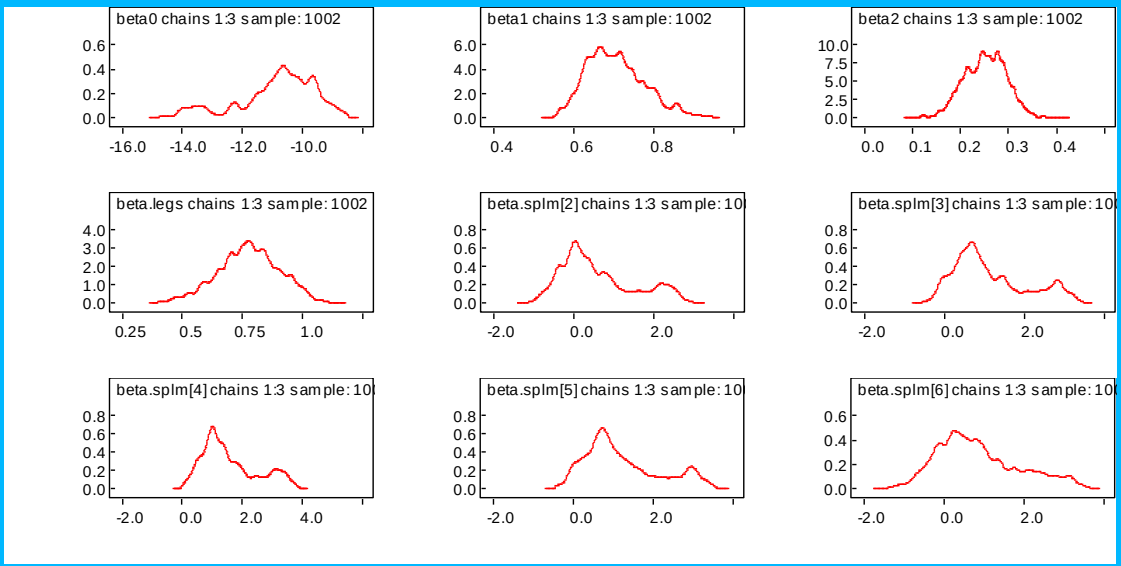

Figure 9 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson regression model was fitted to the Norwegian accident data.

The interpretation of the regression coefficients takes into account the fact that they can be exponentiated and treated as multiplicative effects (see Gelman and Hill, 2007 and Ntzoufras, 2009). As an example, it can be stated that the coefficient of *ln(AADTmaj)*, in

equation 3.1, is the expected difference in the number of injury accidents (on the logarithmic scale) for each additional unitary increase in *ln(AADTmaj)*. Thus, the expected multiplicative increase is the exponential of that coefficient. As with regression models in general, each coefficient is interpreted as a comparison in which one predictor differs by one unit while all the other predictors remain at the same level, which is not necessarily the most appropriate assumption when extending the model to new settings (Gelman and Hill, 2007).

Examination of the coefficient estimates in Table 2 suggests that the expected numbers of accidents on four leg junctions are *a posteriori* expected to have approximately 117% more accidents than a three leg junction with the same *AADTmaj* and *AADTmin* values and speed limit. A unitary increase in either *ln(AADTmaj)* or *ln(AADTmin)* increases the expected number of injury accidents by approximately 103% or 29%, respectively, provided the remaining explanatory variables have constant values. As an example, suppose a junction has a value of *AADTmaj* equal to 13000, *ln(AADTmaj)* is then approximately 9.473. The same junction with 10.473 for *ln(AADTmaj)* (corresponding to 35348 *AADTmaj*) is expected to increase the number of injury accidents by approximately 103%.

A junction with a speed limit of 50km/h is *a posteriori* expected to have around 215% more accidents than a junction with a 60km/h speed limit with the same *AADT* values and number of legs (Table 3).

The expected number of accidents for a period of one year can be obtained by solving the equations shown in Table 3, for the several categories of variables *Legs* and *Speed_Limit*.

Table 3 Expected numbers of accidents for Norwegian junctions, for a one year period, obtained by a Poisson regression model, for a baseline/reference of Number_of_Legs='3' and Speed_Limit='60'.

	Expected Numbers of Accidents
Number_of_Legs='3'	
Speed_Limit	
'40'	$\hat{\mu}_i = 3.523\times10^{-5}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'50'	$\hat{\mu}_i = 6.036\times10^{-5}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'60'	$\hat{\mu}_i = 1.919\times10^{-5}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'70'	$\hat{\mu}_i = 9.245\times10^{-5}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'80'	$\hat{\mu}_i = 6.558\times10^{-5}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'90'	$\hat{\mu}_i = 4.548\times10^{-5}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
Number_of_Legs='4'	
Speed_Limit	
'40'	$\hat{\mu}_i = 7.652\times10^{-5}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'50'	$\hat{\mu}_i = 1.311\times10^{-4}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'60'	$\hat{\mu}_i = 4.168\times10^{-5}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'70'	$\hat{\mu}_i = 2.008\times10^{-4}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'80'	$\hat{\mu}_i = 1.424\times10^{-4}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$
'90'	$\hat{\mu}_i = 9.880\times10^{-5}\times AADTmaj_i^{0.705}\times AADT\min_i^{0.251}$

From observation of Table 3 it can be concluded that from the Poisson regression model the expected number of accidents is greater on 4 leg junctions than on 3 leg ones. It can also be seen that, for a constant value of the variable *Number_of_Legs*, the expected number of

accidents increases with increasing speed limit, up to junctions where the speed limits is 70km/h, then there is a slight decrease for junctions with 80 and 90km/h speed limits. This shows that junctions on roads witjh speed limits of 80 and 90km/h may have better geometric design than junctions on roads with lower speed limits.

3.2.1 Model Checking

According to Gelman *et al*. (2004) the model fits the data when replicated data generated under the model looks similar to the observed data.

Figure 10 contains twenty histograms, with the histogram on the left upper corner (in grey) representing the frequency of the number of observed accidents on each of the 732 junctions (the same as the histogram depicted in Figure 3). The remaining nineteen histograms in Figure 10 were obtained from replicated data (y_{rep}) from the posterior predictive distribution (each denoted Figure 10 by *Acc.rep*). The comparison of the nineteen histograms with the histogram of the observed data shows that most replicates can be considered to be representative of the observed number of accidents. The number of replicated data sets in Figure 10 (i.e. nineteen) was chosen according to examples given by Gelman *et al*. (2004).

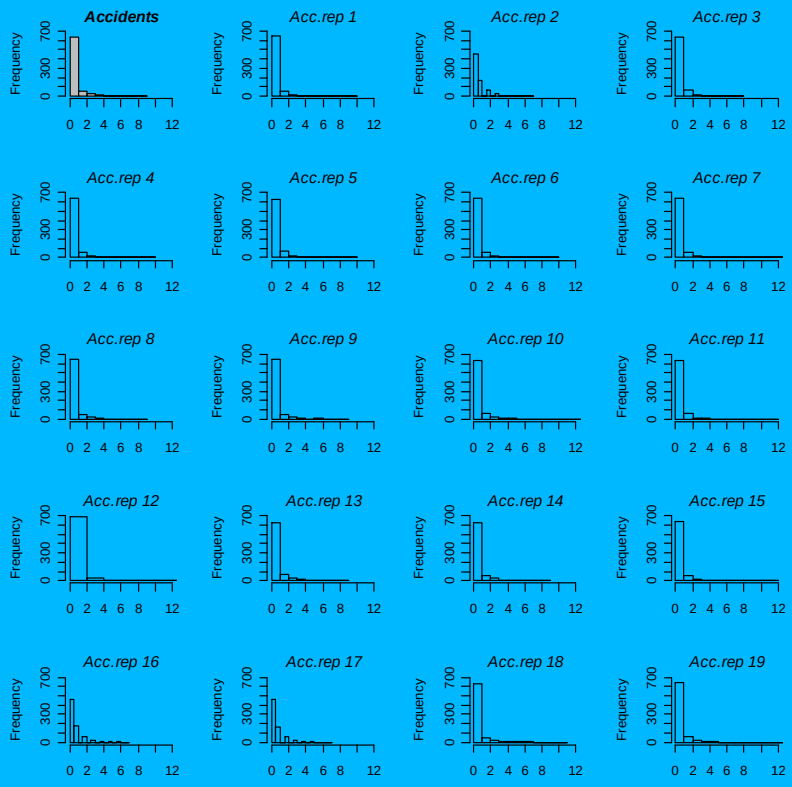

Figure 10 Histogram of the observed number of accidents in Norwegian junctions (left upper corner in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.

Figure 11 shows dot plots, instead of histograms as in Figure 10, of a further batch of 19 replicated data sets formed by the number of accidents per junction. An identical conclusion can be made by observation of the dot plots, as the dot plot of the observed data looks plausible under the posterior predictive distribution. The *x* axis of the dot plots correspond to the junction's indexes and the *y* axis to the number of accidents.

From the interpretation of both figures it can be stated that the data replicated by the model seems to be consistent with the observed data.

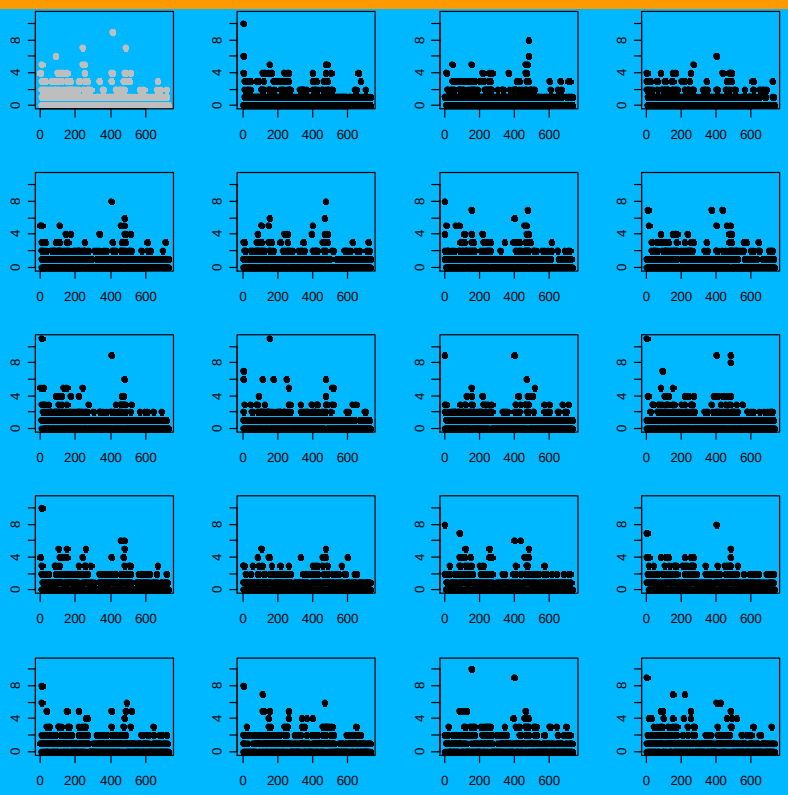

Figure 11 Dot plot of the observed number of accidents in Norwegian junctions (left upper corner in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.

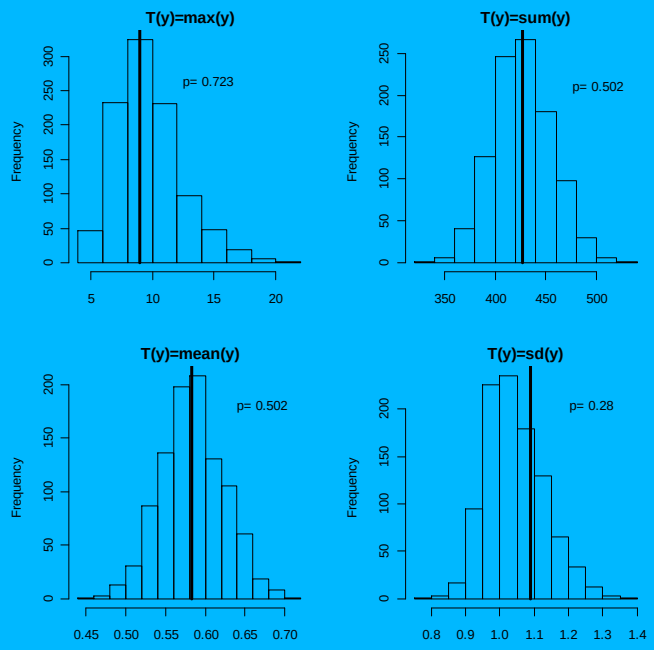

Figure 12 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

Five discrepancy measures were taken into account in order to measure the discrepancy between the model and the data as explained in section 2.4.2 in Chapter 2. The measures referring to the maximum, sum, mean and standard deviation are represented in Figure 12.

By observation of Figure 12 it can be seen that the Poisson regression model captures the variations corresponding to the four measures (estimated probabilities with values lying between 0.1 and 0.9). However, for *T*=max and *T*=sd, the estimated probabilities are far from the ideal 0.5 value (see section 2.4.2).

The effect of the measure of discrepancy suggested by Congdon (2005) to check whether the overdispersion of the data is taken into account by the model (variance over the mean) is displayed in Figure 13.

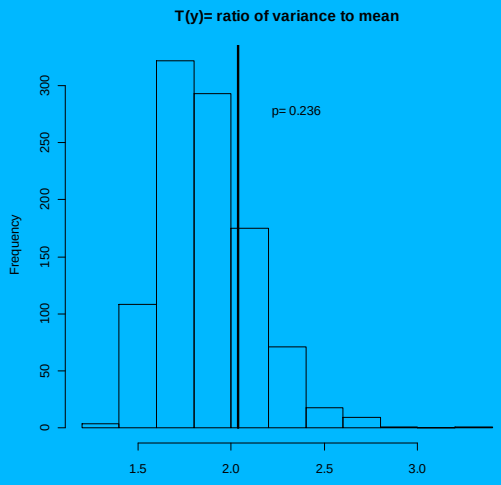

Figure 13 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distributions of the same measure. The *p* is the estimated probability that the measure obtained by the posterior predictive distribution is greater than the one obtained by the observed data.

The estimated probability that the ratio of the variance to the mean in the replicated data is greater than the same ratio calculated from the observed data is equal to 0.236, which is a small value indicating that there is some overdispersion that is not being replicated by the model. However, it still belongs within the suggested limits proposed by Congdon (2005).

This model produces an average deviance, $\overline{D}$, of 1296.270 and an effective model dimension, d_e, of 9.928, giving a DIC (see Equation 2.13 in section 2.4.1 of Chapter 2) of 1306.200. These values are used for model comparison which is discussed in section 3.5.

3.3 Poisson-Gamma hierarchical regression model

The Poisson-Gamma hierarchical regression model was fit to the Norwegian junction data using Equations 2.3 and 2.4 where $\delta = \gamma$ and $\delta \sim Gamma(a,a)$ with a=0.1. The expression given by Equation 3.1 was applied and the β parameters were given *a priori* Normal distributions with mean 0 and precision 0.0001 (variance is equal to the inverse of the precision).

The MCMC algorithm comprised 3 chains and was run for 35000 iterations with 33000 burn-in iterations with a thinning rate of 7. The results thus described were based on a sample with dimension equal to 1002.

From the observations of the Gelman-Rubin plots for the β parameters in Figure 14 it can be concluded that in some parameters the corresponding *Rhat* statistic (the red line) does not seem to converge to 1, which raises some doubts as to the convergence of the corresponding distributions.

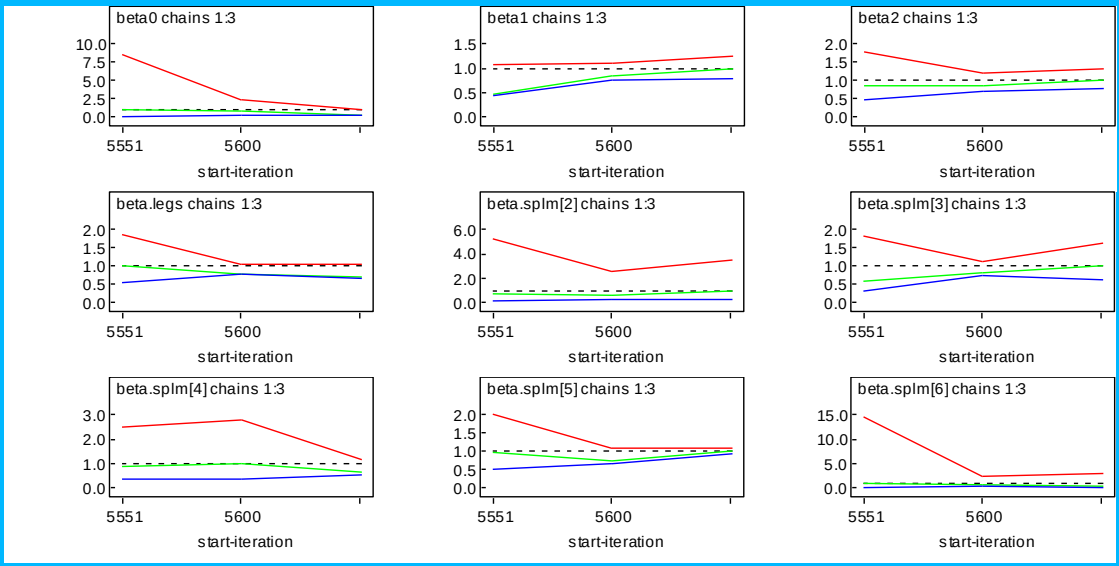

Figure 14 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson-Gamma regression model fitted to the Norwegian junction accident data.

The point estimates for the model obtained are given in Table 4. The impact of both variables *ln(AADTmaj)* and *ln(AADTmin)* remains significant with 95% credible intervals of (0.651, 0.796) and (0.203, 0.395), respectively.

Table 4 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fitted to the Norwegian accident data using 3 leg junctions with a 60km/h speed limit as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-10.130	0.283	3.903E-02	-10.800	-9.771
β_1 (ln(AADTmaj))	0.727	0.039	4.942E-03	0.651	0.796
β_2 (ln(AADTmin))	0.296	0.051	6.458E-03	0.203	0.395
β_3 (Legs='4')	0.844	0.131	1.723E-02	0.545	1.085
β_4 (Speed_Limit='40')	-1.317	0.455	6.231E-02	-2.326	-0.673
β_4 (Speed_Limit='50')	-0.618	0.168	2.255E-02	-0.974	-0.368
β_4 (Speed_Limit='70')	0.381	0.150	2.012E-02	0.114	0.690
β_4 (Speed_Limit='80')	-5.913x10-4	0.097	1.229E-02	-0.180	0.193
β_4 (Speed_Limit='90')	-0.075	0.510	7.064E-02	-1.089.	0.727

From examination of the means in Table 4 (column *mean*) it can be stated that, according to the model, every unitary increase in *ln(AADTmaj)* increases the expected number of accidents by 107%, (when the other variables remain constant). On the other hand, an increase in *ln*(*AADTmin*) increases the expected number of accidents by 34%. See example

on section 3.2.

A 4-leg junction is *a posteriori* expected to have approximately 133% more accidents than a 3-leg junction with the same *ln*(*AADTmaj*), *ln*(*AADTmin*) and speed limit.

A junction with a 50km/h speed limit is *a posteriori* expected to have approximately 46% less accidents than a junction with a 60km/h speed limit and the same values of *ln*(*AADTmaj*), *ln*(*AADTmin*) and number of legs. However, a 70km/h speed limit junction is *a posteriori* expected to have 46% more accidents than a 60km/h speed limit junction with the same values of *AADT* and number of legs.

Junctions with 80km/h and 90km/h are *a posteriori* expected to have approximately less 0.06% and 7%, respectively, expected number of injury accidents than 60km/h junctions (all the other variables remaining constant). This shows that, according to this particular Poisson-Gamma model, junctions with 60km/h and 80km/h have approximately the same number of expected injury accidents.

Figure 15 shows the posterior densities for the β coefficient estimates obtained by the Poisson-Gamma model. The posterior densities of the coefficient estimates seemed to have drifted away from the prior distributions considered, i.e. Normal distributions with mean equal to zero and variance 10000.

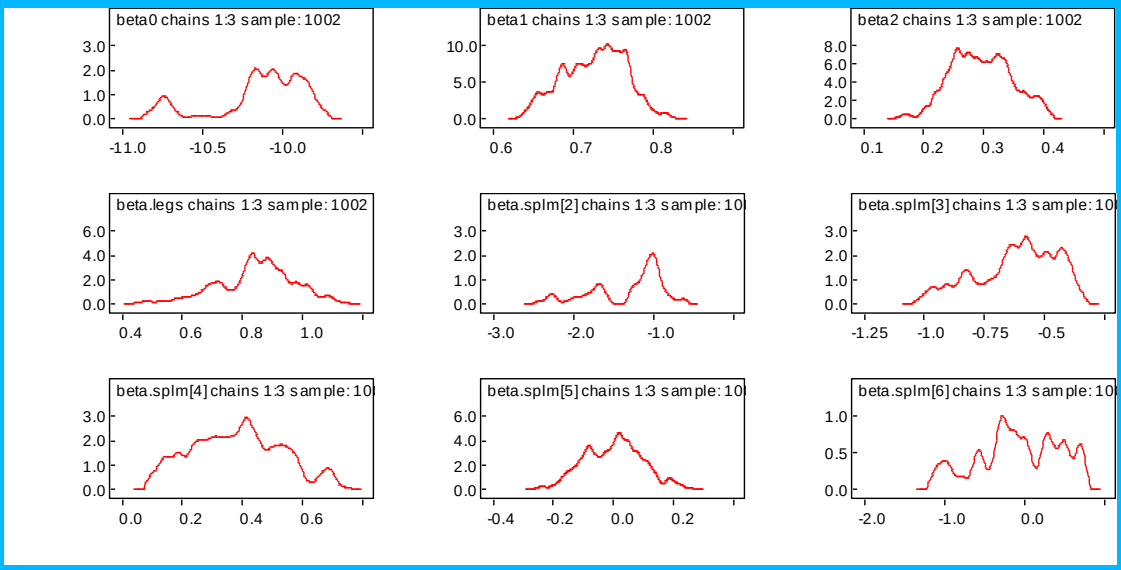

Figure 15 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson-Gamma regression model was fitted to the Norwegian data set.

The expected number of accidents, for a one year period, can be obtained by the equations displayed in Table 5 for junctions with 3 and 4 legs and different speed limits.

Table 5 Expected number of accidents per year for Norwegian junctions, obtained by a Poisson-Gamma regression model using 3 leg junctions with a 60km/h speed limit as baseline.

	Expected Numbers of Accidents
Number_of_Legs='3'	
Speed_Limit	
'40'	$\hat{\nu}_i = 1.067 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'50'	$\hat{\nu}_i = 2.146 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'60'	$\hat{\nu}_i = 3.682 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'70'	$\hat{\nu}_i = 5.827 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'80'	$\hat{\nu}_i = 3.980 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'90'	$\hat{\nu}_i = 3.695 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
Number_of_Legs='4'	
Speed_Limit	
'40'	$\hat{\nu}_i = 2.480 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'50'	$\hat{\nu}_i = 4.990 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'60'	$\hat{\nu}_i = 9.260 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'70'	$\hat{\nu}_i = 1.355 \times 10^{-4} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'80'	$\hat{\nu}_i = 9.254 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$
'90'	$\hat{\nu}_i = 8.593 \times 10^{-5} \times AADTmaj_i^{0.727} \times AADT\,\min_i^{0.296}$

From observation of Table 5 it can be stated that, in general, 3-leg junctions have less expected numbers of accidents than 4-leg ones, when taking into account the same values of *AADTmaj* and *AADTmin* for 3 and 4-leg junctions. The expected number of accidents is lower on junctions with a 50km/h speed limit. As the speed limit increases also increases the expected number of accidents reaching the higher values at junctions with 70km/h speed limit.

Within each type of *Number_of_Legs* the highest speed limits (i.e. 80 and 90km/h) have lower expected number of injury accidents than the 70km/h junctions with junctions with 60km/h having identical expected number of accidents to the 80 and 90km/h speed limit junctions.

Speed limits on Norwegian intersections are established on approaching roads' geometry and traffic characteristics.

3.3.1 Model Checking

Figure 16 contains the histograms of the replicated data together with the histogram of the observed data. The histograms show that the observed data looks plausible under the posterior predictive distribution represented by the histograms of the replicated data.

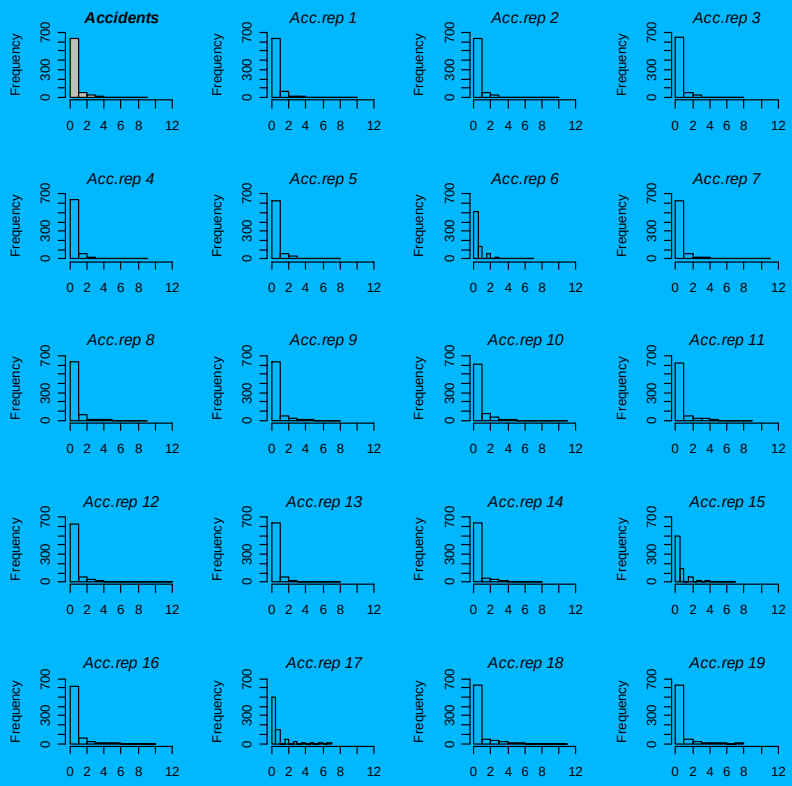

Figure 16 Histogram of the observed number of accidents in Norwegian junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.

Figure 17 shows dot plots instead of histograms for a further set of 19 examples. An identical conclusion is obtained after examination of this figure as that of Figure 16.

The results obtained by the replicated data for the discrepancy measures considered, together with the observed values are plotted on the four graphs displayed in Figure 18. Since all the observed values fall inside the histograms (of replicated data) and the estimated *p*-values are near 0.5 it can be considered that the Poisson-Gamma model adequately captures the variations indicated by the observed data.

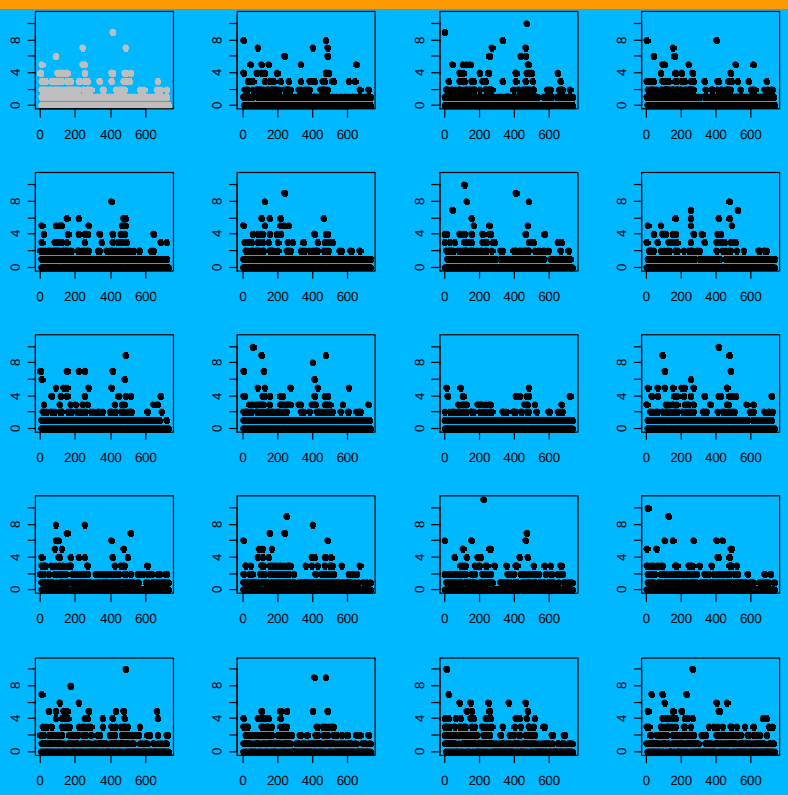

Figure 17 Dot plot of the observed number of accidents in Norwegian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.

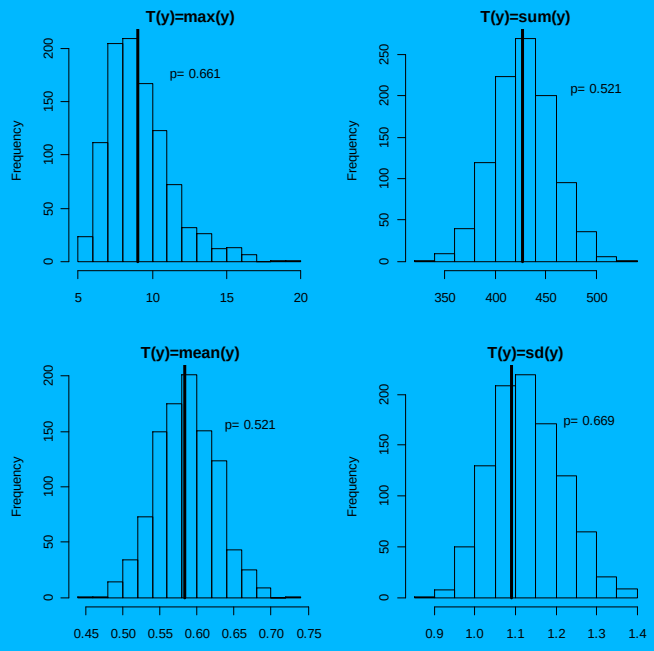

Figure 18 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The effect of the measure of discrepancy given by the ratio of variance over the mean, first suggested by Congdon (2005) to check whether the overdispersion of the data was being taken into account by the model is displayed in Figure 19.

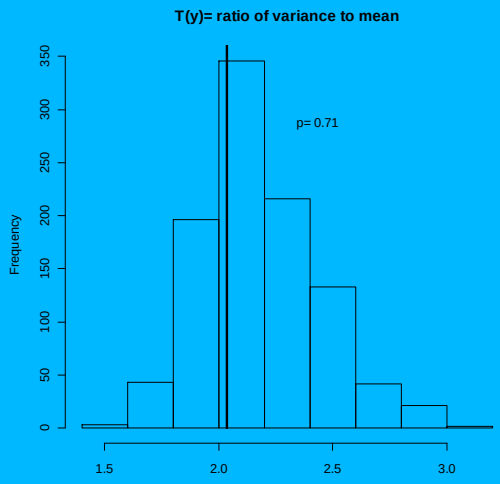

Figure 19 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

The value of 0.710 (reasonably close to 0.5) indicates that the replicated data obtained by the Poisson-Gamma model has a degree of overdispersion similar to the one of the observed data. Therefore, the model is allowing for the overdispersion.

This model produces an average deviance, $\overline{D}$ equal to 1141.200, an effective model dimension of 97.521 and a DIC equal to 1238.720. The dispersion parameter ($1/\delta$) was estimated as 0.384. The comparison of these results for all the fitted models is described in section 3.5.

The expected number of accidents for a one year period for the several types of the categorical variables *Number_of_Legs* and *Speed_Limit* for the minimum, maximum, mean and median profiles of *ln(AADTmaj)* and *ln(AADTmin)* were calculated following the example given by Ntzoufras (2009). In the minimum and maximum profiles, the maximum and minimum values of *ln(AADTmaj)* and *ln(AADTmin)* were considered, respectively, since these variables are positively associated with the number of injury accidents (note the positive parameter estimates β_1 and β_2 in Table 4.

Posterior means and corresponding standard deviations of the obtained expected number of injury accidents are provided in Table 6. The first two rows of Table 6 show the minimum, mean, median and maximum values of variables *ln(AADTmaj)* and *ln(AADTmin)* obtained from the data. The remaining values consist of the posterior means obtained from each profile.

Table 6 Posterior means and corresponding (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the Norwegian accident data.

		Minimum	Mean	Median	Maximum
	ln(AADTmaj)	4.890	7.766	7.726	10.383
	ln(AADTmin)	1.946	5.702	5.617	9.141
Speed_Limit	Number_of_Legs	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
40	3	7.43E-4 (3.596E-04)	0.018 (0.007)	0.017 (0.007)	0.335 (0.136)
	4	0.002 (7.408E-04)	0.041 (0.016)	0.039 (0.015)	0.771 (0.312.)
50	3	0.001 (3.743E-04)	0.033 (0.006)	0.031 (0.006)	0.615 (0.118)
	4	0.003 (7.378E-04)	0.077 (0.012)	0.073 (0.011)	1.423 (0.232)
60	3	0.003 (4.704E-04)	0.061 (0.005)	0.058 (0.005)	1.130 (0.140)
	4	0.006 (0.001)	0.143 (0.022)	0.135 (0.021)	2.645 (0.445)
70	3	0.004 (6.958E-04)	0.090 (0.012)	0.085 (0.011)	1.667 (0.299)
	4	0.009 (0.002)	0.210 (0.039)	0.199 (0.037)	3.910 (0.828)
80	3	0.003 (3.977E-04)	0.061 (0.005)	0.058 (0.005)	1.134 (0.179)
	4	0.006 (0.001)	0.143 (0.022)	0.135 (0.021)	2.656 (0.499)
90	3	0.002 (9.601E-04)	0.063 (0.028)	0.059 (0.027)	1.215 (0.666)
	4	0.006 (0.003)	0.148 (0.070)	0.140 (0.066)	2.861 (1.556)

For a typical (see the values on the *Mean* column in Table 6) three leg junction with a speed limit of 40km/h, one expects 0.018 accidents, while for a four leg junction with the same speed limit 0.041 accidents are expected, both for a one year period.

Note that the worst case scenario (maximum profile) where junctions have 10.383 for log major traffic volume and 9.141 for log minor traffic volume corresponds to an expected number of accidents of 3.910 at a four leg junction with a 70km/h speed limit and 2.861 for a four leg junction with 90km/h speed limit.

The highest expected number of accidents for each of the four scenarios considered

corresponds to four leg junctions with a speed limit of 70km/h.

3.4 Poisson Log-Normal Regression Model

The Poisson Log-Normal regression model was fitted to the data according to Equations 2.5 and 2.6 where parameter α in Equation 2.6 was assigned an *a priori* Gamma(*a*,*a*) distribution with *a*=0.001.

The MCMC algorithm was run with three chains for 35000 iterations with 33000 of those as burn-in with a thinning rate equal to 6. The results and conclusions were drawn from a sample with dimension 1002.

The regression equation is given by Equation 3.1, where the β coefficient parameters were assigned non-informative *a priori* Normal distributions with mean zero and variance 10000.

From the observation of the graphs depicted in Figure 20 concerning the Gelman-Rubin statistics, it is possible to notice that the *Rhat* statistic (red line) tends to 1 as the number of iterations increase and that both *W* and *Vhat* remain relatively constant. Consequently, there are reasons to believe that the iterative simulation converges.

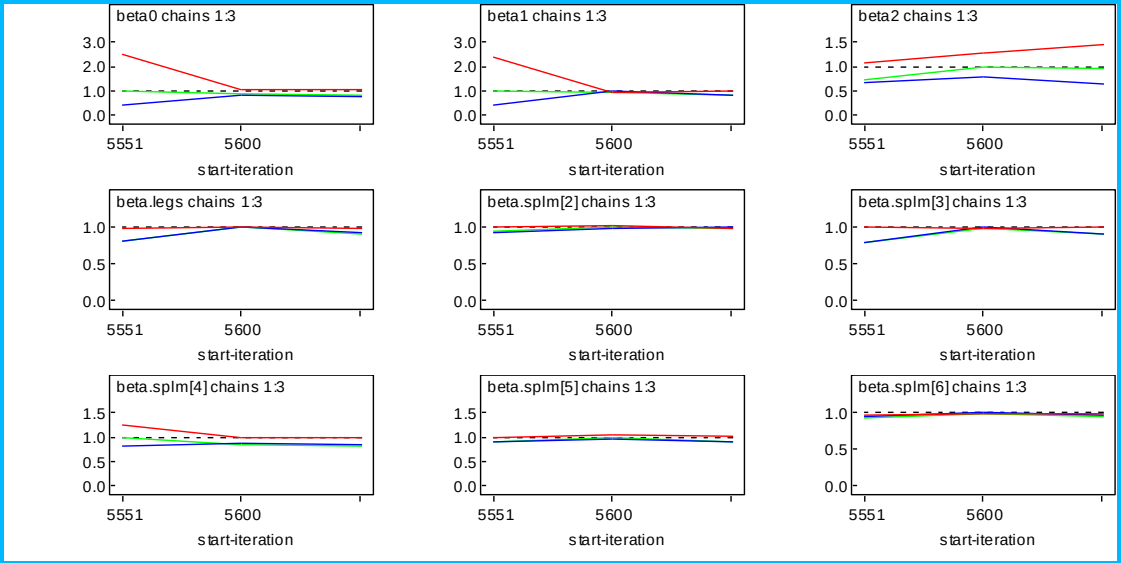

Figure 20 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson Log-Normal regression model fit to the Norwegian accident data.

The posterior means of the parameter estimates, corresponding standard deviations, Monte Carlo (MC) errors and 95% credible intervals are displayed in Table 7.

The MC errors are considerably smaller than the corresponding standard deviations, in particular for the parameter coefficients of the logarithms of the *AADT* volumes indicating that the estimates were calculated with precision (see Ntzoufras, 2009).

Table 7 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the Norwegian accident data using a three leg junction with a 60km/h speed limit as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-10.480	0.678	8.588E-02	-12.010	-9.303
β_1 (ln(AADTmaj))	0.737	0.086	1.068E-02	0.582	0.946
β_2 (ln(AADTmin))	0.301	0.060	6.255E-03	0.182	0.414
β_3 (Legs='4')	0.867	0.158	5.260E-03	0.561	1.168
β_4 (Speed_Limit='40')	-1.563	0.911	2.879E-02	-3.563	-0.154
β_4 (Speed_Limit='50')	-0.575	0.181	5.881E-03	-0.939	-0.239
β_4 (Speed_Limit='70')	0.445	0.193	9.703E-03	0.078	0.813
β_4 (Speed_Limit='80')	0.094	0.150	7.312E-03	-0.186	0.411
β_4 (Speed_Limit='90')	-0.200	0.507	1.963E-02	-1.310	0.795

From examination of the mean posterior estimates values in Table 7 it can be stated that every unit increase in *ln(AADTmaj)* increases the expected number of accidents by approximately 109% (and when the other variables remain constant). An increase in *ln(AADTmin)* increases the expected number of accidents by approximately 35%. See example on section 3.2.

A 4-leg junction is *a posteriori* expected to have approximately 140% more accidents than a 3-leg junction with the same speed limit and values for the two *ln(AADT)*.

A 40km/h junction is *a posteriori* expected to have approximately 79% less accidents than a 60km/h junction (and constant values for the remaining variables).

A junction with 50km/h speed limit is *a posteriori* expected to have 44% less accidents than a junction with the same values of *ln(AADTmaj)*, *ln(AADTmin)* and number of legs but a 60 km/h speed limit (see Table 7). Junctions with 70km/h and 80km/h are expected to have approximately 56% and 10% more injury accidents than a 60km/h junction. Furthermore, a 90km/h junction is expected to have 18% less accidents than a 60km/h junction.

The highest expected number of accidents is found to be on junctions with a 70km/h speed limit, regardless of the number of legs.

The posterior densities of parameter estimates (for which the posterior mean is given in Table 7) are displayed in Figure 21. The mean of the posterior densities has shifted considerably away from zero (the *a priori* mean).

