Underdispersion in count data

Eugenio Bonifazi

Contents

1   Dispersion in count data

1.1   Introduction

In data analysis and in many disciplines, count data have become widely avaliable. The most popular distribution for modelling count data is the Poisson distribution but still, many real data do not adhere to the assumption of equidispersion that is underlined in the Poisson distribution. As a matter of fact, the Poisson distribution:

$$f_x(x) = P(X = x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

has the expected value equal to the variance, $E(X) = Var(X) = \lambda$ with $X \sim Pois(\lambda)$.

However, as mentioned above, in real application count data frequently exhibit underdispersion or overdispersion, which is the most frequent relashionship recognized among real datasets. For this reason overdispersion has been long studied and also its manners are generally well understood, including its causes and various distributions established that allow to model overdispersed data in an appropriate way (like the Negative Binomial). Overdispersion describes higher variance than expected, while underdispersion, on its turn, is related to a deficient variation in count data (Sellers and Morris 2017).

This work focuses on underdispersion and it's constitued by three section: this introductory section on underdispersion; a second part in which we'll present an algorithm that allows you to make bayesian inference on COM-Poisson distribution and regression and, in the end, we'll apply this algorithm to estimate COM-Poisson's posterior parameters as well as regression's coefficients. In the session below a list of distributions, mainly derived from the Poisson, that address underdispersion are presented. This list includes a one parameter distribution, the condensed Poisson, and some distributions with two parameters that give us the capability to estimate a scale and a dispersion parameter separetely (Sellers and Morris 2017 has been taken as reference).

1.2 Distributions that address underdispersion

- **Condensed Poisson**

The condensed poisson (CP) is a one-parameter count data distribution that express data underdispersion. It is used to examine the number of events occurring over a certain period of time. Assuming that the interarrival time follows an $Erlang(2, \lambda)$ then, Y has a $CP(\lambda)$ distribution with probability mass function:

$$P(Y = y \mid \lambda) = \begin{cases} e^{-\lambda}\left(1 + \frac{\lambda}{2}\right) & y = 0 \\ e^{-\lambda}\left(\frac{\lambda^{2y-1}}{2(2y-1)!} + \frac{\lambda^{2y}}{(2y)!} + \frac{\lambda^{2y+1}}{2(2y+1)!}\right) & y = 1, 2, 3, \dots \end{cases}$$

The close forms of mean and variance is:

$$E(Y) = \frac{\lambda}{2} \quad \text{and} \quad Var(Y) = \frac{\lambda + e^{-\lambda}sinh(\lambda)}{4}$$

By definition the variance is less than the mean for all λ.

There are several two-parameter distributions that have the Poisson model as special case and have the ability to model both over- and underdispersion through the associated dispersion parameter. Below some examples.

- **Generalized Poisson**

Famoye and Consul (1995) introduced the generalized Poisson (GP) distribution with pmf:

$$P(Y = y \mid \mu, \theta) = \begin{cases} \frac{\mu(\mu + \theta)^{y-1}e^{-\mu - \theta y}}{y!} & y = 0, 1, 2, \dots \\ 0 \quad \text{for } y > m, \quad \text{when } \theta < 0 \end{cases}$$

where $\mu > 0$ and $max(-1, \mu/4) < \theta \le 1$.

The direct proportionality restriction, $\theta = \phi\mu$, produces a restricted generalized Poisson, $RGP(\mu, \phi)$, with rate paramter μ and dispersion parameter ϕ. A random variable $Y \sim RGP(\mu, \phi)$ has the probability mass function:

$$P(Y = y \mid \mu, \phi) = \begin{cases} \frac{\mu^y(1+\phi y)^{y-1}e^{-\mu(1+\phi y)}}{y!} & y = 0, 1, 2, \dots \\ 0 \quad \text{for } y > m, \quad \text{when } \phi < 0 \end{cases}$$

This distribution has the Poisson as special case, when $\phi = 0$. But, of course, it handles data over- or underdispersion for $\phi > 0$ or $\frac{-2}{\lambda} < \phi < 0$ respectively. The RGP model has a closed form mean and variance:

$$E(Y) = \lambda \quad \text{and} \quad Var(Y) = \lambda(1 + \lambda\phi)^2$$

- **Gamma Count**

Winkelmann (2008) introduced a gamma count distribution based on a difference of incomplete gamma functions. More precisely, $Y \sim GC(\lambda, \alpha)$ has pmf:

$$P(Y = y \mid \alpha, \lambda) = Gamma(\alpha y, \lambda) - Gamma(\alpha y + \alpha, \lambda) \qquad y = 0, 1, 2, \dots$$

where

$$Gamma(\alpha y, \lambda) = \begin{cases} 1 & y = 0 \\ \frac{1}{\Gamma(\alpha y)}\int_0^\lambda u^{\alpha y - 1}e^{-u}du & y = 1, 2, 3, \dots \end{cases}$$

When $\alpha = 1$ we are in the special case of the $Poisson(\lambda)$. This model, on its turn, captures data over- and underdispersion when $0 < \alpha < 1$ or $\alpha > 1$ respectively. The GC has not closed form moments, but we can compute them as:

$$E(Y) = \sum_{y=1}^{\infty} y \cdot Gamma(\alpha y, \lambda)$$

$$Var(Y) = \sum_{y=1}^{\infty} y^2 \cdot [Gamma(\alpha y, \lambda) - Gamma(\alpha y + \alpha, \lambda)] - (E(Y))^2.$$

- **Double Poisson**

By definition, $Y \sim DP(\lambda, \theta)$ is a double Poisson with probability mass function:

$$P(Y = y \mid \lambda, \theta) = c(\lambda, \theta)(\theta^{-1/2} e^{-\theta\lambda}) \left(\frac{e^{-y} y^y}{y!} \right) \left(\frac{e\lambda}{y} \right)^{\theta y} \qquad y = 0, 1, 2, ...$$

where

$$\frac{1}{c(\lambda, \theta)} = \sum_{y=0}^{\infty} (\theta^{-1/2} e^{-\theta\lambda}) \left(\frac{e^{-y} y^y}{y!} \right) \left(\frac{e\lambda}{y} \right)^{\theta y} \approx 1 + \frac{1-\theta}{12\lambda\theta} \left(1 + \frac{1}{\lambda\theta} \right)$$

defines the normalizing constant. The double Poisson, likewise the previous examples, has the $Poisson(\lambda)$ for $\theta = 1$ and models over- or underdispersion for $0 < \theta < 1$ or $\theta > 1$ respectively. It has mean and variance equal to:

$$E(Y) = \lambda \quad \text{and} \quad Var(Y) = \frac{\lambda}{\theta}$$

- **Conway-Maxwell-Poisson**

The Conway-Maxwell-Poisson (COM-Poisson) distribution is a generalization of the Poisson distribution that has pmf:

$$P(Y = y \mid \mu, \nu) = \left(\frac{\mu^y}{y!} \right)^{\nu} \cdot \frac{1}{Z(\mu, \nu)}, \qquad y = 0, 1, 2, ...$$
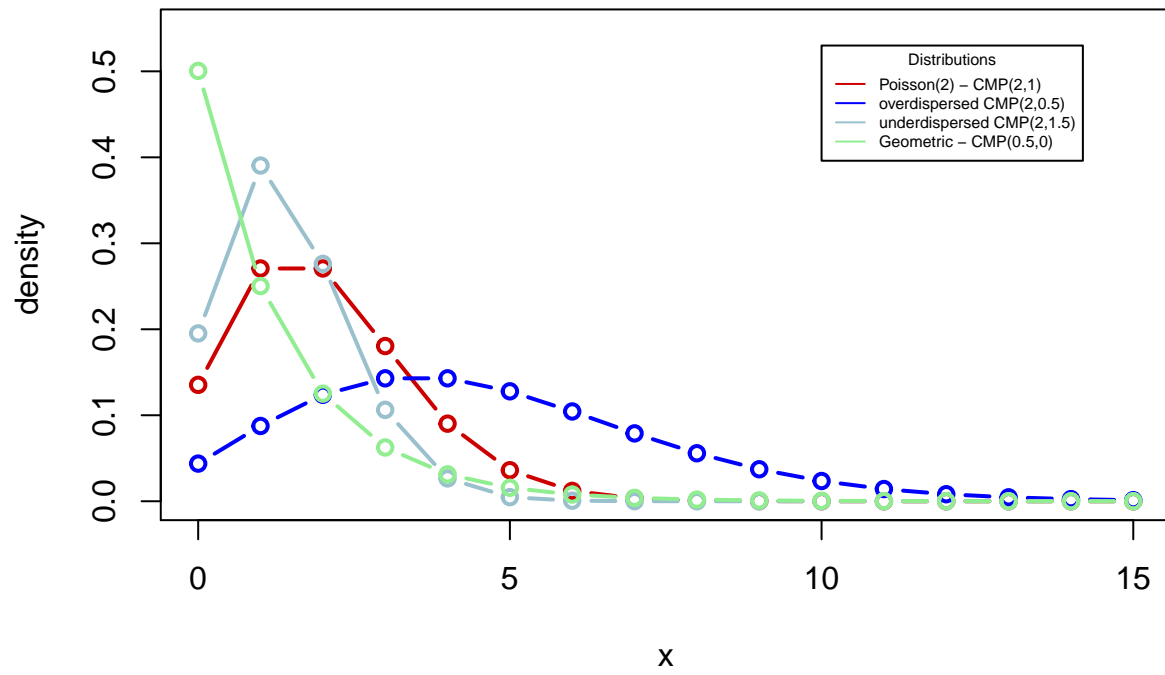
where $Z(\mu, \nu) = \sum_{i=0}^{\infty} \left(\frac{\mu^i}{i!} \right)^{\nu}$ is a normalizing constant. For a given $Y \sim CMP(\mu, \nu)$ we have $\mu = E(Y^{\nu})$ and the dispersion parameter, ν, that expresses the amount of over- or underdispersion, $\nu < 1$ and $\nu > 1$ respectively. On the other hand, $\nu = 1$ implies that equidispersion exists and this entails that, for such value of ν, the COM-Poisson is equal to the Poisson distribution. The Conway-Maxwell-Poisson distribution covers other well known distributions such as gemetric ($\nu = 0$, $\mu < 1$) and Bernoulli ($\nu \to \infty$ with probability $\frac{\mu}{1+\mu}$), some example are shown in a plot at the end of this paragraph.

Taking into account the original parametrization of the COM-Poisson, that can be obtained by replacing μ by $\lambda^{\frac{1}{\nu}}$, the moment generating function of Y is $M_Y(t) = E(e^{Yt}) = \frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)}$, from which we can obtain the moments and, in particular, expected value and variance:

$$E(Y) = \lambda \frac{\partial log Z(\lambda, \nu)}{\partial \lambda} \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}; \quad Var(Y) = \frac{\partial E(Y)}{\partial log\lambda} \approx \frac{1}{\nu} \lambda^{1/\nu}$$

The normalizing constant $Z(\mu, \nu)$ doesn't have a close form and it has to be approximated. Hence not only the regression's coefficients, for both frequentist and Bayesian approaches, but also sampling from the parameter's posterior distribution requires the evaluation of the normalizing constant $Z(\mu, \nu)$. However, there are some methods that allow us to sample from a certain distribution without evaluating the normalizing constant and, in the following paragraph, we'll take into account one of them.

