Rob Flohr: Cursus inleiding Bayesiaanse statistiek

Het wekelijkse aantal verkeersongevallen op de snelweg A.. tussen de plaatsen B. en C. volgt een Poisson-verdeling. Vier studenten gaan de komende acht weken het aantal verkeersongevallen op dat gedeelte van die snelweg vaststellen en de data op een Bayesiaanse manier analyseren.
Op basis van gegevens uit het verleden gaan ze uit van een gemiddeld aantal verkeersongevallen van 2.5 per week met een standaarddeviatie van 1. Ze drukken deze 'prior belief' uit middels een gamma-verdeling.
Nadat de acht weken zijn verstreken blijkt het aantal verkeersongevallen per week als volgt te zijn:

3, 2, 0, 8, 2, 4, 6, 1

Resultaat van een Bayesiaanse data-analyse:
Op basis van gegevens uit het verleden (gemiddeld aantal verkeersongevallen van 2.5 per week met een standaarddeviatie van 1) en van recente data (3, 2, ..., 6, 1) kunnen we afleiden dat de kans 95% is dat het aantal wekelijkse ongevallen zal liggen tussen 2.1 en 4.2.

Hieronder staat de uitwerking:

a)
langs analytische/theoretische weg, op basis van conjugacy, zonder software
b)
met behulp van MCMC (Markov Chain Monte Carlo) simulaties

Aanwijzing:
Om de parameters (α , β) voor een gamma-prior te vinden kun je de volgende formules gebruiken:

$$gemiddelde = \frac{\alpha}{\beta}$$

$$\text{var}\,iantie = \frac{\alpha}{\beta^2}$$

$$\frac{\alpha}{\beta} = 2.5$$

$$\frac{\alpha}{\beta^2} = 1 \rightarrow \alpha = \beta^2 \rightarrow \beta = 2.5 \wedge \alpha = 6.25$$

a)

Aanwijzing:een prior $gamma(\alpha,\beta)$ met y_i successen (n = aantal successen) geeft een posterior $gamma(\alpha + \sum_{i=1}^{n} y_i, \beta + n)$.

In het voorbeeld:
gamma(6.25, 2.5) gecombineerd met de data(successen) (3, 2, 0, 8, 2, 4, 6, 1)
geeft een gamma posterior ((6.25 + som data=6.25+26=32.25), (2.5 + aantal successen = 2.5+8=10.5)), dus gamma(32.25,10.5).
Gemiddelde: 32.25/10.5 = 3.07143, variantie = 32.25 / (10.5)^2 = 32.25/110.25 = 0.292517 -> stand. deviatie = 0.540848

Credible intervals met R:
99% credible interval voor mu: 1.9546 / 4.4676
>qgamma(0.01,32.25,10.5)
1.954566
>qgamma(0.99,32.25,10.5)
4.467603

95% credible interval voor mu: 2.1043 / 4.2186
>qgamma(0.025,32.25,10.5)
2.104253
>qgamma(0.975,32.25,10.5)
4.218608

b)

```
> local({pkg <- select.list(sort(.packages(all.available =
TRUE)),graphics=TRUE)
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
Loading required package: coda
Loading required package: lattice
Loading required package: MASS
##
## Markov Chain Monte Carlo Package (MCMCpack)
## Copyright (C) 2003-2014 Andrew D. Martin, Kevin M. Quinn, and Jong
Hee Park
##
## Support provided by the U.S. National Science Foundation
## (Grants SES-0350646 and SES-0350613)
##
Warning messages:
1: package 'MCMCpack' was built under R version 2.15.3
2: package 'coda' was built under R version 2.15.3
3: package 'lattice' was built under R version 2.15.3
> library(MCMCpack)
> verkeersongevallen=c(3,2,0,8,2,4,6,1)
> posterior=MCpoissongamma(verkeersongevallen,6.25,2.5,5000)
> summary(posterior)

Iterations = 1:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
```

```
   plus standard error of the mean:

          Mean               SD        Naive SE Time-series SE
      3.065579         0.535144        0.007568       0.007568

2. Quantiles for each variable:

 2.5%    25%    50%    75%  97.5%
2.092  2.691  3.031  3.410  4.185

> plot(posterior)
>
```