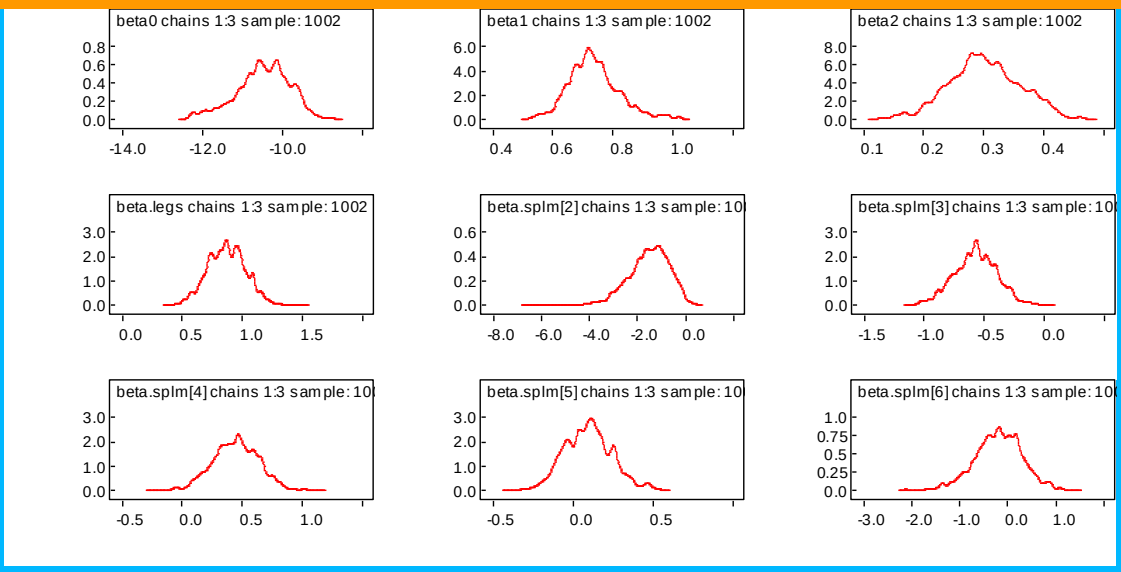

Figure 21 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson Log-Normal regression model was fit to the Norwegian data.

The equations for the expected number of accidents per number of legs and speed limit of the junctions are given in Table 8. From observation of the values in the equations it is possible to deduce that, in general, 3-leg junctions have fewer expected numbers of accidents when compared to 4-leg junctions.

Table 8 Expected number of accidents per year for Norwegian junctions obtained by a Poisson Log-Normal regression model using a three leg junction with a 60km/h speed limit as baseline.

	Expected Numbers of Accidents
Number_of_Legs='3'	
Speed_Limit	
'40'	$\hat{\mu}_i = 5.881\times10^{-6}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'50'	$\hat{\mu}_i = 1.579\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'60'	$\hat{\mu}_i = 2.807\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'70'	$\hat{\mu}_i = 4.381\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'80'	$\hat{\mu}_i = 3.083\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'90'	$\hat{\mu}_i = 2.299\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
Number_of_Legs='4'	
Speed_Limit	
'40'	$\hat{\mu}_i = 1.400\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'50'	$\hat{\mu}_i = 3.758\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'60'	$\hat{\mu}_i = 6.681\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'70'	$\hat{\mu}_i = 1.043\times10^{-4}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'80'	$\hat{\mu}_i = 7.339\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$
'90'	$\hat{\mu}_i = 5.472\times10^{-5}\times AADTmaj_i^{0.737}\times AADT\min_i^{0.301}$

The expected number of accidents increases as the speed limit increases, reaching the highest value on junctions where the speed limit is 70km/h, then it decreases again for junctions with 80km/h and 90km/h. One possible explanation is that 80 and 90km/h limit intersections have more effective traffic channelization than lower speed intersection roads.

3.4.1 Model Checking

Figure 22 contains histograms of modelled data as well as the histogram of the observed data. Figure 23 displays dot plots of the observed number of accidents together with dot plots of replicated accident numbers under the Poisson log-Normal regression model.

From both figures it can be stated that the modelled data appear similar to the observed data suggesting that the Poisson-Log-Normal model performs adequately.

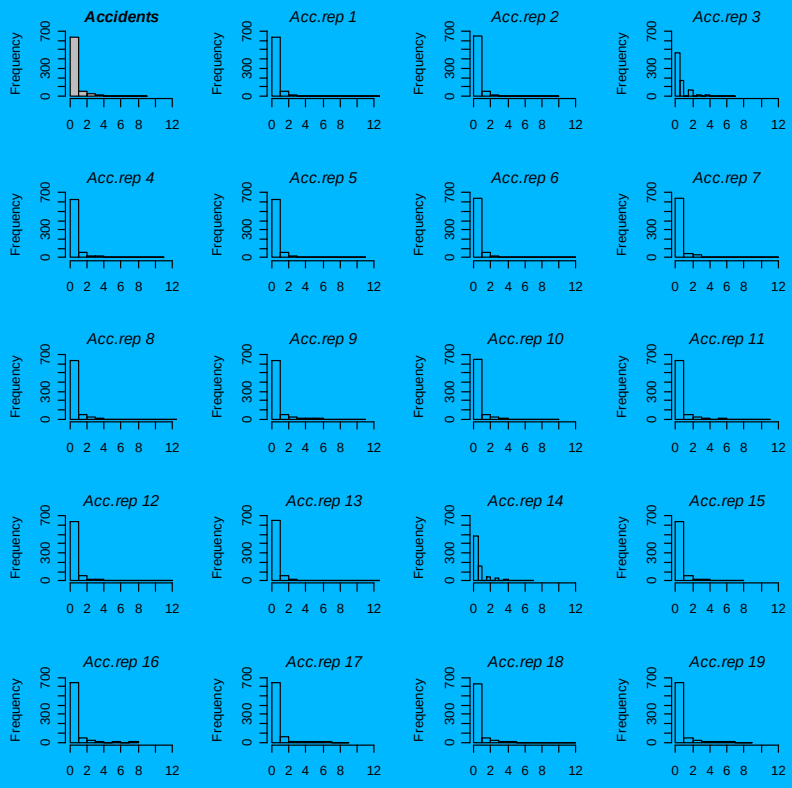

Figure 22 Histogram of the observed number of accidents in Norwegian junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

Figure 23 Dot plot of the observed number of accidents in Norwegian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

The four discrepancy measures concerning the maximum, sum, mean and standard deviation (sd) values are displayed in Figure 24. The probabilities (*p* in the graphs) that the discrepancy measures obtained by the replicated data are greater than the corresponding discrepancy measures from the observed data lie within the acceptable boundaries of 0.1 to 0.9 (Congdon, 2005). The *p* values for the *sum*, *mean* and *sd* measures are also relatively close to the recommended 0.5 value.

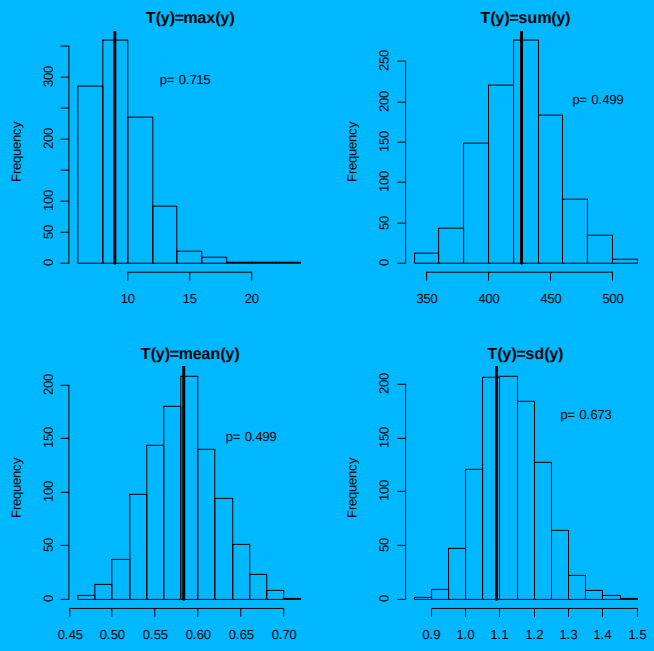

Figure 24 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1287 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fit to the Norwegian data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The values of the discrepancy measured suggested by Congdon (2005) to check whether the overdispersion was taken into account by the model are displayed in Figure 25. The probability that the ratio of the variance over the mean in the replicated data is greater than the corresponding ratio under the observed data is 0.733, which is a satisfactory value, although slightly higher than the ideal 0.5. Nevertheless, it can be concluded that the model is taking into account the overdispersion present in the observed data.

This model produces an average deviance equal to 1149.120, an effective model dimension of 106.579 and a DIC value of 1255.700. The dispersion parameter ($1/\delta$) was estimated as 0.363. The model's comparison will be discussed in section 3.5 of the present chapter.

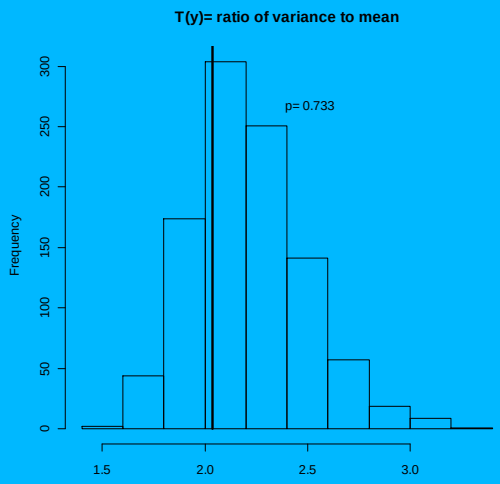

Figure 25 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fit to the Norwegian data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

The posterior means of the expected number of accidents in Norwegian junctions, for a one year period, for four possible scenarios or profiles are given in Table 9, together with their corresponding standard values.

A typical junction with 3 legs and 40km/h speed limit is expected to have 0.014 injury accidents in a one year period. Furthermore, a typical junction with 4 legs and 70km/h speed limit is expected to have 0.172 accidents in the same period of time. These junctions have the higher expected number of accidents followed by junctions with 4 legs with 80km/h and 60km/h speed limit, with 0.120 and 0.109 expected accidents in a one year period.

The worst case scenario (for maximum values of *ln(AADTmaj)* and *ln(AADTmin)* have higher expected number of accidents for 4 leg junctions with 70, 80 and 60km/h (with expected values equal 3.536, 2.491 and 2.253).

Table 9 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the Norwegian accident data.

		Minimum	Mean	Median	Maximum
	ln(AADTmaj)	4.890	7.766	7.726	10.383
	ln(AADTmin)	1.946	5.702	5.617	9.141
Speed_Limit	Number_of_Legs	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
40	3	5.639E-04 (4.943E-04)	0.014 (0.011)	0.013 (0.010)	0.273 (0.218)
	4	0.001 (0.001)	0.033 (0.026)	0.031 (0.025)	0.646 (0.512)
50	3	0.001 (4.196E-04)	0.027 (0.005)	0.026 (0.005)	0.531 (0.117)
	4	0.003 (0.001)	0.065 (0.014)	0.062 (0.013)	1.269 (0.298)
60	3	0.002 (6.605E-4)	0.048 (0.006)	0.046 (0.006)	0.943 (0.198)
	4	0.005 (0.002)	0.116 (0.021)	0.109 (0.020)	2.253 (0.514)
70	3	0.003 (0.001)	0.076 (0.013)	0.072 (0.013)	1.478 (0.338)
	4	0.007 (0.003)	0.182 (0.041)	0.172 (0.039)	3.536 (0.895)
80	3	0.002 (6.663E-04)	0.053 (0.006)	0.050 (0.006)	1.040 (0.244)
	4	0.005 (0.002)	0.127 (0.024)	0.120 (0.022)	2.491 (0.644)
90	3	0.002 (9.634E-04)	0.044 (0.022)	0.042 (0.020)	0.878 (0.512)
	4	0.004 (0.002)	0.105 (0.052)	0.099 (0.049)	2.081 (1.205)

3.5 Discussion

Table 10 shows the resulting fit ($\overline{D}$), complexity (d_e) and overall model choice (DIC) score for the three models considered. In terms of overall model choice the Poisson Log-Normal emerges as the best with the Poisson model performing the worst. The results of d_e suggest that the use of the additive random error in the Poisson Log-Normal model (see section 2.2.2) contributes an extra nine effective parameters (Poisson-Gamma *versus* Poisson Log-Normal model). Taking into account both the DIC and the d_e values the model chosen to best describe the Norwegian junction data is the Poisson-Gamma, even though the results

obtained by these two models do not differ substantially.

Table 10 Comparison of DIC and related statistics for the three models fitted to the Norwegian junction data.

Regression Model	$\bar{D}$	d_e	DIC
Poisson	451.08	57.93	509.00
Poisson-Gamma	1141.200	97.521	1238.720
Poisson Log-Normal	1149.120	106.579	1255.700

From the modelled data it can be concluded that junctions with 4 legs have a higher expected number of accidents than 3-legged ones, as can be seen by the posterior means of the expected values of the number of injury accidents depicted as straight lines in Figure 26 for every type of junction. This finding is not unusual because four legged junctions have more conflict points than three legged ones.

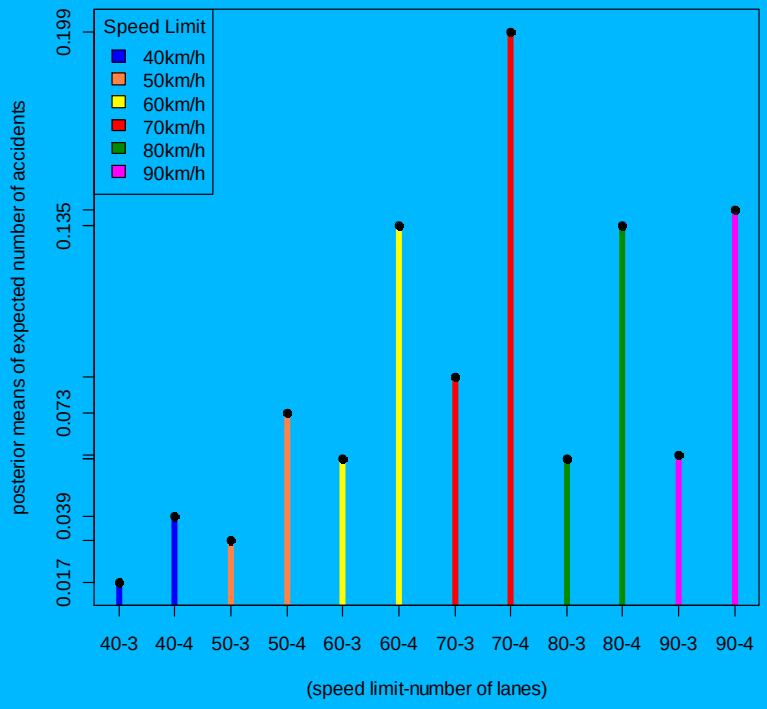

Figure 26 Values of the posterior means of the expected number of accidents for Norwegian junctions classified per speed limit and number of legs (as under column *Mean* in Table 9).

4 Modelling Austrian injury accidents

The present chapter starts by giving a brief description of the data measured at rural road junctions in Austria (section 4.1) and proceeds with the model fit, assessment and checking of three regression models including the Poisson (section 4.2), the hierarchical Poisson-Gamma (section 4.3) and the hierarchical Poisson Log-Normal models (section 4.4) where the response variable was the number of injury accidents occurring within 50 metres from the junctions. The final section (4.5) summarises the results obtained.

4.1 Austrian Junction Data

The data described and analysed in this chapter have been provided by Dr Robert Bauer of the KFV (Austrian Safety Board) and consists on several measurements registered on 213 junctions of the Austrian national road network located in the province of Lower Austria. The data was collected on the four year period ranging from 2007 to 2010. The sample of 213 junctions is based on a list of police recorded accidents on rural junctions ranging from 2003 to 2010. This list was supplemented with information about the respective junction from aerial photos and the respective traffic volumes (values for the most recent year). In order to have also junctions with zero accidents in the sample the range was reduced to years 2007 to 2010. Due to this procedure junctions with zero accidents are most probably under-represented in this sample.

The variables registered included, per junction:

- ***Intersection_code***: gives the junction reference, corresponding to the codes of the road segments that form the junction;
- ***Junction_Type***: a categorical variable referring to the type of the junction, it takes the values 'roundabout', 'X' or 'Y';
- ***Traffic_Control***: a categorical variable indicating the type of traffic control in the junction, it takes the values 'signalised', 'stop' or 'yield';
- ***Accidents***: gives the number of injury accidents;
- ***Serious***: represents the number of accidents which resulted in serious injured victims;
- ***Killed***: gives the number of fatalities;
- ***AADTmaj***: represents the traffic volume entering the major road legs (annual average daily traffic);
- ***AADTmin***: represents the traffic volume entering the minor road legs (annual average daily traffic).

The data can be best described and analysed after several graphical plots have been performed and examined. Figure 27 shows three plots corresponding to the number of accidents (*Accidents*), number of fatalities (*Killed*) and number of serious injury accidents (*Serious*) per junction. From observation of the upper left panel in Figure 27 (regarding variable *Accidents*) it can be easily seen that there are two junctions where fourteen and thirteen injury accidents have occurred (between 2007 to 2010).

The five fatalities occurred in separate five junctions, as can be seen on the upper right panel in Figure 27.

There were two junctions with three and six accidents each that have caused seriously injured victims (lower panel in Figure 27).

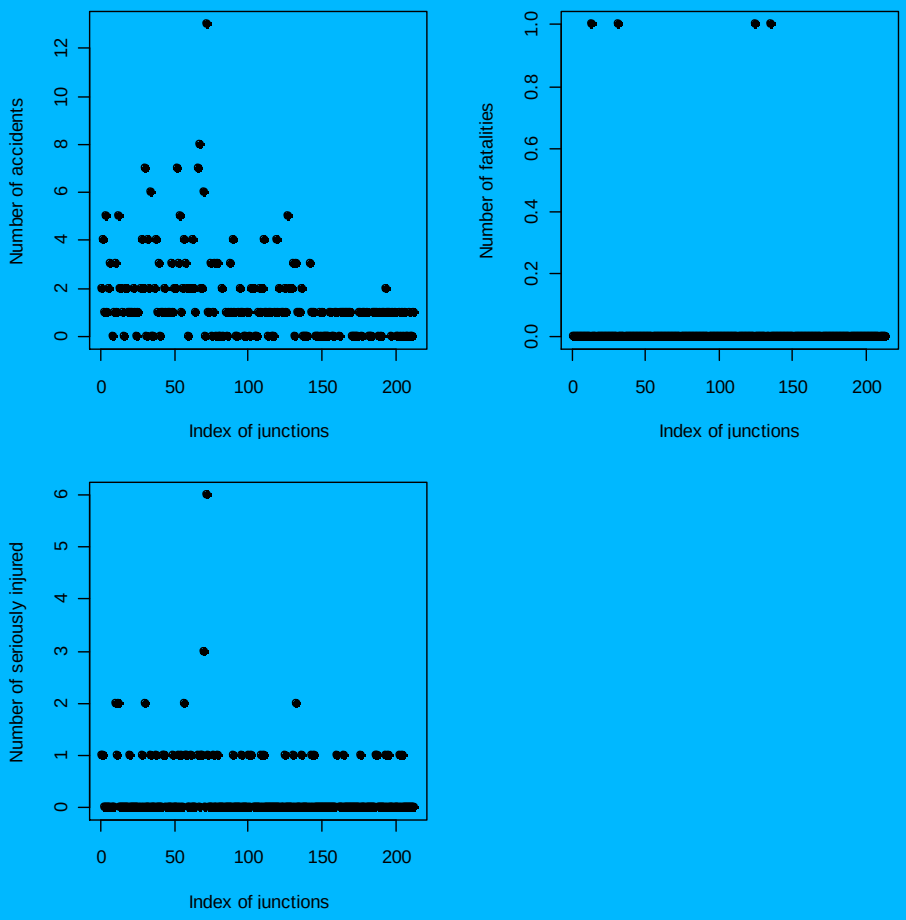

Figure 27 Plots of *Accidents*, *Killed*, and *Serious*, per junction, from upper left to right, respectively, registered from 2007 to 2010 in the Austrian rural road network junctions.

The bar plots shown in Figure 28 consist of sequences of rectangular bars with heights given by the values in each of the three variables (*Accidents*, *Killed* and *Serious*), and therefore corresponding to four years of data collection.

By observation of the upper left panel in Figure 28 (i.e. the bar plot obtained by variable *Accidents*) it can be observed that there are more junctions where one accident has occurred than junctions without accidents (the bar corresponding to zero).

There is the presence of a junction where six accidents (occurred during the four year period) have provoked seriously injured victims (see the lower panel in Figure 28, corresponding to variable *Serious*, and also the lower panel in Figure 27).

In more than half of the overall junctions no fatal injuries have occurred as well as accidents resulting in serious injured victims (this can be seen by the bars corresponding to the value zero for variables *Killed* and *Serious* which have the highest heights).

The maximum number of fatalities per junction was one (see right upper panel in Figure 28).

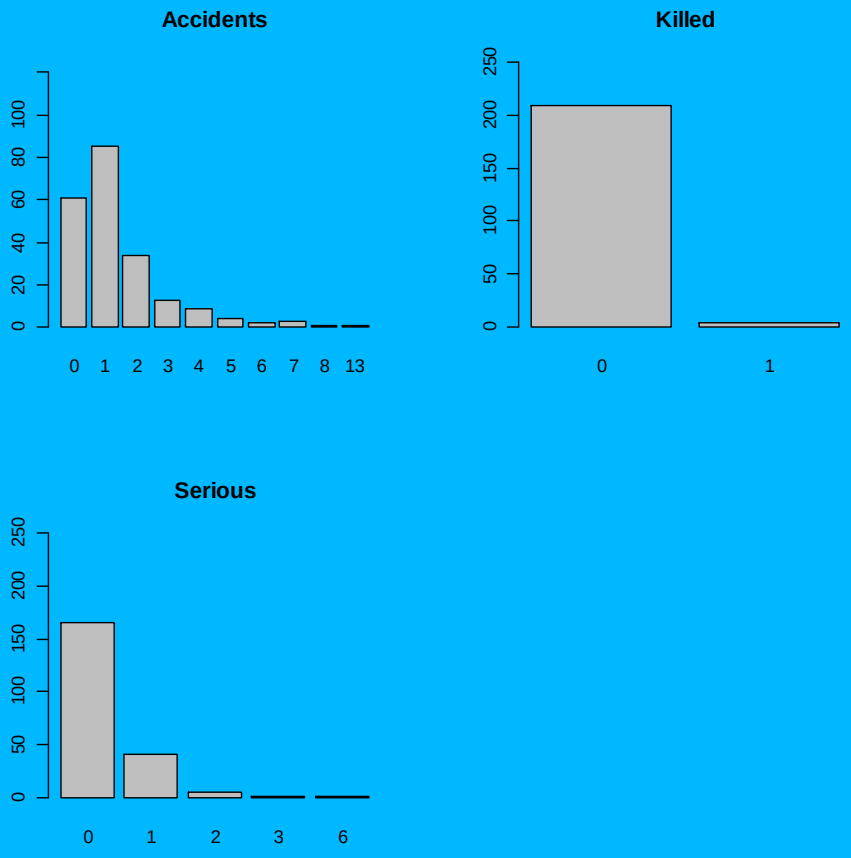

Figure 28 Bar plots giving the frequencies of the total number of accidents, fatalities and seriously injured victims registered from 2007 to 2010 on Austrian junctions from the rural road network.

From the observation of the plots depicting the number of accidents per *AADT* values (Figure 29 with corresponding descriptive statistics shown in Table 11), it can be seen that most accidents occur for values of *AADTmaj* between 567 and 10000 (left plot) and for 560 to around 4000 for *AADTmin* (right plot).

The fitted smooth regression curve employing a regression function developed by Venables and Ripley (2002) could not be applied with the *AADTmin* values (see right panel in Figure 29) due to the high number of equal values.

The junction where 13 accidents were registered in the four year period can be observed on both panels in Figure 29. Due to lack of further information it was decided to maintain that junction in the data set.

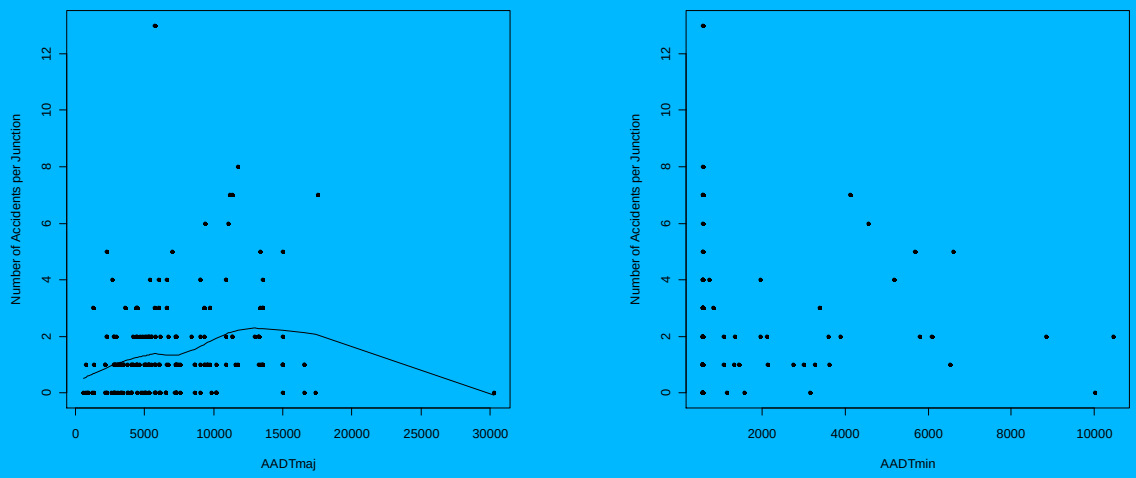

Figure 29 The number of accidents per junction in the Austrian data set against *AADTmaj* and *AADTmin*, and corresponding polynomial fit for the *AADTmaj*, on the left and right panels, respectively.

Figure 30 shows that the majority of the Austrian junctions were of type *Y* and *X* with only thirty nine junctions of type *roundabout* (see left panel in Figure 30). More than 120 junctions had a traffic control of type *yield* with only seven with type *signalised* (see panel on the right).

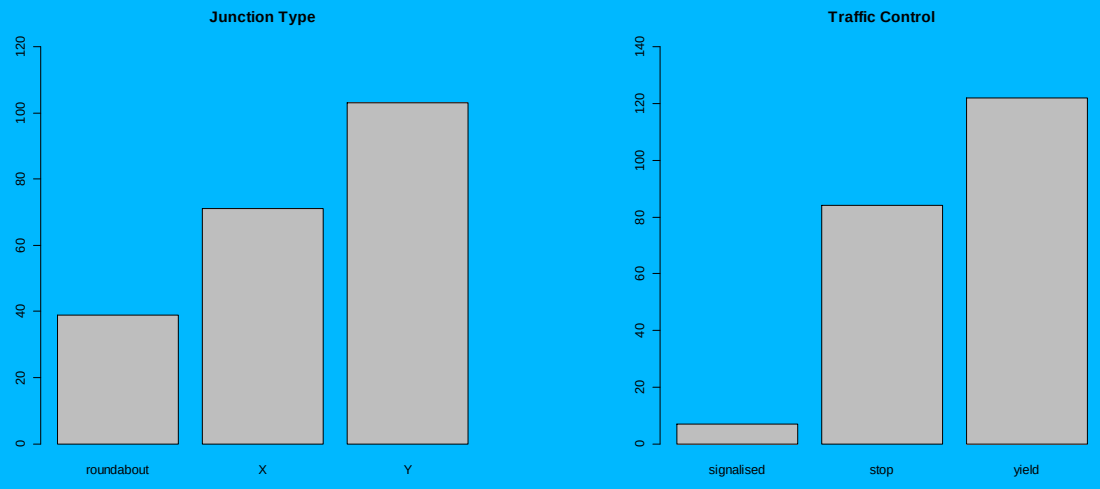

Figure 30 Bar plots giving the frequencies of the number of junctions per categories of variables *Junction_Type* and *Traffic_Control*, registered from 2007 to 2010 on Austrian junctions from the rural road network.

There were 84 junctions with type *stop* as traffic control.

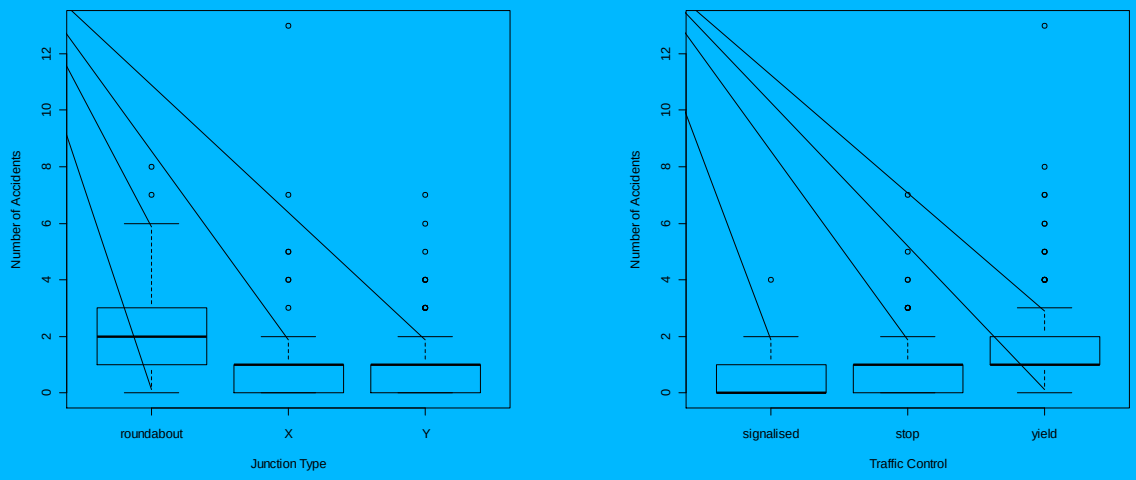

Figure 31 Box plots of the total number of accidents registered in 2007 -2010 at Austrian junctions by group for *Junction_Type* and *Traffic_Control*, on the left and right panels, respectively.

The panels shown in Figure 31 illustrate the distributions of the number of accidents according to the levels of variables *Junction_Type* (left panel) and *Traffic_Control* (right panel). On the panel on the left it can be observed that only the median of junction type *roundabout* differs from the median of the other types.

Table 11 Summary statistics for the variables registered on Austrian junctions from 2007 to 2010.

Variables	minimum	mean	standard deviation	median	maximum
AADTmaj	567	6863.038	4210.441	5759	30278
AADTmin	560	1035.061	1518.719	560	10456
Accidents	0	1.418	1.690	1	13
Serious	0	0.282	0.648	0	6
Fatalities	0	0.019	0.136	0	1

The values in Table 11 concern the five summary statistics of the non-categorical variables measured on the Austrian junctions from 2007 to 2010. The mean number of injury accidents was equal to 1.418.

4.2 The Poisson Regression Model

The Poisson regression model given by Equations 2.1 and 2.2 was fitted to the data with the number of injury accidents (*Accidents*) as the dependent variable and the natural logarithms of both *AADTmaj* and *AADTmin* as well as the categorical variables *Junction_Type* and *Traffic_Control* as explanatory, as shown on Equation 4.1 (corresponding to Equation 2.2).

$$ln(\hat{\mu}_i) = \beta_0 + \beta_1 ln(AADTmaj_i) + \beta_2 ln(AADTmin_i) + \beta_3 Traffic\_Control_i + \beta_4 Junction\_Type_i \quad (4.1)$$

Where $\hat{\mu}_i$ is the expected number of injury accidents for a period of one year. The β

parameters were assigned Normal *a priori* distributions with mean equal to zero and variance 10000. The baseline, or reference, categories for the categorical variables are:

Traffic_Control = '*yield'*;

Junction_Type = *'Y'*.

The MCMC algorithm comprised three chains and was run for 35000 iterations with 33000 as burn-in with a thinning rate equal to 6. The results were drawn from final samples with dimension 1002.

From the observations of the graphs obtained from the Gelman-Rubin statistics for the estimates of all the *β* parameters, and depicted in Figure 32, it can be seen that the *Rhat* statistic (represented by the red line in the graphs) converges to 1 and that both *W* (blue line) and *Vhat* (green line) stabilise as the number of iterations increase. This indicates that there are no reasons to suspect non-convergence of the iterative simulations.

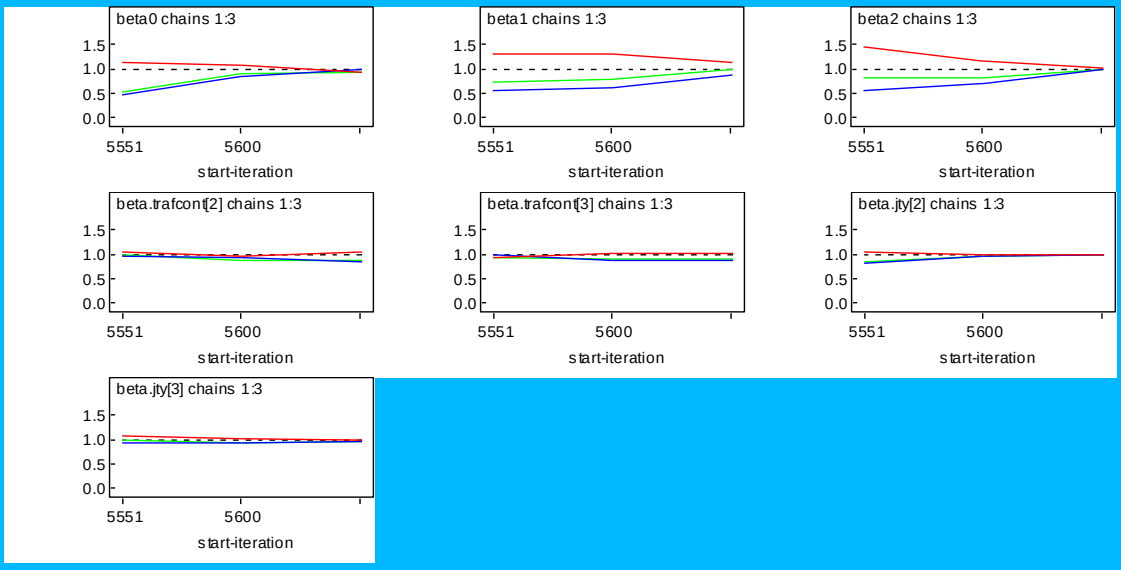

Figure 32 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson regression model fitted to the Austrian junction accident data.

The point estimates for the model represented by Equation 4.1 are given in Table 12. These point estimates are given by the posterior means obtained for the distributions *a posteriori* of each unknown parameter. Table 12 also shows the standard deviations, Monte Carlo errors and 95% credible intervals for the estimates of the coefficient parameters as shown in Equation 4.1.

The 95% credible intervals for the estimates of the continuous independent variable *ln(AADTmaj)* do not include zero thus indicating that this variable is relevant when predicting the expected frequency of accidents. The 95% credible interval for the estimate of *ln(AADTmin)* includes zero, however, from observation of this estimate's densities in Figure 33, it can be seen that the mean shifts away from zero indicating that this variable can still be considered as relevant in the model.

The Monte Carlo errors show relatively low values indicating that the parameters were calculated with accuracy.

Table 12 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson regression model was fitted to the Austrian accident data using a '*Y*' type junction with a '*yield*' traffic control as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-5.804	0.730	9.079E-02	-7.356	-4.569
β_1 (ln(AADTmaj))	0.440	0.071	8.578E-03	0.302	0.585
β_2 (ln(AADTmin))	0.141	0.081	9.565E-03	-0.022	0.300
β_3 (Traffic_Control='sign')	-1.001	0.436	1.665E-02	-1.939	-0.225
β_3 (Traffic_Control='stop')	-0.388	0.141	5.026E-03	-0.676	-0.109
β_4 (Junction_Type='roundabout')	0.285	0.172	8.497E-03	0.077	0.628
β_4 (Junction_Type='X')	0.213	0.145	4.993E-03	-0.063	0.502

The coefficient estimates in Table 12 suggest that the expected number of accidents on junctions with traffic control of type '*stop*' are *a posteriori* expected to have around approximately 32% less accidents than a junction with traffic control of type '*yield*' but the same values of *ln(AADTmaj)* and *ln(AADTmin)* and junction type. However, junctions with traffic control of type '*signalised*' are *a posteriori* expected to have approximately 63% less accidents than a junction with traffic control '*yield*' for the same values of the remaining explanatory variables.

Also, according to the Poisson regression model, junctions with type '*roundabout*' and '*X*' are expected to have around 33% and 24% more accidents than junctions with type '*Y*', for the same values of the remaining explanatory variables.

The posterior densities of the parameter estimates are shown in Figure 33. It shows that the mean values of these distributions have moved away from zero (the mean value assumed *a priori*).

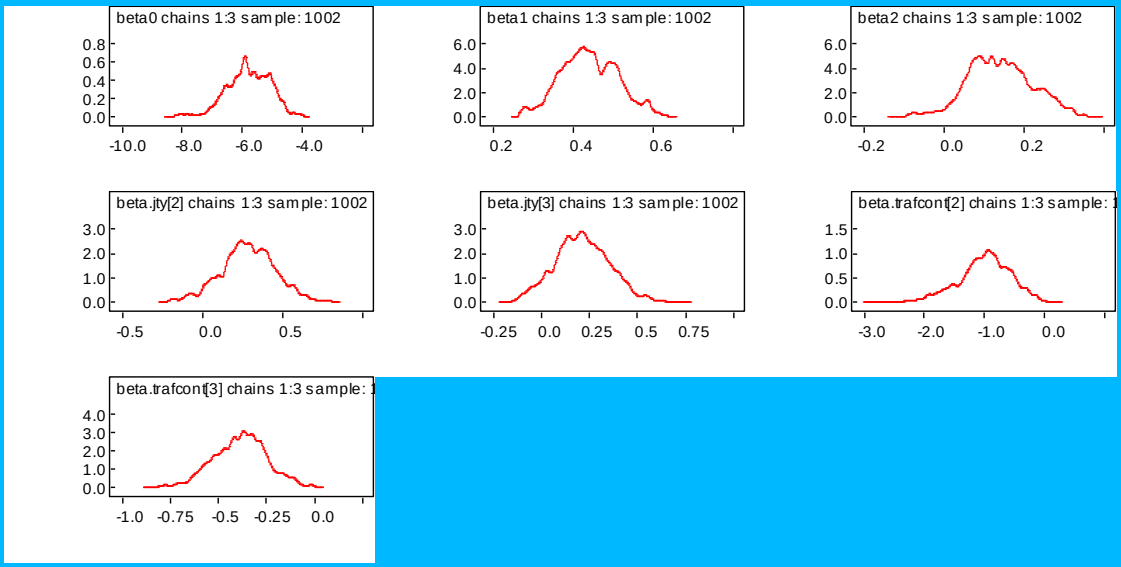

Figure 33 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson regression model was fit to the Austrian data.

The expected numbers of accidents, for a one year period, at junctions in Lower Austria are

given by the equations shown in Table 13, for the various categories of the explanatory variables *Traffic_Control* and *Junction_Type*.

It can be observed that '*signalised*' junctions produce the least number of predicted accidents, followed by '*stop*' and '*yield*' controlled junctions, this last with the higher number of expected accidents.

The type of junction with higher expected number of accidents is '*yield*'. The number of accidents decreases for '*roundabout*' and '*X*', with '*Y*' having the lowest value.

For each category of *Traffic_Control* the type of junction with the lowest expected number of accidents is *Y*.

Intersections with three approaches (legs) have the lowest predicted number of accidents in all the intersection control categories modelled.

Table 13 Expected number of accidents per year for Austrian junctions obtained by a Poisson regression model using a '*Y*' type junction with a '*yield*' traffic control as baseline.

	Expected Numbers of Accidents
Traffic_Control='signalised'	
Junction_Type 'X' 'Y'	 $\hat{\mu}_i = 1.371\times10^{-3} \times AADTmaj_i^{0.440} \times AADT\ \min_i^{0.140}$ $\hat{\mu}_i = 1.110\times10^{-3} \times AADTmaj_i^{0.440} \times AADT\ \min_i^{0.140}$
Traffic_Control='stop'	
Junction_Type 'X' 'Y'	 $\hat{\mu}_i = 2.533\times10^{-3} \times AADTmaj_i^{0.440} \times AADT\ \min_i^{0.140}$ $\hat{\mu}_i = 2.047\times10^{-3} \times AADTmaj_i^{0.440} \times AADT\ \min_i^{0.140}$
Traffic_Control='yield'	
Junction_Type 'roundabout' 'X' 'Y'	 $\hat{\mu}_i = 4.011\times10^{-3} \times AADTmaj_i^{0.440} \times AADT\ \min_i^{0.140}$ $\hat{\mu}_i = 3.732\times10^{-3} \times AADTmaj_i^{0.440} \times AADT\ \min_i^{0.140}$ $\hat{\mu}_i = 3.017\times10^{-3} \times AADTmaj_i^{0.440} \times AADT\ \min_i^{0.140}$

4.2.1 Model Checking

The histograms and dot plots in Figures 34 and 35 illustrate the observed and modelled accident data at Austrian junctions over the four year period. The replicated data sets were obtained from the Poisson regression model given by the equation in Table 13.