Particular cases of COM−Poisson

Figure showing density plots of particular cases of COM−Poisson distributions. Legend (Distributions): Poisson(2) − CMP(2,1) (red); overdispersed CMP(2,0.5) (blue); underdispersed CMP(2,1.5) (light blue); Geometric − CMP(0.5,0) (green).

2 Exchange algorithm

This section exhibits a method for sampling from the COM-Poisson distribution without having to evaluate the normaliziong constant. We consider the exchange algorithm proposed by Møller et al. (2006) and used by C. Chanialidis et al. (2018). This algorithm is a version of the Metropolis-Hastings algorithm designed for cases in which the likelihood function contains an intractable normalizing constant that is a function of the parameters. This algorithm is presented below and it is also compared with the standard Metropolis-Hastings.

2.1 Exchange algorithm vs Metropolis Hastings

The exchange algorithm has the same procedure of the Metropolis Hastings and they are both Monte-Carlo Markov Chain based; however they differ for the structure of the acceptance probability, but let's proceed step by step. As we know, having a sampling model with probability mass function $p(y \mid \theta)$ conditioned on a parameter θ that has prior distribution $\theta \sim p(\theta)$, the **Metropolis Hastings algorithm** proceeds as follows:

1. Initiate $\theta^{(0)} \sim p(\theta)$ Then, **for** each iteration $i = 1, 2, ..., n$:
2. Generate $\theta^\star \sim p(\theta)$
3. Take $r = min\{a, 1\}$ with

$$a = \frac{p(y \mid \theta^\star)p(\theta^\star)}{p(y \mid \theta^{i-1})p(\theta^{i-1})}$$

4. Generate $u \sim Unif(0, 1)$, if $u < r$:

$$\theta^{(i)} = \theta^\star$$

else:

$$\theta^{(i)} = \theta^{i-1}$$

This is the standard Metropolis Hastings. If we are dealing with a sampling model with probability mass function $p(y \mid \theta) = \frac{q_\theta(y)}{Z(\theta)}$ in which $q_\theta(y)$ is the unnormalized probability density and the normalisation constant $Z(\theta) = \sum_y = q_\theta(y)$ or $Z(\theta) = \int q_\theta(y)dy$ is unknown, the acceptance probability becomes:

$$a = \frac{\left(\prod_i \frac{q_{\theta^\star}(y^\star)}{Z(\theta^\star)} \right)p(\theta^\star)}{\left(\prod_i \frac{q_\theta(y)}{Z(\theta)} \right)p(\theta)}$$

The idea behind the exchange algorithm is to enlarge the state of the Markov chain to include, beside the parameter θ, an auxiliary variable $y^\star$ defined in the same sample space of $y = (y_1, ...y_n)$. For each update of the MCMC, first a candidate parameter $\theta^\star$ is generated from the proposal distribution of the parameter θ (that is supposed to be a simmetric distribution); then the auxiliary data $y^\star$ are drown from the sampling model $p(y^\star \mid \theta^\star)$, conditional on the parameter value (C. Chanialidis et al. 2018). Hence, in the exchange algorithm we have:

$$a = \frac{p(y \mid \theta^\star)p(\theta^\star)p(y^\star \mid \theta)}{p(y \mid \theta)p(\theta)p(y^\star \mid \theta^\star)}$$

$$= \frac{\left[\prod_i \frac{q_{\theta^\star}(y_i)}{Z(\theta^\star)} \right]p(\theta^\star)\left[\prod_i \frac{q_\theta(y^\star)}{Z(\theta)} \right]}{\left[\prod_i \frac{q_\theta(y_i)}{Z(\theta)} \right]p(\theta^\star)\left[\prod_i \frac{q_{\theta^\star}(y^\star)}{Z(\theta^\star)} \right]}$$

$$= \frac{\left[\prod_i q_{\theta^\star}(y_i) \right]p(\theta^\star)\left[\prod_i q_\theta(y^\star) \right]}{\left[\prod_i q_\theta(y_i) \right]p(\theta)\left[\prod_i q_{\theta^\star}(y^\star) \right]}$$

Remarking the difference between the two algorithms we can see how the acceptance ratio a for the standard Metropolis-Hastings involves the ratio of the normalization constants $\frac{Z(\theta)}{Z(\theta^\star)}$, while in the exchange algorithm the ratio $\frac{Z(\theta)}{Z(\theta^\star)}$ cancels out and it is replaced by $\frac{q_\theta(y_i^\star)}{q_{\theta^\star}(y_i^\star)}$; this suggests that the latter can be thought as an importance sampling estimate of the former.

2.2   The COM-Poisson case

In this part we'll take into account more closely the exchange algorithm in the case considered in this work, i.e. COM-Poisson's parameters estimate, giving some more specifications about the model and the distributions involved in the procedure. First we have to specify the distribution of the unnormalized probability densities that are nothing more than the COM-Poisson density without the normalizing constant:

$$q_{\theta_i}(y_i) = \left(\frac{\mu_i^{y_i}}{y_i!}\right)^{\nu_i}, \quad q_{\theta_i^\star}(y_i) = \left(\frac{(\mu_i^\star)^{y_i}}{y_i!}\right)^{\nu_i^\star}$$

$$q_{\theta_i}(y_i^\star) = \left(\frac{\mu_i^{y_i^\star}}{y_i^\star!}\right)^{\nu_i}, \quad q_{\theta_i^\star}(y_i^\star) = \left(\frac{(\mu_i^\star)^{y_i^\star}}{y_i^\star!}\right)^{\nu_i^\star}$$

Of course these probability densities come from the same distribution but they refer to all the possible values this distribution assumes in the exchange algorithm's acceptance probability that are, from left to right, respect to: un-updated values of the parameter and the vector of data, updated parameters and observed data, un-updated parameters and simulated data $(y^\star)$ and updated parameters, $\theta^\star$, and simulated data, respectively.

As we previously said, with the exchange algorithm we want to estimate the posterior distribution via MCMC sampling without computing the likelihood function (that is computationally hard because of the normalisation constant). We clarified the part related to the likelihood, namely the COM-Poisson probability distribution, and now is the turn the parameters' prior distribution. Indeed the Conway-Maxwell-Poisson distribution depends on two parameters, μ and ν. Assuming $\mu \sim p(\mu)$ and $\nu \sim p(\nu)$ and the indipendence of the two parameter's prior distributions, we finally have:

$$a = \frac{\left[\prod_i q_{\theta^\star}(y_i)\right] p(\mu^\star) p(\nu^\star) \left[\prod_i q_\theta(y^\star)\right]}{\left[\prod_i q_\theta(y_i)\right] p(\mu) p(\nu) \left[\prod_i q_{\theta^\star}(y^\star)\right]}$$

where $\theta = (\mu, \nu)$.

3 Applications

The last part of this work is composed by two sections: in the first one we'll provide two applications of the exchange algorithm, the former applying the algorithm on random generated data and the latter using real data; in the second section we'll apply the exchange algorithm in order to estimate regression's coefficients. For the second analysis I used the R package *combayes* implemented by Chanialidis (Chanialidis 2017), still following the idea of the exchange algorithm but, in this case, used to compute an estimation of the regression's coefficients (we'll go into the details further).

3.1 Posterior COM-Poisson's parameters estimate

The implementation of the algorithm proposed below has been designed just for the case of a $CMP(\mu, \nu)$ likelihood distribution with normal prior on the parameters' distribution, that are: $\mu \sim N(\mu_{n_1}, \tau_{n_1}^2)$ and $\nu \sim N(\mu_{n_2}, \tau_{n_2}^2)$; according with the Metropolis Hastings' hypothesis of symmetric prior distribution. The functions a_ratio and q are the acceptance probability and the unnormalized density, respectively:

$$a = \frac{\left[\prod_i q_{\theta^\star}(y_i)\right] p(\mu^\star) p(\nu^\star) \left[\prod_i q_\theta(y^\star)\right]}{\left[\prod_i q_\theta(y_i)\right] p(\mu) p(\nu) \left[\prod_i q_{\theta^\star}(y^\star)\right]}, \quad q(y \mid \mu, \nu) = \left(\frac{\mu^y}{y!}\right)^\nu$$

Below the R functions' code.

```
#   exchange.alg <- function(nsim, y, likelihood_par=list(n, mu, nu),
#                       prior.mu_par=list(mu, sd), prior.nu_par=list(mu, sd)){
#
#   acc = 0
#   mu = rep(NA,nsim)
#   nu = rep(NA,nsim)
#   yres = rep(NA,nsim)
#   yres[1] = rcomp(likelihood_par$n, likelihood_par$mu, likelihood_par$nu)
#   mu_cand = 0
#   nu_cand = 0
#   mu[1] = mu_cand
#   nu[1] = nu_cand
#
#   for(i in 2:nsim){
#      mu_cand = rnorm(n=1, prior.mu_par$mu, prior.mu_par$sd)
#      nu_cand = rnorm(n=1, prior.nu_par$mu, prior.nu_par$sd)
#      while (mu_cand<=0){
#         mu_cand = rnorm(n=1, prior.mu_par$mu, prior.mu_par$sd)
#      }
#      while (nu_cand<=0){
#         nu_cand <- rnorm(n=1, prior.nu_par$mu, prior.nu_par$sd)
#      }
#
#      y_star = rcomp(likelihood_par$n, likelihood_par$mu, likelihood_par$nu)
#
#      a = a_ratio(y, y_star, mu[i-1], nu[i-1], mu_cand, nu_cand,
#                  prior.mu_par=prior.mu_par, prior.nu_par=prior.nu_par)
#
#
#      r = min(exp(a),1)
#      u = runif(1)
```

```
#
#      if(u<r){
#         mu[i] = mu_cand
#         nu[i] = nu_cand
#         yres[i] = y_star
#         acc = acc + 1/nsim
#      }
#      else{
#         mu[i] = mu[i-1]
#         nu[i] = nu[i-1]
#         yres[i] = yres[i-1]
#      }
#   }
#   result = list(mu_acc = mu, nu_acc = nu, yres = yres, acc = acc*100)
#   return(result)
# }

#   a_ratio <- function(y, y_star, mu, nu, mu_cand, nu_cand,
#   prior.mu_par=list(mu, sd), prior.nu_par=list(mu, sd)){
#
#     a = sum(log(q(y, mu_cand, nu_cand))+log(q(y_star, mu, nu))
#             + log(dnorm(mu_cand, prior.mu_par$mu, prior.mu_par$sd))
#             + log(dnorm(nu_cand, prior.nu_par$mu,prior.nu_par$sd))
#             -log(q(y, mu, nu)) - log(q(y_star, mu_cand, nu_cand))
#             - log(dnorm(mu, prior.mu_par$mu, prior.mu_par$sd))
#             - log(dnorm(nu, prior.nu_par$mu,prior.nu_par$sd)))
#   }
#
# q <- function(y, mu, nu) (mu^y/factorial(y))^nu
```

3.1.1   Simulated data

In this first example we have a vector of data, $y = y_1, ..., y_n$ generated from a COM-Poisson $Y \sim CMP(2.5, 5)$ of size $n = 1500$. This is an underdispersed sample; indeed, we have $\bar{y} = 1.12$ and $Var(y) = 0.61$. We want to see if we can estimate the true posterior prameters of the Conway-Maxwell-Poisson, μ and ν, via exchange algorithm choosing suitable priors on the parameters' distributions.

The algoritm is applied three times with different priors on the parametars:

1. We choose completely uninformative priors, i.e.:

$$\mu \sim N(0, 100), \quad \nu \sim N(1, 100)$$

2. Normal priors centered on the true-value of the sampling distribution, with small variabilily with respect to it:
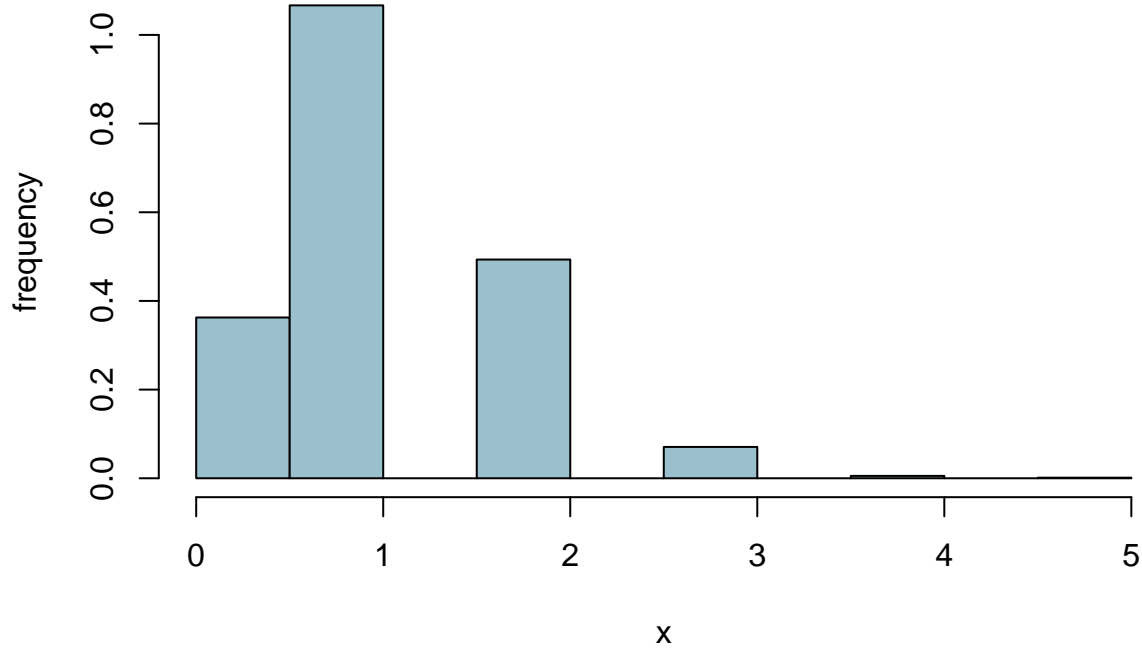
$$\mu \sim N(2.5, 1), \quad \nu \sim N(5, 1)$$

3. Normal priors centered on the true-value of the sampling distribution but with a greater variance with respect to the previous case:

$$\mu \sim N(2.5, 4), \quad \nu \sim N(5, 4)$$

An histogram of the simulated COM-Poisson data is shown below.
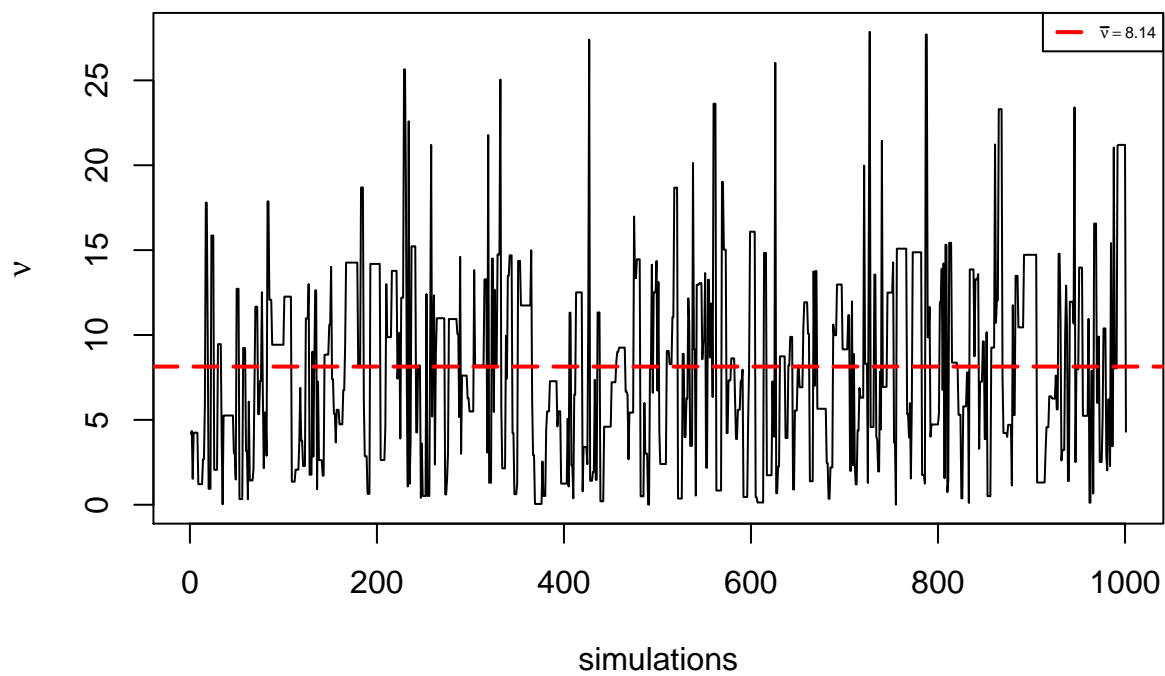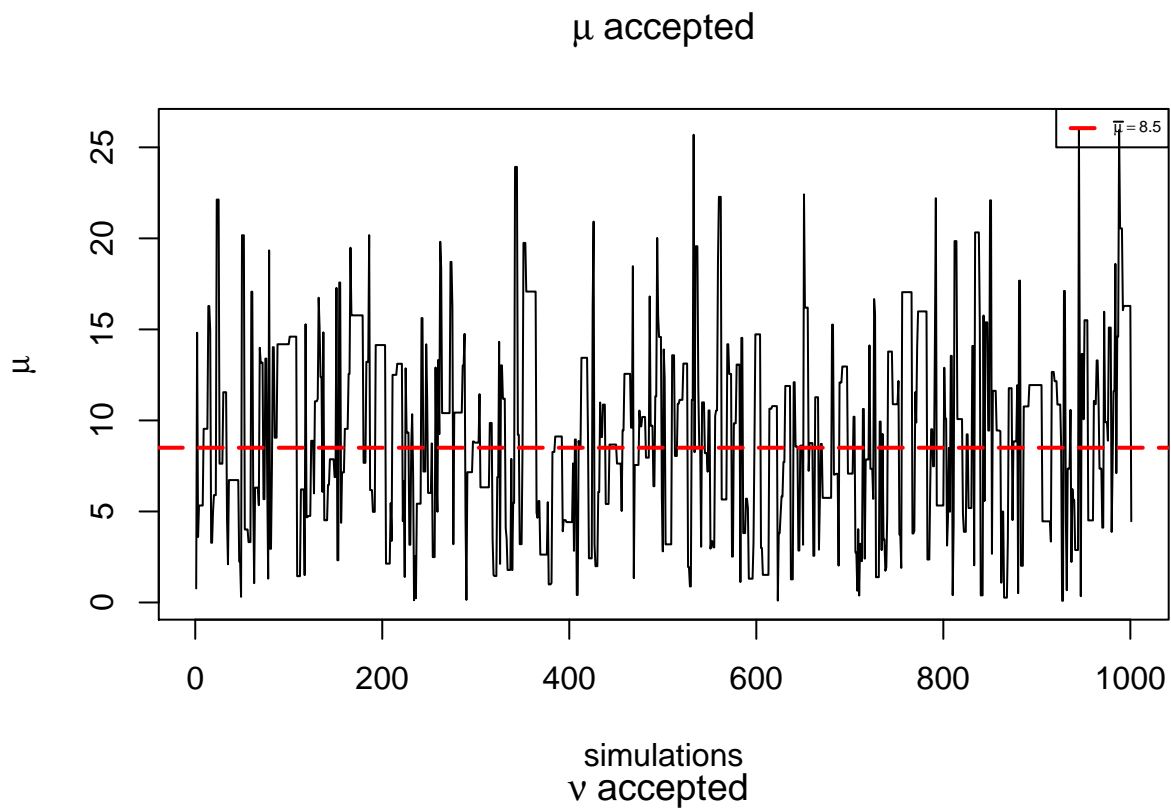
8

Random sample from Y~CMP(2.5,5)