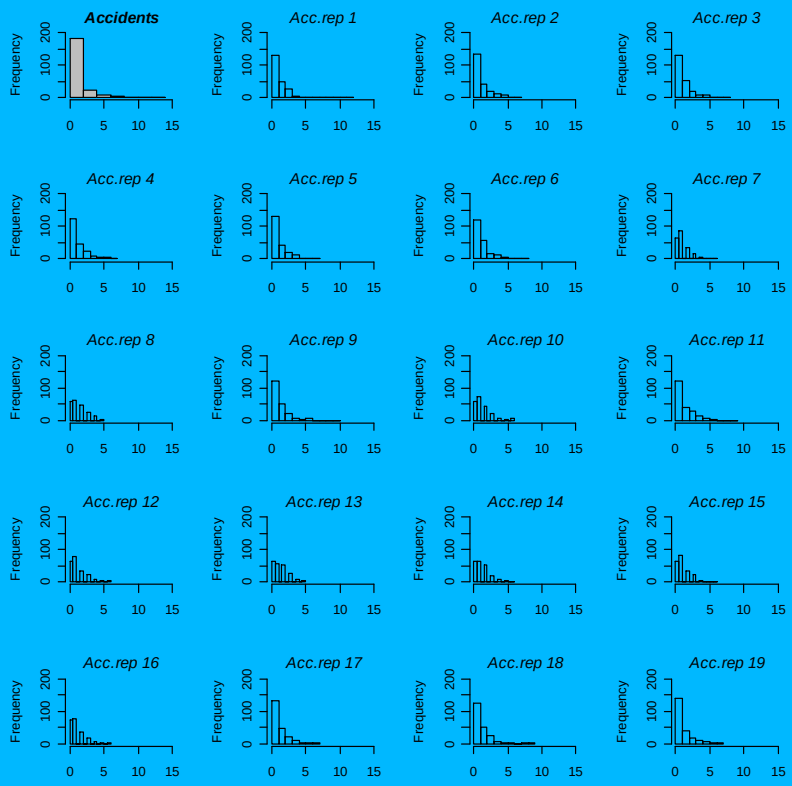

Figure 34 Histogram of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.

The *x* axis of the dotplots in Figure 35 corresponds to the junction's indexes and the *y* axis to the number of accidents.

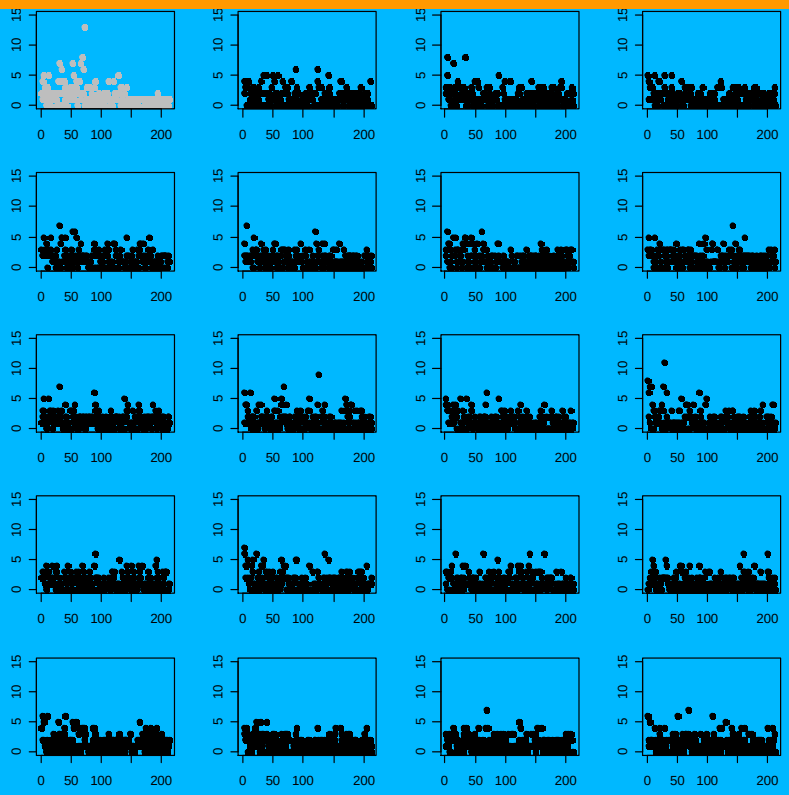

Figure 35 Dot plot of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.

Both figures reveal that the replicated data does generally not resemble the observed data (graphs in grey). In particular, observing the dot plots in Figure 35 it can be noticed that the predicted data for junctions indexed as number 1 to around junction 100 do not have the high values of number of injury accidents that are present in the observed data (grey dot plot at the top left panel).

The four discrepancy measures are plotted in Figure 36. Observation of this figure together with the estimated probabilities that the replicated data is greater than the observed discrepancy shows that the Poisson model does not seem to be able to capture the maximum value and the standard deviation of the observed data.

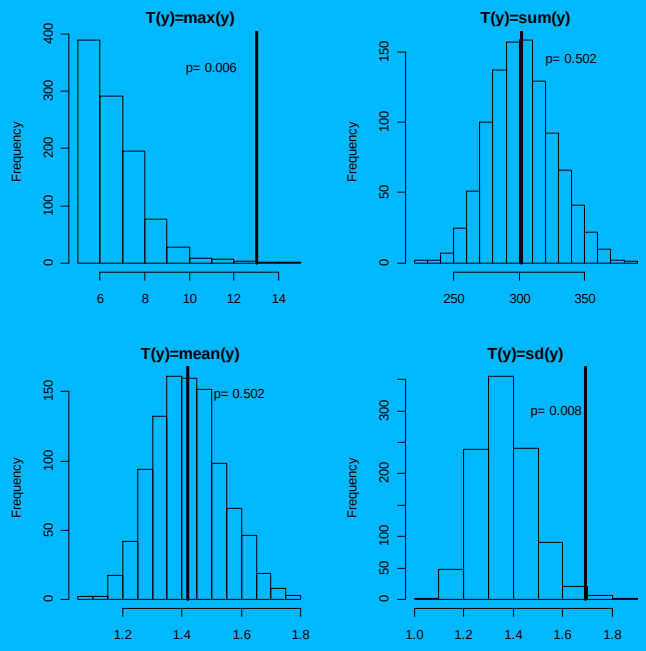

Figure 36 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson regression model fit to the Austrian data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The effect of the measure of discrepancy suggested by Congdon (2005) to help checking whether the overdispersion of the data was being taken into account by the model (variance over the mean) is displayed in Figure 37.

The estimated probability that the ratios of the variance to the means in the replicated data is greater than the same ratio calculated from the observed data, is equal to 0.003, which is a very small value, indicating that there is overdispersion that is not being replicated by the Poisson model. Consequently, there are strong reasons to believe that the Poisson regression model is not appropriate to describe the number of accidents in the Austrian junction data.

This model produces an average deviance, $\overline{D}$, equal to 678.869, an effective model dimension of 6.758 and a DIC equal to 685.627.

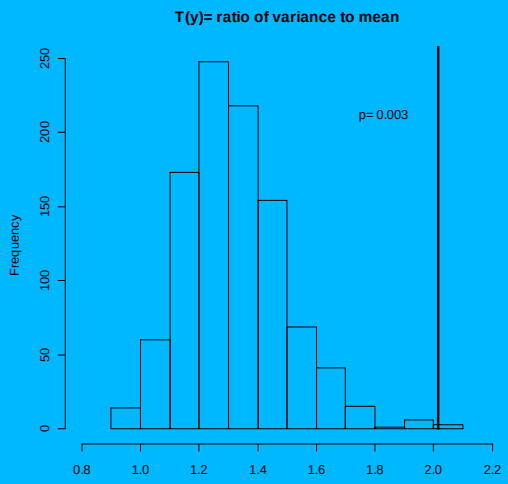

Figure 37 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson regression model fit to the Austrian data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

4.3 The Poisson-Gamma hierarchical regression model

The Poisson-Gamma hierarchical regression model was fitted to the Austrian junction data using Equations 2.3 and 2.4 with $\delta = \gamma$ and $\delta \sim Gamma(a,a)$ with a=0.1. Equation 4.1 was applied and the β parameters were given *a priori* Normal distributions with mean 0 and precision 0.0001 (the variance is equal to the inverse of the precision).

The MCMC algorithm comprised 3 chains and was run for 35000 iterations with 33000 burn-in iterations with a thinning rate of 6. The results thus described were based on a sample with dimension equal to 1002.

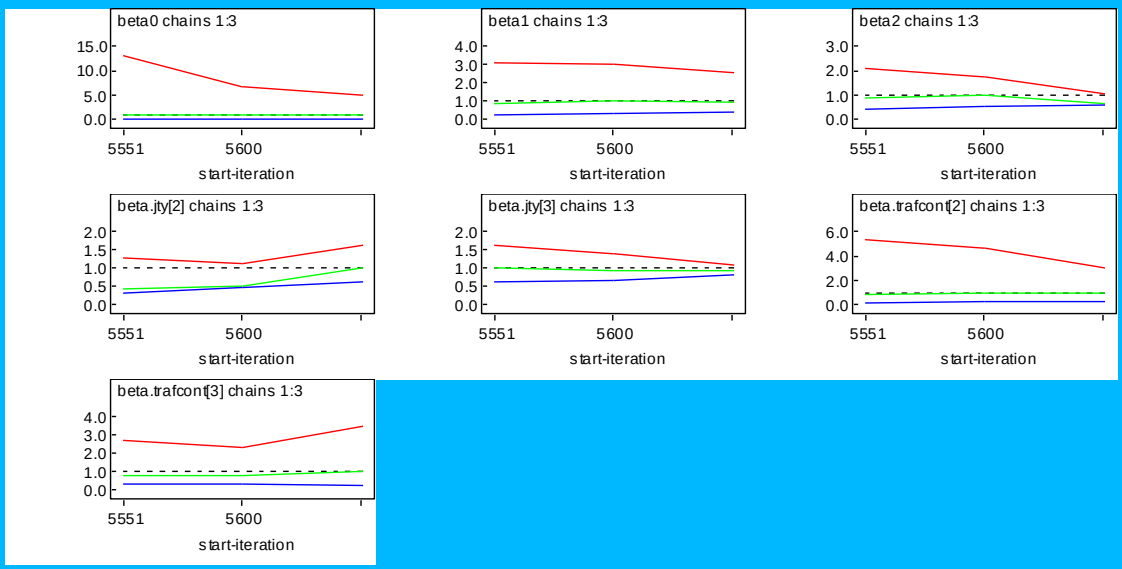

Figure 38 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson-Gamma regression model fitted to the Austrian junction accident data.

The Gelman-Rubin plots for the β parameters are displayed in Figure 38. The *Rhat* statistic (represented by the red line) tends to 1 for all estimated parameters apart from the β corresponding to factor '*stop*' of variable *Traffic_Control*, which decreased up to around iteration 5600 but then increased after that, as well as the estimate of category '*roundabout*' for *Junction_Type*.

Since there are no reasons to doubt the non-convergence of the majority of the parameters it is assumed that the MCMC algorithm has converged to each corresponding stationary distributions.

The Monte Carlo standard errors are relatively low values indicating that the parameters were calculated with accuracy (see Ntzoufras, 2009).

Point estimates, and corresponding standard deviations, Monte Carlo standard errors and 95% credible intervals, obtained after the Poisson-Gamma model was fit to the data, are displayed in Table 14. From the examination of the point mean estimates it can be stated that, according to this particular Poisson-Gamma model, every unit increase in *ln(AADTmaj)* increases the expected frequency of accidents by approximately 18% (and assuming that all the other explanatory variables remain constant). An increase in one unit of *ln(AADTmin)* increases the expected frequency of accidents in only 5%. See example on section 3.2.

Table 14 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fit to the Austrian accident data using a *'Y'* type junction with a *'yield'* traffic control as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-2.751	0.810	1.116E-01	-3.705	-1.454
β_1 (ln(AADTmaj))	0.168	0.108	1.470E-02	-0.013	0.344
β_2 (ln(AADTmin))	0.045	0.087	1.161E-02	-0.132	0.211
β_3 (Traffic_Control='sign')	-0.659	0.484	6.651E-02	-1.347	0.203
β_3 (Traffic_Control='stop')	-0.446	0.205	2.760E-02	-0.744	-0.060
β_4 (Junction_Type='roundabout')	0.332	0.145	1.899E-02	0.130	0.697
β_4 (Junction_Type='X')	0.128	0.126	1.653E-02	-0.134	0.334

Signalised and stop controlled junctions have a lower predicted number of accidents (48% and 36%) when compared to '*yield*' controlled junctions (and the remaining explanatory variables are kept constant). *roundabout*' and '*X*' are *a posteriori* expected to have a higher expected accidents of around 39% and 14%, respectively, when compared with equivalent junctions of type '*Y*' (being the other explanatory variables with constant values).

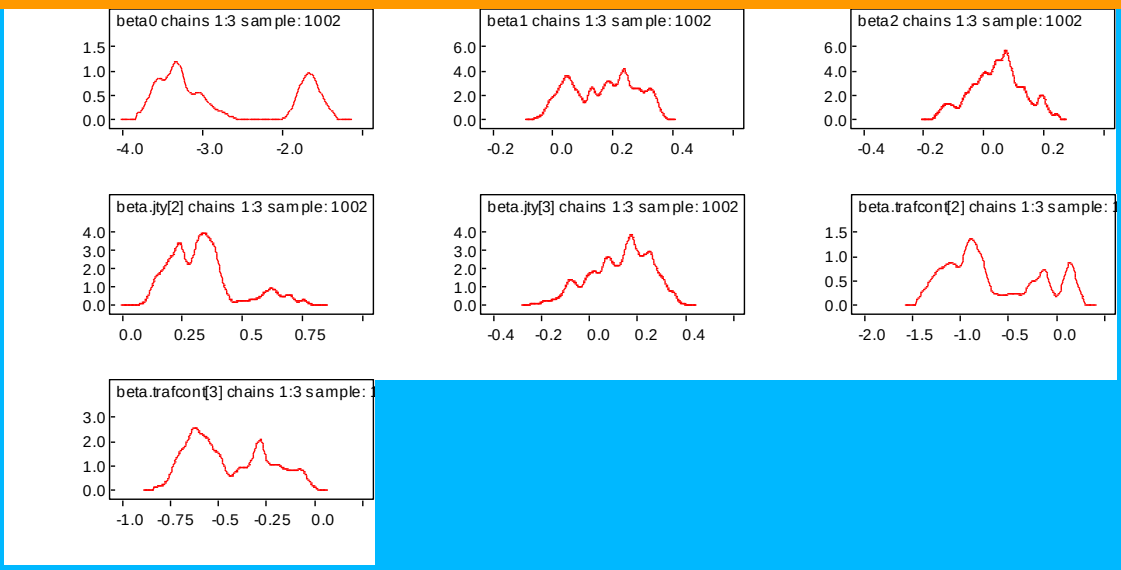

Figure 39 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson-Gamma regression model was fit to the Austrian data.

The posterior densities of the estimated β coefficient parameters are displayed in Figure 39. Several spikes can be observed on all parameter densities.

The expected frequency of accidents for a one year period is given by the equations displayed in Table 15.

Table 15 Expected number of accidents per year for Austrian junctions obtained by a Poisson-Gamma regression model using a *'Y'* type junction with a *'yield'* traffic control as baseline.

	Expected Numbers of Accidents
Traffic_Control='signalised'	
Junction_Type	
'X'	$\hat{\mu}_i = 3.757 \times 10^{-2} \times AADTmaj_i^{0.168} \times AADT\ \min_i^{0.045}$
'Y'	$\hat{\mu}_i = 3.305 \times 10^{-2} \times AADTmaj_i^{0.168} \times AADT\ \min_i^{0.045}$
Traffic_Control='stop'	
Junction_Type	
'X'	$\hat{\mu}_i = 4.649 \times 10^{-2} \times AADTmaj_i^{0.168} \times AADT\ \min_i^{0.045}$
'Y'	$\hat{\mu}_i = 4.090 \times 10^{-2} \times AADTmaj_i^{0.168} \times AADT\ \min_i^{0.045}$
Traffic_Control='yield'	
Junction_Type	
'roundabout'	$\hat{\mu}_i = 8.903 \times 10^{-2} \times AADTmaj_i^{0.168} \times AADT\ \min_i^{0.045}$
'X'	$\hat{\mu}_i = 7.261 \times 10^{-2} \times AADTmaj_i^{0.168} \times AADT\ \min_i^{0.045}$
'Y'	$\hat{\mu}_i = 6.389 \times 10^{-2} \times AADTmaj_i^{0.168} \times AADT\ \min_i^{0.045}$

The junctions with signalised traffic control have the lowest expected frequencies of accidents than any of the other three categories. The level of *Traffic_Control* with the highest expected frequencies of injury accidents is the '*yield*'.

The junctions with type '*X*' have the highest expected accident frequency of the categorical variable *Junction_Type*. They are followed by category '*Y*' being this one the type where less number of accidents is expected.

When considering traffic control equal to '*yield*', the '*roundabouts*' have higher expected number of injury accidents than types '*X*' and '*Y*'.

4.3.1 Model Checking

Graphs of replicated data in the form of histograms and dot plots are displayed in Figure 40 and Figure 41, respectively. Both the histograms and the dot plots show that the observed data seems plausible under the posterior predictive distribution data represented by the histograms and dot plots of the replicated data.

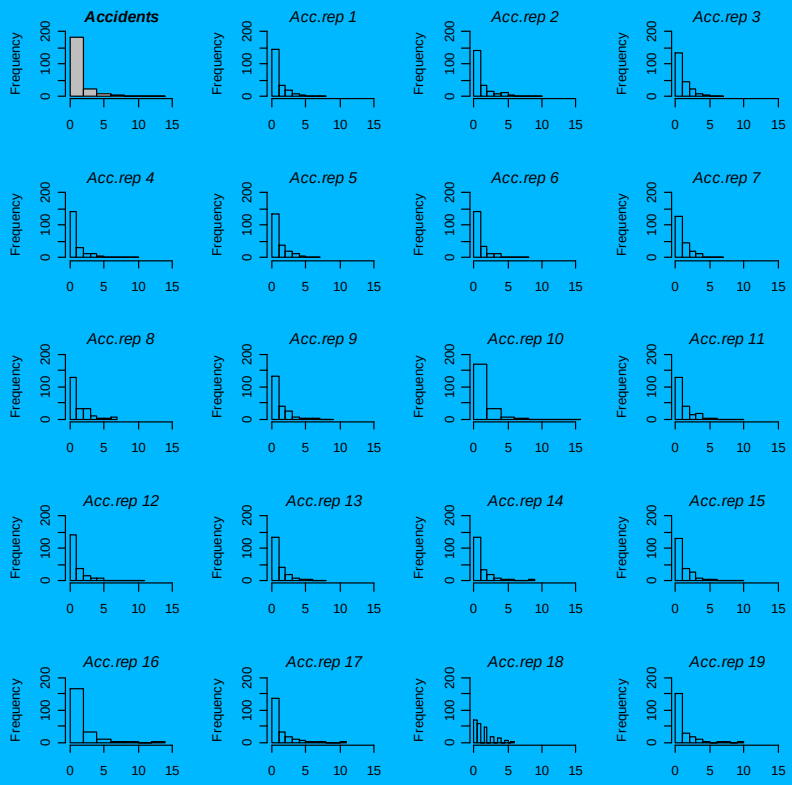

Figure 40 Histogram of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Gamma regression model.

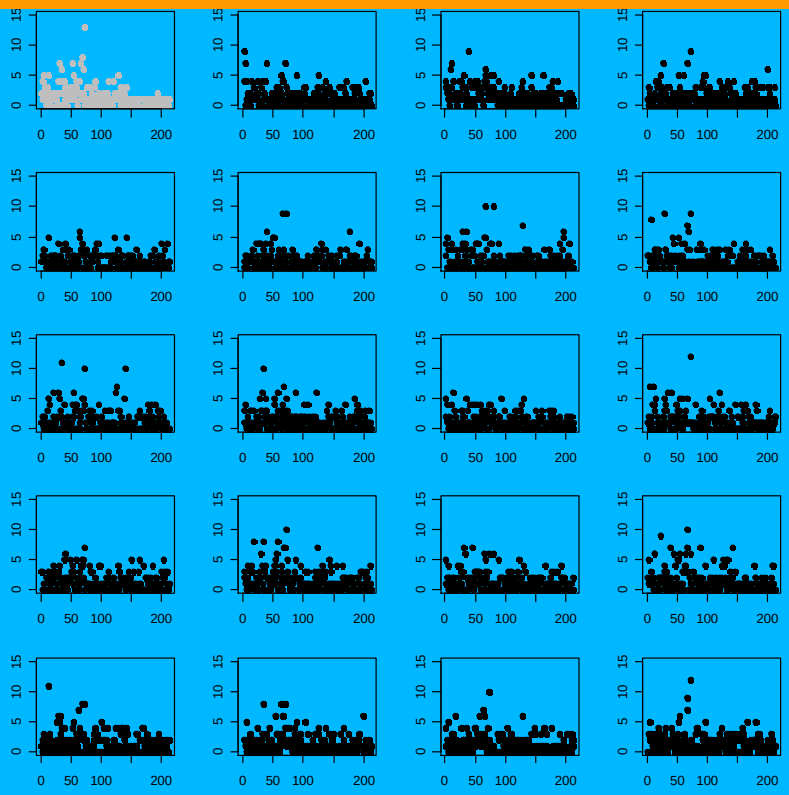

Figure 41 Dot plot of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.

The dot plots displayed in Figure 41 show that for junctions indexed as 1 to around 100 the model seemed able to replicate higher values of accident occurrences, just as it was observed in the real data (grey dot plot at top left in Figure 41).

The four discrepancy measures; *max*, *sum*, *mean* and *sd* are displayed in Figure 42. The posterior probability of the discrepancy measures obtained from the replicated data being greater than the corresponding discrepancy measures resulting from the observed data show better results for the Poisson-Gamma regression model than they did for the Poisson regression (see Figure 36).

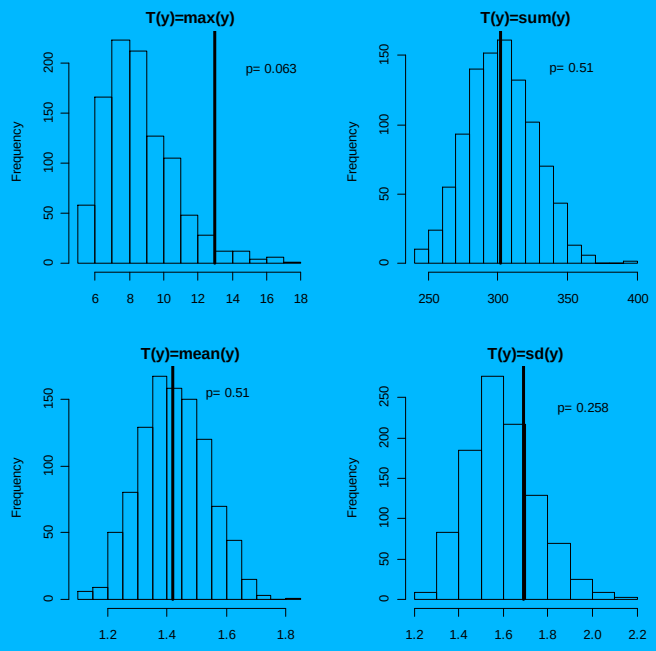

Figure 42 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fit to the Austrian data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

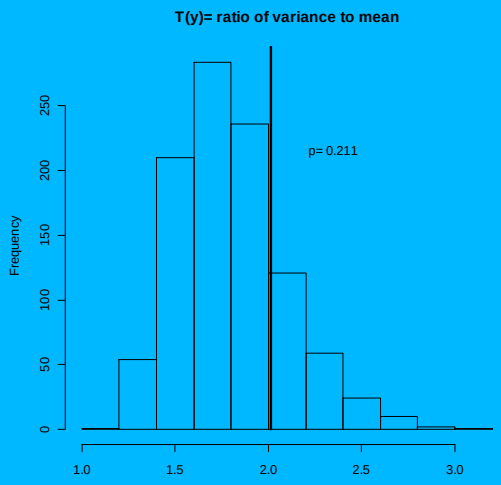

Figure 43 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fit to the Austrian data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

Figure 43 shows the graph of the discrepancy measure obtained by the variance to mean ratio, giving an estimated posterior probability equal to 0.211 which indicates that the Poisson-Gamma model does not seem to be taking the observed data overdispersion into account, even though the estimated probability is within the satisfactorily limits proposed by Congdon (2005) of 0.1 to 0.9.

This model produces an average deviance, $\bar{D}$, of 585.535 and an effective model dimension, d_e, of 63.177, giving a DIC of 648.711.

The dispersion parameter ($1/\delta$ in Equation 2.4) was estimated as 0.364.

The expected accident frequency for the four junction types and three traffic controls of junctions for the minimum, mean, median and maximum profiles have been calculated. In these profiles the minimum, mean, median and maximum values of *ln(AADTmaj)* and *ln(AADTmin)* were considered, since these variables are positively associated with the number of injury accidents.

The posterior means and the corresponding standard deviations of these profiles are provided in Table 16.

Table 16 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the Austrian accident data.

		Minimum	Mean	Median	Maximum
	ln(AADTmaj)	6.340	8.643	8.659	10.318
	ln(AADTmin)	6.328	6.575	6.323	9.255
Junction_Type	Traffic_Control	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
roundabout	yield	0.358 (0.103)	0.577 (0.134)	0.579 (0.135)	0.840 (0.258)
X	signalised	0.166 (0.086)	0.286 (0.195)	0.287 (0.196)	0.436 (0.358)
	stop	0.183 (0.037)	0.298 (0.055)	0.299 (0.055)	0.437 (0.127)
	yield	0.295 (0.098)	0.472 (0.116)	0.473 (0.116)	0.682 (0.212)
Y	signalised	0.145 (0.072)	0.248 (0.161)	0.249 (0.162)	0.377 (0.296)
	stop	0.161 (0.035)	0.263 (0.050)	0.263 (0.050)	0.384 (0.111)
	yield	0.259 (0.083)	0.414 (0.096)	0.415 (0.096)	0.598 (0.175)

For a typical junction (under column *Mean*) with type '*roundabout*' and '*yield*' traffic control it is expected 0.577 accidents, while for a junction of type '*Y*' and with '*signalised*' for traffic control the corresponding number of injury accidents is about 0.248 (the lowest value).

The worst case scenario (maximum profile) where junctions with 10.318 of *ln(AADTmaj)* and a *ln(AADTmin)* of 9.255 corresponds to an expected number of 0.840 and 0.682 injury accidents for '*roundabout*' yield junctions and '*X*' yield junctions, respectively.

4.4 The Poisson Log-Normal regression model

The Poisson Log-Normal regression model was fit to the data according to Equations 2.5 and 2.6 in Chapter 2, where the parameter α in Equation 2.6 had a Gamma(*a*,*a*) *a priori* distribution with *a*=0.001.

The MCMC algorithm was run with three chains for 35000 iterations with 33000 as burn-in with a thinning rate equal to 6. The results were drawn from samples with dimension 1002.

The graphs depicted in Figure 44 show the three Gelman-Rubin statistics obtained for the estimates of the β parameters. It can be observed in all of the graphs that the *Rhat* statistic (red line) converges to the value one as the iterations increase. The other two statistics (*W* and *Vhat*) remain constant. Therefore, there are no reasons to suspect the non-convergence of the iterative simulation.

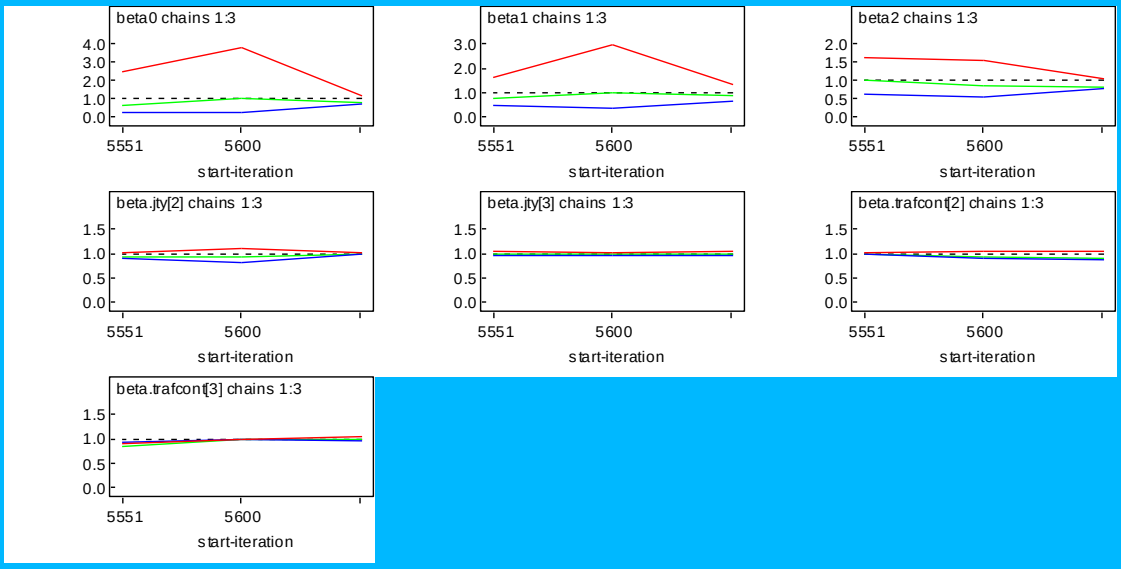

Figure 44 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson Log-Normal regression model fitted to the Austrian junction accident data.

The posterior means of the parameter estimates, corresponding standard deviations, Monte Carlo standard errors and 95% credible intervals are displayed in Table 17.

Table 17 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the Austrian accident data using a '*Y*' type junction with a '*yield*' traffic control as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-6.021	1.558	2.081E-01	-8.897	-3.180
β_1 (ln(AADTmaj))	0.439	0.163	2.173E-02	0.139	0.754
β_2 (ln(AADTmin))	0.145	0.100	1.224E-02	-0.038	0.336
β_3 (Traffic_Control='sign')	-0.932	0.536	2.406E-02	-2.061	0.080
β_3 (Traffic_Control='stop')	-0.333	0.177	7.421E-03	-0.674	-0.008
β_4 (Junction_Type='roundabout')	0.331	0.221	1.320E-02	-0.128	0.758
β_4 (Junction_Type='X')	0.184	0.173	6.893E-03	-0.149	0.512

The MC errors have small values indicating that the parameter estimates were calculated with accuracy. The 95% credible intervals for some of the parameter estimates contain zero, however, from the observation of the estimates densities in Figure 45 it can be concluded that the mean is away from zero (this former value lying usually on the tail of the densities).

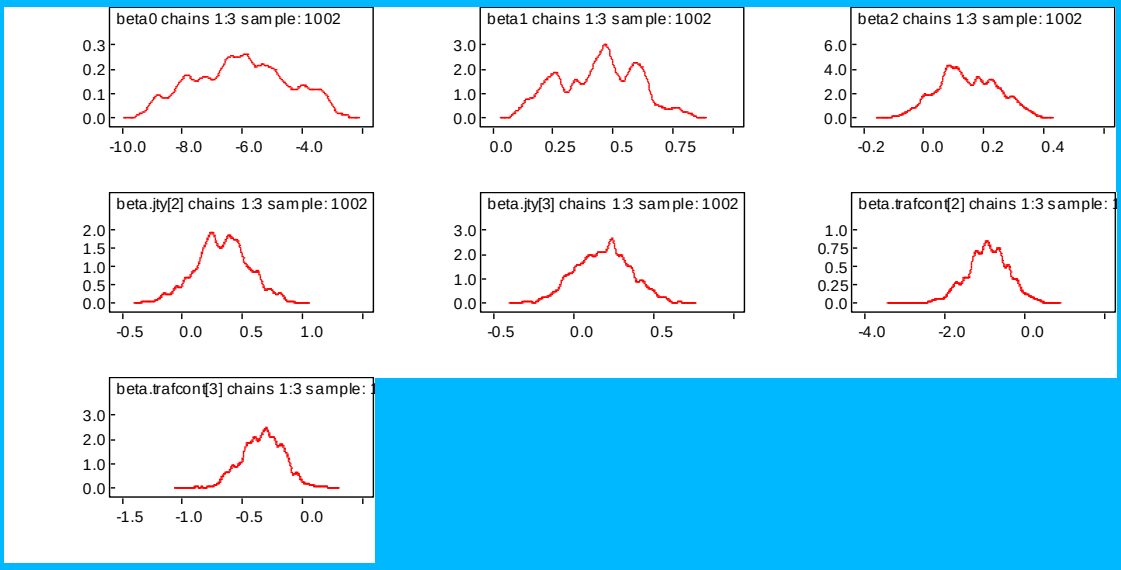

Figure 45 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson Log-Normal regression model was fit to the Austrian data.

From the examination of the mean posterior estimates in Table 17 it can be stated that every unitary increase in *ln(AADTmaj)* increases the expected number of accidents by approximately 55%, provided the other explanatory variables remain constant. An increase in one unit of *ln(AADTmin)* increases the expected frequency of injury accidents in around 16%. See example on section 3.2.

A junction with '*signalised*' and '*stop*' as traffic control are *a posteriori* expected to have approximately 61% and 28% fewer accidents, respectively, than a junction with '*yield*' for traffic control (with the other explanatory variables remaining constant).

A junction with type '*roundabout*' and '*X*' are expected to have an increase in the expected number of accidents by 39% and 20%, respectively, when compared with a junction with type '*Y*' (provided the other variables remain constant).

The equations giving the expected frequency of injury accidents, per year, per junction type and traffic control are given in Table 18.

Table 18 Expected number of accidents per year for Austrian junctions obtained by a Poisson Log-Normal regression model using a '*Y*' type junction with a '*yield*' traffic control as baseline.

	Expected Numbers of Accidents
Traffic_Control='signalised'	
Junction_Type 'X' 'Y'	 $\hat{\mu}_i = 1.150 \times 10^{-3} \times AADTmaj_i^{0.439} \times AADT\ \min_i^{0.145}$ $\hat{\mu}_i = 9.566 \times 10^{-4} \times AADTmaj_i^{0.439} \times AADT\ \min_i^{0.145}$
Traffic_Control='stop'	
Junction_Type 'X' 'Y'	 $\hat{\mu}_i = 2.093 \times 10^{-3} \times AADTmaj_i^{0.439} \times AADT\ \min_i^{0.145}$ $\hat{\mu}_i = 1.741 \times 10^{-3} \times AADTmaj_i^{0.439} \times AADT\ \min_i^{0.145}$
Traffic_Control='yield'	
Junction_Type 'roundabout' 'X' 'Y'	 $\hat{\mu}_i = 3.381 \times 10^{-3} \times AADTmaj_i^{0.439} \times AADT\ \min_i^{0.145}$ $\hat{\mu}_i = 2.919 \times 10^{-3} \times AADTmaj_i^{0.439} \times AADT\ \min_i^{0.145}$ $\hat{\mu}_i = 2.428 \times 10^{-3} \times AADTmaj_i^{0.439} \times AADT\ \min_i^{0.145}$

The junctions with yield traffic control have higher expected frequencies of accidents than any of the other three categories, being the traffic control equal to '*signalised*' the one whose junctions have lower expected accident frequencies.

The junctions with type '*roundabout*' have the higher expected accident frequency of the categorical variable *Junction_Type*. They are followed by factors '*X*' and '*Y*' being the latter the type where less number of accidents is expected.

4.4.1 Model Checking

Figure 46 contains nineteen histograms of replicated data under the Poisson Log-Normal model as well as the histogram of the observed number of injury accidents. In general, it can be stated that the model under consideration is able to replicate conveniently the observed data.

The dot plots of a further set of nineteen replicated data sets are displayed in Figure 47. Again there are no reasons to suspect that the model does not replicate conveniently the observed data (*Accidents*).

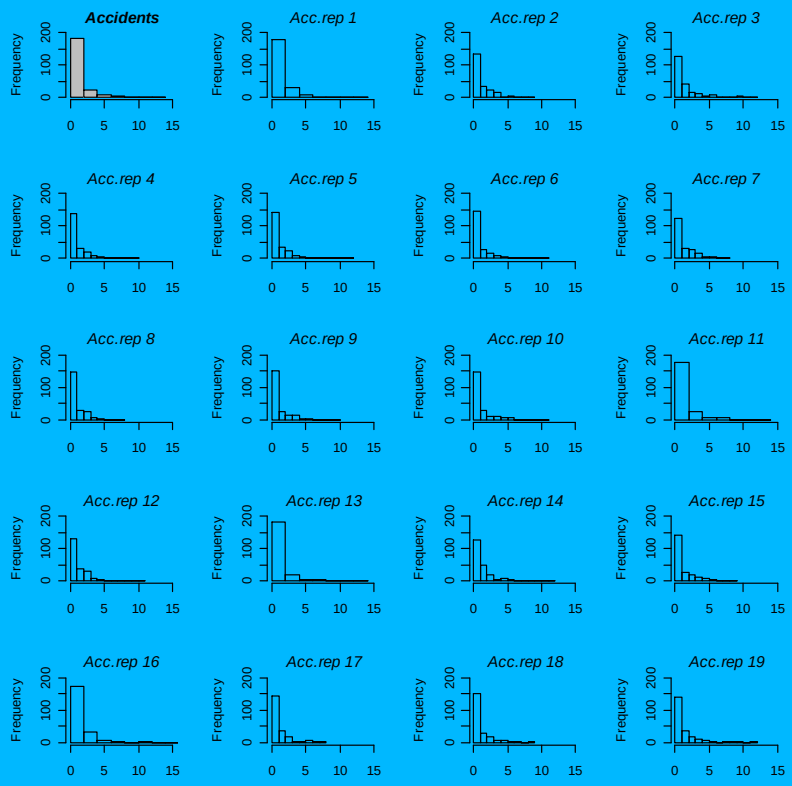

Figure 46 Histogram of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

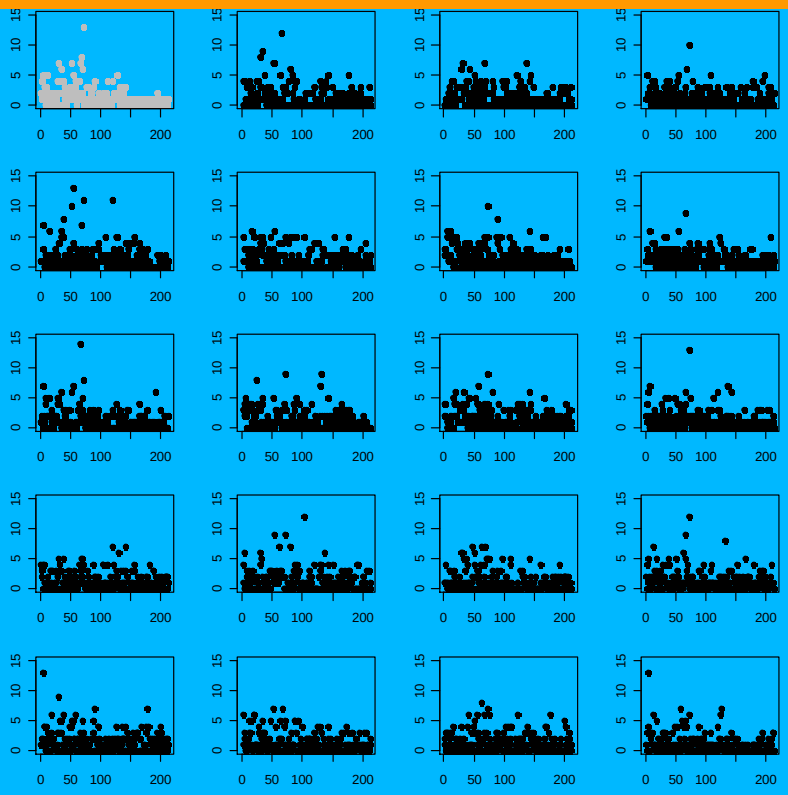

Figure 47 Dot plot of the observed number of accidents in Austrian junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

The four discrepancy measures concerning the maximum, sum, mean and standard deviation (sd) values are displayed in Figure 48.

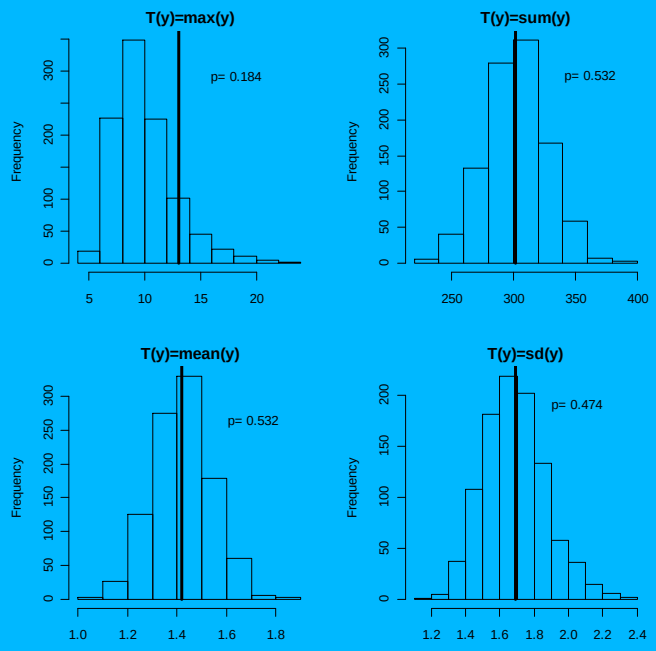

Figure 48 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fit to the Austrian data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The probabilities (denoted by *p* in the graphs) that the discrepancy measures obtained by the replicated data are greater than the corresponding discrepancy measures from the observed data lie within the satisfactory boundaries of 0.1 to 0.9 as suggested by Congdon (2005). The *p* values are also near 0.5 for all the discrepancies apart from the discrepancy measuring the maximum value in the data.