In the table you can detect the estimated values for μ and ν and the percentage of accepted values for each of the three experiments. In the first case we can see that we have a relatively high percentage of accepted values but, we can see at the same time, that the Monte Carlo Markov chain moves around values quite far away from the true ones. This beacause prior distributions are completely non informative, centered in zero and one, respectively, and with a very high variance; in other words it seems that in the chain we often reach points with an higher density with respect to the previous one (that makes a proposed value the right candidate to be accepted) but with a so high variability that does not allow the algorithm to catch the true distribution's parameters. In the second and third case we have estimated values closer to the true ones and this is clearly due to the more informative prior distributions; we have, indeed, both priors centered on the true values of the distribution's parameters. However, the two cases present different percentages of accepted values: this can be explained by the different priors' variances. In the former we have a lower variance ($\sigma^2 = 1$ instead of $\sigma^2 = 4$) and this implies that the proposed values tend to be accepted and the MCMC converges earlier than in the latter case and, hence, it's harder to have new viable candidates. Anyway, the higher variability of the last simulation seems to allow us in finding the closest values to the true ones.
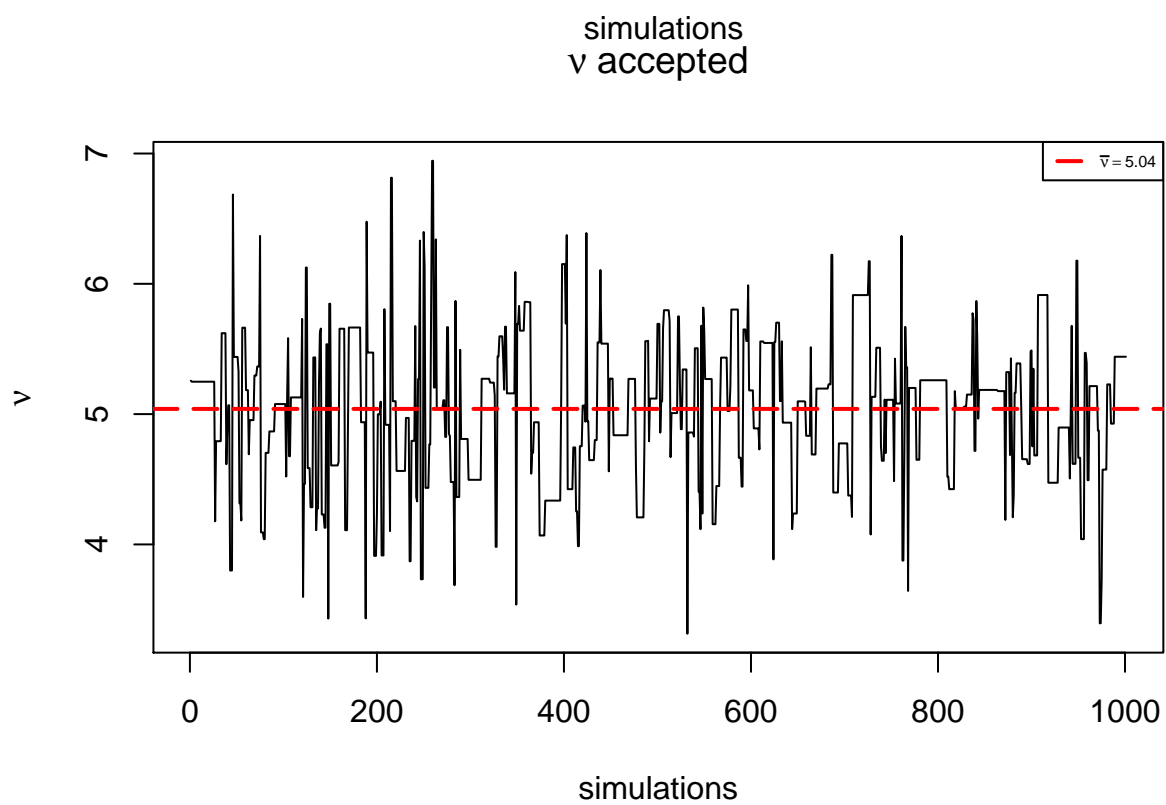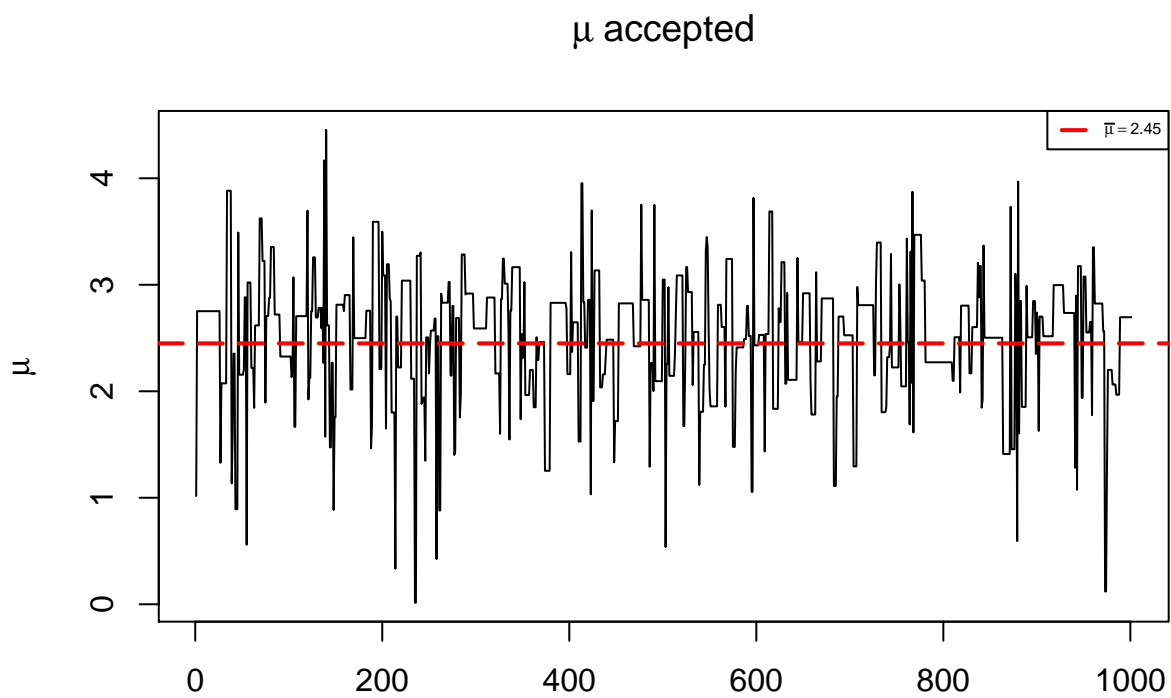
	MCMC results		
	first trial	second trial	third trial
mu	8.50	2.45	2.71
nu	8.14	5.04	5.17
accepted %	41.55	28.64	39.44

Below we have the traceplots of the first 1000 simulations for μ and ν related to the three different experiments; the flat periods in the black line represent not accepted candidates.
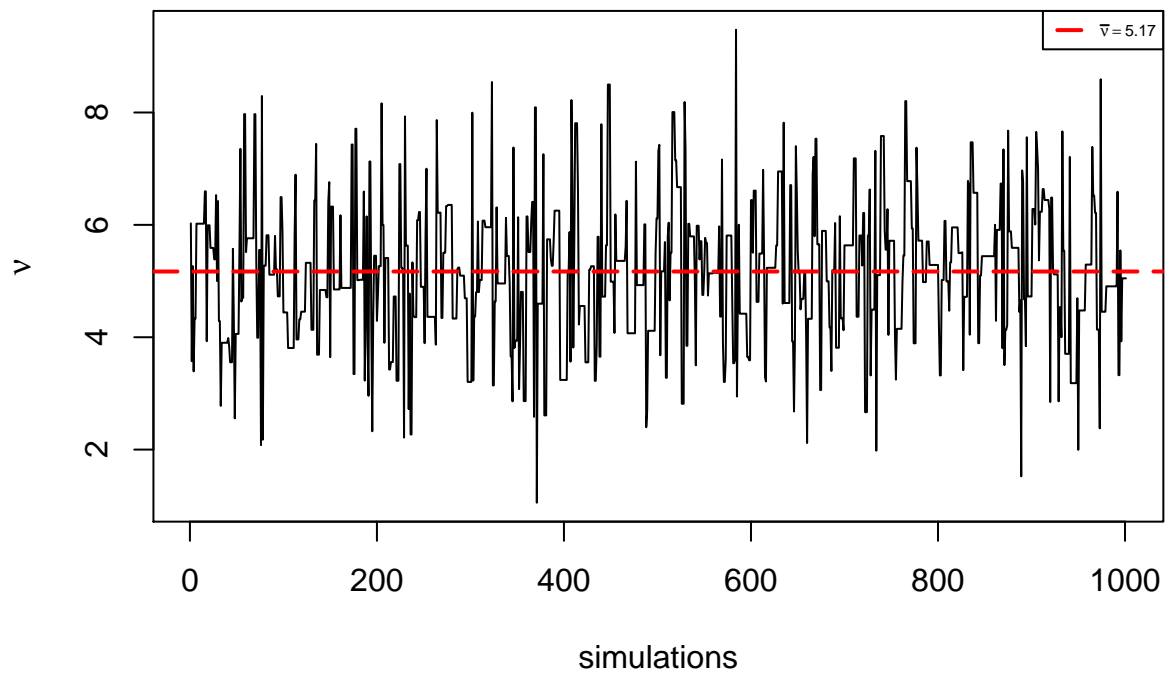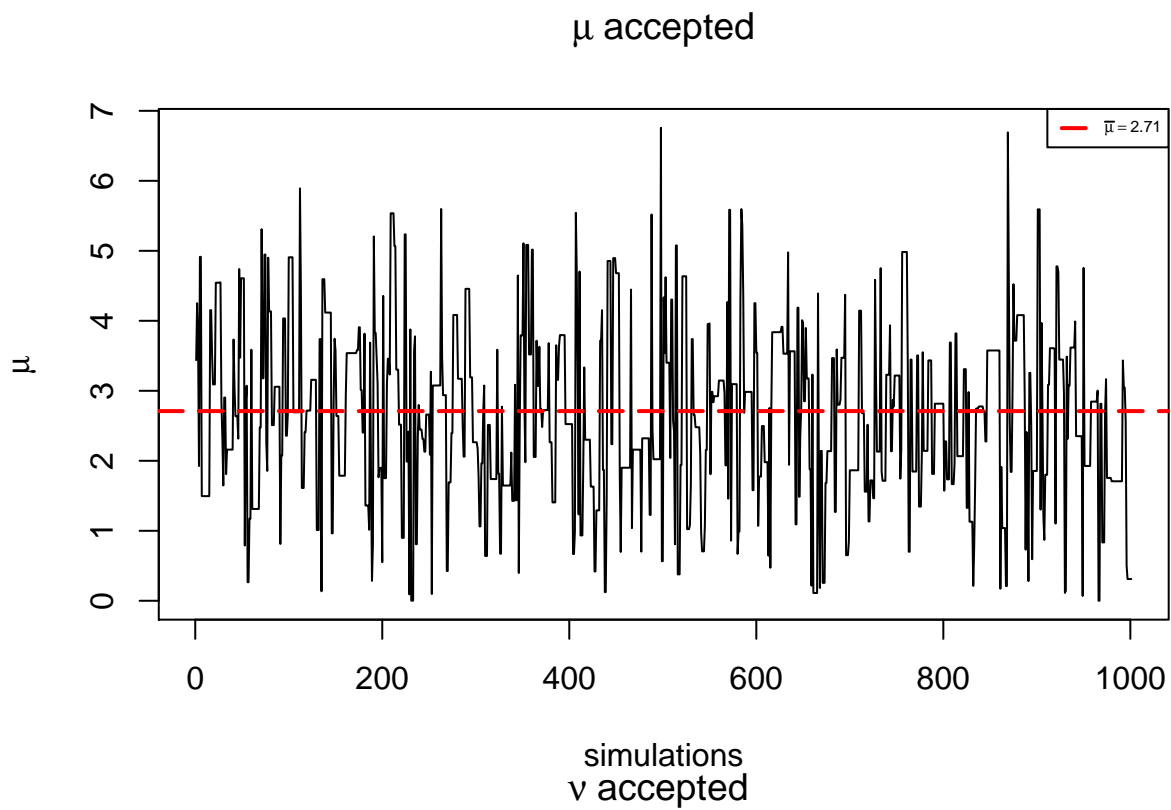
μ accepted

simulations

ν accepted

simulations

μ accepted

simulations

ν accepted

simulations

third simulation:

μ accepted

ν accepted

3.1.2 Fertility data

In both this section and the following one we are using data coming from the dataset "Fertility Data" used by Winkelmann (1995) and avaliable in the R package *Countr*.

The sample is formed by observation of 1243 women aged 44 or older in 1985 who answered the questions relevant to the analysis conducted by the German Socio-Economic Panel. The count variable children is the number of births per woman and is characterised by small underdispersion: the variance of the number of births is less than the mean, 2.33 versus 2.38. There are 8 possible explanatory variables: 3 continuous and 5 categorical (Kharrat and Boshnakov 2017). In this section we'll only use the "number of births" variable, assuming that these data come from a COM-Poisson. Since we want to estimate the posterior parameters and we don't know from which specific CMP data comes from, we have to make some assumption on it. Taking into account the histogram of birth data below, we can see that $\mu = 3$ and $\nu = 1.8$ can be acceptable values for our likelihood distribution.