The values of the discrepancy measure used to check whether the overdispersion is taken into account by the Poisson Log-Normal model are displayed in Figure 49. The probability that the ratios of the variance over the mean each replicated data set is greater than the same ratio calculated under the observed data, is 0.455, which is a reasonable value by being close to 0.5. It can be concluded that the model is taking the overdispersion into account.

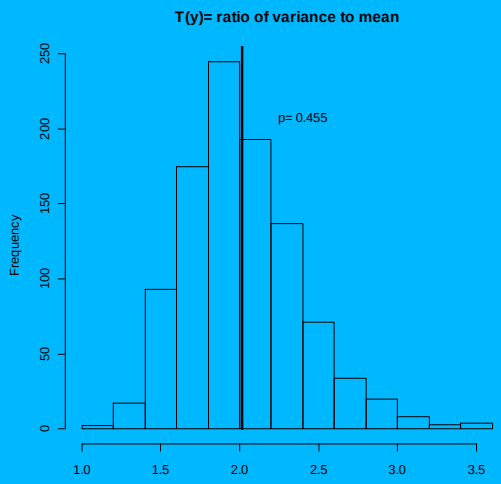

Figure 49 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fit to the Austrian data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

This model produces an average deviance, $\overline{D}$, of 578.013 and an effective model dimension, d_e, of 66.310, giving a DIC of 644.323.

The expected accident frequency for the three traffic control levels and four junction types for the minimum, mean, median and maximum profiles, obtained by the posterior means, are depicted in Table 19. In these profiles the minimum, mean, median and maximum values of *ln(AADTmaj)* and *ln(AADTmin)* were acknowledge since these variables are positively associated with the number of injury accidents.

Table 19 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the Austrian accident data.

		Minimum	Mean	Median	Maximum
	ln(AADTmaj)	6.340	8.643	8.659	10.318
	ln(AADTmin)	6.328	6.575	6.323	9.255
Junction_Type	Traffic_Control	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
roundabout	yield	0.153 (0.073)	0.542 (0.126)	0.547 (0.127)	1.529 (0.634)
X	signalised	0.059 (0.044)	0.211 (0.128)	0.213 (0.130)	0.597 (0.425)
	stop	0.093 (0.040)	0.338 (0.089)	0.342 (0.091)	0.977 (0.473)
	yield	0.128 (0.052)	0.473 (0.130)	0.477 (0.132)	1.379 (0.722)
Y	signalised	0.050 (0.039)	0.176 (0.111)	0.178 (0.112)	0.495 (0.353)
	stop	0.078 (0.034)	0.281 (0.073)	0.284 (0.074)	0.807 (0.376)
	yield	0.107 (0.043)	0.391 (0.099)	0.395 (0.101)	1.135 (0.564)

For a typical roundabout junction (under column *Mean*) with type '*yield*' as traffic control it is expected 0.542 accidents, while for a junction of type '*Y*' and with '*signalised*' for traffic control the corresponding number of injury accidents is about 0.176.

The worst case scenario (maximum profile) where junctions with 10.318 as *ln(AADTmaj)* and 9.255 of *ln(AADTmin)* corresponds to an expected number of 1.529 and 1.379 injury accidents for '*roundabout*' yield junctions and '*X*' yield junctions, respectively.

4.5 Discussion

The Table 20 shows the resulting fit ($\overline{D}$), complexity (d_e) and overall model choice (DIC) score for the models fit to the Austrian junction data.

Table 20 Comparison of DIC and related statistics for the three models fitted to the Austrian junction data.

Regression Model	$\overline{D}$	d_e	DIC
Poisson	678.869	6.758	685.627
Poisson-Gamma	585.535	63.177	648.711
Poisson Log-Normal	578.013	66.310	644.323

The Poisson-Gamma and the Poisson Log-Normal emerge as the models with the lower DIC values. The results of d_e suggest that the use of the multiplicative frailty component of the Poisson-Gamma model (see section 2.2.1) contributes to a decrease in around 3 effective parameters. However, taking into account the values in the second and third rows of Table 20 it seems that the choice model is irrelevant as far as the DIC values are concerned. Therefore, since the Poisson Log-Normal model captures in a more satisfactorily way the data discrepancies (see Figure 48 and Figure 49) than the Poisson-Gamma model (see Figure 42 and Figure 43) and that the posterior means in Table 19 have smaller standard deviations values than the corresponding means in Table 16, it was decided to assume that the Poisson Log-Normal model provided a better fit of the Austrian junction data.

Consequently, according to the Poisson Log-Normal, and observing the values in Figure 50, it can be concluded that the highest expected number of accidents occured on junctions with type '*roundabout*'. This is an unexpected result as it has been found that roundabouts are especially safer than unsignalised junctions; nevertheless the fact that the Austrian data has zero accident junctions under represented might explain this result.

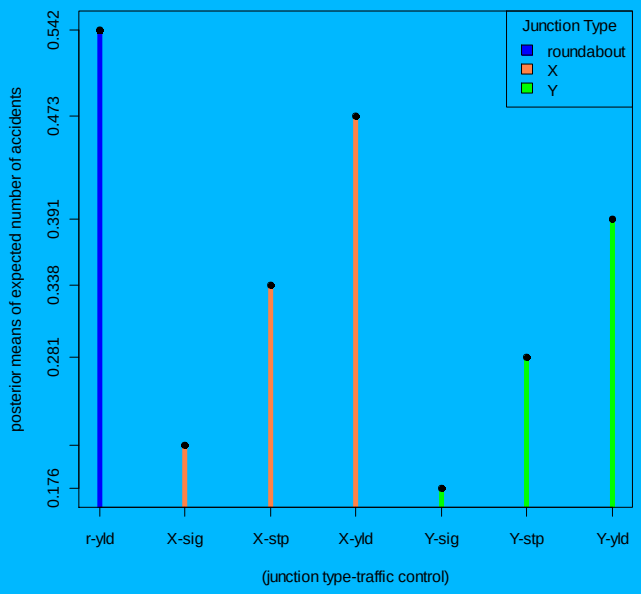

Figure 50 Values of the posterior means of the expected number of accidents for Austrian junctions classified per junction type and traffic control (as under column *Mean* in Table 19).

The type of junctions with the highest expected number of injury accidents is the '*roundabout*', they are followed, by decreasing expected number of accidents, by junctions of type 'X' and 'Y'.

Within each of X and Y types of junctions the ones with higher expected values of accidents have yield as traffic control followed, by decreasing order of expected number of accidents, by the control categories of '*stop*' and '*signalised*'.

Overall, the safer junctions are the ones which are signalised and of type '*Y*'.

5 Modelling Portuguese injury accidents

This chapter describes the data collected on Portuguese rural road networks junctions (section 5.1), proceeds with the description of the fit, assessment and checking of three Bayesian regression models including the Poisson (section 5.2), the hierarchical Poisson-

Gamma (section 5.3) and the hierarchical Poisson Log-Normal model (section 5.4). The final section (5.5) summarises the results obtained for the chosen model.

5.1 Portuguese Junction Data

The data described in this section consists of several measurements registered on 257 junctions belonging to the Portuguese rural road network. The data was collected over a five year period ranging from 2003 to 2007. Due to the low number of cases, staggered intersections and intersections with more than 4 approaches were removed from the sample.

The variables registered included, per junction:

- ***District_name***: gives the name of the Portuguese district where the junction is situated;
- ***Junction_Type***: a categorical variable referring to the type of the junction, it takes the values 'roundabout' or 'intersection' (with three or four legs);
- ***Number_of_Legs***: a categorical variable indicating the number of legs in the junction, it takes the values '3', or '4';
- ***Accidents***: gives the number of injury accidents;
- ***KSI_Acc***: represents the number of accidents which resulted in killed or serious injured victims;
- ***Killed_Acc***: gives the number of accidents that resulted in at least one fatality;
- ***AADTmaj***: represents the major traffic entering volume (annual average daily traffic);
- ***AADTmin***: represents the minor traffic entering volume (annual average daily traffic).

The dot plots displayed in Figure 51 show, in the *y* axis, the number of injury accidents (*Accidents*), accidents involving fatalities (*Killed_Acc*) and accidents involving fatalities or serious injury victims (*KSI_Acc*), per junction (the *x* axis gives the index of each junction) over the five year period of time. It can be observed the existence of one junction in which it has occurred 17 accidents (see upper left panel in Figure 51). There are also two junctions where two fatal accidents have occurred (upper right panel) and two junctions where three accidents produced killed or seriously injured victims (lower left panel).

Another way of displaying this information is through the employment of bar plots which are shown in Figure 52. Where the *y* axis corresponds to the frequencies of each bar which in turn corresponds to 0, 1, 2, etc. number of each type of accidents.

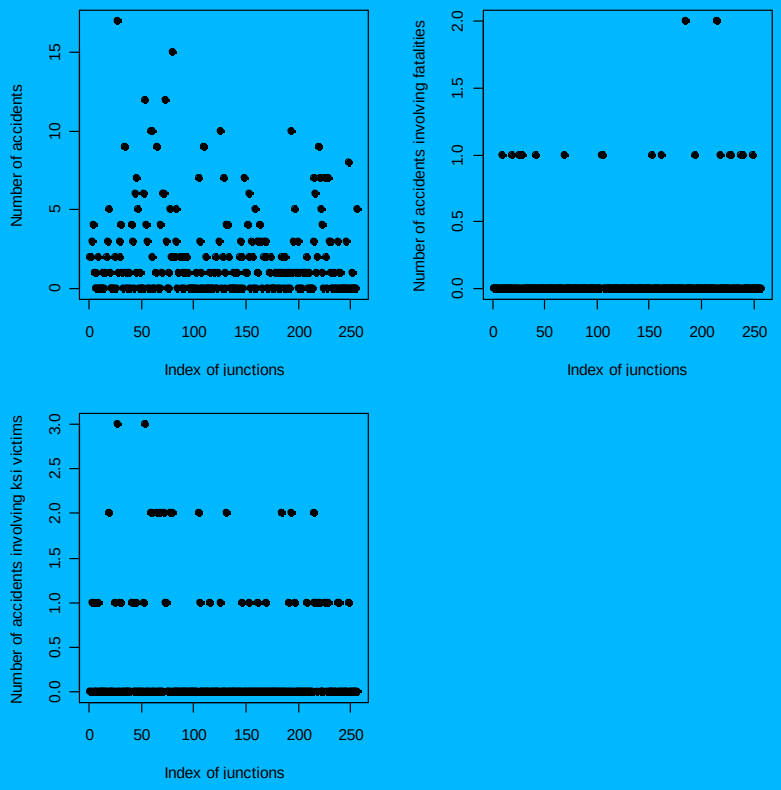

Figure 51 Plots of the total number of *Accidents*, *Killed_Acc*, and *KSI_Acc*, per junction, from upper left to right, respectively, registered from 2003 to 2007 in Portuguese rural road network junctions.

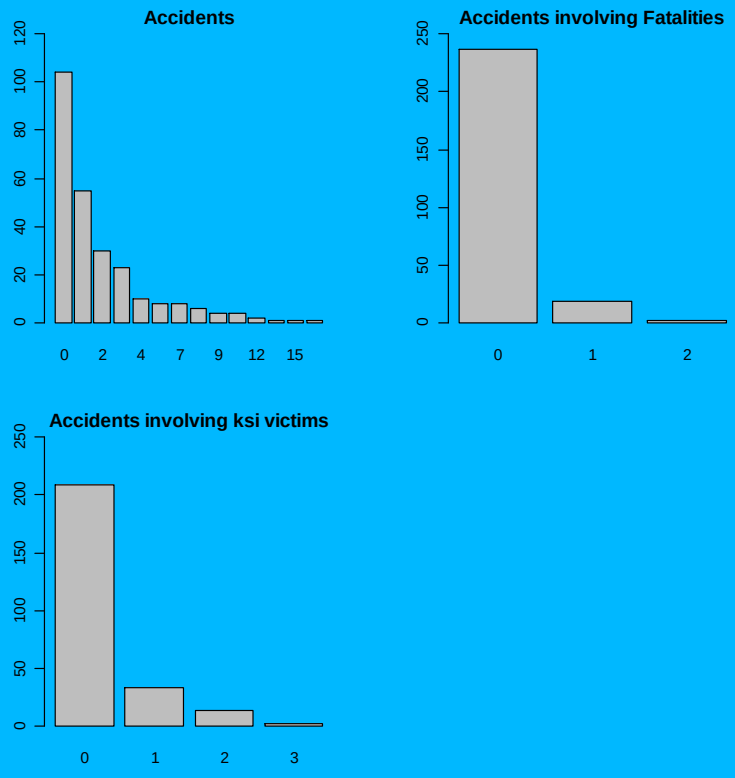

Figure 52 Bar plots of the frequencies of *Accidents*, number of accidents involving fatalities and number of accidents involving killed and seriously injured victims registered from 2003 to 2007 on Portuguese junctions from the rural road network.

The values of *AADTmaj* and *AADTmin* plotted against the total number of accidents are shown in Figure 53. The great majority of accidents occur at values of *AADTmaj* between 2500 and 12000 (Figure 53, left panel) and for 243 to around 5000 for *AADTmin* (Figure 53, right panel).

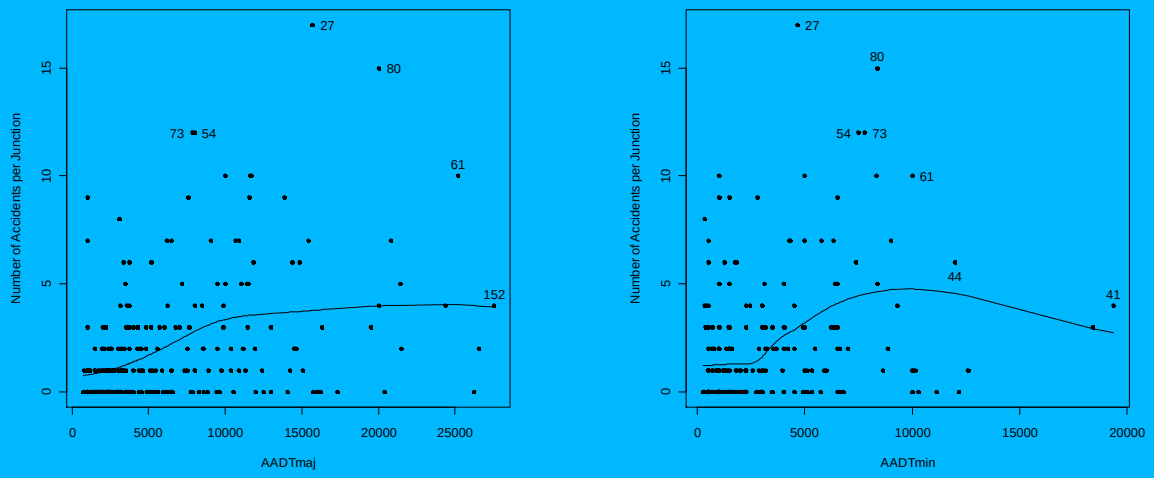

Figure 53 The number of accidents per junction in the Portuguese data set against *AADTmaj* and *AADTmin*, and corresponding polynomial fits, on the left and right panels, respectively.

The bar plots and the box plots of the two categorical variables are displayed in Figure 54 and Figure 55, respectively.

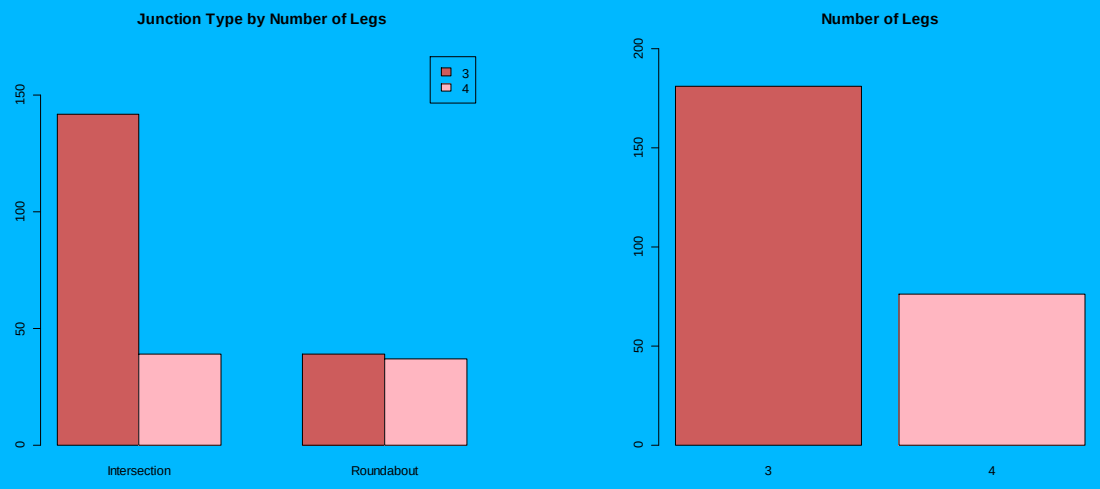

Figure 54 Bar plots giving the frequencies of the number of junctions per categories of variables *Junction_Type* and *Number_of_Legs*, registered from 2003 to 2007 on Portuguese junctions from the rural road network.

There are 76 junctions of type '*roundabout*' and 181 of type '*intersection*'. There are also 181 junctions with three legs and 76 with four legs. Nevertheless, there are 39 roundabouts with three legs and 37 with four legs. There are 142 junctions with type 'intersection' having three legs and 39 with four legs.

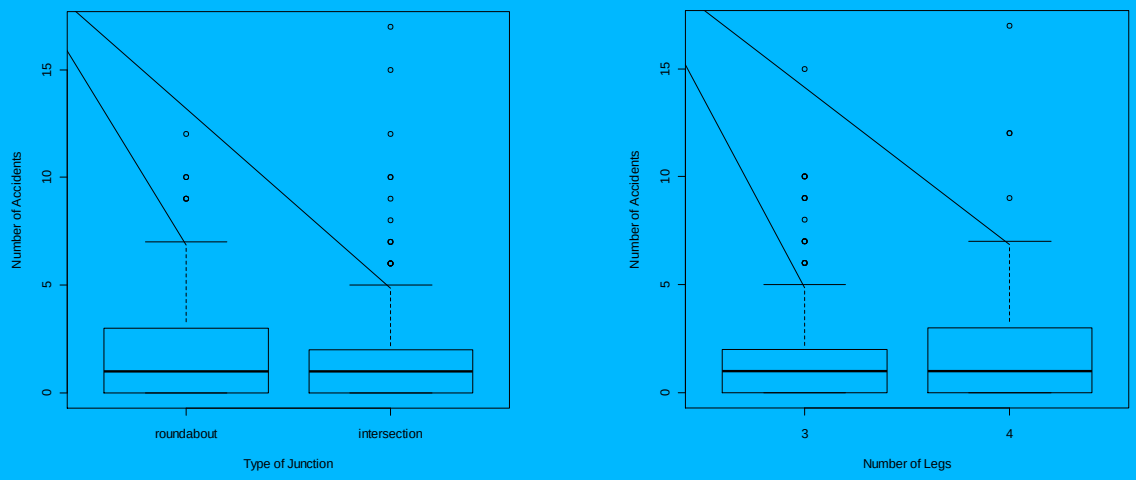

Figure 55 Box plots of the number of accidents in the Portuguese junctions by group for *Junction_Type* and *Number_of_Legs*, on the left and right panels, respectively.

From the observation of the box plots in Figure 55 it seems that there are not many differences between the median accident counts for the two types of junctions and also number of legs. The distributions of the injury accident counts for each type and number of legs are skewed to the right.

Table 21 contains summary statistics of the variables measured over the sample of the Portuguese junctions.

Table 21 Summary statistics for the variables registered on Portuguese junctions from 2003 to 2007.

Variables	minimum	mean	standard deviation	median	maximum
AADTmaj	743	6225.172	5547.6	3956	27530
AADTmin	243	2894.6	3020.7	1606	19359
Number_of_Legs	3	-	-	-	4
Accidents	0	1.930	2.783	1	17
Fatality_Acc	0	0.086	0.307	0	2
KSI_Acc	0	0.253	0.581	0	3

5.2 The Poisson regression model

The Poisson regression model given by Equations 2.1 and 2.2 was fitted to the data with the number of injury accidents (*Accidents*) as the dependent variable and the logarithms of both *AADTmaj* and *AADTmin*, *Number_of_Legs* and *Junction_Type* as explanatory variables as shown in Equation 5.1.

$$ln(\hat{\mu}_i) = \beta_0 + \beta_1 ln(AADTmaj_i) + \beta_2 ln(AADTmin_i) + \beta_3 Number\_of\_Legs_i + \beta_4 Junction\_Type_i \quad (5.1)$$

The parameter $\hat{\mu}$ gives the expected number of injury accidents for a period of one year. The β parameters were assigned Normal *a priori* distributions with mean 0 and variance 10000. The baseline, or reference, categories for the categorical variables are:

Number_of_Legs = 3

Junction_Type = '*intersection'*.

The MCMC algorithm comprised three chains and was run for 35000 iterations of which 33000 were considered burn-in, resulting in a thinning rate equal to 6. The results were drawn from samples with dimension 1002.

The plots of the Gelman-Rubin statistics for the Poisson regression parameters are shown Figure 56. There is no evidence to suspect non-convergence to the pretended density for each parameter as the *Rhat* statistic tends to one and the remaining parameters stabilise as the number of iterations increase.

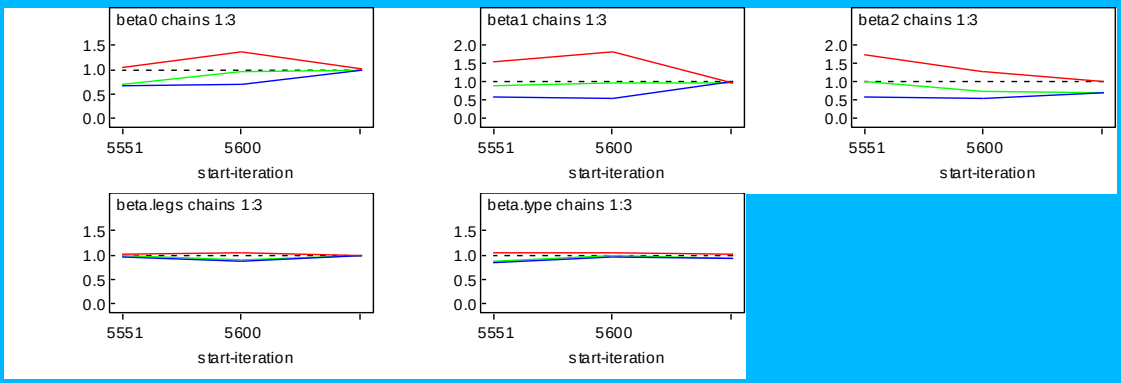

Figure 56 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson regression model fitted to the Portuguese junction accident data.

The point estimates for the regression model as in Equation 5.1 are given in Table 22. These point estimates are given by the posterior means obtained for the distributions *a posteriori* of each parameter.

Table 22 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson regression model was fit to the Portuguese accident data using a 3-leg of 'intersection' type junction as baseline.

Parameters	mean	s.d.	MC error	2.5%	97.5%
β_0	-6.774	0.403	4.307E-02	-7.609	-6.058
β_1 (ln(AADTmaj))	0.575	0.067	7.989E-03	0.454	0.701
β_2 (ln(AADTmin))	0.119	0.061	7.332E-03	-0.002	0.239
β_3 (Number_of_Legs='4')	0.062	0.103	4.294E-03	-0.135	0.252
β_4 (Junction_Type='roundabout')	-0.216	0.104	3.922E-03	-0.422	-0.014

The Monte Carlo errors show relatively small values indicating that the parameter estimates were calculated with accuracy. Even though the 95% credible interval for the estimate of β_2 contains zero, the observation of the density for this estimated coefficient (see Figure 57) shows that zero falls in the tail of the distribution. Consequently, we can assume that *ln(AADTmin)* is relevant in the model.

The coefficient estimates in Table 22 suggest that an increase in one unit of *ln(AADTmaj)* and *ln(AADTmin)* increases the expected accident frequencies by approximately 78% and

13%, respectively. The expected number of accidents on Portuguese junctions with 4 legs is expected to have around 6% more accidents than a junction with 3 legs while keeping the values of the other variables constant.

Junctions of type 'roundabout' are expected to have around 19% fewer accidents than junctions of type '*intersection*' (provided the remaining variables are kept constant).

The posterior parameter estimates densities are shown in Figure 57. By examination of the graphs it can be concluded that the posterior densities of the coefficient estimates have shifted from the *a priori* distribution which was a Normal with mean zero and variance 10000. The means are now away from zero and the variances are much smaller.

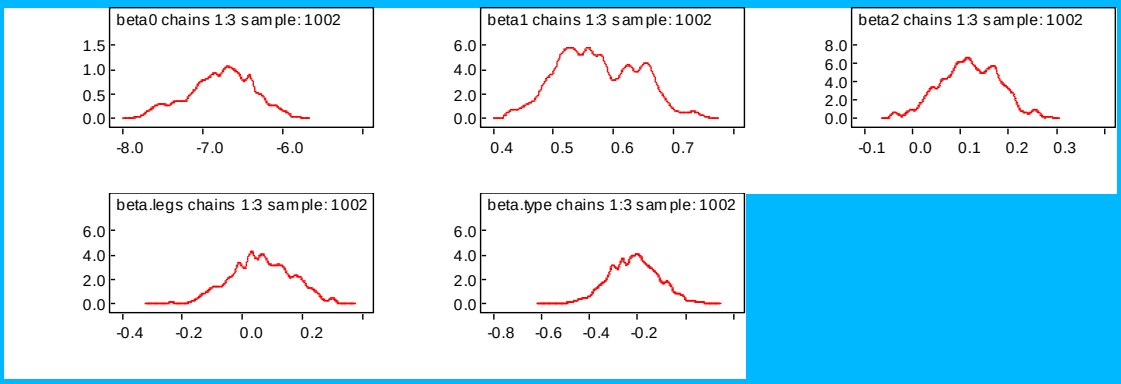

Figure 57 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson regression model was fit to the Portuguese data.

The predicted number of accidents after applying the Poisson regression model (Equation 5.1) to the data is given by the equations shown in Table 23. Regardless of the number of legs, the junctions with type '*roundabout*' have the smaller number of expected injury accidents. Overall, the expected number of accidents is lower at 3 leg junctions when compared to 4 leg junctions.

Table 23 Expected number of accidents per year for Portuguese junctions obtained by a Poisson regression model using a 3-leg 'intersection' type junction as baseline.

	Expected Numbers of Accidents
Number_of_Legs='3'	
Junction_Type 'roundabout' 'intersection'	 $\hat{\mu}_i = 9.206\times10^{-4}\times AADTmaj_i^{0.574}\times AADT\ \min_i^{0.119}$ $\hat{\mu}_i = 1.143\times10^{-3}\times AADTmaj_i^{0.574}\times AADT\ \min_i^{0.119}$
Number_of_Legs='4'	
Junction_Type 'roundabout' 'intersection'	 $\hat{\mu}_i = 1.202\times10^{-3}\times AADTmaj_i^{0.574}\times AADT\ \min_i^{0.119}$ $\hat{\mu}_i = 1.216\times10^{-3}\times AADTmaj_i^{0.574}\times AADT\ \min_i^{0.119}$

5.2.1 Model Checking

The histograms of the data replicated by the Poisson regression model (Figure 58) show that in most cases the model failed to replicate the higher frequencies of zero accidents as well

as some few junctions with higher frequencies of accidents (as the histogram of the observed data shows).

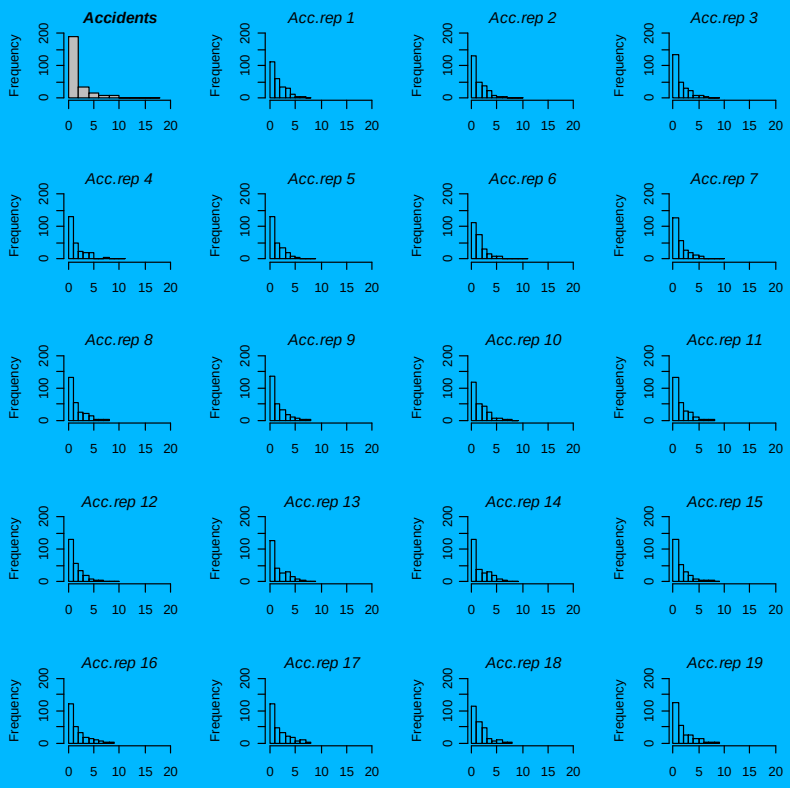

Figure 58 Histogram of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.

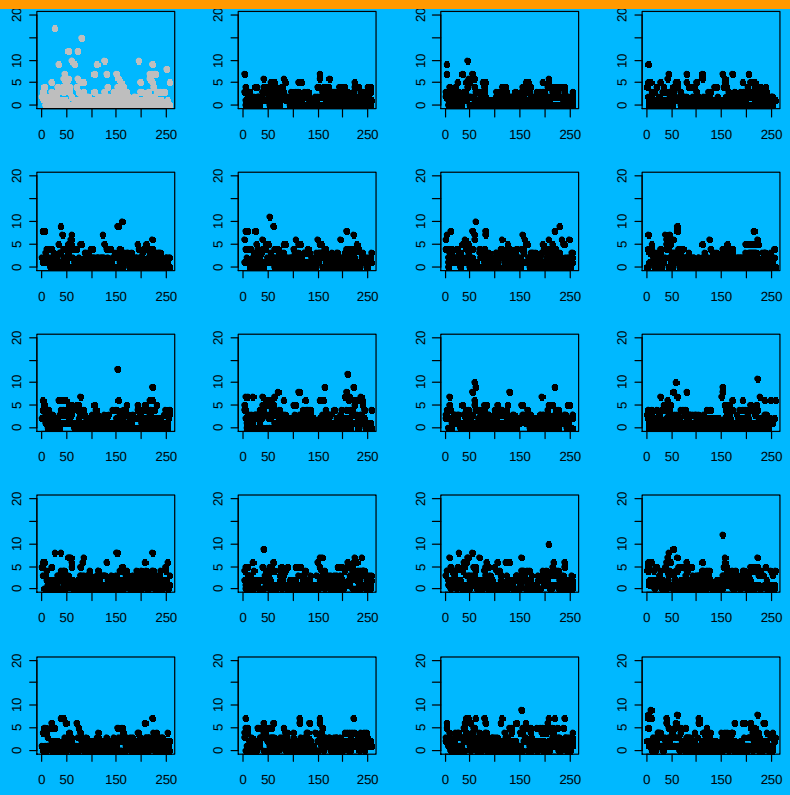

Figure 59 Dot plot of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.

It is evident from the dot plots that the number of modelled accidents (see Figure 59) remains constant throughout the junction set. This does not happen with the observed data where there are a few junctions which have higher number of injury accidents than the others. It can be concluded that the Poisson regression model does not seem to replicate the data properly.

The four discrepancy measures plotted in Figure 60, together with the probabilities estimate that the replicated discrepancy is greater than the observed discrepancy, shows that the Poisson model does not seem to be able to capture the maximum value and the overall standard deviation of the data (*p*-values of zero or near zero).

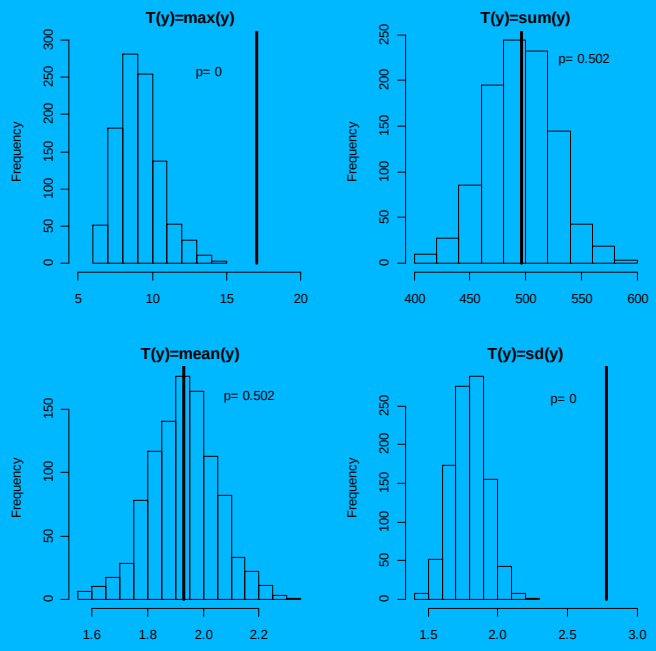

Figure 60 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson regression model fitted to the Portuguese data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The measure of the discrepancy used to check whether the model is taking the data's overdispersion into account is displayed in Figure 61. The estimated probability that the ratio of the variance to the mean in the replicated data is greater than the same ratio calculated from the observed data is equal to zero, strongly indicating that there is overdispersion present in the data. The Poisson regression model is not able to detect this and therefore cannot accurately replicate the observed data. Consequently, it leads to the conclusion that the Poisson model does not seem appropriate to represent the Portuguese junction data.

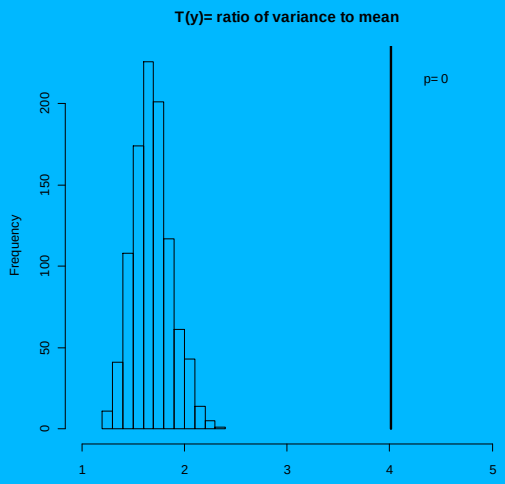

Figure 61 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson regression model fitted to the Portuguese data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

This model produces an average deviance, $\overline{D}$, of 1115.730 and an effective model dimension, d_e, of 4.527, giving a DIC of 1120.260. These results will be used for model comparison which is described in section 5.5.

5.3 The Poisson-Gamma hierarchical regression model

The Poisson-Gamma regression model was fitted using the Equations 2.3 and 2.4 with $\delta = \gamma$ and δ ~Gamma(a,a) where a=0.1. Equation 5.1 was applied and the β parameters were given *a priori* Normal distributions with mean zero and variance 10000.

The MCMC algorithm comprised three chains and was run for 35000 iterations with 33000 burn-in and resulting in a thinning rate equal to 6. The results were obtained from samples with size 1002.

The graphs of the Gelman-Rubin statistics for the estimates of the model parameter estimates are displayed in Figure 62 and in all of them it can be observed that the Rhat statistic (red line) converges to the value 1 and the remaining ones stabilize around 1. This indicates that there seems to be no reasons to doubt that non-convergence has occurred.

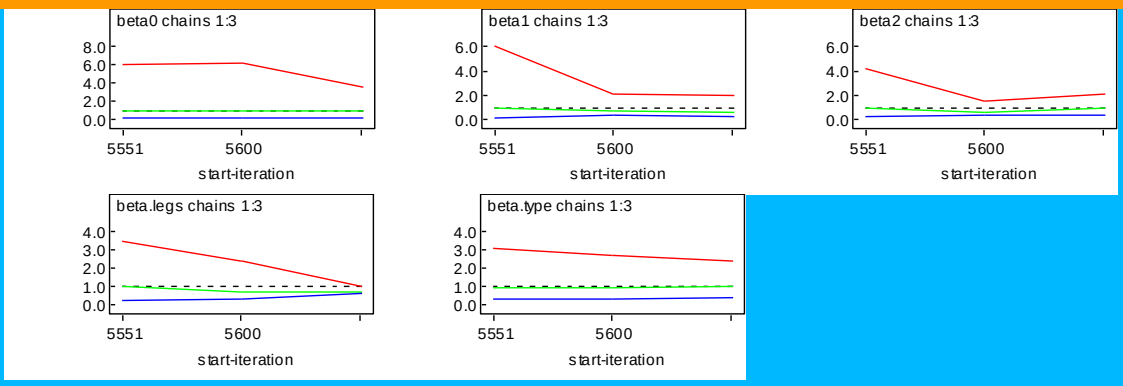

Figure 62 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson-Gamma regression model fitted to the Portuguese junction accident data.

Apart from the more dubious case of the estimate of the β_2 coefficient, there seems to be no reason to doubt the non-convergence of the algorithm.

The point estimates and further statistics are displayed in Table 24. The credible interval for the estimates of β_2, β_3 and β_4 contain zero but fall within the tail of the distribution and the acceptable limits as can be seen in Figure 63.

Table 24 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fit to the Portuguese accident data using a 3-leg 'intersection' type junction as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-5.917	0.395	5.409E-02	-6.373	-5.046
β_1 (ln(AADTmaj))	0.558	0.083	1.127E-02	0.420	0.697
β_2 (ln(AADTmin))	0.022	0.083	1.130E-02	-0.123	0.170
β_3 (Number_of_Legs='4')	0.255	0.164	2.197E-02	-0.092	0.543
β_4 (Junction_Type=´roundabout')	-0.296	0.194	2.606E-02	-0.604	0.057

From the examination of the point mean estimates it can be stated that, according to this particular Poisson-Gamma model, every unitary increase in *ln(AADTmaj)* increases the expected frequency of accidents in approximately 75% (and assuming that all the other explanatory variables remain constant). An increase in one unit of *ln(AADTmin)* increases the expected frequency of accidents in around only 2%. See example on section 3.2.

Junctions with four legs have approximately 29% higher expected number of injury accidents when compared with junctions with three legs (provided the remaining variables have been kept constant). Junctions of type '*roundabout*' have approximately 26% lower expected accident frequencies when compared with junctions of type '*intersection*'.

The posterior densities of the β estimated coefficient parameters are displayed in Figure 63. The densities have deviated considerably from the Normal *a priori* distribution.

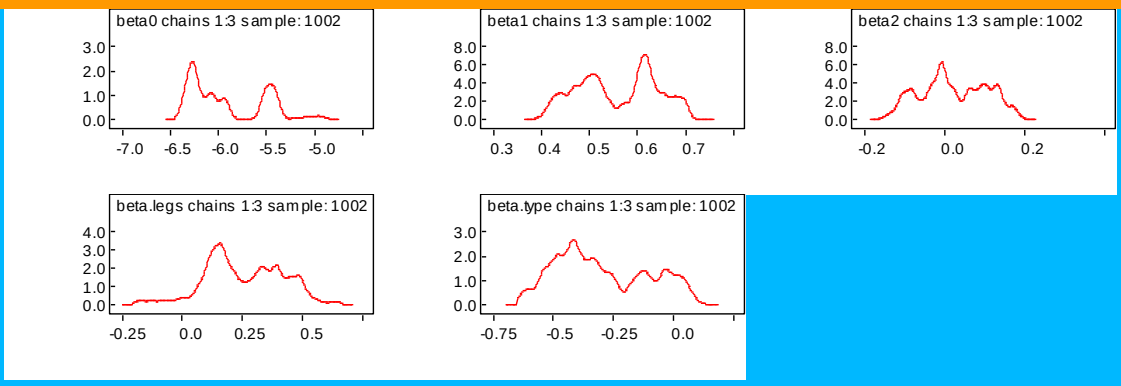

Figure 63 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson-Gamma regression model was fit to the Portuguese data.

The expected accident frequencies for a one year period are given by equations displayed in Table 25 for the various numbers of legs and junction's types.

Table 25 Expected number of accidents per year for Portuguese junctions obtained by a Poisson-Gamma regression model using a 3-leg 'intersection' type junction as baseline.

	Expected Numbers of Accidents
Number_of_Legs='3'	
Junction_Type 'roundabout' 'intersection'	$\hat{\mu}_i = 2.003\times10^{-3}\times AADTmaj_i^{0.558}\times AADT\ \min_i^{0.022}$ $\hat{\mu}_i = 2.694\times10^{-3}\times AADTmaj_i^{0.558}\times AADT\ \min_i^{0.022}$
Number_of_Legs='4'	
Junction_Type 'roundabout' 'intersection'	$\hat{\mu}_i = 2.586\times10^{-3}\times AADTmaj_i^{0.558}\times AADT\ \min_i^{0.022}$ $\hat{\mu}_i = 3.476\times10^{-3}\times AADTmaj_i^{0.558}\times AADT\ \min_i^{0.022}$

In general, the model detected that the junctions with 4 legs have a slightly higher expected injury accident frequency than junctions with 3 legs. Regardless of the number of legs, the junctions with type '*intersection*' have higher numbers of expected accidents than junctions with type '*roundabout*'. In all cases 3-leg junctions have fewer predicted accidents than 4 leg junctions.

5.3.1 Model Checking

The graphs of nineteen replicated data sets are depicted as histogram in Figure 64 and, for a further set of 19 replicated data in Figure 65, as dot plots. They all show that the observed data looks plausible amongst the replicated data (see Gelman *et al*., 2004) and that the Poisson-Gamma model seems to be able to adequately replicate most of the features of the observed data.

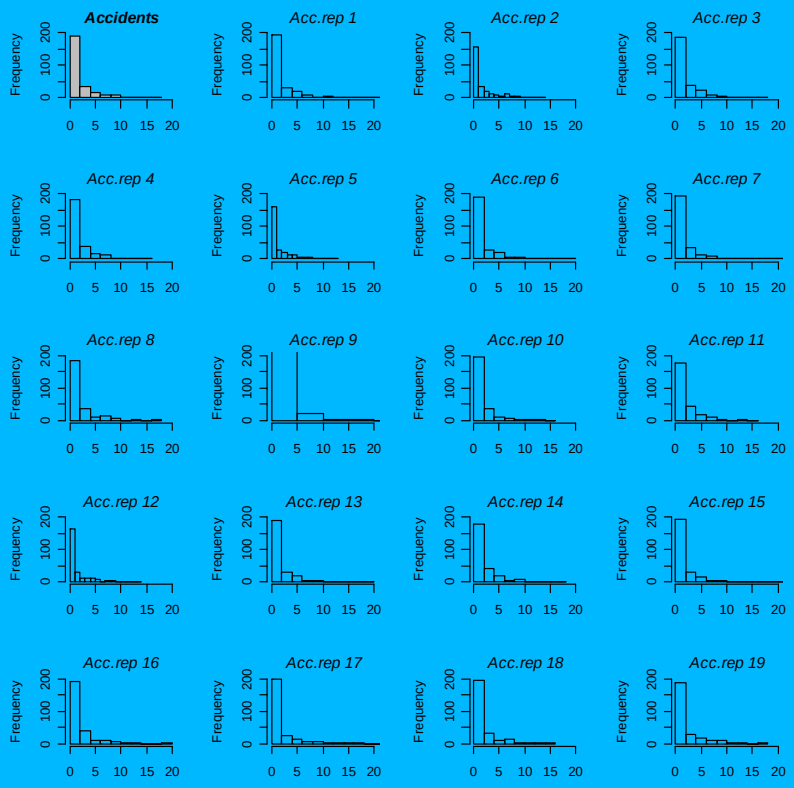

Figure 64 Histogram of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.

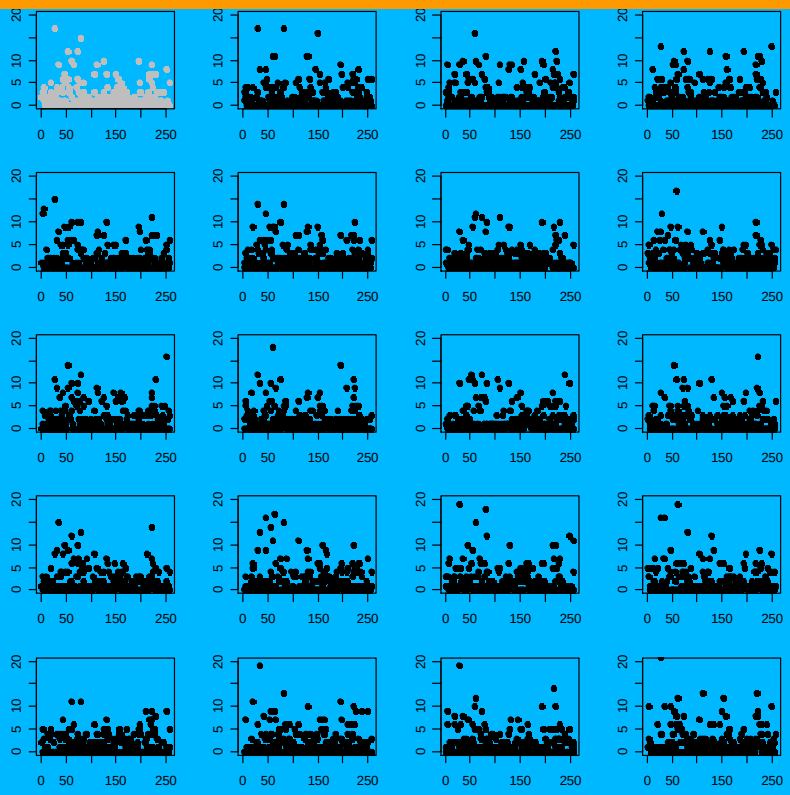

Figure 65 Dot plot of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.

The discrepancy measures *max*, *sum*, *mean* and *sd* calculated from the observed data and replicated data are displayed in Figure 66. It can be seen that the observed values lie inside the histograms of the replicated data and that the corresponding probabilities are all close to 0.5 which indicates that the Poisson-Gamma model is replicating the data properly and capturing the variations that were calculated. An equivalent conclusion can be drawn when considering the ratios of the variances over the mean to check whether the model takes into account the data overdispersion with a *p*-value of 0.503 (see Figure 67).

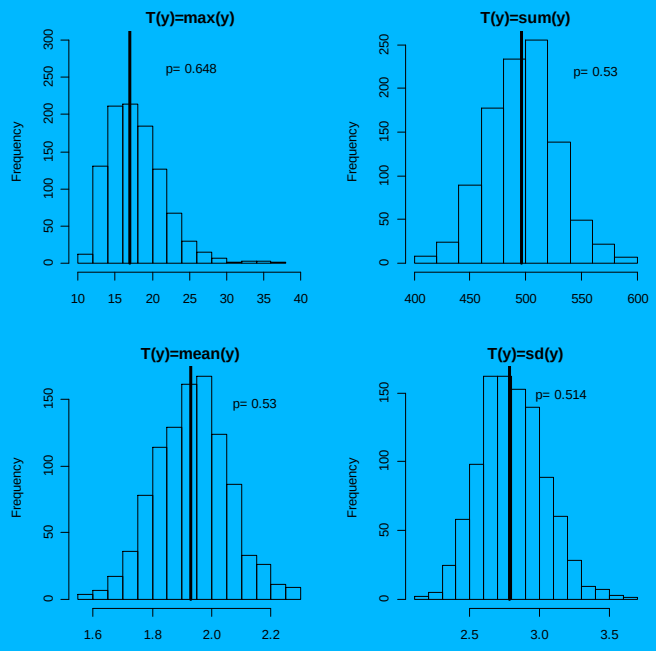

Figure 66 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fit to the Portuguese data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

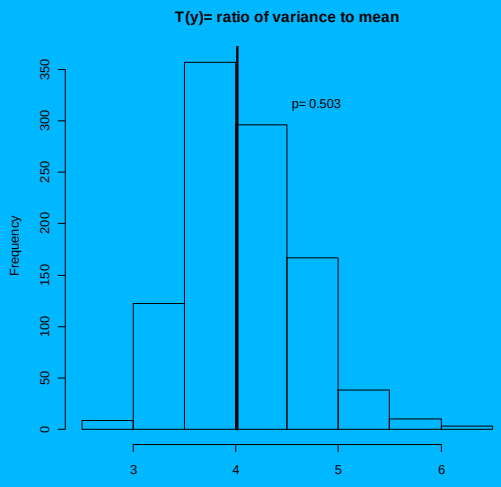

Figure 67| Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fit to the Portuguese data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

This model produces an average deviance, $\bar{D}$, of 672.284 and an effective model dimension, d_e, of 113.426, giving a DIC of 785.710. The dispersion parameter ($1/\delta$ in Equation 2.4) was estimated as 1.155. This suggests that the Poisson-Gamma model adequately replicates the Portuguese junction data.

The posterior means and corresponding standard deviations for the expected numbers of accidents for a one year period are given in Table 26 for the minimum, mean, median and maximum profiles, and for the various types of junctions and number of legs.

Table 26 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the Portuguese accident data.

		Minimum	Mean	Median	Maximum
	ln(AADTmaj)	6.611	8.370	8.283	10.223
	ln(AADTmin)	5.493	7.506	7.382	9.871
Junction_Type	Number_of_Legs	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
roundabout	3	0.094 (0.025)	0.258 (0.050)	0.246 (0.049)	0.763 (0.134)
	4	0.120 (0.025)	0.330 (0.045)	0.314 (0.043)	0.980 (0.139)
intersection	3	0.123 (0.014)	0.343 (0.033)	0.326 (0.031)	1.027 (0.181)
	4	0.160 (0.026)	0.447 (0.078)	0.425 (0.074)	1.343 (0.318)

For a typical 4 leg junction with type '*intersection*' one expects 0.447 accidents in one year, while a typical '*roundabout*' junction with three legs is expected to have 0.258 accidents per year.

The worst case scenario (corresponding to the maximum profile) where junctions have *AADTmaj* and *AADTmin* of approximately 27529 and 19360, respectively, corresponds to an expected number of 1.343 accidents for junctions of type *'intersection'* and four legs.

5.4 The Poisson Log-Normal regression model

The Poisson Log-Normal regression model was fitted to the data according to Equations 2.5 and 2.6 (in Chapter 2) where the parameter α in Equation 2.6 had an *a priori* Gamma distribution with parameter *a* equal to 0.001.

The MCMC algorithm was run with three chains for 35000 iterations of which 33000 were burn-in and a thinning rate equal to 6. The results and conclusions were drawn from a sample with dimension 1002.

The Gelman-Rubin statistics were plotted for the estimates of the regression coefficients and are displayed in Figure 68. From the observation of those plots it can be stated that there are no reasons to believe that the algorithm does not converge, as for most of the parameter estimates considered the *Rhat* statistic converges to 1 and the remaining statistics stabilise as the number of iterations increase.

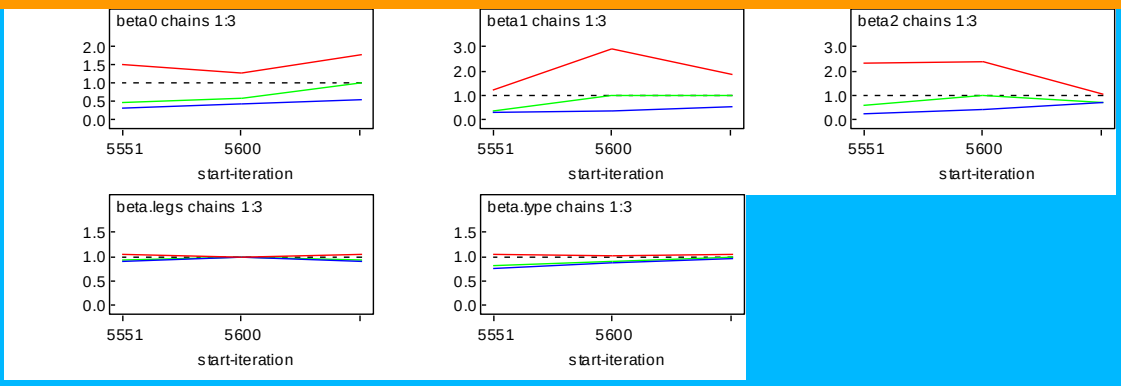

Figure 68 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters from the Poisson Log-Normal regression model fitted to the Portuguese junction accident data.

The posterior means of the parameter estimates, corresponding standard deviations, Monte Carlo errors and 95% credible intervals are shown in Table 27.

From the values in Table 27 it can be stated that every unitary increase in *ln(AADTmaj)* increases the expected number of accidents by approximately 84%, provided the other explanatory variables remain constant. An increase in one unit of *ln(AADTmin)* increases the expected frequency of injury accidents by around 15%. See example on section 3.2.

A junction with four legs is *a posteriori* expected to have approximately 9% more accidents than a junction with three legs (with the remaining explanatory variables constant).

A junction of type '*roundabout*' is *a posteriori* expected to have approximately less 26% injury accidents than a junction of type '*intersection*'.

Table 27 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the Portuguese accident data using a 3-leg of 'intersection' type junction as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-7.723	1.291	1.724E-01	-9.843	-4.673
β_1 (ln(AADTmaj))	0.609	0.202	2.726E-02	0.172	0.892
β_2 (ln(AADTmin))	0.142	0.148	1.998E-02	-0.075	0.474
β_3 (Number_of_Legs='4')	0.083	0.212	1.304E-02	-0.365	0.469
β_4 (Junction_Type=´roundabout')	-0.302	0.221	1.432E-02	-0.769	0.119

The MC errors have comparatively smaller values than the standard deviations thus indicating that the parameter estimates were calculated with accuracy.

The 95% credible intervals for some parameter estimates contain zero, however, from the observation of the parameter estimate densities in Figure 69 it is evident that zero lies on the tails of the densities (i.e. quite away from the mean), with the exception of the estimate corresponding to β_3 (coefficient for variable *Number_of_Legs*='4') which may suggest that this categorical variable is not relevant in the model. However, it was decided to keep this variable so that comparisons between models for different countries could be made.

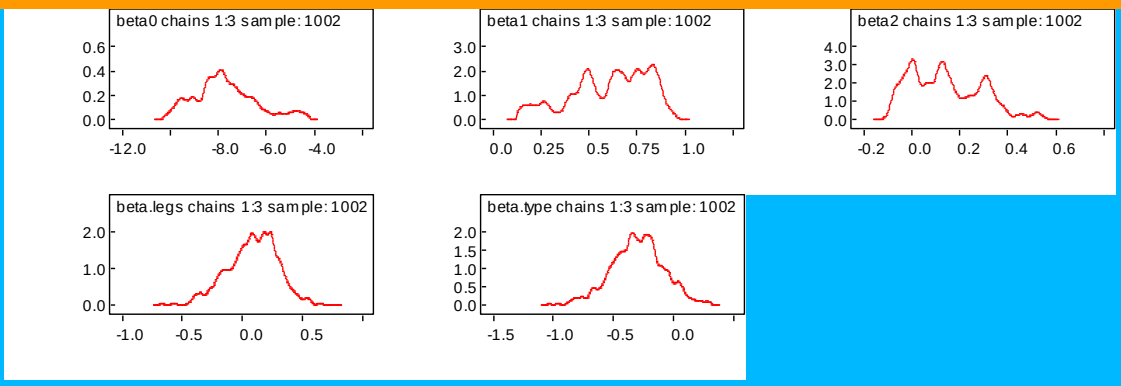

Figure 69 Posterior densities of the coefficients corresponding to the beta parameters obtained after a Poisson Log-Normal regression model was fit to the Portuguese data.

The equations giving the expected injury accident frequency, per year and per number of legs and junction type, are given in Table 28.

Table 28 Expected number of accidents per year for Portuguese junctions obtained by a Poisson Log-Normal regression model using a 3-leg of 'intersection' type junction as baseline.

	Expected Numbers of Accidents
Number_of_Legs='3'	
Junction_Type 'roundabout' 'intersection'	$\hat{\mu}_i = 3.274\times10^{-4}\times AADTmaj_i^{0.609}\times AADT\ \min_i^{0.142}$ $\hat{\mu}_i = 4.427\times10^{-4}\times AADTmaj_i^{0.609}\times AADT\ \min_i^{0.142}$
Number_of_Legs='4'	
Junction_Type 'roundabout' 'intersection'	$\hat{\mu}_i = 3.559\times10^{-4}\times AADTmaj_i^{0.609}\times AADT\ \min_i^{0.142}$ $\hat{\mu}_i = 4.812\times10^{-4}\times AADTmaj_i^{0.609}\times AADT\ \min_i^{0.142}$

Overall, the junctions with 3 legs have only a slightly lower expected accident frequency than junctions with 4 legs. Regardless of the number of legs, junctions of type '*intersection*' have higher expected accident frequency than junctions of type '*roundabout*'.

5.4.1 Model Checking

The observation of both Figure 70 and Figure 71 show that the various histograms and dot plots of replicated data resemble the observed data and hence it can be stated that the Poisson Log-Normal model seems to be able to adequately replicate the observed data.

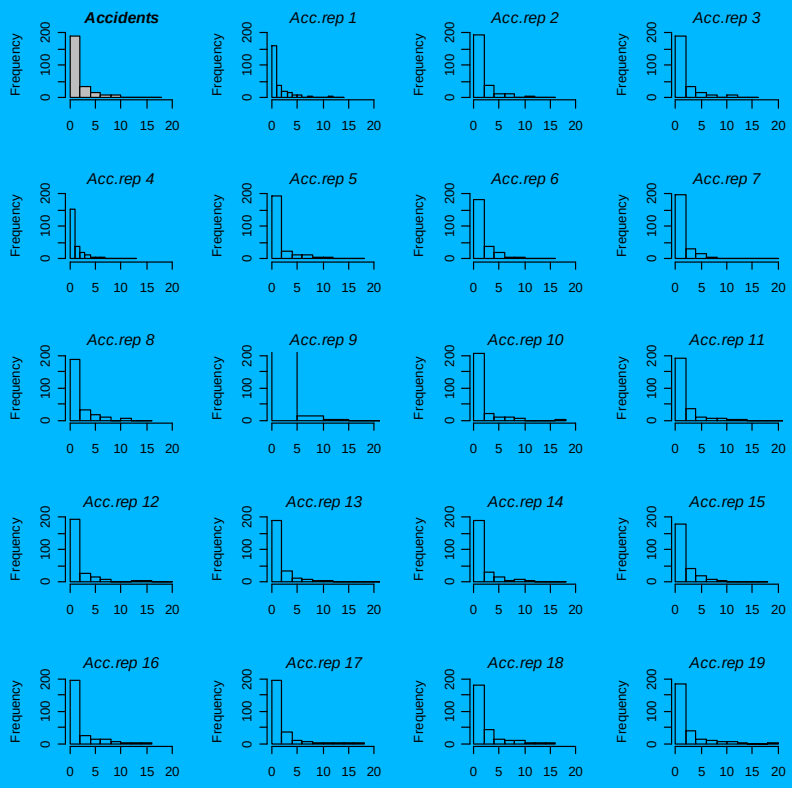

Figure 70 Histogram of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

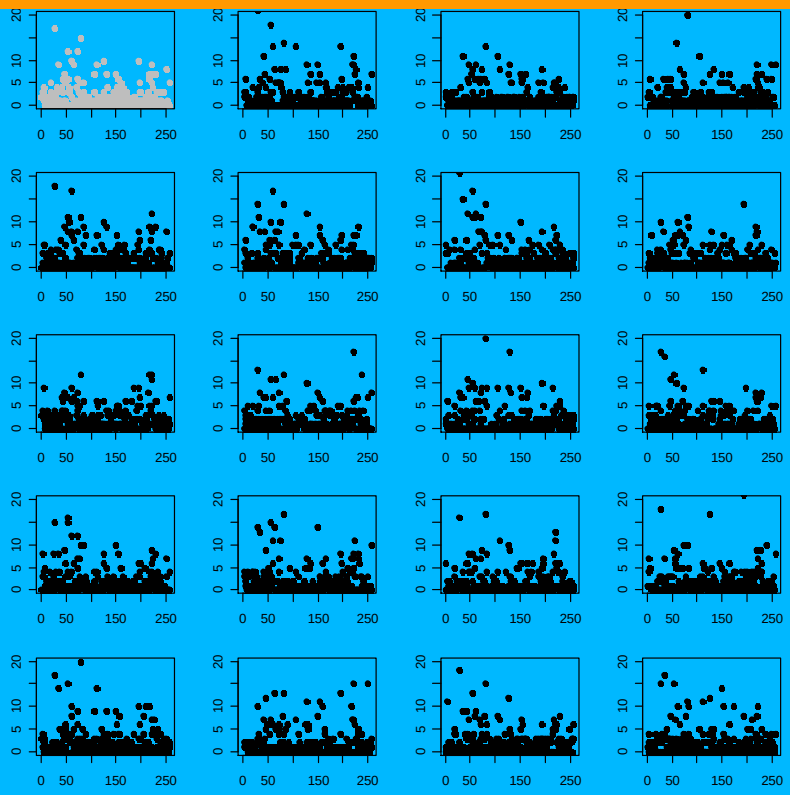

Figure 71 Dot plot of the observed number of accidents in Portuguese junctions (left upper corner, in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

The Poisson Log-Normal model also seems to be able to capture variations in the Portuguese junction data as exemplified by the four discrepancy measures whose values are displayed in Figure 72. In all of the four discrepancy measures considered, the histograms of the frequencies of the measured discrepancies in various sets of replicated data surround the observed value (represented by a vertical line in the plots). The probability that the measures obtained with the replicated data are greater than the observed measure is near 0.5 for nearly all of the four discrepancy measures, therefore indicating that the model in question captures these variations of the observed data reasonably well.

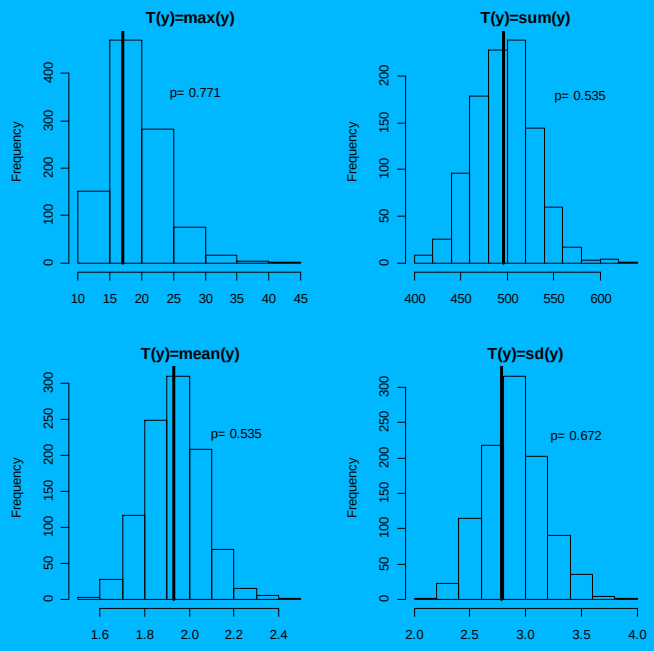

Figure 72 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fit to the Portuguese data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The plot depicted in Figure 73 shows that the Poisson log-Normal regression model applied takes the overdispersion of the data into account by being able to replicate it.

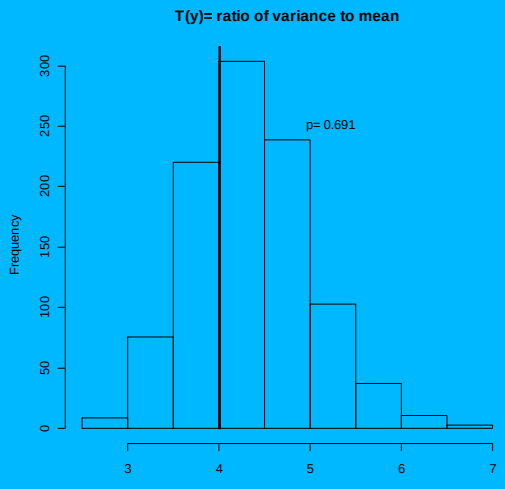

Figure 73 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fit to the Portuguese data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

This model produces an average deviance, $\bar{D}$, of 691.848 and an effective model dimension, d_e, of 131.586, giving a DIC of 823.434. The dispersion parameter ($1/\alpha$ in Equation 2.6) was estimated as 1.091. These quantities are used for model comparison, which is discussed in more detail in section 5.5.

The expected accident frequency for the three types of junctions and for three and four approaches, for the minimum, mean, median and maximum profiles of the explanatory variables are displayed in Table 29. For a typical junction with type '*roundabout*' and 3 approaches the predicted accident frequency for a one year period is 0.160 injury accidents. For a junction with 4 legs (*intersection)* 0.233 accidents per year are predicted.

Table 29 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the Portuguese accident data.

		Minimum	Mean	Median	Maximum
	ln(AADTmaj)	6.611	8.370	8.283	10.223
	ln(AADTmin)	5.493	7.506	7.382	9.871
Junction_Type	Number_of_Legs	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
roundabout	3	0.044 (0.020)	0.160 (0.035)	0.149 (0.034)	0.702 (0.211)
	4	0.048 (0.023)	0.174 (0.040)	0.162 (0.039)	0.756 (0.198)
intersection	3	0.057 (0.019)	0.212 (0.027)	0.198 (0.026)	0.957 (0.306)
	4	0.063 (0.023)	0.233 (0.046)	0.218 (0.044)	1.042 (0.336)

The worst case scenario (corresponding to the maximum profile) is obtained for the maximum values of *ln(AADTmaj)* and *ln(AADTmin)* and corresponds to an expected number of injury accidents of 1.042 in a one year period in junctions of type '*intersection*' with 4 legs.

5.5 Discussion

Table 30 contains the resulting fit ($\bar{D}$), complexity (d_e) and overall model choice (DIC) score for the three models fitted to the Portuguese data.

Table 30 Comparison of DIC and related statistics for the three models fitted to the Portuguese junction data.

Regression Model	$\bar{D}$	d_e	DIC
Poisson	1115.730	4.527	1120.260
Poisson-Gamma	672.289	113.426	785.710
Poisson Log-Normal	691.848	131.586	823.434

The Poisson regression model was already seen not to be appropriate to model the

Portuguese junction data as it does not replicate conveniently the observed data and does not take the overdispersion into account.

The Poisson-Gamma emerges as the model with the lower DIC value (see value in Table 30). It has also the smaller d_e (effective parameter dimension) when comparing with the Poisson Log-Normal model. Since the Poisson-Gamma model obtained also the best results when checking for the data discrepancy measures it was decided to present its results as the overall conclusions of the Portuguese junction's analysis.

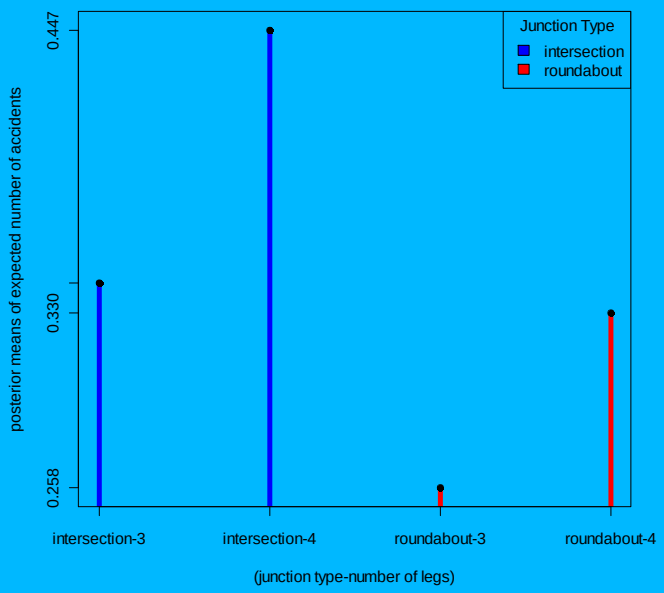

Figure 74 Values of the posterior means of the expected number of accidents for Portuguese junctions classified per junction type and number of legs (as under column *Mean* in Table 26).

Therefore, from the observation of the plot displayed in Figure 74 (whose values were taken from the fourth column in Table 26) it can be stated that the highest value of the expected numbers of injury accidents are obtained on 4 leg *intersection type* junctions. The lowest value is obtained on 3 leg *roundabout* junctions. This result is the opposite of the one obtained with the Austrian data, being the former a more expected result and in line with international research findings. The equations giving the number of expected accident frequencies for the Poisson-Gamma model are displayed in Table 25.

6 Modelling Austrian, Norwegian and Portuguese injury accidents

This chapter describes the combined analysis of the set of data formed after joining the junction data sets from the three countries (Austria, Norway and Portugal) described in previous chapters. The junctions analysed and described in chapters 3, 4 and 5 were joined together with their corresponding number of accidents and *AADT* values and described in section 6.1.

The assessment of the results obtained for the Poisson, Poisson-Gamma and Poisson Log-Normal regression models which were fitted to the new set of aggregated data is described in this chapter in section 6.2, 6.3 and 6.4. Section 6.5 summarises the results and

conclusions obtained.

6.1 Aggregated Junction Data

The aggregated junction data was formed by joining the rural junction data sets from Austria, Norway and Portugal and the new aggregated data is formed by 1208 junctions. The variables that could be aggregated from all three countries included the following:

- ***Accidents***: gives the number of injury accidents;
- ***AADTmaj***: represents the major entering volume traffic (annual average daily traffic);
- ***AADTmin***: represents the minor entering volume traffic (annual average daily traffic).

The plots shown in Figure 75 depict the number of accidents per junction and per country (left panel in Figure 75 where the three colours represent the three countries) and the frequency of the total number of accidents (panel on the right).

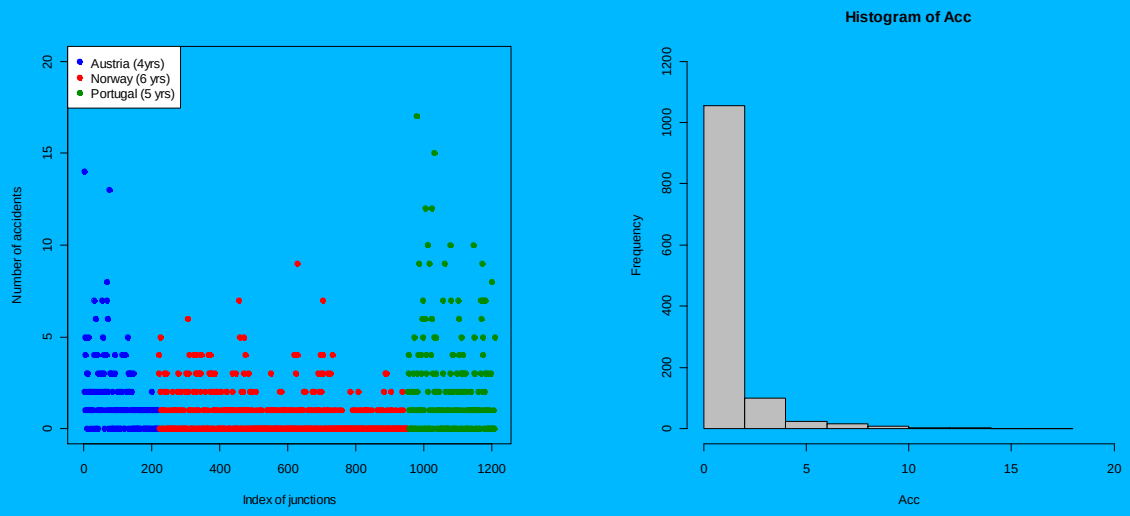

Figure 75 Plot of *Accidents*, per junction, registered in the aggregated rural road network junctions (on the left) and the histogram giving the frequency of the number of accidents (on the right).

Note that the data were measured on different time periods for each country but are now aggregated. However, the analysis described in this chapter takes the period of time of the measurements into account so that the results correspond to a one year period.

The plots shown in Figure 76 represent the graphs of *AADTmaj* and *AADTmin* plotted against the number of accidents on the left and right, respectively. It is evident to notice that some junctions have higher values of either AADTmaj or AADTmin, which has no doubtly influenced the fitted smooth regression shown on the plots. However, since the aim of these smooth regressions was to help in visualising these particular graphs it was decided not to pursue further investigations on those junctions.

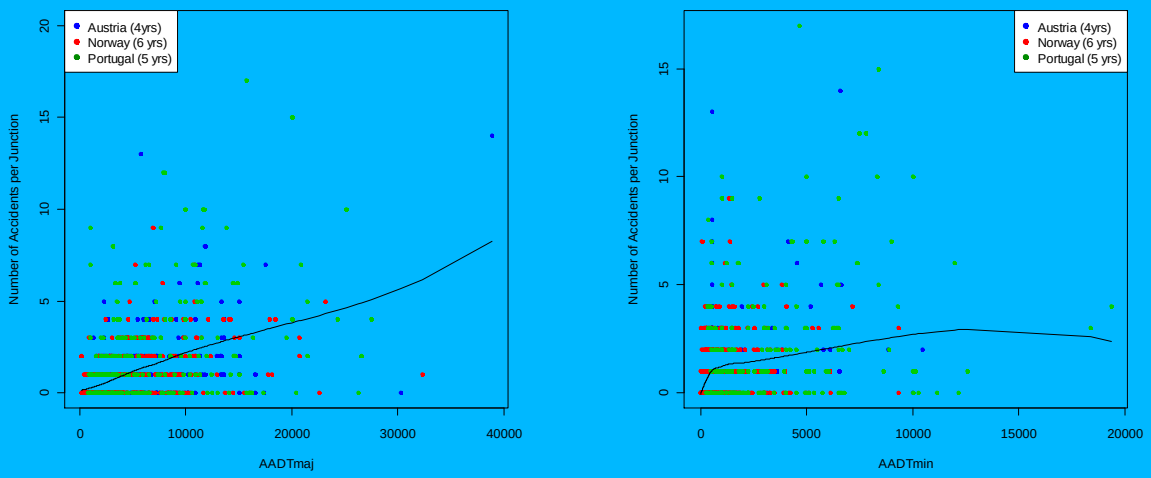

Figure 76 The number of accidents on the aggregated data set against *AADTmaj* and *AADTmin*, on the left and right panels, respectively, and corresponding polynomial fits.

The fitted smooth regression curve increases with increasing *AADTmaj* but tends to stabilise as *AADTmin* increases. The great majority of accidents occur for values of *AADTmaj* between 133 and 15000 and 7 to 7000 for *AADTmin*.

Table 31 displays summary descriptive statistics for the variables belonging to the aggregated data set.

Table 31 Summary statistics for the variables registered on the aggregated junctions.

Variables	minimum	mean	standard deviation.	median	maximum
AADTmaj	133	4778.19	4530.90	3293	38875
AADTmin	7	1197.93	1946.28	560	19359
Accidents	0	1.029	1.825	0	17

6.2 The Poisson regression model

The Poisson regression model given by Equations 2.1 and 2.2 in chapter 2 was fitted to the data with *Accidents* as the dependent variable and the natural logarithms of *AADTmaj* and *AADTmin* as independent variables as is shown in Equation 6.1.

$$ln(\hat{\mu}_i) = \beta_0 + \beta_1 ln(AADTmaj_i) + \beta_2 ln(AADTmin_i) \quad (6.1)$$

The β parameters were assigned Normal *a priori* distributions with mean zero and variance 10000. The MCMC algorithm consisted on three chains and was run for 9000 iterations with 6000 as burn-in and a thinning rate equal to 9. The results are obtained from samples with a dimension of 1002.

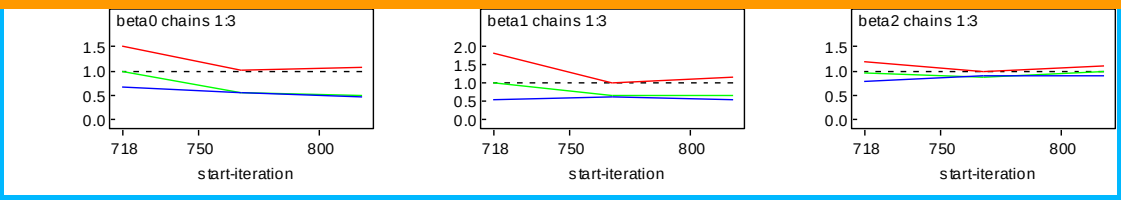

Figure 77 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters obtained after the Poisson regression model was fit to the aggregated data set.

From observing the plots of the Gelman-Rubin statistics displayed in the graphs in Figure 77, it is reasonable to conclude that there are no reasons to doubt the non-convergence of the MCMC algorithm.

The point estimates of the Poisson regression model are given in Table 32. The 95% credible intervals for the estimates of the non-categorical independent variables do not include zero, therefore indicating that the variables *ln(AADTmaj)* and *ln(AADTmin*) have a relevant effect when predicting the number of accidents.

Table 32 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson regression model was fit to the aggregated data set.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-9.165	0.312	3.453E-02	-9.785	-8.577
β_1 (ln(AADTmaj))	0.651	0.041	4.553E-03	0.580	0.734
β_2 (ln(AADTmin))	0.305	0.024	2.026E-03	0.258	0.355

The posterior densities of the parameter estimates are given in Figure 78. The mean values have shifted away from the *a priori* mean value of zero for the three estimates.

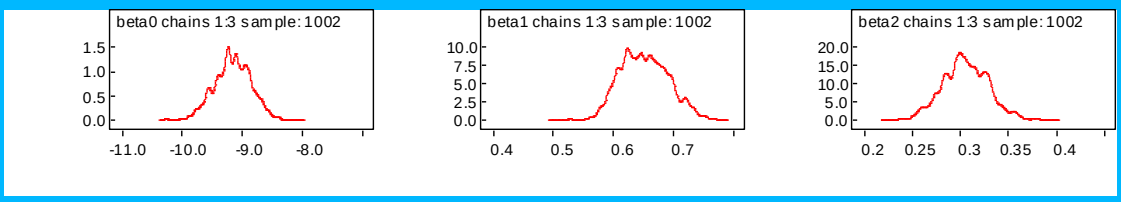

Figure 78 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson regression model was fit to the aggregated junction data.

The equation for the expected number of accidents for a one year period in the aggregated data set is given by:

$$\hat{\mu} = 1.046 \times 10^{-4} \times AADTmaj^{0.651} \times AADT\ \min^{0.305} \qquad (6.2)$$

Every increase in a unit of *ln(AADTmaj)* increases the expected number of accidents by 92% (provided the *ln(AADTmin)* is constant), whereas an increase in a unit of *ln(AADTmin*) increases the expected number of accidents by approximately 36%. For the meaning of unitary increase see example in section 3.2.

6.2.1 Model Checking

Figure 79 contains the histogram of the observed data (in the upper left panel) and nineteen histograms of data replicated by the Poisson model obtained. From the observation of these histograms it can be seen that the replicated data does not assume high values for the accidents, i.e., values between 10 and 17 (which were present in the observed data and are represented by a thin horizontal line in the grey histogram).

This fact is better observed in the plots displayed in Figure 80. The observed data (upper left plot in grey) has considerable higher numbers of accidents in the first and last group of variables, which the Poisson model does not seem to be replicating.

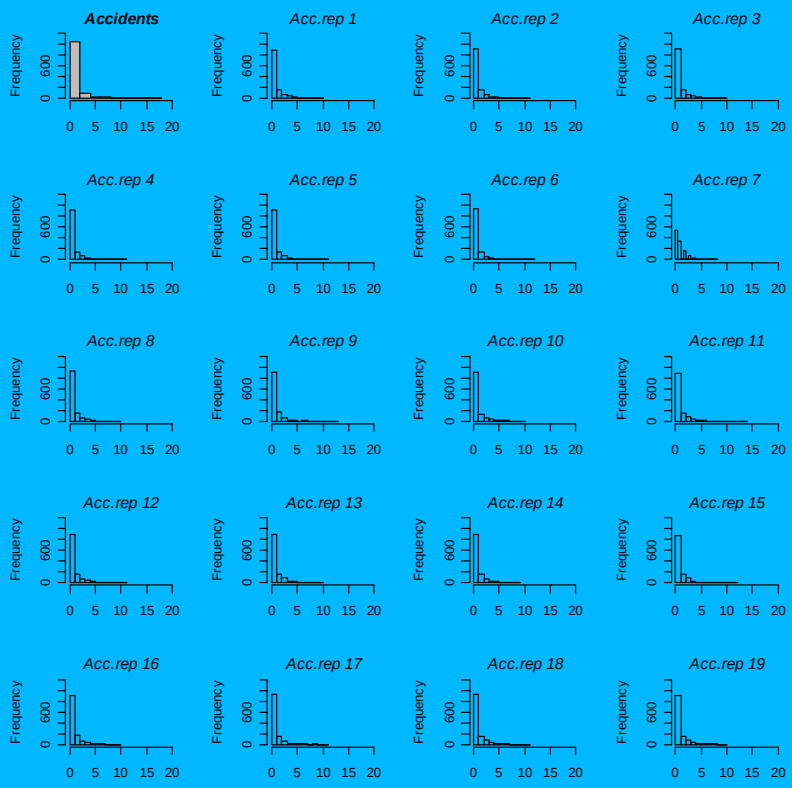

Figure 79 Histogram of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 histograms of replicated data set ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.