Histogram of birth
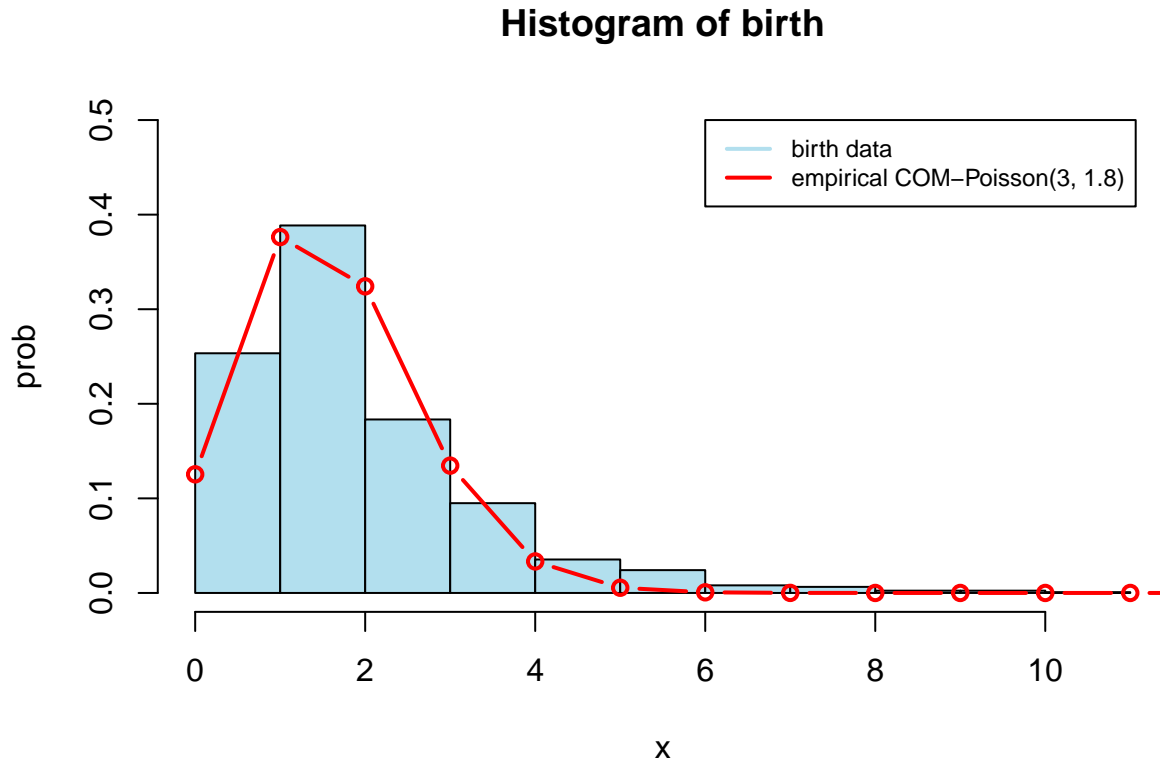
At this point, as the previous case, the exchange algorithm is applied three times using different priors on the parameters' distributions. In this case, since we are assuming that data, $y = y_1, ..., y_n$, come from a $CMP(3, 1.8)$, at each step of the MCMC the vector $y^\star = y_1^\star, ..., y_n^\star$ is generated from this probability distribution. Let's take a closer look to the applied priors.

1. We choose uninformative standard Normal priors, i.e.:

$$\mu \sim N(0, 1), \quad \nu \sim N(0, 1)$$

2. Normal priors centered on the true-value of the sampling distribution, with small variabilily with respect to it:

$$\mu \sim N(3, 1), \quad \nu \sim N(1.8, 1)$$

3. Normal priors centered on the true-value of the sampling distribution but with a greater variance with respect to the previous simulation:

$$\mu \sim N(3,4), \quad \nu \sim N(1.8,4)$$

Here the estimated posterior parameters via exchange algorithm.

	MCMC results		
	first trial	second trial	third trial
mu	0.54	3.43	4.14
nu	0.34	1.99	2.57
accepted %	4.78	21.06	29.19

The *MCMC results* table shows us different outcomes for the three exmeriments. Indeed, for the uninformative priors case, we have a very low percentage of accepted candidates (displayed also in the trace plots at the end of the paragraph) and unsuitable estimated values of the posterior parameters.

Taking into account the second and the third simulation, we observe how the percentage of accepted values of candidates $\mu^\star$,$\nu^\star$ increases with a greater variance set for the priors. The estimated values of the parameters are $(\bar{\mu} = 3.4, \bar{\nu} = 2)$ and $(\bar{\mu} = 4.1, \bar{\nu} = 2.6)$, respectively. We can say that the estimated parameters have greater values than the ones we assumed for the likelihood. The second estimate seems to be the one that better fits data distribution, while the last one over-estimates the modal value.
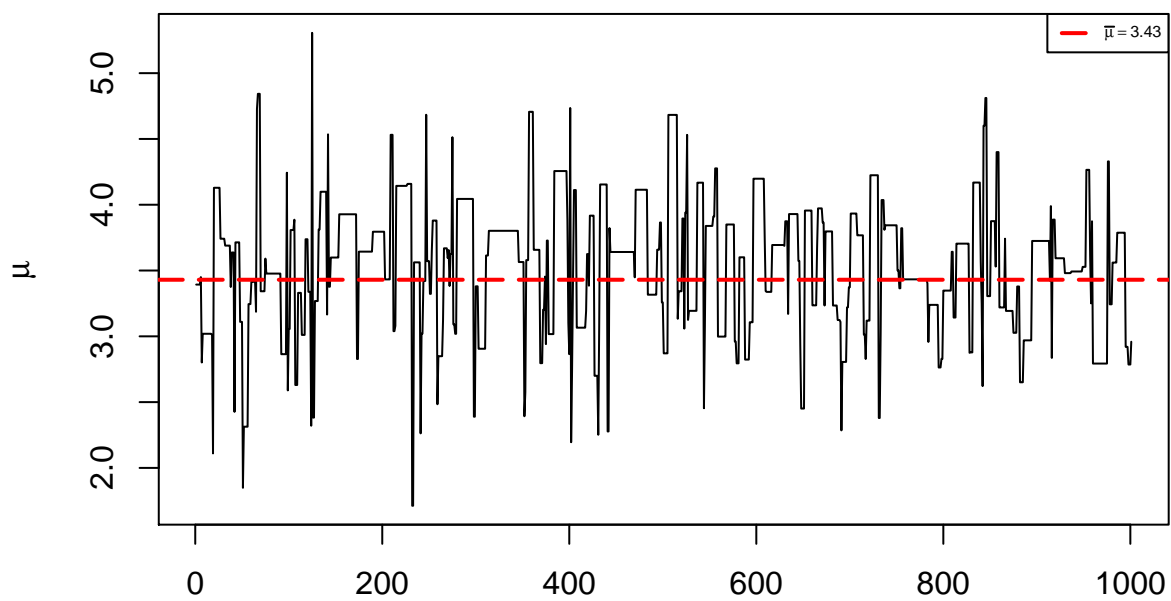
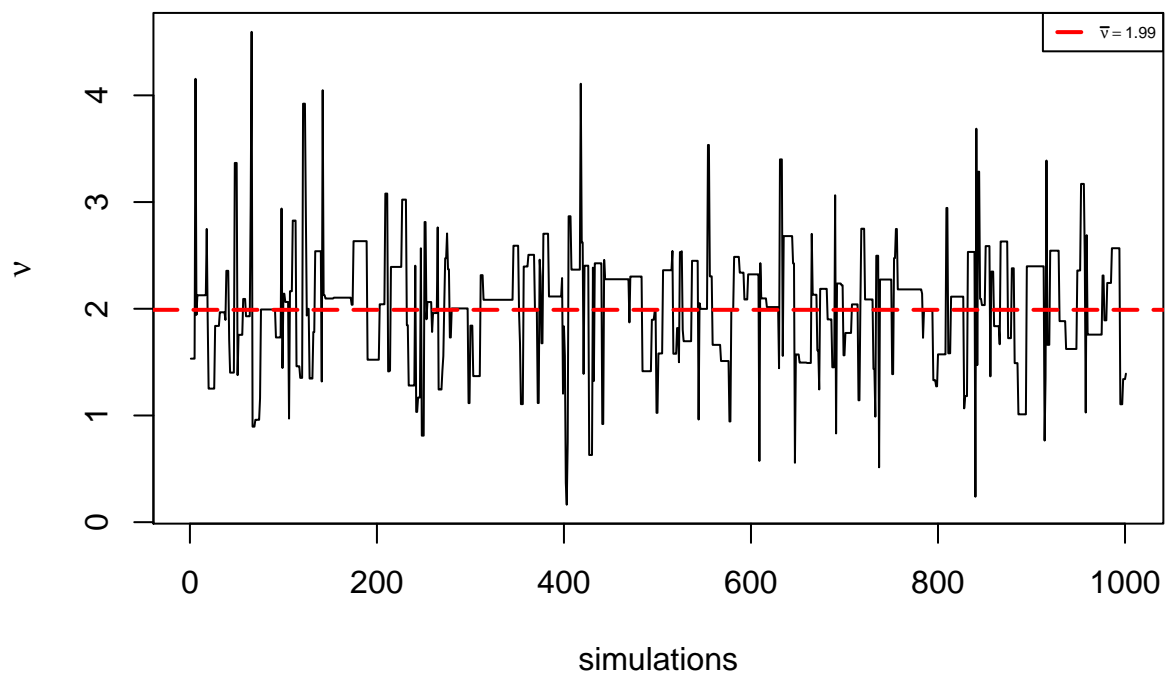## Birth data and estimated COM–Poisson distributions
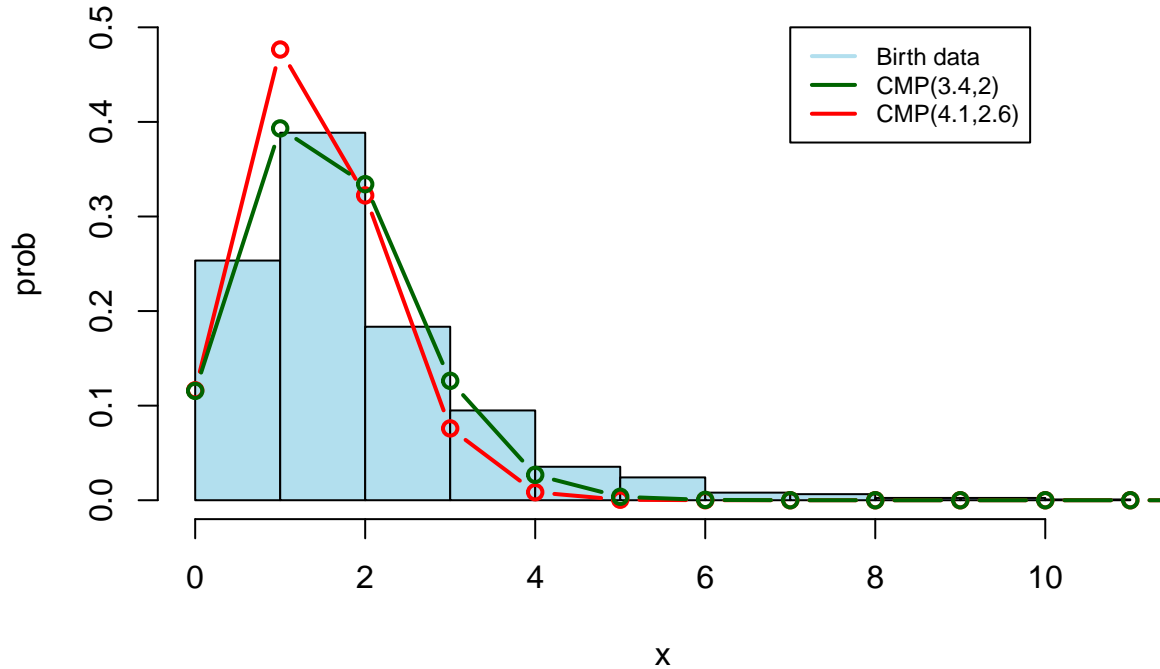
first simulation:

μ accepted

simulations

ν accepted

simulations

second simulation:

μ accepted

simulations
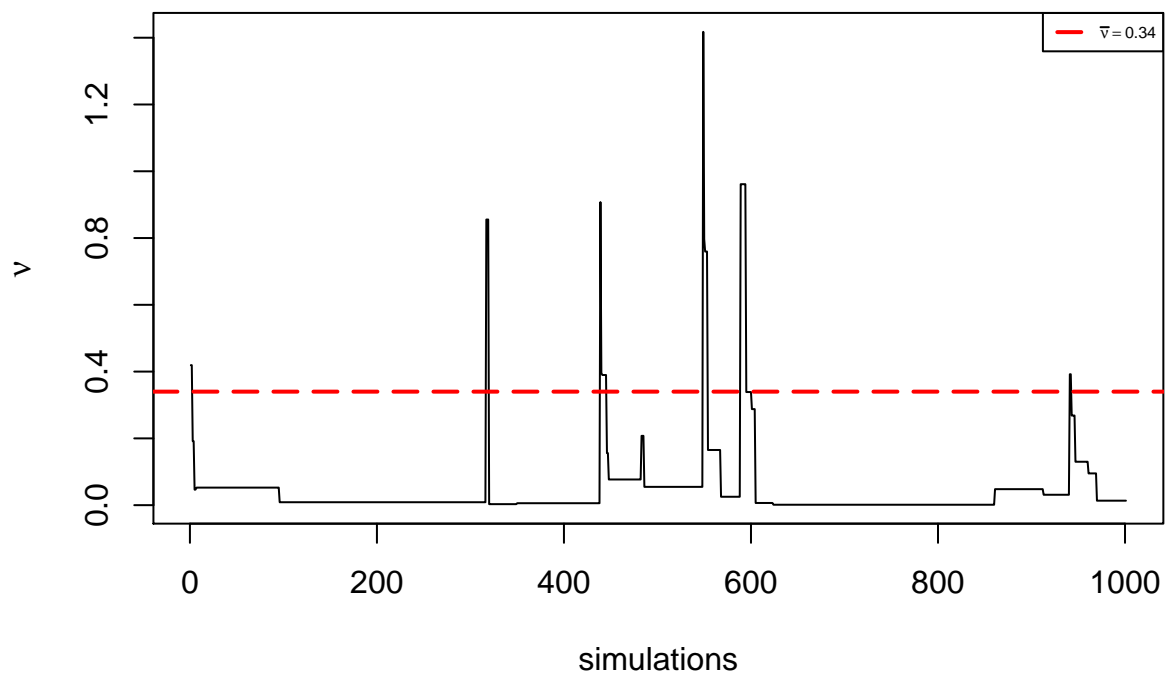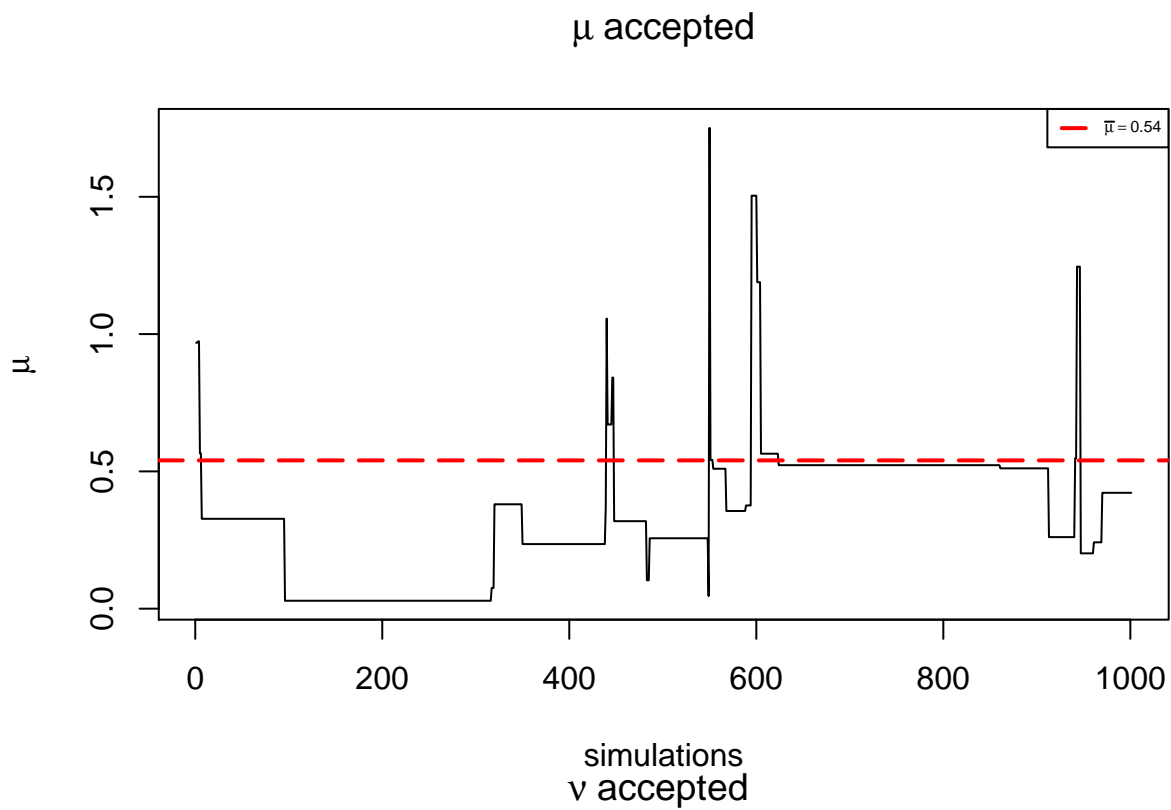
ν accepted
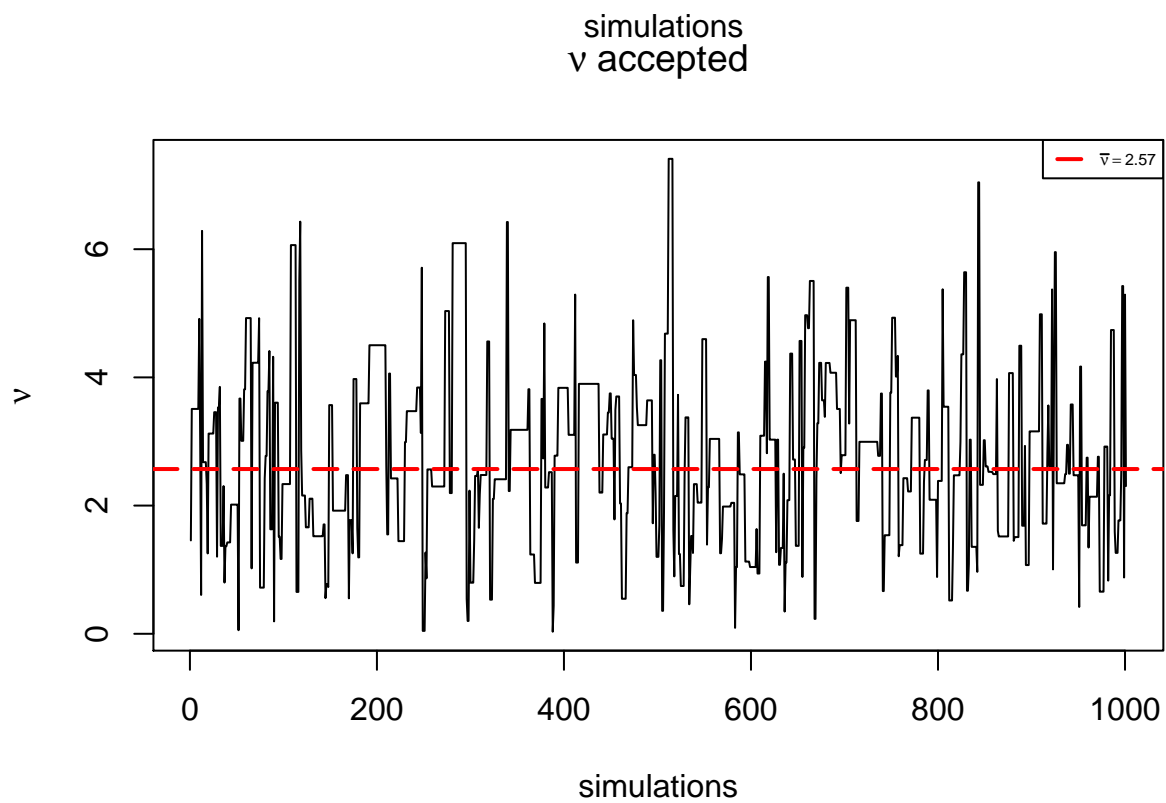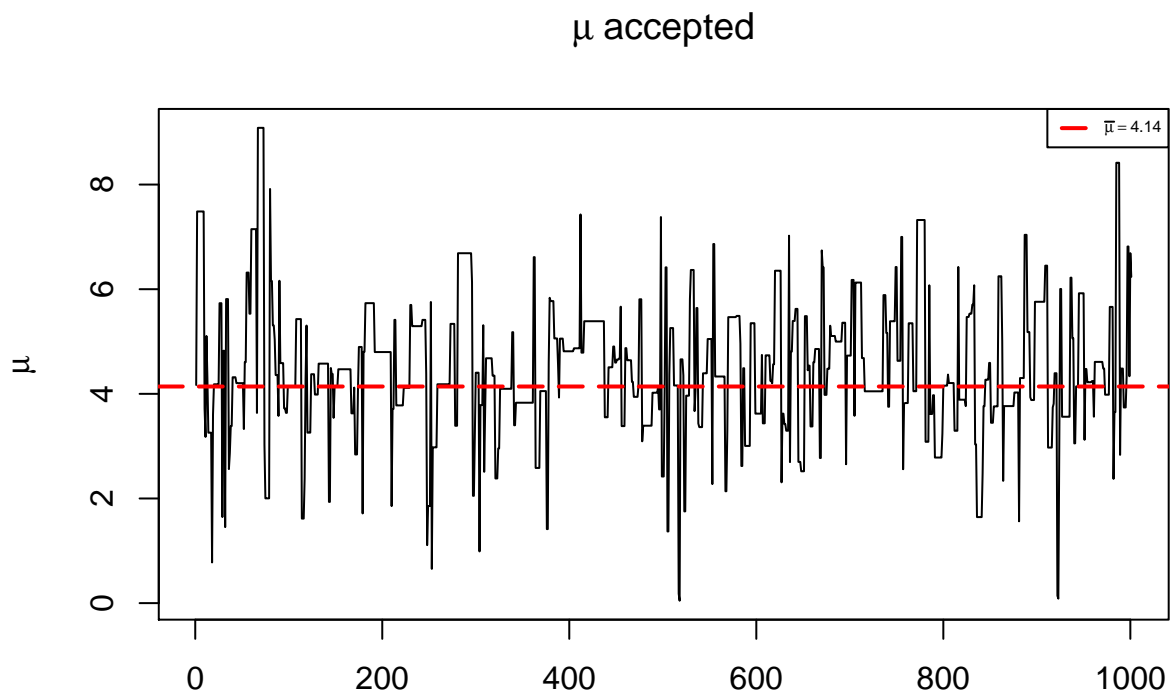

simulations

third simulation:

μ accepted

simulations

ν accepted

simulations

3.2 COM-Poisson regression

This last part regards a regression analysis based on the algorithm proposed by C. Chanialidis et al. (2018) and implemented in the *cmpoisreg* R function (Chanialidis (2017)). The COM-Poisson model that we are taking into account is defined by Guikema and Goffelt (2008) and it has the following form:

$$P(Y_i = y_i \mid \mu_i, \nu_i) = \left(\frac{\mu_i^{y_i}}{y_i!} \right)^{\nu_i} \frac{1}{Z(\mu_i, \nu_i)},$$

$$log(\mu_i) = x_i^T \beta \implies E[Y_i] \approx \exp\{x_i^T \beta\},$$
$$log(\nu_i) = -x_i^T \delta \implies V[Y_i] \approx \exp\{x_i^T \beta + x_i^T \delta\}$$

where Y is the dependent random variable that has to be modeled, while β and δ are the regression coefficients for the centring and the shape parameters.