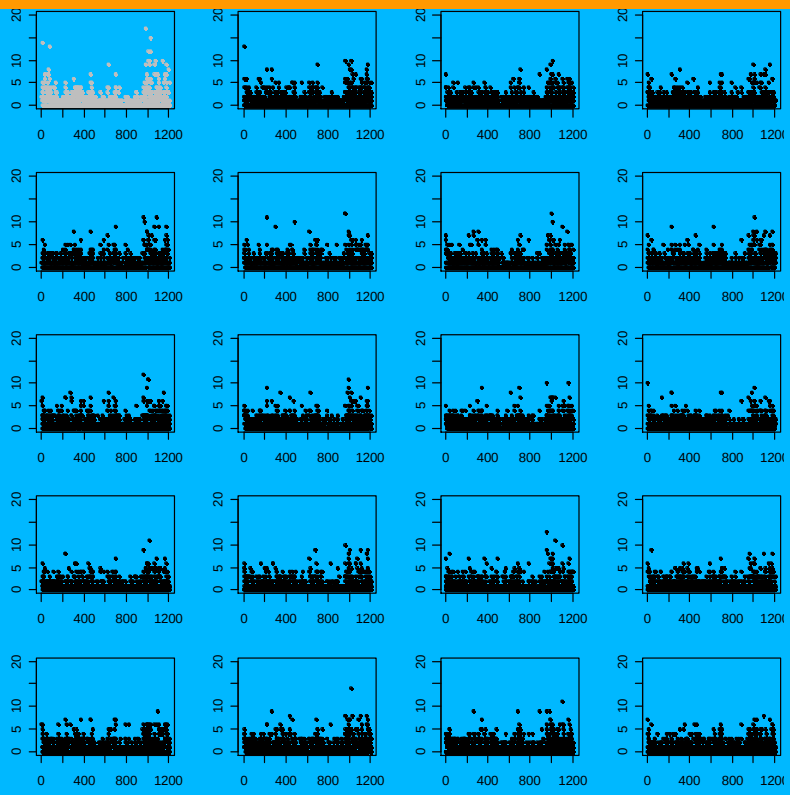

Figure 80 Dot plot of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson regression model.

The number of accidents replicated seemed to vary in a constant way in the 0 to 10 range (*y*-axis).

The four plots of the discrepancy measures are displayed in Figure 81. It can be seen that the Poisson regression model seems to capture the variations corresponding to the *sum* and *mean* values (with estimated probabilities of 0.5), but does not capture the *maximum* and *standard deviation* values of the observed data.

The measure of discrepancy suggested by Congdon (2005) to check whether overdispersion is taken into account is displayed in Figure 82.

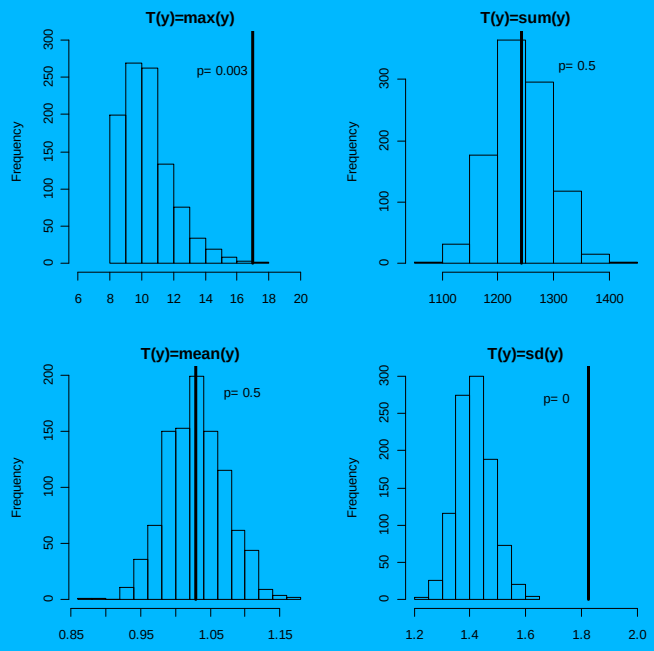

Figure 81 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson regression model fitted to the aggregated data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The Poisson model does not seem to be taking the overdispersion of the observed data into account as the ratio of the variance to the mean in the observed data (given by the straight line) is considerably greater than any of the ratios obtained by the 1002 replicated sets of data (given by the histogram) with an estimated probability equal to zero.

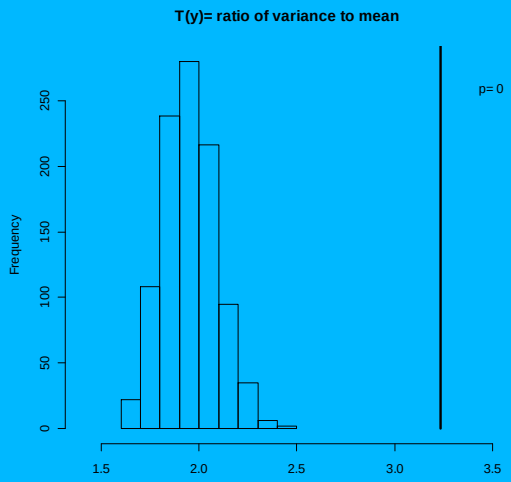

Figure 82 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson regression model fitted to the aggregated data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

Consequently, the Poisson regression model does not seem to be adequate to model the aggregated junction data.

This model produces an average deviance, $\overline{D}$, of 3340.570 and an effective model dimension, d_e, of 3.059, giving a DIC of 3343.630.

6.3 The Poisson-Gamma hierarchical regression model

The Poisson-Gamma hierarchical regression model was fit to the aggregated data set according to Equations 2.3 and 2.4 in chapter 2. The parameters γ and δ were such that $\delta = \gamma$ and $\delta \sim Gamma(a,b)$ with a=1 and b=0.01. The expression given by Equation 6.1 was applied and the β parameters were given *a priori* Normal distributions with mean 0 and precision 0.0001 (variance is equal to the inverse of the precision).

The MCMC algorithm comprised 3 chains and was run for 35000 iterations of which 20000 were burn-in and a thinning rate of 45 resulting in samples with dimension 1002.

The Gelman-Rubin graphs in Figure 83 show evidence to believe the convergence of the MCMC algorithm.

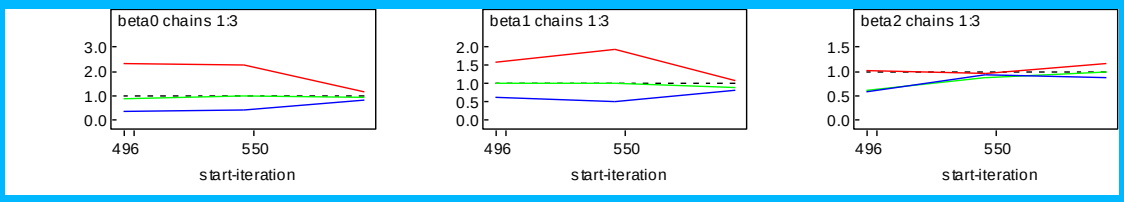

Figure 83 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters obtained after the Poisson-Gamma hierarchical regression model was fit to the aggregated data set.

The point estimates for the Poisson-Gamma fit are given in Table 33. The impact of both independent variables is significant as the 95% credible intervals do not contain the value zero.

Table 33 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fit to the aggregated data set.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-9.891	0.406	5.406E-02	-10.710	-9.218
β_1 (ln(AADTmaj))	0.689	0.051	6.436E-03	0.586	0.777
β_2 (ln(AADTmin))	0.369	0.037	3.678E-03	0.305	0.450

The posterior densities for the β coefficient estimates are displayed in Figure 84.

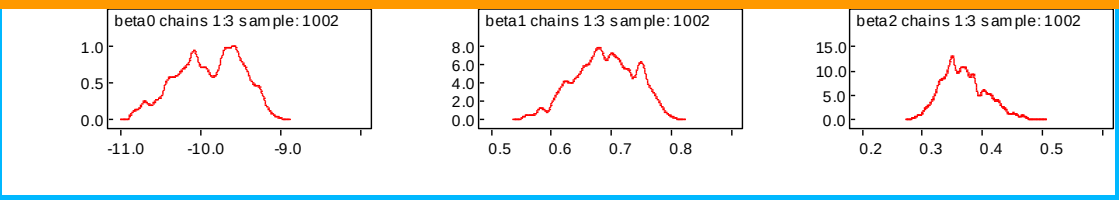

Figure 84 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson-Gamma hierarchical regression model was fit to the aggregated junction data.

The equation giving the expected accident frequency for a one year period in junctions from the aggregated data set is:

$$\hat{\mu} = 5.063 \times 10^{-5} \times AADTmaj^{0.689} \times AADT\ \min^{0.369} \quad (6.3)$$

It can be stated that an increase in one unit in *ln(AADTmaj)* (and keeping *ln(AADTmin)* constant) increases the expected number of accidents by 99%. The same increase in *ln(AADTmin)*, with constant *ln(AADTmaj)* increases the expected number of accidents by 45%. See explanation of unit increase in section 3.2.

6.3.1 Model Checking

A comparisson of the predicted and observed accident data using the Poisson Gamma regression are shown in the histograms in Figure 84. These reveal that the modelled data are comparable to the observed data. The model seems to replicate observed data in especially the higher ranges much better than the Poisson model (see section 6.2).

The same conclusion can be drawn from the observation of the dot plots in Figure 86, where it is possible to observe that the modelled data sets have higher values for the first and last sets of junctions with spikes in the middle, therefore replicating realistically the observed data.

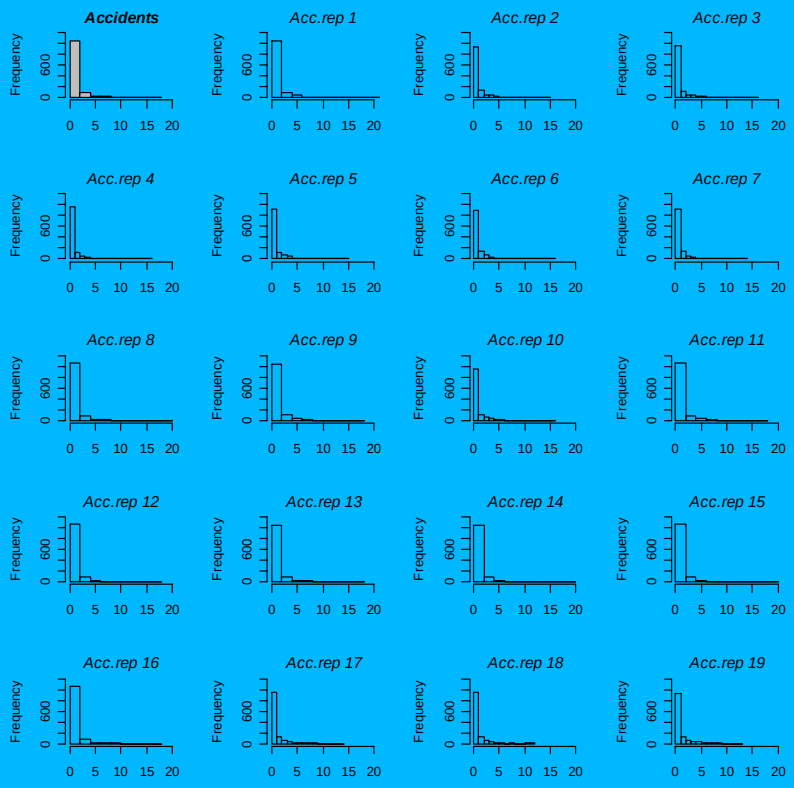

Figure 85 Histogram of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 histograms of replicated data set ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma hierarchical regression model.

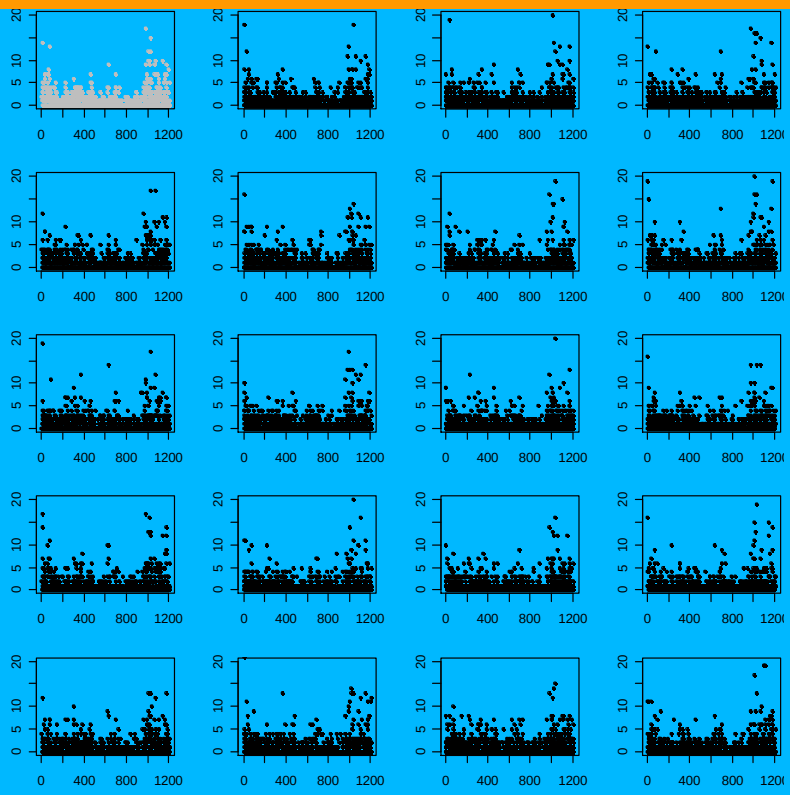

Figure 86 Dot plot of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma hierarchical regression model.

The plots of the discrepancy measures are displayed in Figure 87.

All four variations seem to be captured by the Poisson-Gamma model. The same conclusion can be drawn from the observation of the discrepancy measure employed to check whether the model is taking the data overdispersion into account and displayed in Figure 88.

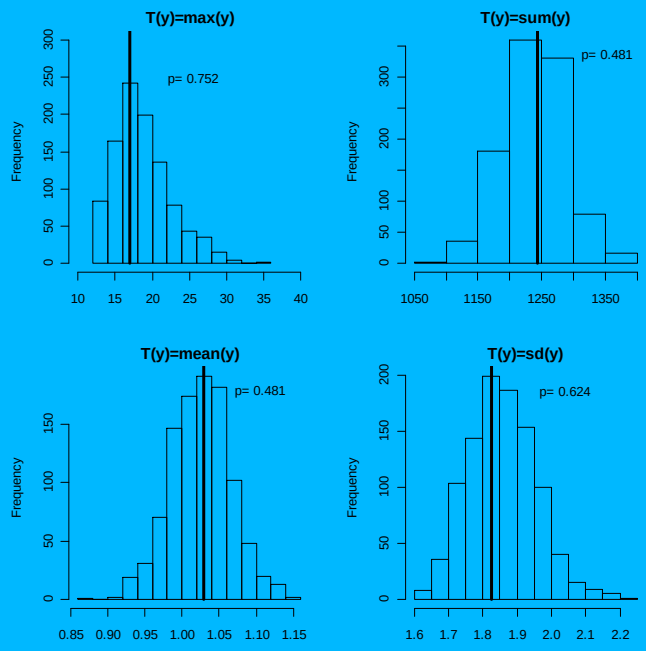

Figure 87 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fitted to the aggregated data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The estimated probability that the ratio of the variance to the mean in the replicated data is greater than the equivalent ratio calculated from the observed data is now equal to 0.650 and this is well within the limits stated by Congdon (2005).

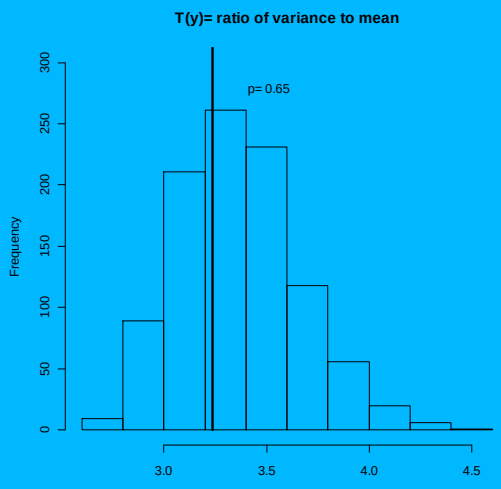

Figure 88 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fit to the aggregated data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

This model produces an average deviance, $\bar{D}$, of 2406.570 and an effective model dimension, d_e, of 336.781, giving a DIC of 2743.350. The estimated dispersion parameter ($1/\delta$) was equal to 0.842. These values characterise the model, being some of them used for model comparison (discussed in section 6.5).

The expected numbers of accidents for a one year period for the aggregated junctions for the minimum, mean, median and maximum profiles were calculated. The posterior means and the corresponding standard deviations of these profiles are provided in Table 34.

Table 34 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the aggregated accident data.

	Minimum	Mean	Median	Maximum
ln(AADTmaj)	4.890	8.054	8.100	10.568
ln(AADTmin)	1.946	6.244	6.328	9.871
	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
Expected Number of Accidents	0.003 (5.975E-04)	0.130 (0.006)	0.139 (0.006)	2.827 (0.378)

A typical junction belonging to the aggregated data set is expected to have 0.130 accidents in a one year period.

6.4 The Poisson Log-Normal regression model

The Poisson Log-Normal regression model was fitted to the aggregated data according to Equations 2.5 and 2.6 (Chapter 2) where the parameter α in Equation 2.6 followed a Gamma(a,b) *a priori* distribution with a=1 and b=0.01.

The MCMC algorithm was run with three chains for 35000 iterations with 20000 as burn-in with a thinning rate of 45. The results were drawn from samples with dimension 1002.

The Gelman-Rubin statistics corresponding to the estimated coefficient parameters from Equation 6.1 are plotted against the iterations and displayed in Figure 89. From observation of those graphs it is concluded that there are no reasons to suspect of non-convergence of the iterative simulation.

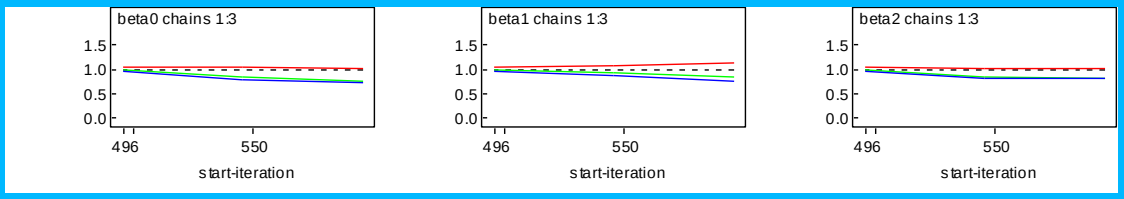

Figure 89 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters obtained after the Poisson Log-Normal hierarchical regression model was fitted to the aggregated data set.

The posterior means of the parameter estimates, corresponding standard deviations, Monte Carlo standard errors and 95% credible intervals are displayed in Table 35. The MC errors have small values indicating that the parameter estimates were calculated with precision. The 95% credible intervals do not contain zero for any parameter estimate, meaning that the

corresponding variable is relevant in the model.

Table 35 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fitted to the aggregated data set.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-10.080	0.424	3.277E-02	-10.930	-9.223
β_1 (ln(AADTmaj))	0.672	0.056	4.833E-03	0.555	0.783
β_2 (ln(AADTmin))	0.363	0.035	2.141E-03	0.294	0.431

The mean posterior estimates indicate that the every increase in *ln(AADTmaj)* increases the expected number of accidents by approximately 96%, provided *ln(AADTmin)* remains constant. An increase in one unit of *ln(AADTmin)* increases the expected frequency of accidents in around 44%.

The parameters estimate densities are displayed in Figure 90.

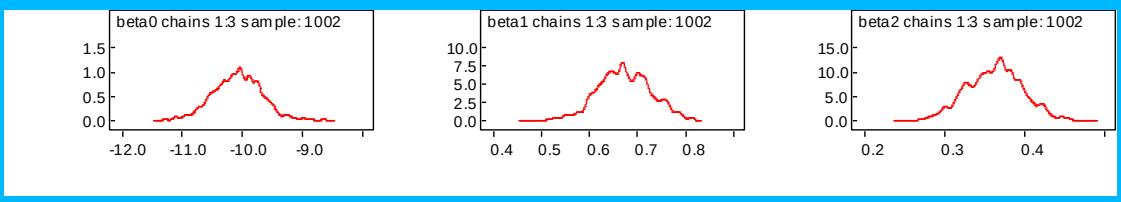

Figure 90 Posterior densities of the coefficients corresponding to the beta parameters obtained after the Poisson Log-Normal hierarchical regression model was fitted to the aggregated junction data.

The equation giving the expected number of accidents for a one year period for the aggregated data is the following:

$$\hat{\mu} = 4.191 \times 10^{-5} \times AADTmaj^{0.672} \times AADT\ \min^{0.363} \qquad (6.3)$$

6.4.1 Model Checking

Figure 91 contains histograms from a sample of the Poisson Log-Normal replicated data. In general these histograms resemble the histogram of the observed data, which indicates that the model seems able to replicate the observed data, also in the higher ranges.

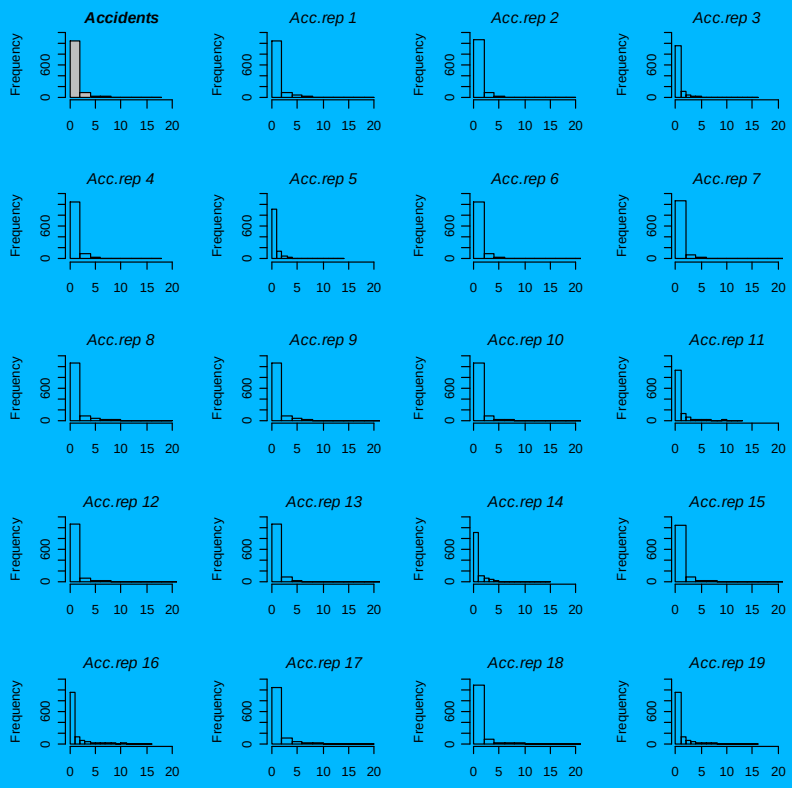

Figure 91 Histogram of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 histograms of replicated data set ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal hierarchical regression model.

An equivalent conclusion can be drawn from observing the dot plots in Figure 92 which were obtained from a further sample of nineteen replicated data sets.

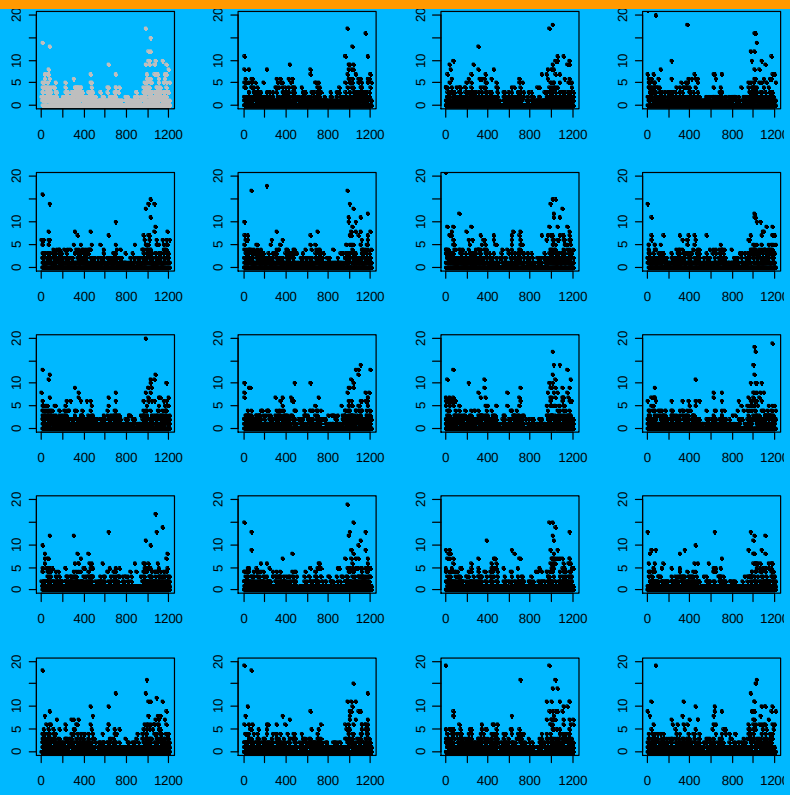

Figure 92 Dot plot of the observed number of accidents in the aggregated data set (first row on the left in grey) and 19 dot plots from replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

All the replicated data sets mimick the number of accidents on the first and third batch of junctions (they are slightly higher than the number of accidents of the junctions in between) and therefore reproducing the observed data (which has higher numbers of accident occurrences in Austria and Portugal and fewer in Norway).

The discrepancy measures (Figure 93) show that the model also captures the variations that these measures indicate. The probabilities that the discrepancy measures obtained by the replicated data are greater than the corresponding discrepancy measure from the observed data lie within the satisfactory boundaries of 0.1 to 0.9 as suggested by Congdon (2005). The *p* values are also near 0.5 for all the discrepancies.

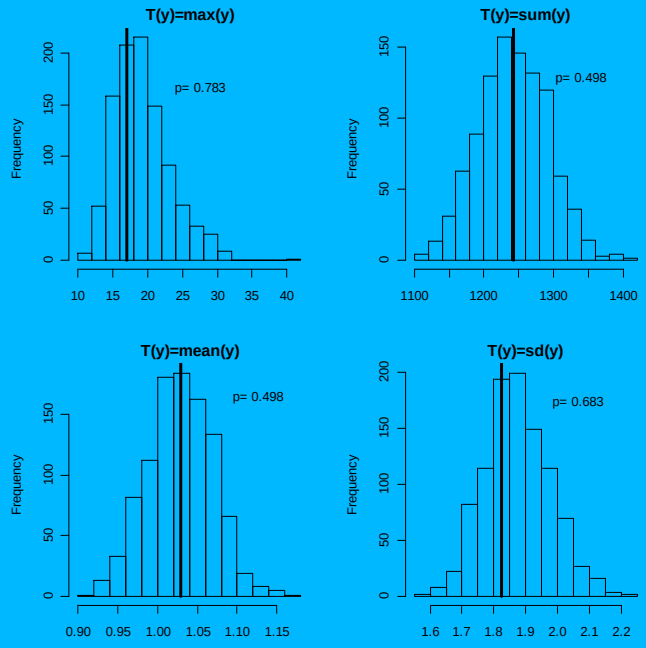

Figure 93 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fitted to the aggregated data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The verification of the measure of discrepancy suggested by Congdon (2005) to check whether the model takes the data overdispersion into account is depicted in Figure 94. It can be concluded that the Poisson Log-Normal is taking the overdispersion of the aggregated data into account.

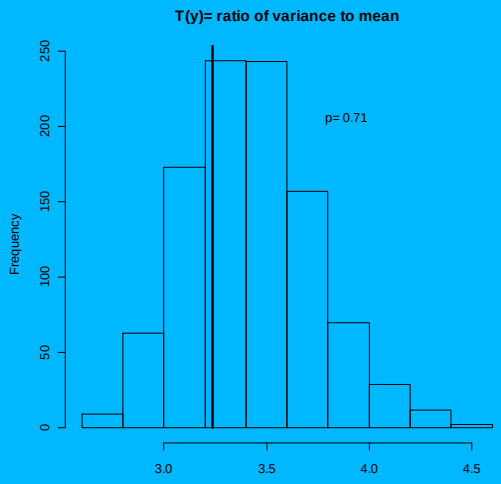

Figure 94 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fitted to the aggregated data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

This model produces an average deviance, $\bar{D}$, of 2437.680 and an effective model dimension, d_e, of 373.785, giving a DIC of 2811.460. The comparisons between the models will be discussed in section 6.5.

Table 36 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the aggregated accident data.

	Minimum	Mean	Median	Maximum
ln(AADTmaj)	4.890	8.054	8.100	10.568
ln(AADTmin)	1.946	6.244	6.328	9.871
	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
Expected Number of Accidents	0.002 (4.712E-04)	0.090 (0.005)	0.096 (0.005)	1.837 (0.226)

The posterior means and corresponding standard deviations for the minimum, mean, median and maximum profiles of *ln(AADTmaj)* and *ln(AADTmin)* for the aggregated set of junctions are provided in Table 36. A typical junction is expected to have 0.090 accidents during a one year period. For a maximum profile of *ln(AADTmaj)* and *ln(AADTmin)* any given junction is expected to have 1.837 accidents per year.

6.5 Discussion

The values of DIC and d_e shown in Table 37 for the three models considered in this chapter suggest that the model exhibiting the better fit and having a smaller degree of parsimony is the Poisson-Gamma regression model. Since this model also provided good results when the procedures for model checking were calculated (see Figure 85 to Figure 88) it can therefore be chosen as the model that best fits the aggregated data for the three countries.

Table 37 Comparison of DIC and related statistics for the three models fitted to the aggregated junction data.

Regression Model	$\bar{D}$	d_e	DIC
Poisson	3341	3.059	3344
Poisson-Gamma	2407	336.781	2743
Poisson Log-Normal	2438	373.785	2811

7 Modelling Austrian, Norwegian and Portuguese injury accidents on non-roundabout junctions

This chapter describes the combined analysis of the set of data formed after joining the three junction data sets (excluding the roundabouts) from Austria, Norway and Portugal. The analysis and the results obtained for the Poisson-Gamma and Poisson Log-Normal models are described in sections 7.2 and 7.3, respectively. In this chapter, and the following, it was decided to discard the Poisson model as this model does not seem to give appropriate fits on the accident data collected in the various countries analysed.

Section 7.4 finalises with a discussion.

7.1 Aggregated Junction Data (excluding roundabouts)

The data analysed in this chapter consists of 1087 junctions on the rural road network where 174 belong to Austria, 732 to Norway and 181 to Portugal.

The variables considered (note that there are no signalised roundabouts in Portugal nor data on traffic control measurements from Norway) included, per junction:

- ***Accidents***: gives the number of injury accidents;
- ***AADTmaj***: represents the major traffic entering volume (annual average daily traffic);
- ***AADTmin***: represents the minor traffic entering volume (annual average daily traffic).
- ***Number_of_Legs***: a binary (categorical) variable indicating whether the junction was formed by "3" or "4" legs;
- ***Country***: a categorical variable indicating the country where the junction was registered ("1" if Austria, "2" if Norway and "3" if Portugal).

The graphs in Figure 95 show the plot of variable *Accidents* per junction and per country (with different colours) on the left and the histogram of the same variable on the right.

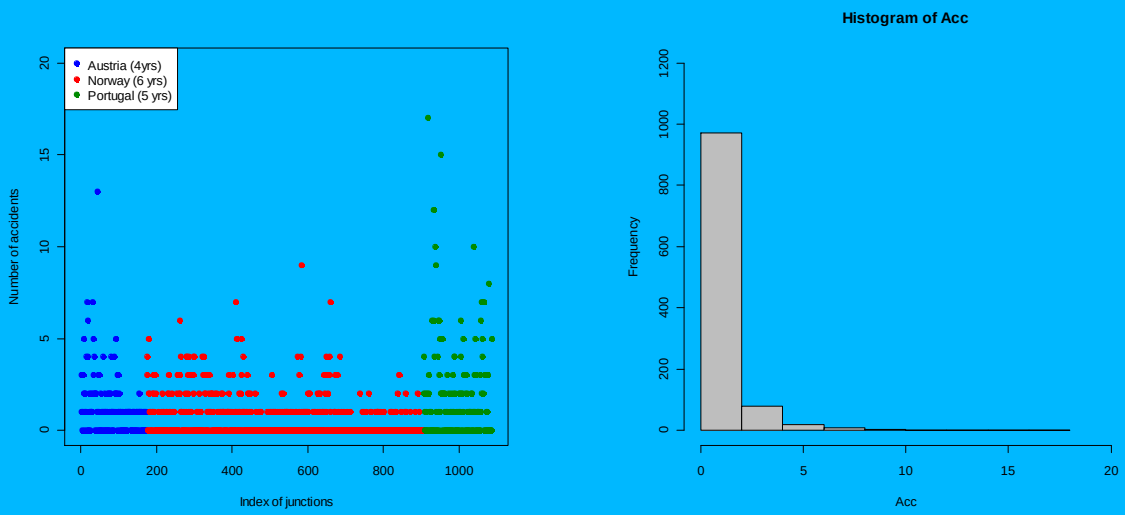

Figure 95 Plot of *Accidents*, per junctions, registered in the aggregated rural road network junctions excluding the roundabouts (on the left) and the histogram of the frequency of Accidents (on the right).

The number of injury accidents plotted against *AADTmaj* and *AADTmin*, together with a polynomial fit is depicted in the two graphs in Figure 96. The majority of the injury accidents occur for values of between 133 to 10000 vehicles per day for *AADTmaj* and 7 to 3000 vehicles per day for *AADTmin.*

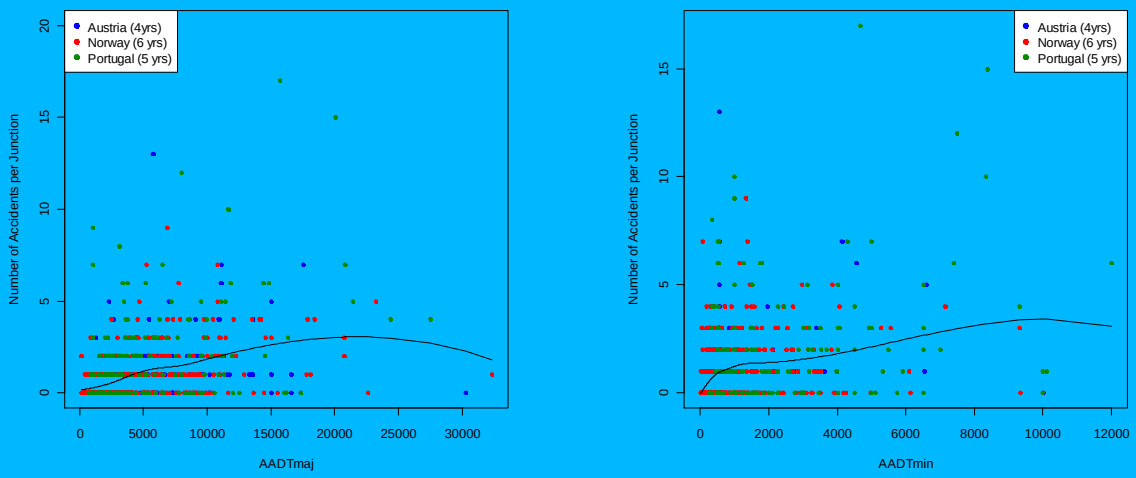

Figure 96 The number of accidents per junction in the aggregated data set, excluding the roundabouts, against *AADTmaj* and *AADTmin*, and corresponding polynomial fits, on the left and right panels, respectively.

From Figure 97 it can be seen that the great majority of the non-roundabout junctions have three approaches (i.e. legs). The box plots on the panel on the right show an increase on the number of accidents on four legged junctions when compared with three legged ones.

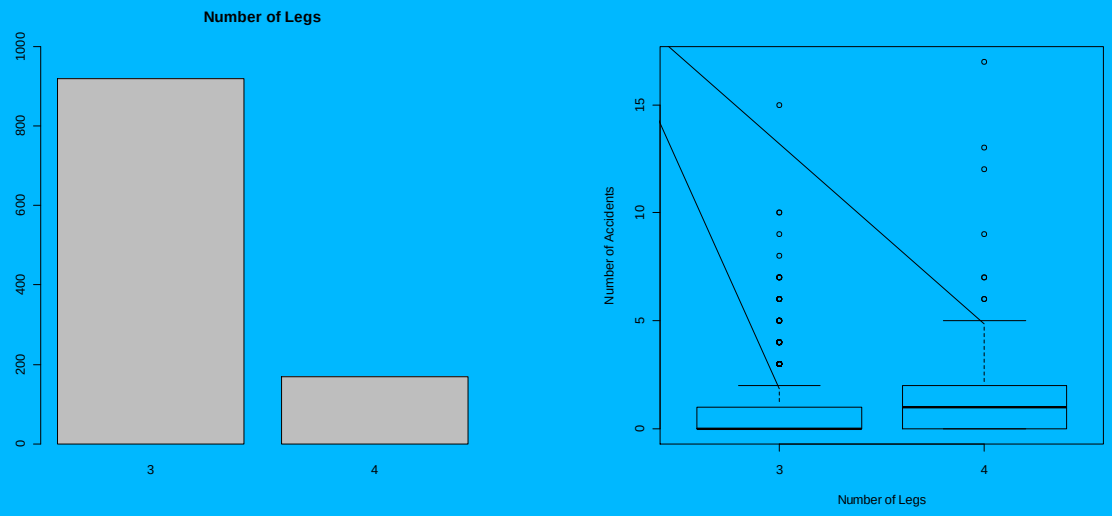

Figure 97 Bar plot giving the number of junctions with 3 and 4 legs (on the left panel) and box plot of the number of accidents per number of legs (right panel) for the aggregated data excluding the roundabouts.

The box plots of the number of accidents per country as shown in Figure 98 suggest that highest accident median value is taken by the Portuguese junctions followed by Austrian and Norway.

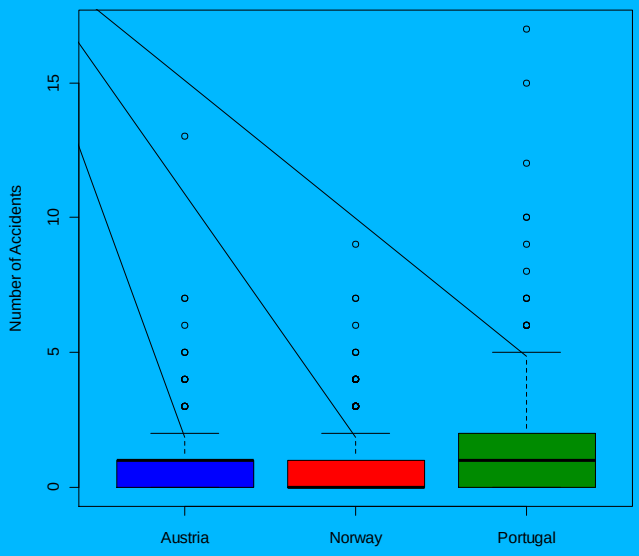

Figure 98 Box plot of the number of accidents in the aggregated set excluding the roundabouts by *Country*.

The values in Table 38 were obtained for the three variables and also for the number of accidents per country.

Table 38 Summary statistics for the variables registered on the aggregated junctions excluding roundabouts.

Variables	minimum	mean	standard deviation	median	maximum
AADTmaj	133	4349.817	4085.678	3024	32311
AADTmin	7	946.159	1441.086	540	11993
Accidents	0	0.892	1.623	0	17
Accidents (Austria)	0	1.218	1.598	1	13
Accidents (Norway)	0	0.584	1.090	0	9
Accidents (Portugal)	0	1.829	2.691	1	17

7.2 The Poisson-Gamma hierarchical regression model

The Poisson-Gamma model as given by Equations 2.3 and 2.4 with with $\delta = \gamma$ and $\delta \sim Gamma(a,a)$ with a=0.01.

$$ln(\hat{\mu}_i) = \beta_0 + \beta_1 ln(AADTmaj_i) + \beta_2 ln(AADTmin_i) + \beta_3 Number\_of\_Legs_i + \beta_4 Country_i \quad (7.1)$$

Equation 7.1 was applied and the β parameters were given *a priori* Normal distributions with mean 0 and precision 0.0001 (the variance is equal to the inverse of the precision). The baseline model was taken to be a three legged Norwegian junctions as these had the higher sample size (as compared to the samples formed by Austria and Portugal).

The MCMC algorithm comprised 3 chains and was run for 35000 iterations with 20000 burn-in iterations with a thinning rate of 45. The results thus described were based on a sample with dimension equal to 1002.

The Gelman-Rubin statistics for the beta parameters plotted in Figure 99 show that there seems that there are reasons to believe that the densities of the first two coefficient estimates have not converged. Several simulations were made with higher number of iterations and several different initial values were also considered. The resulting simulations obtained similar Gelman-Rubin graphs as the ones represented in Figure 99, therefore indicating that there were some difficulties in attaining convergence for the two initial coefficient parameters.