For the COM-Poisson regression model, the acceptance ratio considered for the exchange algorithm becomes:

$$a = \frac{\left[\prod_i q_{\theta^\star}(y_i) \right] p(\beta^\star) p(\delta^\star) \left[\prod_i q_\theta(y^\star) \right]}{\left[\prod_i q_\theta(y_i) \right] p(\beta) p(\delta) \left[\prod_i q_{\theta^\star}(y^\star) \right]}$$

where $\theta = (\beta, \delta)$. In the sampler, two different kinds of moves are applied, in order to reduce the correlation between successive samples of the regression coefficients β and δ. Each sweep of the MCMC sampler performs these two moves in a sequence. The first proposes a move from β to $\beta^\star$ and afterwards from δ to $\delta^\star$. The second proposes a move from (β_i, δ_i) to $(\beta_i^\star, \delta_i^\star)$ for $i = 1, .., p$, where p is the number of indipendent variables (C. Chanialidis et al. 2018). The prior distributions for β and δ are both $N(\mu_p, \Sigma)$, where μ is a p-dimentional vector of zeroes and $\Sigma = I_p \cdot 100$.

The covariates are the following:

- *German*: equal 1 if the woman is German, else 0;
- *schooling*: measured as years of schooling;
- *vocational education*: Equals 1 if the woman had vocational training, else 0;
- *University*: Equals 1 if the woman had university degree, else 0;
- *age*: age of the woman at the time of the survey;
- *rural*: Equals 1 if the woman lives in a rural area, else 0;
- *age at marriage*: Age of the woman at the time of marriage;
- *Religion*: The woman's religious denomination (Catholic, Protestant, Muslim).

The variable *Religion* has been transformed into three binary variables corresponding to Catholic, Protestant and Muslim religions, respectively.

In the following table you can see the percentage of the candidates accepted by the algorithm for both first and second move in the simulation (singular and mixed updates). The table a very low percentage of accepted β, less than 10%, while the percentage of accepted δ is slighly greater (around 22%). On the contrary, when we move from (β_i, δ_i) to $(\beta_i^\star, \delta_i^\star)$, the exchange algorithm accepts more than the half of the candidates.

accepted candidates	
	accepted (%)
beta	7.96
delta	21.70
mixed	56.43

In the table below the estimated coefficients are given and they are also shown with 50% and 90% inner intervals in the following plots. At the end it's also given a diagnostic view of both the coefficients' distributions, that roughly respect normality, and the trace plots of the simulations.

	regression coefficients	
	beta	delta
intercept	1.4212	-0.2062
German	-0.1414	-0.3206
schooling	0.0042	0.0251
vocational education	-0.0685	-0.2556
University	-0.1529	0.4073
age	-0.0169	0.2858
rural	0.0572	0.0482
age at marriage	-0.0688	-0.1757
Catholic	-0.5604	-0.0181
Muslim	-0.3741	0.2939
Protestant	-0.4414	0.1327
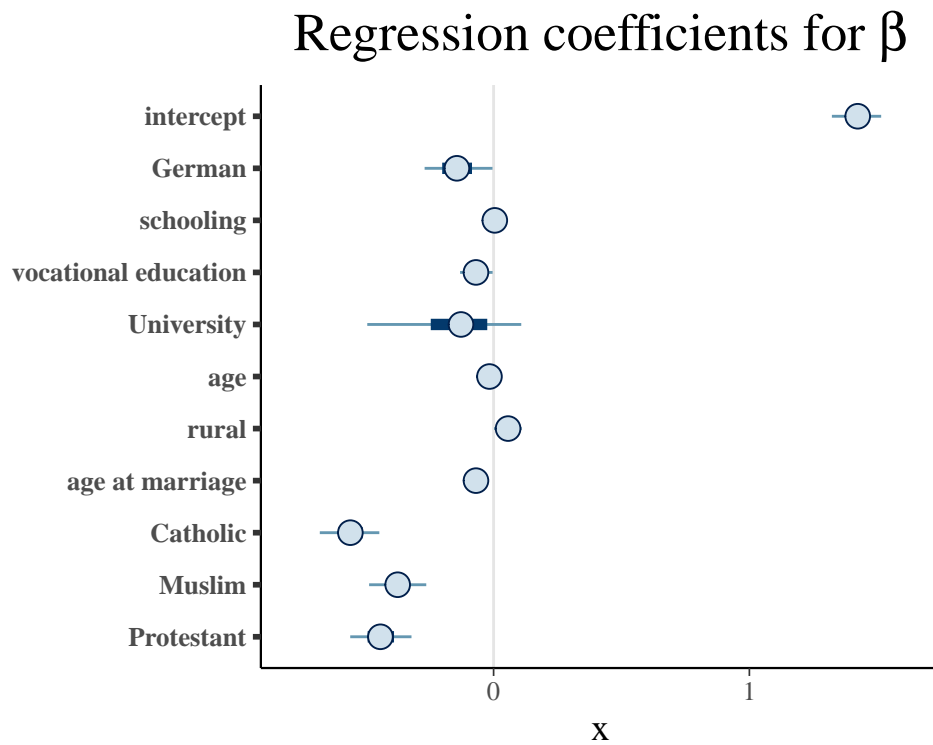
Regression coefficients for β