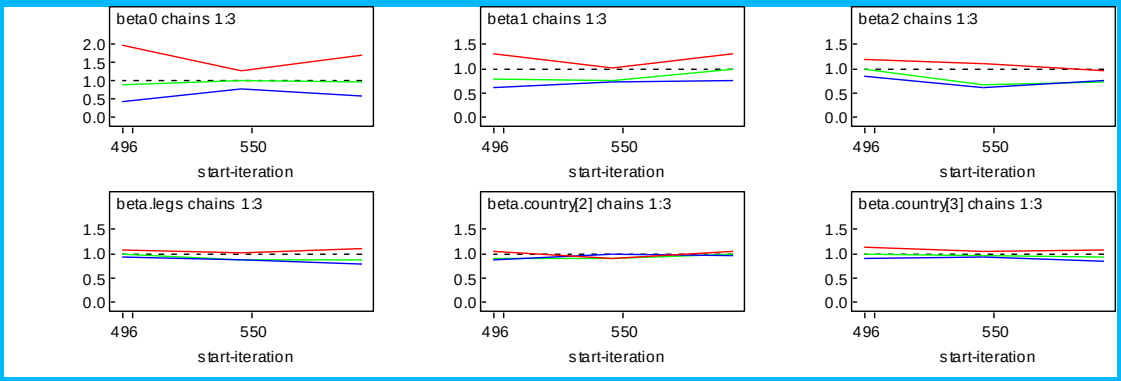

Figure 99 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters, from the Poisson-Gamma regression model fitted to the aggregated junction data excluding the roundabouts.

Point estimates, corresponding standard deviations, Monte Carlo errors and 95% credible intervals obtained after the Poisson-Gamma regression model was fitted to the data are given in Table 39 for the regression parameters. The corresponding posterior densities for the parameters are given in the plots depicted in Figure 100.

Table 39 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fit to the aggregated accident data (excluding roundabouts) using 3-leg Norwegian junctions as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-8.790	0.421	5.646E-02	-9.689	-8.088
β_1 (ln(AADTmaj))	0.594	0.060	7.566E-03	0.487	0.729
β_2 (ln(AADTmin))	0.256	0.045	4.936E-03	0.156	0.339
β_3 (Number_of_Legs='4')	0.449	0.112	1.073E-02	0.223	0.668
β_4 (Country=´Austria')	0.580	0.129	1.374E-02	0.335	0.854
β_4 (Country='Portugal')	0.751	0.134	1.458E-02	0.485	1.003

From the examination of the point mean estimates it can be stated that, according to this particular Poisson-Gamma model, every unitary increase in *ln(AADTmaj)* (see example in section 3.2) increases the expected frequency of accidents by approximately 81% (assuming all other variables remain constant). A unit increase in *ln(AADTmin)* increases the expected frequency of accidents in around 29%.

Junctions with four legs have an expected number of accidents around 57% higher than junctions with three legs (see values in Table 39). Respectively, Austrian and Portuguese junctions have an expected 79% and 112% more accidents than similar Norwegian junctions.

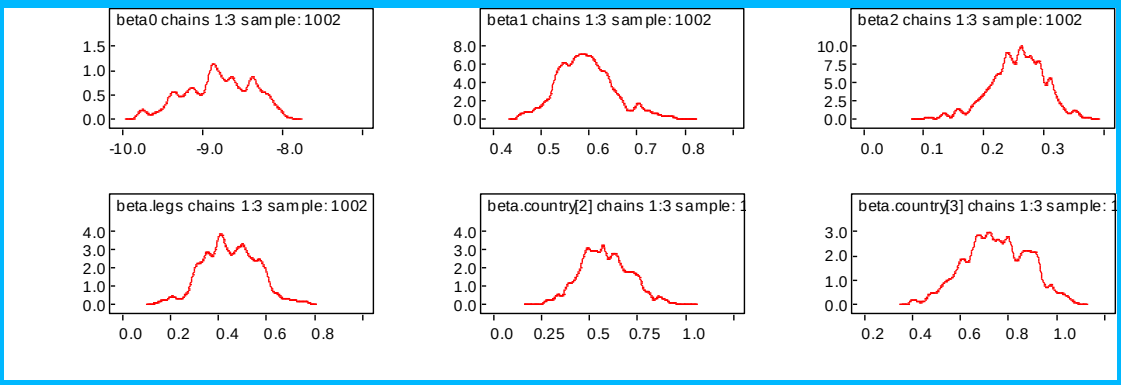

Figure 100 Posterior densities of the beta parameter estimates obtained after a Poisson-Gamma regression model was fitted to the aggregated data set excluding roundabouts.

The expected frequency of accidents for a one year period is given by the equations shown in Table 40.

Table 40 Expected number of accidents per year for the aggregated junction data set (omitting roundabouts) obtained by a Poisson-Gamma regression model using 3-leg Norwegian junctions as baseline.

	Expected Numbers of Accidents
Number_of_Legs='3'	
Country	
'Austria'	$\hat{\mu}_i = 2.720 \times 10^{-4} \times AADTmaj_i^{0.594} \times AADT\ \min_i^{0.256}$
'Norway'	$\hat{\mu}_i = 1.523 \times 10^{-4} \times AADTmaj_i^{0.594} \times AADT\ \min_i^{0.256}$
'Portugal'	$\hat{\mu}_i = 3.227 \times 10^{-4} \times AADTmaj_i^{0.594} \times AADT\ \min_i^{0.256}$
Number_of_Legs='4'	
Country	
'Austria'	$\hat{\mu}_i = 4.261 \times 10^{-4} \times AADTmaj_i^{0.594} \times AADT\ \min_i^{0.256}$
'Norway'	$\hat{\mu}_i = 2.385 \times 10^{-4} \times AADTmaj_i^{0.594} \times AADT\ \min_i^{0.256}$
'Portugal'	$\hat{\mu}_i = 5.055 \times 10^{-4} \times AADTmaj_i^{0.594} \times AADT\ \min_i^{0.256}$

Overall, junctions with three legs have lower expected number of accidents than the junctions with four legs. The country with the least expected number of accidents is Norway whereas Portugal has the highest accident frequency.

7.2.1 Model Checking

The graphs of replicated data in form of histograms and dot plots are depicted in Figure 101 and Figure 102, respectively.

Both the histograms and the box plots show that the posterior predictive distribution data replicated by the Poisson-Gamma model adequately models the observed data.

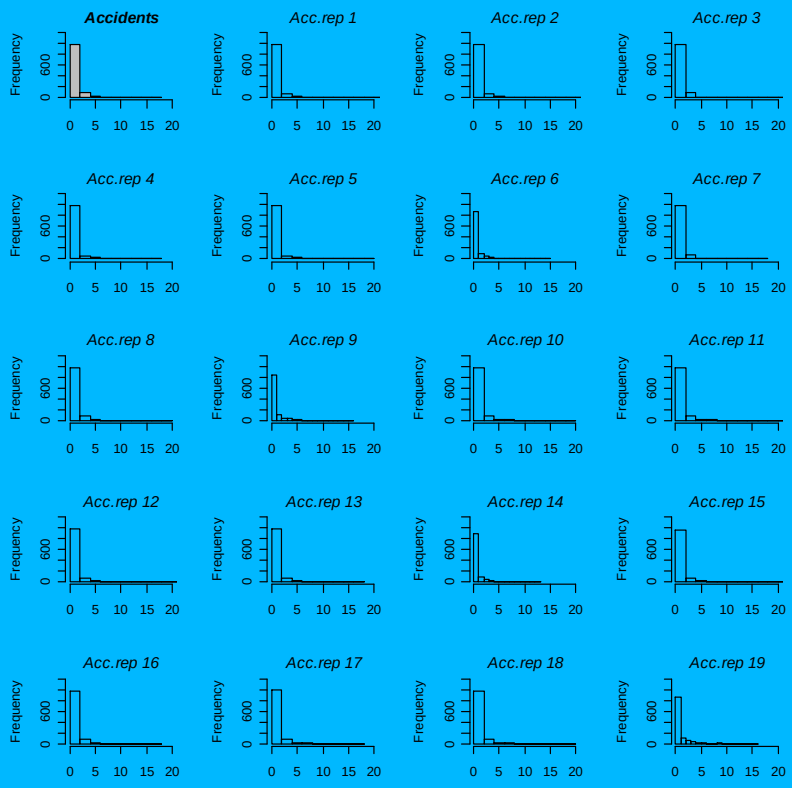

Figure 101 Histogram of the observed number of accidents in the aggregated set, excluding roundabouts (left upper corner) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.

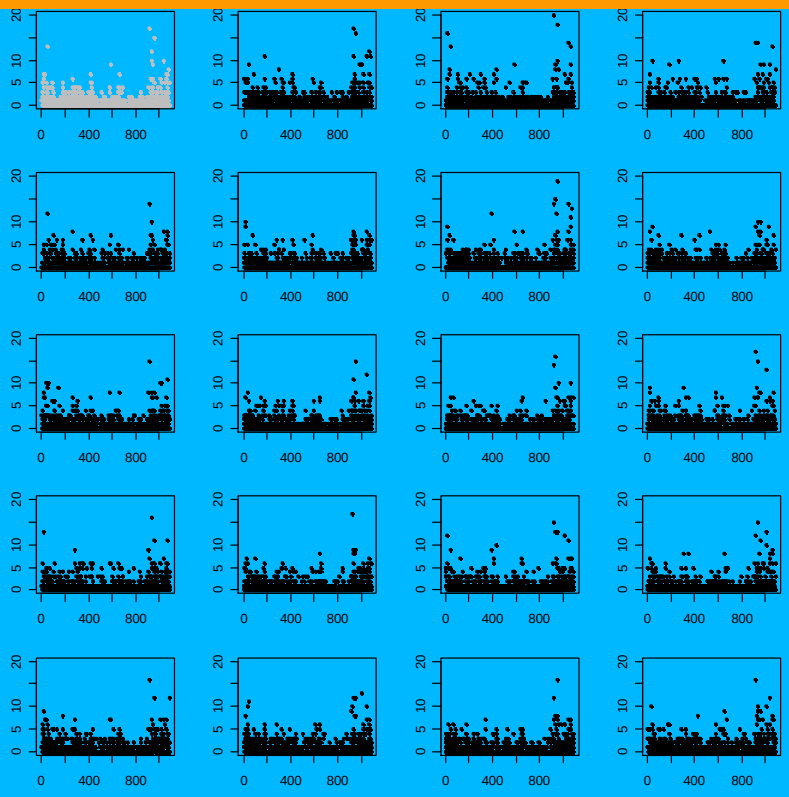

Figure 102 Dot plot of the observed number of accidents in the aggregated set, excluding roundabouts (left upper corner) and 19 dot plots of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.

The posterior probability of the discrepancy measures obtained from the replicated data is greater than the corresponding discrepancy measures resulting from the observed data which indicates a good fit as can be seen in the plots in Figure 103.

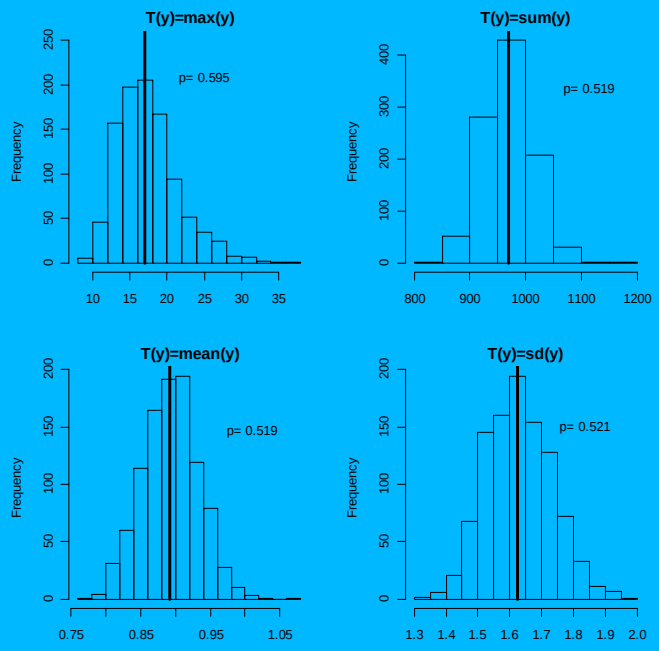

Figure 103 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fitted to the aggregated data excluding roundabouts. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

Figure 104 shows the graph of the discrepancy measure obtained by the variance to mean ratio, giving an estimated posterior probability equal to 0.522 which indicates that the Poisson-Gamma model seems to be taking the overdispersion into account.

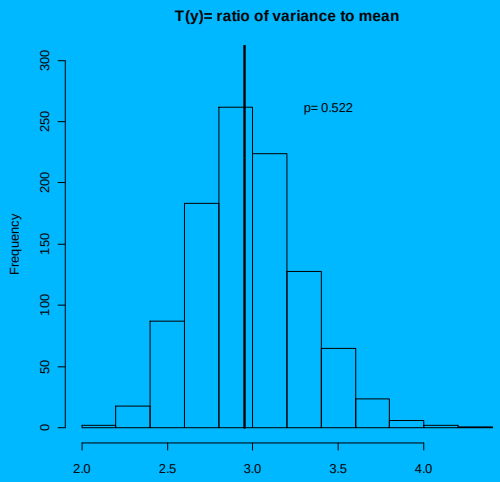

Figure 104 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fitted to the aggregated data omitting the roundabouts, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

This model produces an average deviance, $\bar{D}$, of 2047.810 and an effective model dimension, d_e, of 258.418, giving a DIC of 2306.230. These values are used for model comparison which is described in section 7.4.

The dispersion parameter ($1/\delta$ in Equation 2.4) was estimated as 0.692.

The expected accident frequency for the four junction types and three traffic controls of junctions excluding roundabouts, for the minimum, mean, median and maximum profiles were calculated for the corresponding values of *ln(AADTmaj)* and *ln(AADTmin)*. The posterior means and corresponding standard deviations are provided in Table 41.

Table 41 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the aggregated accident data (omitting roundabouts).

		Minimum	Mean	Median	Maximum
	ln(AADTmaj)	4.890	7.971	8.014	10.383
	ln(AADTmin)	1.946	6.095	6.292	9.392
Country	Number_of_Legs	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
Austria	3	0.008 (0.002)	0.149 (0.018)	0.161 (0.019)	1.456 (0.218)
	4	0.013 (0.003)	0.233 (0.027)	0.252 (0.029)	2.283 (0.356)
Norway	3	0.005 (9.147E-04)	0.083 (0.006)	0.090 (0.006)	0.816 (0.128)
	4	0.007 (0.002)	0.130 (0.014)	0.141 (0.016)	1.285 (0.242)
Portugal	3	0.010 (0.003)	0.177 (0.020)	0.190 (0.021)	1.725 (0.248)
	4	0.016 (0.004)	0.277 (0.038)	0.299 (0.040)	2.715 (0.475)

For a typical Norwegian junction (under the column *Mean*) with three legs one expects 0.083 accidents, while for Austrian and Portuguese junctions with the same number of legs 0.149 and 0.177 accidents, respectively are expected.

7.3 The Poisson Log-Normal hierarchical regression model

The Poisson Log-Normal model was fitted to the data according to Equations 2.5 and 2.6 in Chapter 2, where the parameter α in Equation 2.6 had a Gamma(*a*,*b*) *a priori* distribution with *a*=1 and *b*=0.01. The MCMC algorithm was run with three chains for 35000 iterations with 20000 as burn-in with a thinning rate of 45 iterations, resulting in samples with dimension 1002.

The graphs depicted in Figure 105 show the three Gelman-Rubin statistics obtained for the estimates of the β parameters. It can be observed that the *Rhat* statistic converges to 1 in all the parameters as the number of iterations increase. Therefore, there are no reasons to

suspect non-convergence of the iterative simulation.

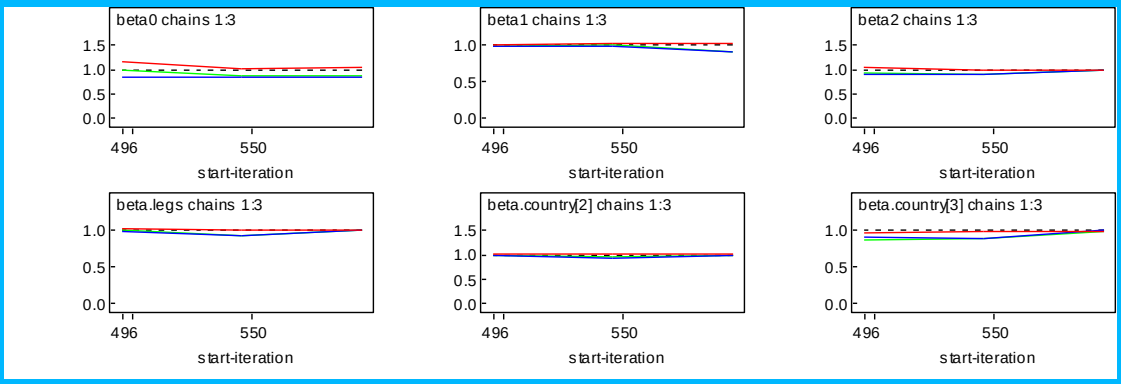

Figure 105 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters, from the Poisson Log-Normal regression model fitted to the aggregated junction data excluding the roundabouts.

The posterior means of the parameter estimates, corresponding standard deviations, Monte Carlo standard errors and 95% credible intervals are displayed in Table 42. The corresponding estimated posterior densities are displayed in the graphs in Figure 106.

Table 42 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the aggregated accident data (omitting roundabouts) using 3-leg Norwegian junctions as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-9.268	0.484	4.421E-02	-10.230	-8.377
β_1 (ln(AADTmaj))	0.605	0.063	5.482E-03	0.486	0.734
β_2 (ln(AADTmin))	0.272	0.053	3.533E-03	0.167	0.371
β_3 (Number_of_Legs='4')	0.392	0.112	3.594E-03	0.176	0.613
β_4 (Country='Austria')	0.600	0.116	3.651E-03	0.368.	0.816
β_4 (Country='Portugal')	0.677	0.118	4.919E-03	0.447	0.902

From the examination of the mean posterior estimates in Table 42 it can be stated that every unitary increase in *ln(AADTmaj)* increases the predicted number of accidents by approximately 83%, provided the other variables remain constant. An increase in one unit of *ln(AADTmin)* increases the accident frequency in around 31%. A typical four leg junction is *a posteriori* expected to have 48% more accidents than a three leg junction (with the other explanatory variables constant).

Austrian and Portuguese junctions are *a posteriori* expected to have 82% and 97% more injury accidents than a Norwegian junction with the same values of *ln(AADTmaj)*, *ln(AADTmin)* and number of legs.

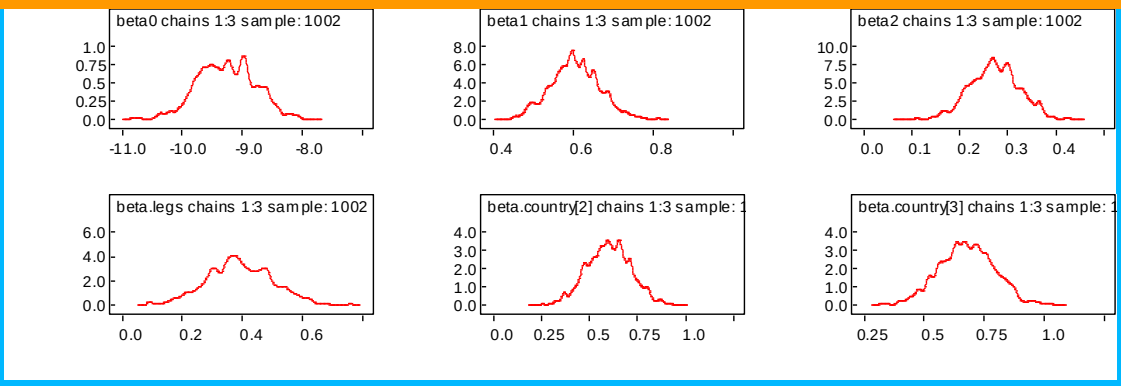

Figure 106 Posterior densities of the beta parameter estimates obtained after a Poisson Log-Normal regression model was fitted to the aggregated data set, excluding roundabouts.

The equations giving the expected frequency of injury accidents, per year, per number of legs and country are given in Table 43.

Table 43 Expected number of accidents per year for the aggregated junction data set (omitting roundabouts) obtained by a Poisson Log-Normal regression model using 3-leg Norwegian junctions as baseline.

	Expected Numbers of Accidents
Number_of_Legs='3'	
Country 'Austria' 'Norway' 'Portugal'	 $\hat{\mu}_i = 1.719\times10^{-4}\times AADTmaj_i^{0.605}\times AADT\ \min_i^{0.272}$ $\hat{\mu}_i = 9.436\times10^{-5}\times AADTmaj_i^{0.605}\times AADT\ \min_i^{0.272}$ $\hat{\mu}_i = 1.857\times10^{-4}\times AADTmaj_i^{0.605}\times AADT\ \min_i^{0.272}$
Number_of_Legs='4'	
Country 'Austria' 'Norway' 'Portugal'	 $\hat{\mu}_i = 2.543\times10^{-4}\times AADTmaj_i^{0.605}\times AADT\ \min_i^{0.272}$ $\hat{\mu}_i = 1.396\times10^{-4}\times AADTmaj_i^{0.605}\times AADT\ \min_i^{0.272}$ $\hat{\mu}_i = 2.747\times10^{-4}\times AADTmaj_i^{0.605}\times AADT\ \min_i^{0.272}$

7.3.1 Model Checking

Figure 107 and Figure 108 contain a collection of histograms and dot plots of data replicated by the model as well as the histogram and dot plot (on the upper left corner) of the observed data, i.e. the number of injury accidents (*Accidents*). Overall, from the examination of these figures it can be stated that the model under consideration is able to replicate the observed data adequately.

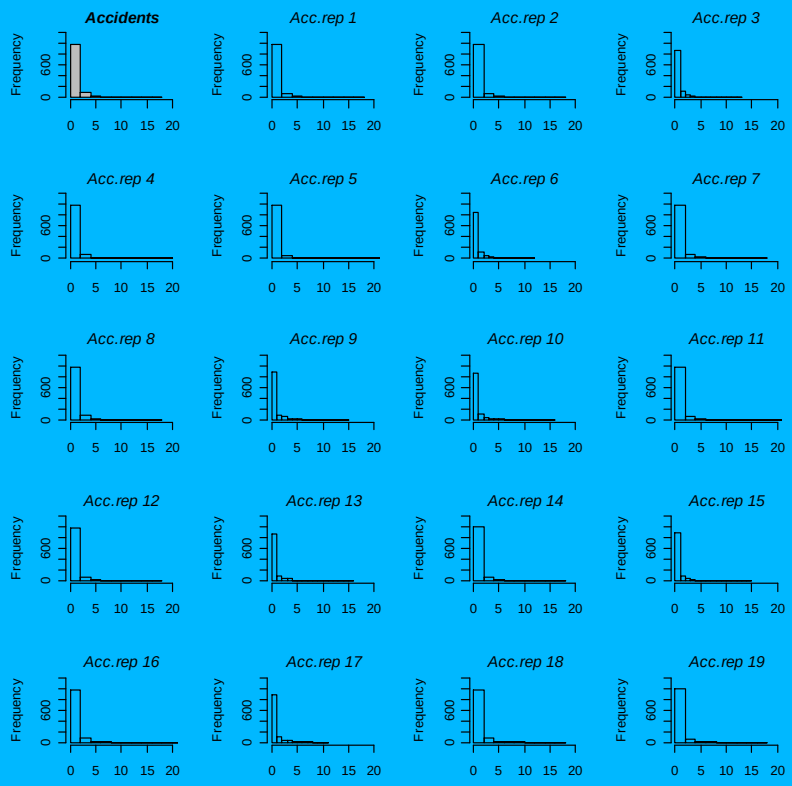

Figure 107 Histogram of the observed number of accidents in the aggregated set, excluding roundabouts (left upper corner) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

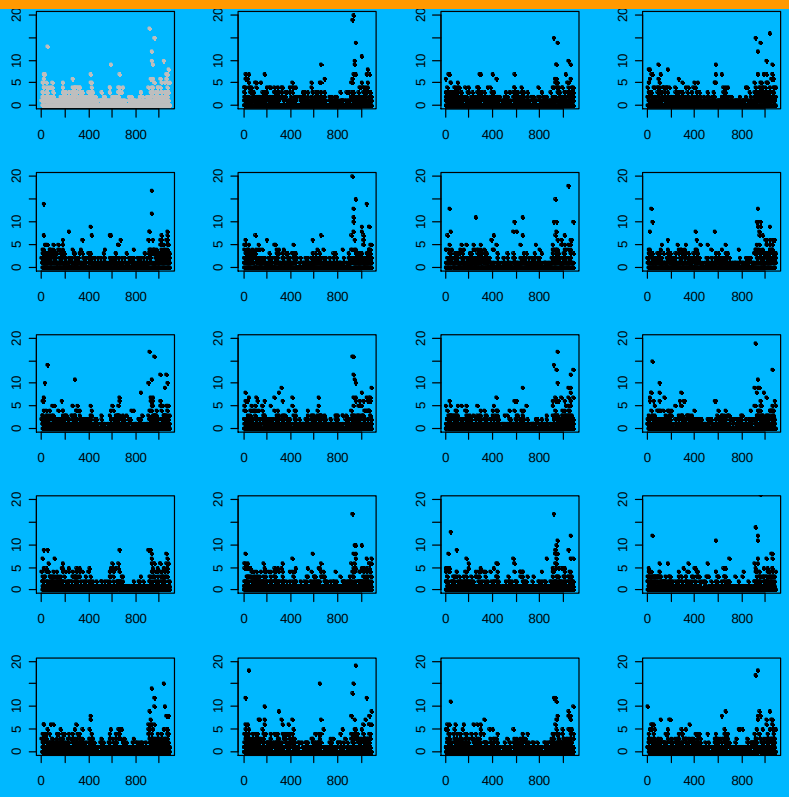

Figure 108 Dot plot of the observed number of accidents in the aggregated set, excluding roundabouts (left upper corner) and 19 dot plots of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

The results of the four discrepancy measures are displayed in Figure 109. The probabilities that the discrepancy measures obtained by the replicated data are greater than the corresponding discrepancy measures from the observed data lie within the satisfactory boundaries of 0.1 to 0.9 (according to Congdon, 2005) and indeed are all near the ideal value of 0.5, indicating that the model seemed to be able to replicate these particular discrepancies.

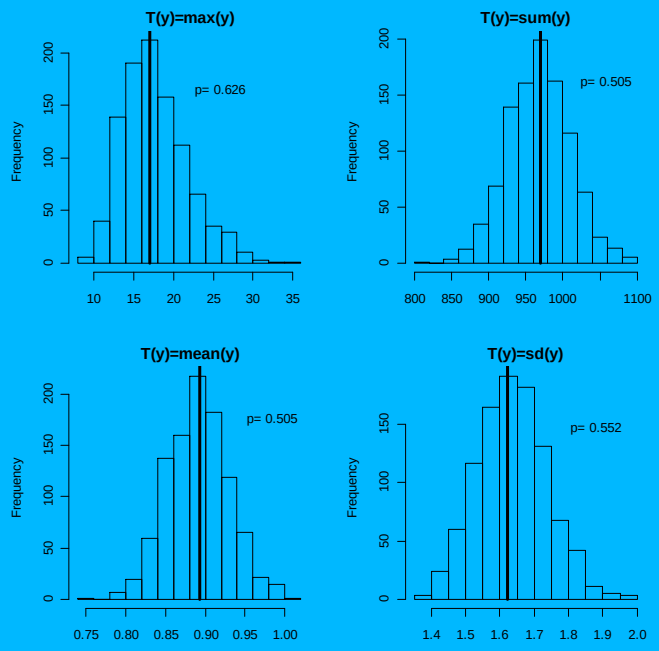

Figure 109 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fit to the aggregated data excluding roundabouts. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The same conclusion can be drawn when examining the ratio of variance to mean in Figure 110.

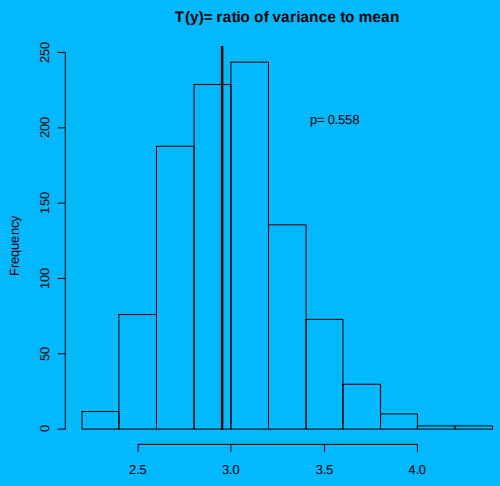

Figure 110| Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fitted to the aggregated data omitting the roundabouts, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

This indicates that the Poisson Log-Normal model is also able to detect and replicate the

data's overdispersion.

This model produces an average deviance, $\bar{D}$, of 2076.530 and an effective model dimension, d_e, of 281.447, giving a DIC of 2357.980. The dispersion parameter ($1/\alpha$) was estimated as 0.610. These values will be used in the model comparison described in section 7.4.

The expected accident frequencies for the three countries per number of legs for four profiles are displayed in Table 44.

Table 44 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the aggregated accident data (omitting roundabouts).

		Minimum	Mean	Median	Maximum
	ln(AADTmaj)	4.890	7.971	8.014	10.383
	ln(AADTmin)	1.946	6.095	6.292	9.392
Country	Number_of_Legs	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
Austria	3	0.006 (0.001)	0.113 (0.013)	0.123 (0.014)	1.208 (0.223)
	4	0.009 (0.002)	0.168 (0.020)	0.182 (0.021)	1.782 (0.302)
Norway	3	0.003 (7.943E-04)	0.062 (0.005)	0.067 (0.005)	0.662 (0.113)
	4	0.005 (0.001)	0.092 (0.011)	0.099 (0.012)	0.979 (0.171)
Portugal	3	0.006 (0.002)	0.122 (0.014)	0.133 (0.014)	1.298 (0.202)
	4	0.009 (0.003)	0.182 (0.027)	0.197 (0.028)	1.923 (0.314)

A typical Norwegian junction with three legs is expected to have 0.062 accidents while a junction with four legs is expected to have 0.092 accidents. In the worst case scenario, for maximum values of *ln(AADTmaj)* and *ln(AADTmin)* a Portuguese four leg junction is expected to have1.923 accidents per year. Typical Austrian junctions with three legs are expected to have 0.113 accidents per year.

7.4 Discussion

Table 45 shows the resulting fit, complexity and overall model choice (DIC) score for the models fitted to the aggregated junction data (but excluding roundabouts).

Table 45 Comparison of DIC and related statistics for the three models fitted to the aggregated junction data (excluding roundabouts).

Regression Model	$\overline{D}$	d_e	DIC
Poisson-Gamma	2048	258	2306
Poisson Log-Normal	2077	281	2358

The Poisson-Gamma emerges as the model with the lower DIC value. Since this model adequately captures the data discrepancies, including overdispersion, it can be concluded that it models the aggregated junction data (without the roundabouts) well.

Figure 111 shows the posterior means of the expected number of accidents for mean profile values of *ln(AADTmaj)* and *ln(AADTmin)* obtained by the Poisson-Gamma model (as depicted in Table 44).

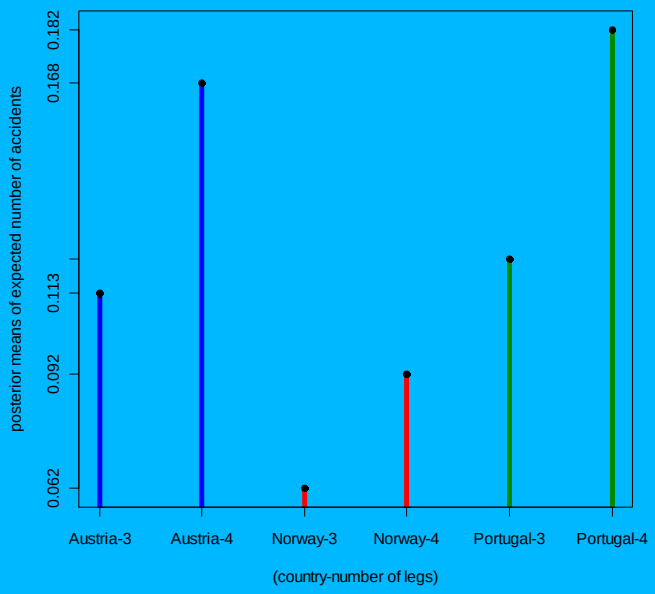

Figure 111 Values of the posterior means of the expected number of accidents for the aggregated data set excluding the roundabouts, classified per country and number of legs (as under column *Mean* in Table 44).

Junctions with four legs have higher values for the expected number of injury accidents regardless of the country, than do junctions with three legs.

8 Modelling Austrian, Dutch and Portuguese injury accidents on roundabout Junctions

The present chapter describes the combined analysis of the set of data formed after joining the Austrian, Dutch and Portuguese roundabout junction data sets.

8.1 Austrian, Dutch and Portuguese Roundabout Data

The data described in this section consists of several measurements registered on 142

roundabouts from the Austrian, Dutch and Portuguese rural road network, being 39 roundabouts from Austria, 27 from the Netherlands and 76 from Portugal.

Per junction, the variables considered (note that there were no measurements for the numbers of legs in the Austrian data) included:

- ***Accidents***: gives the number of injury accidents;
- ***AADTmaj***: represents the traffic volume entering the major road legs (annual average daily traffic);
- ***AADTmin***: represents the traffic volume entering the minor road legs (annual average daily traffic).
- ***Country***: a categorical variable indicating the country where the junction was registered ("1" if Austria, "2" if Holland and "3" if Portugal).

The dot plot showed on the left panel of Figure 112 shows the numbers of accidents per junction for the three countries. The histogram displayed on the right panel of the same figure gives the accident frequencies per intervals.

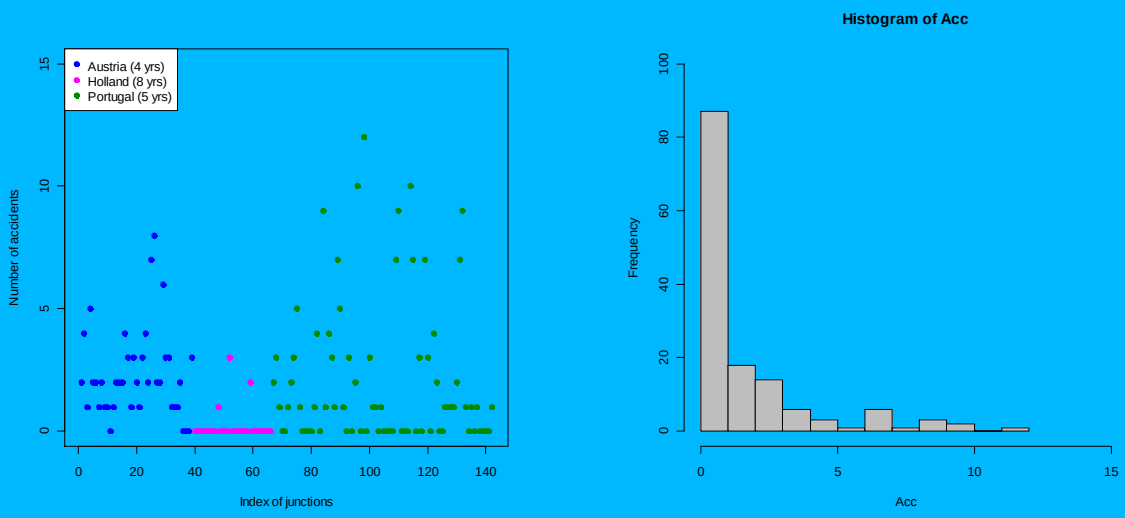

Figure 112 Plot of *Accidents*, per junctions, registered in the Austrian, Dutch and Portuguese roundabout rural road network junction data (on the left) and the histogram of the frequency of *Accidents* (on the right).

The data sets corresponding to the three countries were joined even though each set represent counts measured over different time periods. The analysis performed, including the modelling, took the three time periods into account so that the final results concerned a period of one year.

The two graphs depicted in Figure 113 represent the plots of *AADTmaj* (left panel) and *AADTmin* (right panel) plotted against the number of injury accidents, In both graphs was included a fitted smooth regression curve (represented by the solid curves). The great majority of accidents occur for values of *AADTmaj* between 1000 and 10000 and between 500 and 5000 for *AADTmin*.

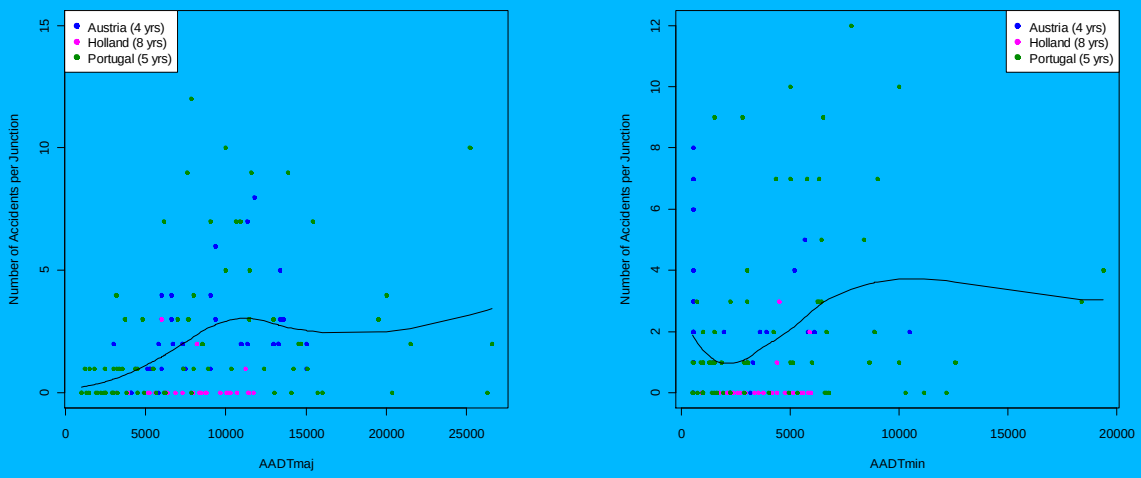

Figure 113 The number of accidents at roundabouts for Austrian, Dutch and Portuguese data setsagainst *AADTmaj* and *AADTmin*, and corresponding polynomial fits, on the left and right panels, respectively.

Figure 114 shows the box plots of the number of accidents per country.

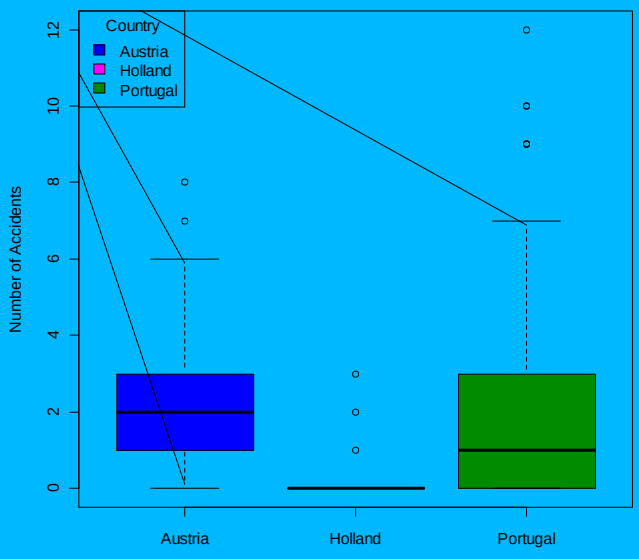

Figure 114 Box plot of the number of accidents in the aggregated set including only roundabouts (Austria, Holland and Portugal) by *Country*.

It can be observed that, due to the small sample size of Dutch roundabouts the corresponding box plot was reduced to its median (horizontal line) and three points. It can also be seen that the Austrian median number of injury accidents in roundabouts is higher than the corresponding value in Portuguese roundabouts.

The summary statistics for some of the variables in the data set are displayed in Table 46

Table 46 Summary statistics for the variables registered on the aggregated roundabout junctions from Austria Holland and Portugal.

Variables	minimum	mean	standard deviation	median	maximum
AADTmaj	1000	8428.310	5180.784	7534.500	26565
AADTmin	500	3597.662	3404.899	2722.438	19359
Accidents	0	1.838	2.525	1	12

8.2 The Poisson-Gamma hierarchical regression model

The Poisson-Gamma hierarchical regression model was fitted to the aggregated roundabout data set according to Equations 2.3 and 2.4 in Chapter 2. The parameters γ and δ where such that $\delta = \gamma$ and $\delta \sim Gamma(1,a)$ with a=0.01. The expression given by Equation 8.1 was applied in Equation 2.3.

$$ln(\hat{\mu}_i) = \beta_0 + \beta_1 ln(AADTmaj_i) + \beta_2 ln(AADTmin_i) + \beta_3 Country_i \quad (8.1)$$

The β parameters were given *a priori* Normal distributions with mean equal to 0 and variance 10000. The MCMC algorithm comprised three chains and was run for 35000 iterations of which 33000 were burn-in with a thinning rate of 6, resulting in samples of size 1002.

The baseline, or reference, category was:

Country = Portugal.

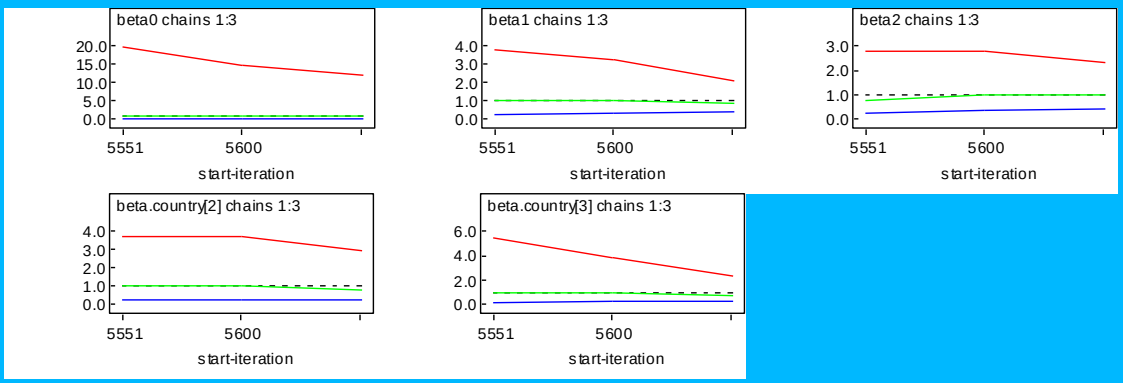

Figure 115 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters, from the Poisson-Gamma regression model fitted to the aggregated junction roundabout data.

The Gelman-Rubin statistics are plotted in the graphs depicted in Figure 115. There seems to be no reason to doubt non-convergence in any of the parameters concerned.

The point estimates for the Poisson-Gamma fit are given in Table 47. The densities of the parameter estimates are displayed in Figure 116. They show several spikes leading to believe that convergence was not properly attained for those regressor parameter estimators, especially for β_0.

Since the prior beliefs about the beta parameters were ´vague', diffuse and in the limit uninformative, the posterior densities will be dominated by the likelihood (i.e. the data

contains much more information than the prior about the parameters). According to Jackman (2009), in the limiting case of an uninformative prior, the only information about the parameters is that in the data, and the posterior has the same shape as the likelihood function. In this particular case the information about the parameters was taken mostly from the initial values considered.

Table 47 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson-Gamma regression model was fit to the aggregated roundabout accident data Portuguese junctions as baseline.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-4.776	1.120	1.543E-01	-6.207	-3.143
β_1 (ln(AADTmaj))	0.403	0.091	1.230E-02	0.226	0.539
β_2 (ln(AADTmin))	0.044	0.093	1.253E-02	-0.100	0.202
β_3 (Country=´Austria')	0.287	0.239	3.247E-02	-0.125	0.701
β_4 (Country=´Holland')	-2.970	0.375	5.167E-02	-3.675	-2.407

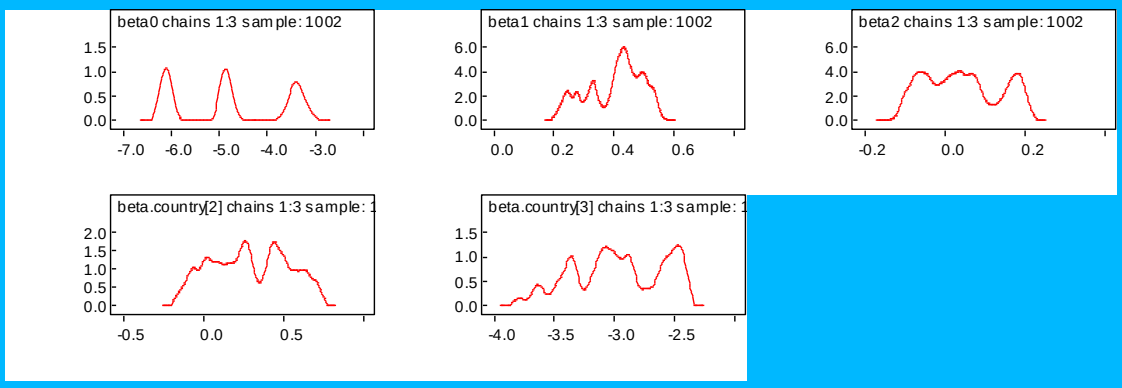

Figure 116 Posterior densities of the beta parameter estimates obtained after a Poisson-Gamma regression model was fitted to the roundabout aggregated data set.

From observation of the values of the mean estimates in Table 47 it can be stated that one unitary increase in *ln(AADTmaj)* (see example in section 3.2 for the meaning of unitary increase) increases the expected number of injury accidents by 50% (when the remaining variables have constant values). The same increase in *ln(AADTmin)* increases the expected number of accidents in only 4%. Austria and Dutch roundabouts are *a posteriori* expected to have approximately 33% more and 95% less accidents, respectively, than a Portuguese roundabout when the other variables remain constant.

Table 48 Expected number of accidents per year obtained by a Poisson-Gamma regression model, for the aggregated roundabout junction data set Portuguese junctions as baseline.

Country	Expected Numbers of Accidents
'Austria'	$\hat{\mu}_i = 1.123 \times 10^{-2} \times AADTmaj_i^{0.403} \times AADT\min_i^{0.044}$
'Holland'	$\hat{\mu}_i = 4.323 \times 10^{-4} \times AADTmaj_i^{0.403} \times AADT\min_i^{0.044}$
'Portugal'	$\hat{\mu}_i = 8.426 \times 10^{-3} \times AADTmaj_i^{0.403} \times AADT\min_i^{0.044}$

The equations giving the expected accident frequencies for a one year period in junctions from the aggregated data set are displayed in Table 48, for the three countries. The country with the lowest expected number of injury accidents in roundabouts is the Netherlands and the country with the highest is Austria.

8.2.1 Model Checking

The replicated numbers of injury accidents displayed as histograms and dot plots are shown in Figure 117 and Figure 118, respectively. Overall, it can be stated that this particular Poisson-Gamma model seems able to replicate the observed data reasonably well.

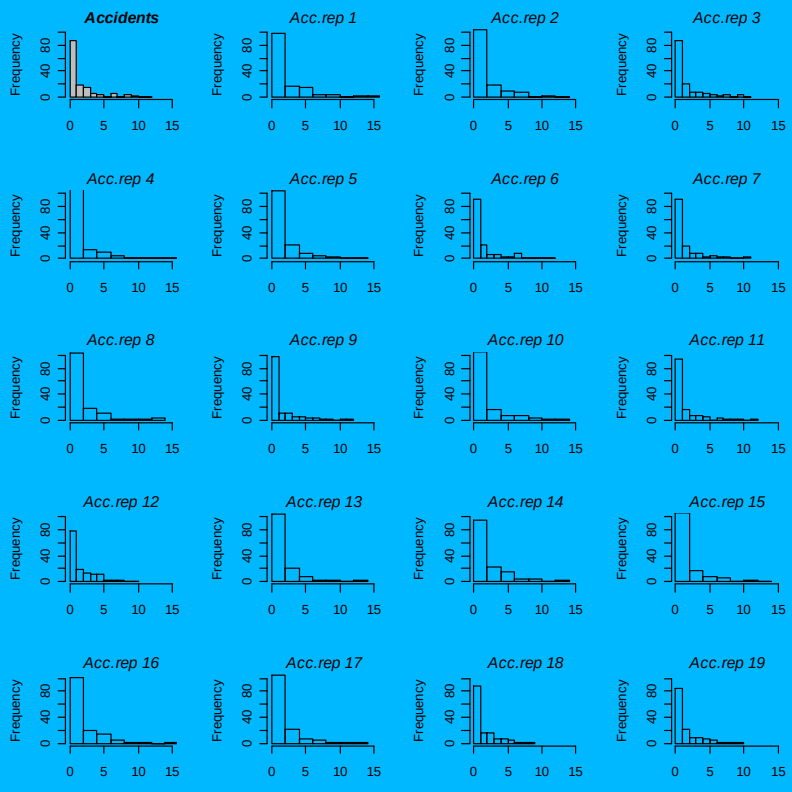

Figure 117 Histogram of the observed number of accidents in the aggregated roundabout data set (left upper corner) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.

The plots of the discrepancy measures are displayed in Figure 119. All the four measures seem to be captured by the Poisson-Gamma model in a satisfactory way (p is close to 0.5).

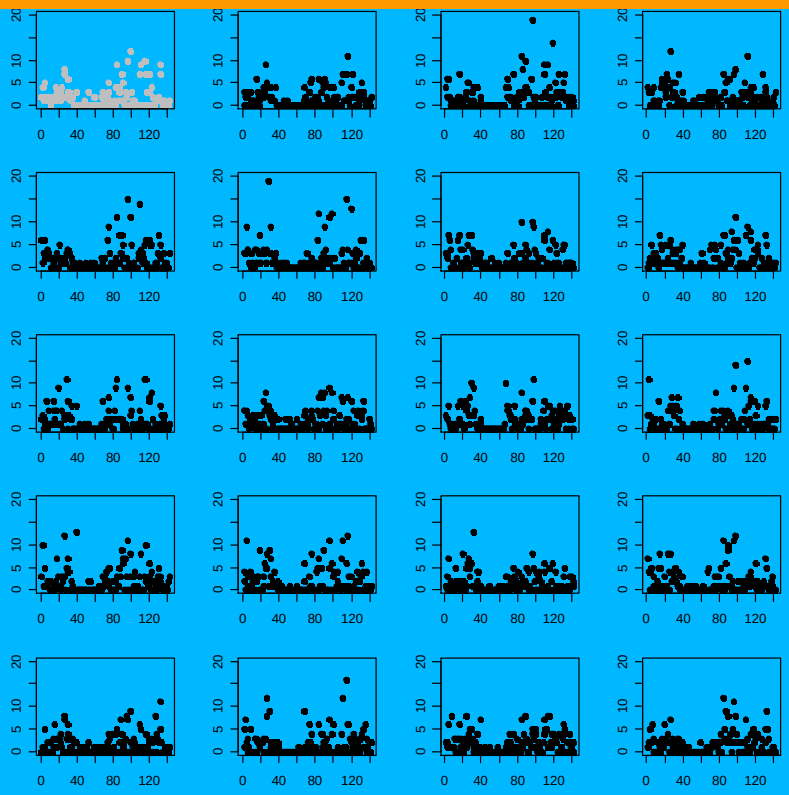

Figure 118 Dot plot of the observed number of accidents in the aggregated set, including roundabouts only (left upper corner) and 19 dot plots of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson-Gamma regression model.

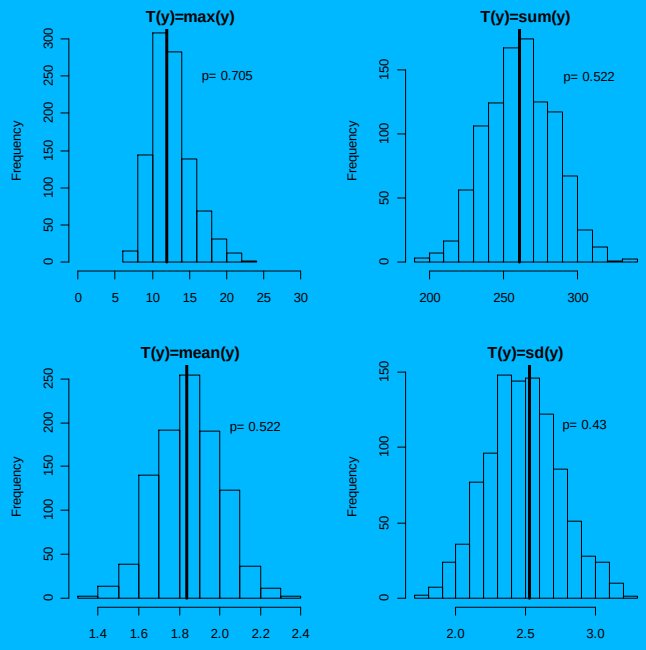

Figure 119 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson-Gamma regression model fit to the aggregated roundabout data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

The same conclusion can be drawn from the observation of the discrepancy measure employed to check whether the model is taking the overdispersion into account. This measure is displayed in Figure 120. It can be concluded that the Poisson-Gamma model is taking the observed data overdispersion into account.

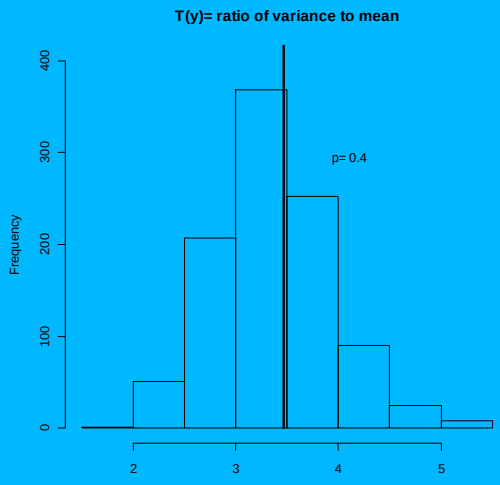

Figure 120 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson-Gamma regression model fit to the aggregated roundabout data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

This model produces an average deviance, $\overline{D}$, of 370.101 and an effective model dimension, d_e, of 55.441, giving a DIC of 425.542. The dispersion parameter ($1/\delta$ in Equation 2.4) was estimated as 0.758. Some of these values will be used for model comparison which is discussed in section 8.4.

The expected numbers of accidents for a one year period for the aggregated roundabout junctions for the minimum, mean, median and maximum profiles were calculated and presented in Table 49.

Table 49 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson-Gamma regression model for the aggregated roundabout accident data.

	Minimum	Mean	Median	Maximum
ln(AADTmaj)	6.908	8.828	8.927	10.190
ln(AADTmin)	6.215	7.731	7.909	9.871
	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
Austria	0.247 (0.068)	0.561 (0.099)	0.589 (0.109)	1.113 (0.398)
Holland	0.009 (0.002)	0.022 (0.006)	0.023 (0.007)	0.045 (0.022)
Portugal	0.187 (0.054)	0.418 (0.059)	0.438 (0.058)	0.803 (0.160)

Given the traffic volumes, a typical Austrian roundabout junction is expected to have 0.561 accidents in one year whereas a typical Dutch roundabout is expected to have only 0.022 and a Portuguese 0.418 in the same period of time.

8.3 The Poisson Log-Normal hierarchical regression model

The Poisson Log-Normal model was fitted to the data according to Equations 2.5 and 2.6 in Chapter 2, where the parameter α in Equation 2.6 had a Gamma(*a*,*b*) *a priori* distribution with *a*=1 and *b*=0.01. The MCMC algorithm was run with three chains for 35000 iterations with 33000 as burn-in with a thinning rate of 6 iterations, resulting in samples with dimension 1002.

The graphs depicted in Figure 121 show the three Gelman-Rubin statistics obtained for the estimates of the β parameters. It can be observed that the *Rhat* statistic converges to 1 in all the parameters as the number of iterations increase. Therefore, there are no reasons to suspect non-convergence of the iterative simulation.

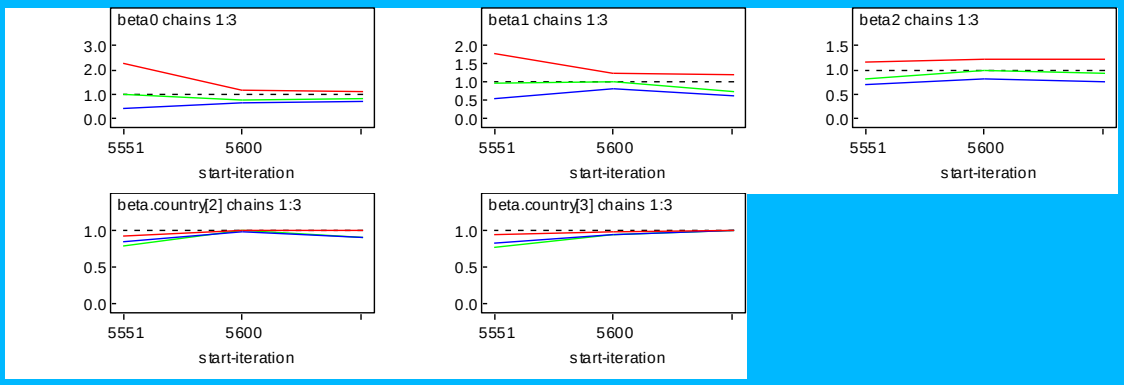

Figure 121 Plots of the Gelman-Rubin statistics corresponding to three Markov chains, for the beta coefficient parameters, from the Poisson Log-Normal regression model fitted to the aggregated junction roundabout data.

The posterior means of the parameter estimates, corresponding standard deviations, Monte Carlo standard errors and 95% credible intervals are displayed in Table 50 and the parameter estimates densities are shown in the graphs of Figure 122.

Table 50 Point estimates, standard deviations, MC errors and 95% credible intervals for the coefficients of the parameters obtained after a Poisson Log-Normal regression model was fit to the aggregated roundabout accident data.

Parameters	mean	s.d.	MC errors	2.5%	97.5%
β_0	-10.380	1.188	1.564E-01	-12.480	-7.703
β_1 (ln(AADTmaj))	0.923	0.148	1.968E-02	0.694	1.221
β_2 (ln(AADTmin))	0.103	0.106	1.333E-02	-0.107	0.303
β_3 (Country=´Austria')	0.521	0.258	1.735E-02	0.013	1.026
β_4 (Country=´Holland')	-2.817	0.487	1.447E-02	-3.776	-1.879

The mean posterior estimates indicate that every unitary increase in *ln(AADTmaj)* and

ln(AADTmin) increases the expected number of injury accidents by approximately 152% and 11%, respectively, when all the other variables remain constant (see example in section 3.2). An Austrian roundabout is expected to have approximately 68% more accidents than a Portuguese roundabout and a Dutch roundabout is expected to have 94% less accidents than a Portuguese one.

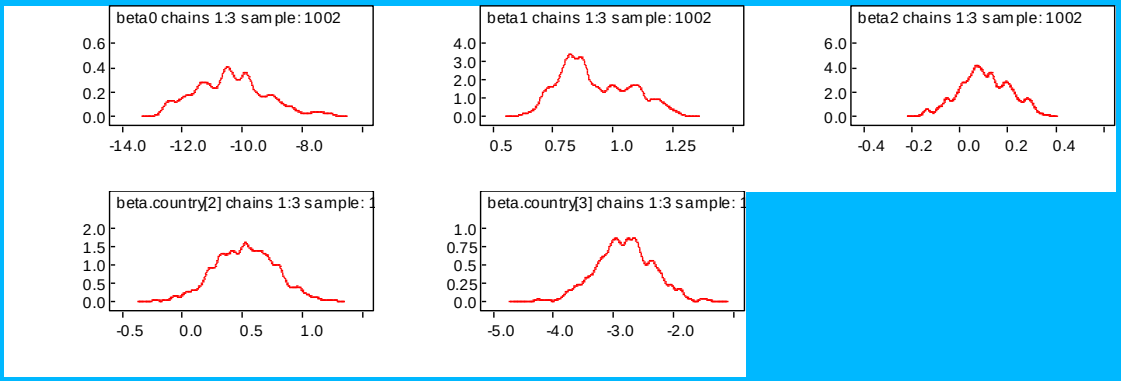

Figure 122 Posterior densities of the beta parameter estimates obtained after a Poisson Log-Normal regression model was fitted to the roundabout aggregated data set.

The equations giving the expected number of injury accidents on roundabouts, per country and per year, are displayed in Table 51. The lowest expected number of injury accidents is to be found at Dutch roundabouts followed by Portuguese and Austrian roundabouts.

Table 51 Expected number of accidents per year obtained by a Poisson Log-Normal regression model, for the aggregated roundabout junction data set.

Country	Expected Numbers of Accidents
'Austria'	$\hat{\mu}_i = 5.207 \times 10^{-5} \times AADTmaj_i^{0.923} \times AADT\ \min_i^{0.103}$
'Holland'	$\hat{\mu}_i = 1.848 \times 10^{-6} \times AADTmaj_i^{0.923} \times AADT\ \min_i^{0.103}$
'Portugal'	$\hat{\mu}_i = 3.092 \times 10^{-5} \times AADTmaj_i^{0.923} \times AADT\ \min_i^{0.103}$

8.3.1 Model Checking

Figure 123 and Figure 124 show histograms and dot plots, respectively, of replicated data by the Poisson Log-Normal model together with the histogram and dot plot of the observed data (i.e. the number of injury accidents). From observation of those figures it can be stated that the model seems to replicate the data satisfactorily.

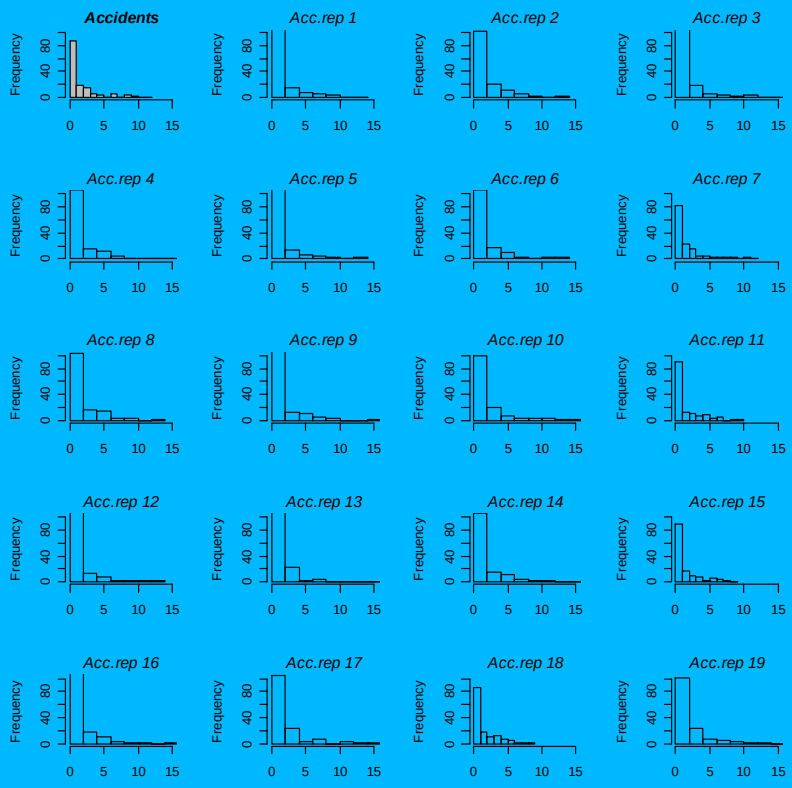

Figure 123 Histogram of the observed number of accidents in the aggregated roundabout data set (left upper corner) and 19 histograms of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

The four discrepancy measures whose results are displayed in Figure 125 show that the model also seems to capture the variations that these measures indicate.

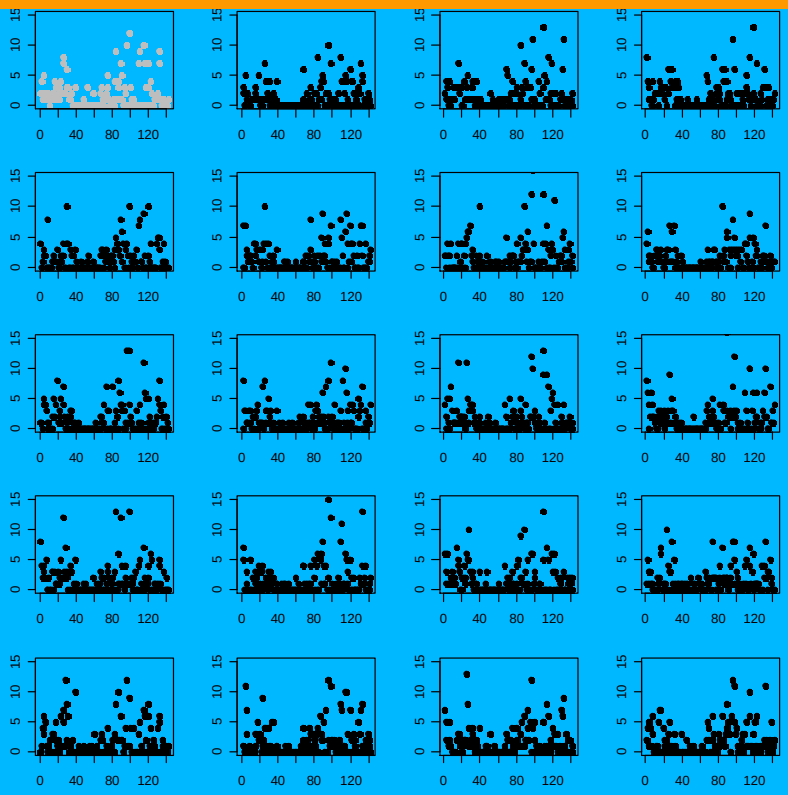

Figure 124 Dot plot of the observed number of accidents in the aggregated set, including roundabouts only (left upper corner) and 19 dot plots of replicated data sets ("Acc.rep") of the numbers of accidents obtained by the posterior predictive distribution according to the Poisson Log-Normal regression model.

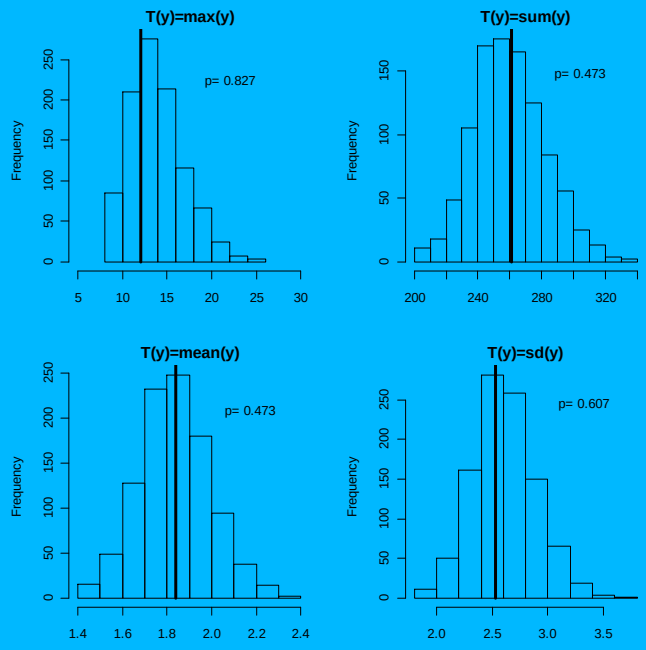

Figure 125 Observed values of four discrepancy measures (vertical lines) compared with histograms of 1002 simulations from the posterior predictive distributions of the same measures obtained by the Poisson Log-Normal regression model fitted to the aggregated roundabout data. The discrepancy measures *T* are: maximum, sum, mean and standard deviation (sd). The *p* is the estimated probability that the measures obtained by the posterior predictive distributions are greater than the ones obtained by the observed data.

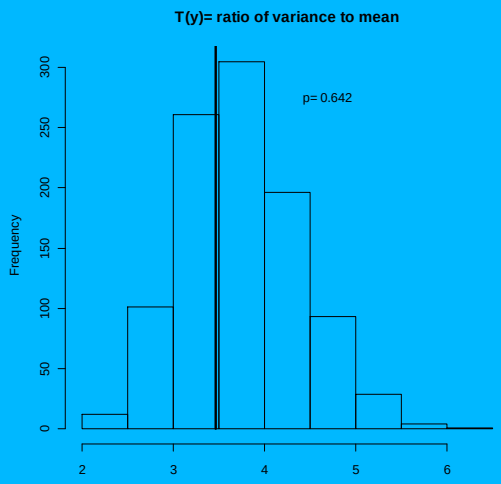

Figure 126 Observed values of a discrepancy measure (vertical line) corresponding to the variance over the mean, compared with a histogram of 1002 simulations from the posterior predictive distribution from the Poisson Log-Normal regression model fitted to the aggregated roundabout data, for the same measure. The *p* gives the estimated probability that the measure obtained by the posterior predictive distributions is greater than the one obtained by the observed data.

The measure of discrepancy suggested by Congdon (2005) to check whether the model takes the data overdispersion into account is depicted in Figure 126. By observation of this figure it can be concluded that the Poisson Log-Normal model is recognising the overdispersion present in the observed data.

This model produces an average deviance, $\overline{D}$, of 367.824 and an effective model dimension, d_e, of 61.481, giving a DIC of 429.305. The dispersion parameter ($1/\alpha$ in Equation 2.6) was estimated as 0. 761.

The posterior means and corresponding standard deviations for the minimum, mean, median and maximum profiles of *ln(AADTmaj)* and *ln(AADTmin)* for the roundabouts from Austria, Holland and Portugal obtained by the Poisson Log-Normal model are displayed in

Table 52.

Given the traffic volumes tested, a typical Austrian roundabout is expected to have 0.409 injury accidents in a one year period whereas a Dutch roundabout is expected to have 0.016 accidents in the same period.

Table 52 Posterior means (standard deviations) of expected number of accidents for minimum, mean, median and maximum profiles obtained by the Poisson Log-Normal regression model for the aggregated roundabout accident data.

	Minimum	Mean	Median	Maximum
ln(AADTmaj)	6.908	8.828	8.927	10.190
ln(AADTmin)	6.215	7.731	7.909	9.871
	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
Austria	0.062 (0.020)	0.409 (0.086)	0.457 (0.098)	1.843 (0.563)
Holland	0.002 (0.001)	0.016 (0.008)	0.018 (0.009)	0.070 (0.035)
Portugal	0.036 (0.012)	0.241 (0.039)	0.269 (0.043)	1.072 (0.249)

8.4 Discussion

The values of $\overline{D}$, d_e and DIC shown in Table 53 for the two models considered in the present chapter suggest that, according with these values, there seems to be no great difference between the fit of the two models.

Table 53 Comparison of DIC and related statistics for the three models fitted to the aggregated roundabout junction data.

Regression Model	$\overline{D}$	d_e	DIC
Poisson-Gamma	370.101	55.441	425.542
Poisson Log-Normal	367.824	61.481	429.305

When taking into account the discrepancy measures obtained by the two models it can be seen that the Poisson-Gamma obtains slightly better results than the Poisson Log-Normal model. Consequently, the Poisson-Gamma model was chosen to represent the aggregated roundabout data set.

9 Conclusions

The aim of the study described in this report was to obtain accident prediction models for junctions situated on the rural road networks of four European countries with the employment of Bayesian statistical methods and techniques.

As part of Workpackage 4 of the RISMET project a framework was created consisting of a set of several variables to be measured and collected on rural junctions belonging to the countries of several of the RISMET project partners. The objective was to obtain a data set formed by data from several European countries that shared the same registered variables. Consequently, an extensive statistical analysis was conducted with the corresponding results discussed for each country as well as allowing for cross country comparisons.

The data analysed consisted of measurements taken at junctions belonging to the rural road

networks of Austria, Holland, Norway and Portugal. Although different time periods were reported for each country (i.e. different numbers of years of measurements for each country), the results were aggregated to represent a period of one year. Injury accident prediction models were obtained for each of the following countries: Austria, Norway and Portugal. The data sets of these three countries together with the Dutch data were also aggregated and analysed as one set taking an explanatory variable indicating the country.

The Dutch data set consisted of a collection of approximately 500 junctions of which only some 50 had traffic volumes on all approaches. Because of this small sample these could not be used for obtaining individual accident prediction models. Consequently, it was decided to employ the data from Holland on the aggregated set formed by the other European countries.

9.1 Model and Model Development

Three regression models were fitted to each data set; they consisted on the Poisson, Poisson-Gamma and the Poisson Log-Normal models. Each model was assessed taking into account measurements of fit and adequacy. The model providing the best fit and therefore, the most appropriate to model the injury accidents in each set of data were then identified.

In each of the three models the dependent variable was taken to be the number of injury accidents registered per junction over a period of time being the explanatory variables the major and minor annual average daily traffic volumes (AADT), the type of junction, number of legs, the traffic control and the speed limit. Due to some differences in measuring and gathering the data within each country, as well as availability of the data already collected, separate and different models were fitted to the country specific data. .

All models were fitted with vague or non-informative prior and hyper-prior distributions. The posterior distributions and the parameter estimates were obtained using Monte Carlo methods and algorithms for sampling from arbitrary densities by implementing MCMC for effective Bayesian computation via the freely available, general purpose computer program for Bayesian statistica inference WinBUGS. Other analyses and several graphs and plots were produced with the R software (The R project for Statistical Computing). The MCMC simulations were obtained with three sequences of Markov chains. The starting values were chosen to be wide apart in the parameter space. The convergence was monitored and verified by observation of the graphs of the parameter's Gelman-Rubin statistics.

The models were validated and checked using posterior predictive values and discrepancy measures that reflect the data attributes and features that should, afterwards, be reflected on the model's replicated data. The discrepancy measurements considered were the maximum, mean, median and standard deviation values as well as the ratio of the variance over the mean of the number of injury accidents to check whether the model took the data's overdispersion into account.

After the data for the various countries were analysed (described in full detail in Chapters 3 to 8) it was concluded that, of all three models considered, the Poisson regression model was found to be the least appropriate for modelling the accidents occurring at junctions of rural road networks in Austria, Norway, Portugal or even the aggregated data set. The main reason is that this regression model does not allow for overdispersion. The hierarchical Poisson-Gamma and Poisson Log-Normal models both provided better results and either of them can be considered appropriate for modelling the accident data of the analysed countries. However, this document proposes a final model from which inferences can be made, the choice which is based on the deviance information criterion (DIC) and posterior predictive checks. Based on that criterion, the Poisson-Gamma regression model is the most suited for modelling accidents at junctions.

9.2 Primary Conclusions per Country

The expected injury accident frequencies were found to increase with increasing values of entering major and minor traffic volumes for all models considered, the increase in the former being more relevant for the increase in accident expectancy. The report also includes the values of the expected number of accidents for each type of junction considered, for the minimum, maximum, mean and median profiles of both major and minor annual average daily traffic volumes for each country and for the aggregated data set for all the countries.

Overall, and for each analysis performed, it can be concluded that the occurrence of injury accidents on Norwegian junctions depends on the number of legs forming the junction (3 or 4) and also on the speed limit of the approaching roads. Junctions with 4 legs have approximately 140% higher accident expectancy than 3 leg junctions, provided the other explanatory variables remain constant. It was also found that junctions with 70km/h approach speed limit have the highest accident expectancy followed by junctions with 60, 80 and 90km/h speed limits (which have similar accident expectancy between them). One possible explanation is that 80 and especially 90km/h limit intersections are located on high standard roads and might have more effective traffic channelization and junction signing than lower speed limit junctions.

The analysis of Austrian junctions found that the type of junction (*roundabout*, *X* or *Y*), as well as the traffic control employed (*stop*, *signalised* or *yield*) affects the injury accident expectancy. Roundabouts have the higher accident expectancy followed by junctions of type *X*. The categories of traffic control by decreasing order of accident expected frequency are: *yield*, *stop* and *signalised*.

Portuguese junctions, like the Norwegian, have higher accident expectancy on 4 leg intersections than on 3 legged ones, approximately 29% more, provided the remaining variables keep constant values. The types of junction (*intersection* and *roundabout*) by decreasing order of accident expectancy are: *intersection* and *roundabouts*. Consequently, Portuguese roundabouts are expected to have lower numbers of accidents as the other types, which did not seem to happen for Austrian roundabouts where this type of junction was the one with higher accident expectancy.

The analysis of the data set consisting on the aggregated non-roundabout junctions from Austria, Norway and Portugal showed that the expected accident frequency is higher (approximately 48% more) on 4 leg junctions than on 3 legged ones (provided the remaining variables were kept constant). The country indicator variable influences the injury accident frequencies. The countries expected accident frequencies by decreasing order are: Portugal, Austria and Norway, with Portugal and Austria non-roundabouts expected to have approximately 82% and 97%, respectively, higher accident expectancies than a Norwegian non-roundabout (all other variables remaining constant).

The results obtained from the analysis of the aggregated set of Austrian, Dutch and Portuguese roundabout junctions showed that an unitary increase in the logarithm of major and minor annual average daily traffic volume increases the expected number of accidents by approximately 50% and 4%, respectively, and that Austrian and Dutch roundabouts are expected to have approximately 33% more and 95% less number of injury accidents, respectively, than a Portuguese roundabout, provided all other variables remain constant. As an example, suppose a junction from this particular aggregated set has a value of *AADTmaj* equal to 8500, consequently *ln*(*AADTmaj*) is equal to approximately 9.048. The same junction with 10.048 for *ln*(*AADTmaj*), i.e. a unitary increase (corresponding to an *AADTmaj* value of 23109.52) is expected to increase the number of injury accidents by approximately 50%, provided all other variables suffer no change.

9.3 General

As a result of all the work that was performed within this particular task it is recommended to

use model fits using a Bayesian approach as these have the advantage of taking into account previous recorded information when modelling road accident traffic events, in particular in rural intersections. In addition, the use of Bayesian techniques has the advantage of providing not only estimates for the regression coefficients but also a density function for those coefficients. Nevertheless, the model assessment should also imply the use of the appropriate Bayesian techniques as exemplified in the analysis described in this deliverable.

As future work, it would be very useful and interesting to study with more emphasis and detail the regression model forms that could be more appropriate to each country and also to the several aggregated country data sets.

The wealth of data collected and gathered within the RISMET project tasks allows for further analysis and consequent conclusions to be drawn, amongst others, on the several types of accidents and victims injuries of the several European countries and consequent cross-country comparisons.

Sources

Cameron, A.C. and Trivedi, P.K. (1998) – Regression Analysis of Count Data. Cambridge University Press.

Carlin, B. and Louis, T. (2009) – Bayesian Methods for Data Analysis. 3rd Edition, Chapman & Hall/CRC, Texts in Statistical Sciences, Florida.

Congdon, P. (2005) – Bayesian Models for Categorical Data. Wiley Series in Probability and Statistics. Wiley, Chichester, U.K.

Congdon, P. (2006) – Bayesian Statistical Modeling. Wiley Series in Probability and Statistics, 2nd edition, Wiley, Chichester, U.K.

Congdon, P. (2010) – Applied Bayesian Hierarchical Methods. Chapman & Hall/CRC.

Eenink, R., Reurings, M., Elvik, R., Cardoso, J., Wichert, S. and Stefan, C. (2007) – Accident Prediction Models and Road Safety Impact Assessment: Results of the pilot studies. RIPCORD-ISEREST report.

Elvik, R. (2010) – Assessment and applicability of evaluation tools: current practise and state of the art in Europe. RISMET Deliverables 4 and 5.

Fahrmeir, L. and Osana, (2006) – Structured additive regression for overdispersed and zero-inflated count data. Applied Stochastic Models in Business and Industry. 22, pp: 351-369.

Gelfand, A. (1996) – Model determination using sampling-based methods. In Markov Chain Monte Carlo in Practise, Gilks, W., Richardson, S. and Spiegelhalter, D. (eds). Chapman & Hall: London, 145-161.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004) – Bayesian Data Analysis. Texts in Statistical Science, 2nd edition, Chapman & Hall, London, U.K.

Gelman, A. and Hill, J. (2007) –Data Analysis Using Regression and Multilevel/Hierarchical Models. (Analytical Methods for Social Research), Cambridge University Press, NY.

Hauer, H. (1997) – Observational before-after studies in road safety: estimating the effect of highway and traffic engineering measures on road safety. Pergamon, Elsevier Science Ltd., Oxford, UK.

Hauer, E. (2002) – Estimating safety by the empirical Bayes method: a tutorial. Transportation Research Record. 1784: 27-31.

Jackman, S. (2009) – Bayesian Analysis for the Social Sciences. Wiley Series in Probability and Statistics. Wiley, Chichester, U.K.

Lord, D. (2006) – Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis and Prevention, vol. 38, No. 4, pp. 751-766.

Lord, D. & Miranda-Moreno, L. (2008) – Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-Gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science. 46 (5): 751-770.

Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. (2000) – WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10, pp. 325-337.

Miaou, S.-P., Lord, D. (2003) – Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes. Transport. Res. Rec., 1840, pp. 31-40.

Ntzoufras, I. (2009) – Bayesian Modeling Using WinBUGS. Wiley Series in Computational Statistics, Wiley, New Jersey.

Park, B. and Lord, D. (2008) – Adjustment of the maximum likelihood estimate of the negative binomial dispersion parameter. Transportation Research Record. 2061, pp. 9-19.

Park, B. & Lord, D. (2009) – Application of finite mixture models for vehicle crash data analysis. Accident Analysis and Prevention. 41 (4): 683-691.

R Development Core Team (2011) – R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Schermers, G. and Elvik, R. (2009) – Safety at the Heart of Road Design. RISMET Description of Work.

Smith, A.F.M. and Gelfand, A.E. (1992) – Bayesian Statistics Without Tears: A Sampling-Resampling Perspective. The American Statistician, Vol. 46, No. 2, pp. 84-88.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002) – Bayesian Measures of Model Complexity and Fit. Journal of the Royal Statistical Society, B, 64, Part 4, pp. 583-639.

Spiegelhalter, D.J., Thomas, D., Best, N.G. and Lunn, D. (2003) – WinBUGS Version 1.4 User Manual. MRC Biostatistics Unit, Cambridge, U.K. http://www.mrc-cam.ac.uk/bugs.

Sturtz, S., Ligges, U. and Gelman, A. (2005) – R2WinBUGS: A Package for Running WinBUGS from R. Journal of Statistical Software, 12 (3), pp. 1-16.

Venables, W.N. and Ripley, B.D. (2002) – Modern Applied Statistics with S. Fourth Edition. Springer-Verlag, New York.

Zhang, Y., Ye, Z. & Lord, D. (2007) – Estimating the dispersion parameter of the negative binomial distribution for analyzing crash data using a bootstrapped maximum likelihood method. Transportation Research Record. 2019: 15-21.