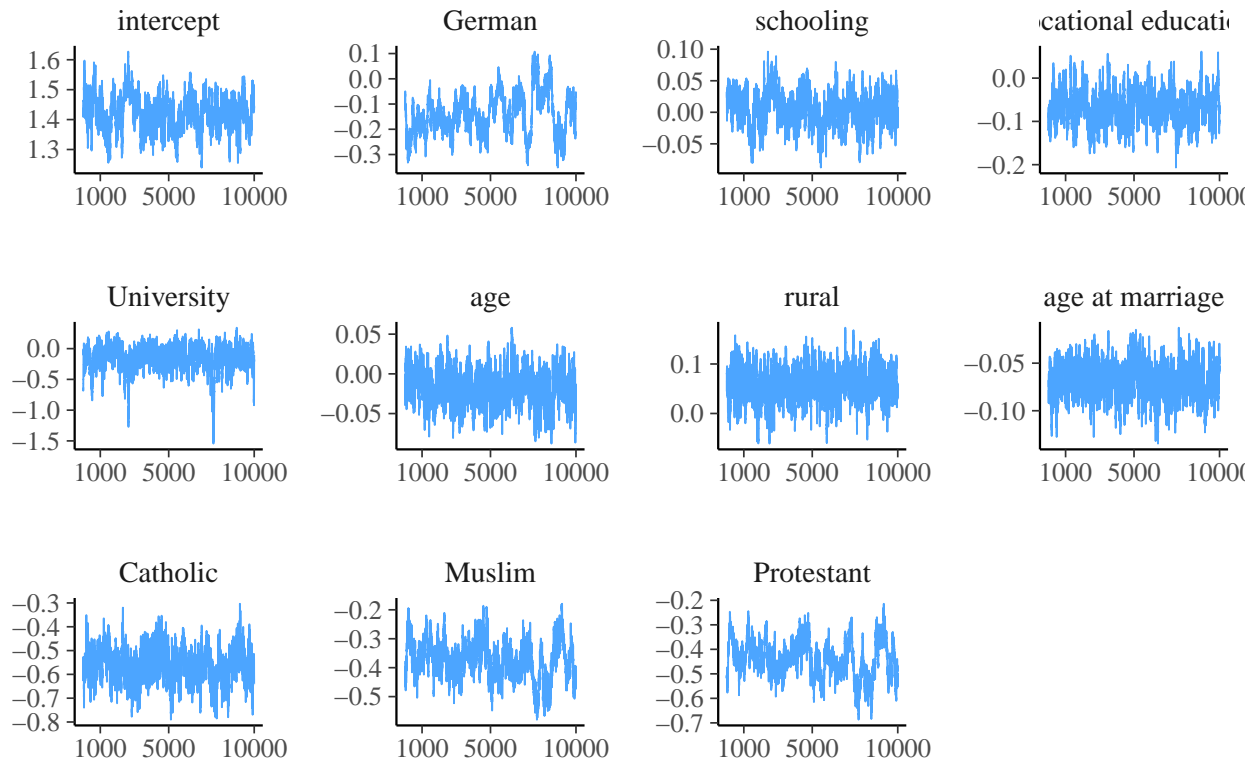# Regression coefficients for δ
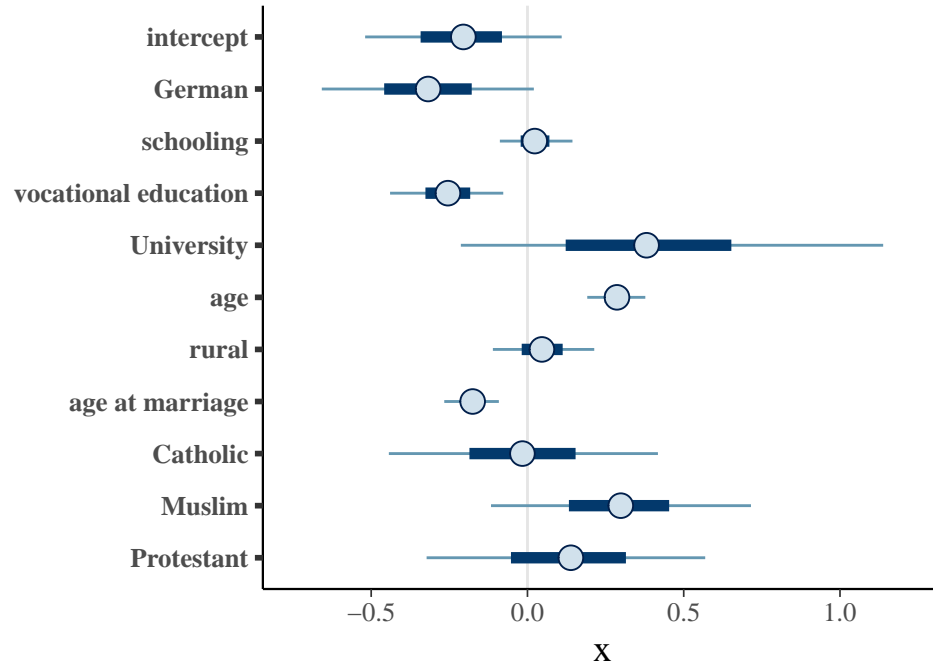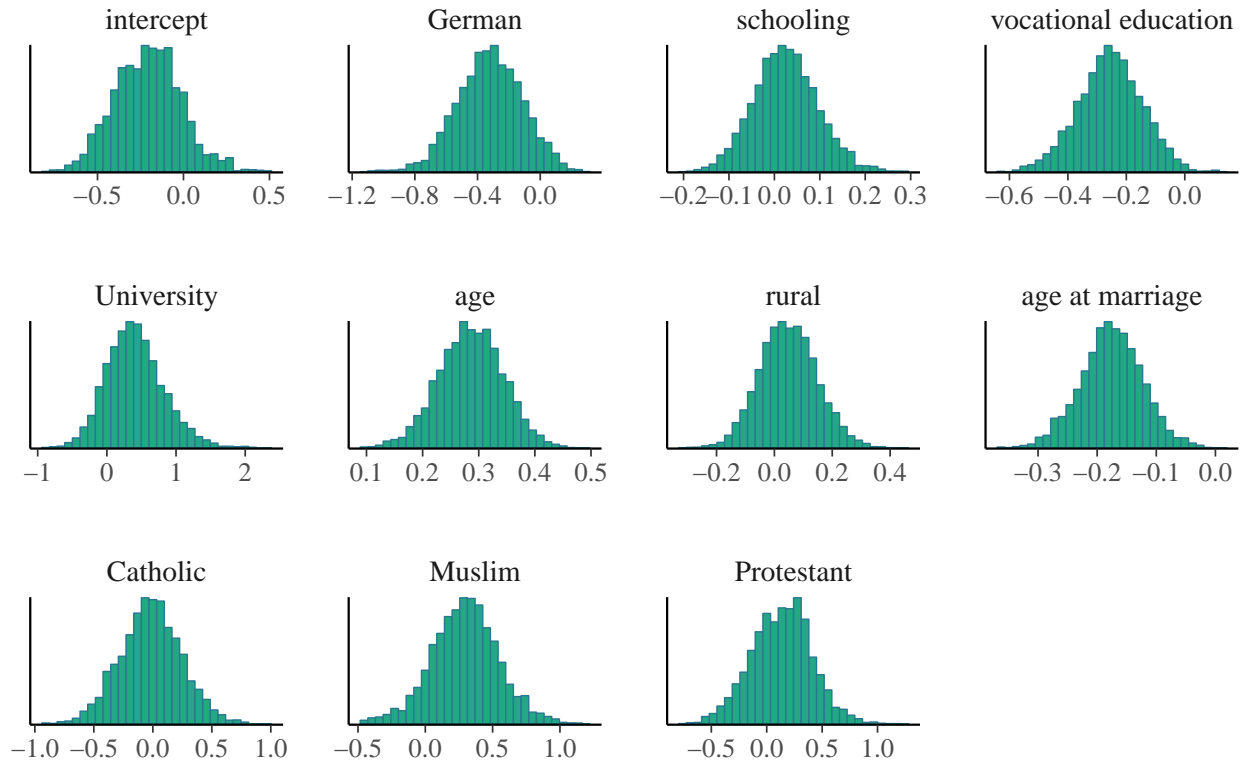
Marginal posterior distribution of β coefficients
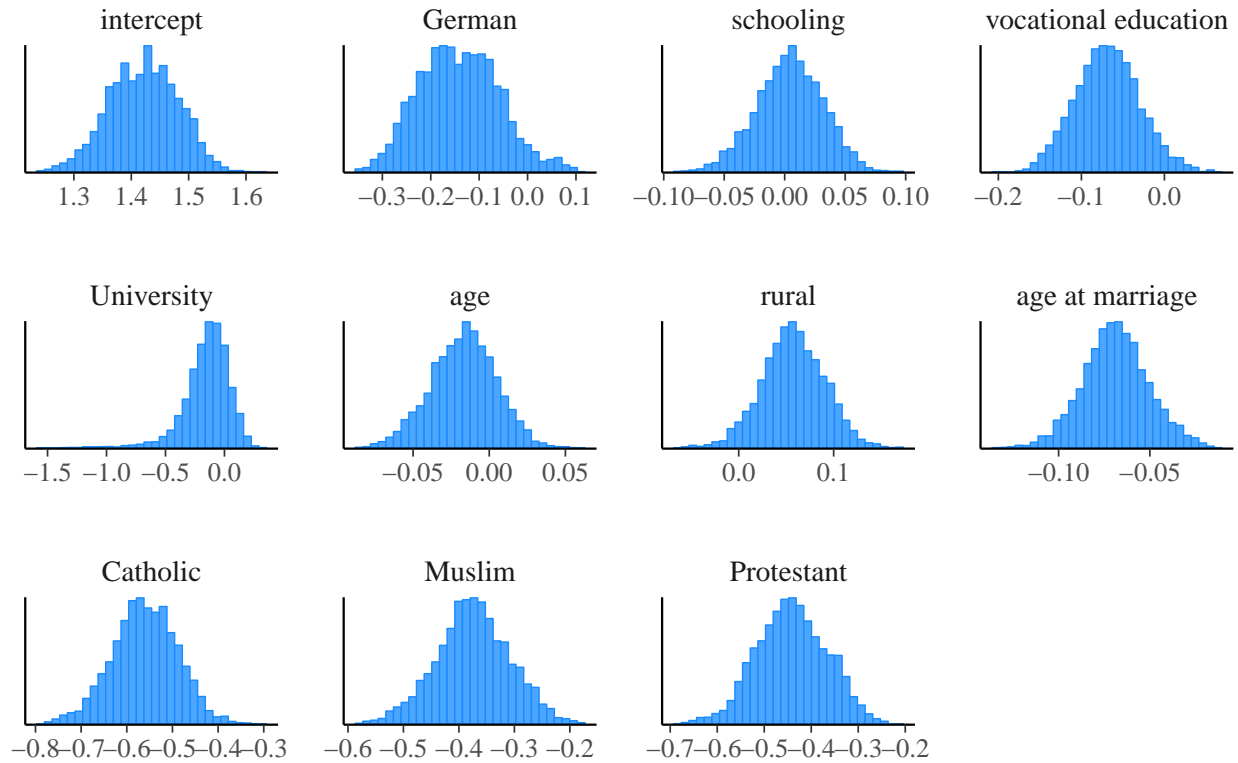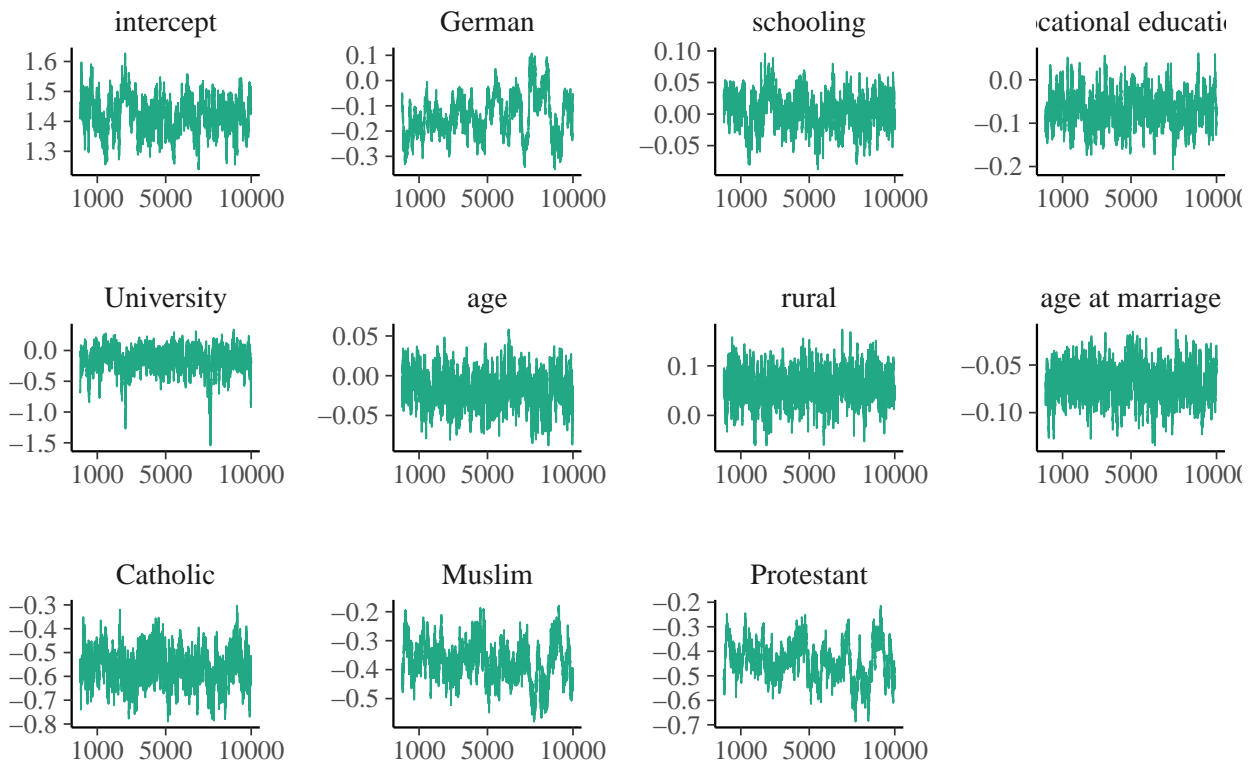

Trace plots of β coefficients

Marginal posterior distribution of δ coefficients

Trace plots of δ coefficients

References

Chanialidis. 2017. "Combayes R Package, Github." https://github.com/cchanialidis/combayes.

Chanialidis, Charalampos, Ludger Evers, Tereza Neocleous, and Agostino Nobile. 2018. "Efficient Bayesian Inference for Com-Poisson Regression Models." *Statistics and Computing* 28 (3). Springer: 595–608.

Famoye, Felix, and PC Consul. 1995. "Bivariate Generalized Poisson Distribution with Some Applications." *Metrika* 42 (1). Springer: 127–38.

Guikema, Seth D, and Jeremy P Goffelt. 2008. "A Flexible Count Data Regression Model for Risk Analysis." *Risk Analysis: An International Journal* 28 (1). Wiley Online Library: 213–23.

Kharrat, Tarak, and Georgi N. Boshnakov. 2017. "Fertility Data, Extensive Analysis." https://cran.r-project.org/web/packages/Countr/vignettes/exampleFertility.pdf.

Møller, Jesper, Anthony N Pettitt, Robert Reeves, and Kasper K Berthelsen. 2006. "An Efficient Markov Chain Monte Carlo Method for Distributions with Intractable Normalising Constants." *Biometrika* 93 (2). Oxford University Press: 451–58.

Sellers, Kimberly F, and Darcy S Morris. 2017. "Underdispersion Models: Models That Are 'Under the Radar'." *Communications in Statistics-Theory and Methods* 46 (24). Taylor & Francis: 12075–86.

Sellers, Kimberly F, Sharad Borle, and Galit Shmueli. 2012. "The Com-Poisson Model for Count Data: A Survey of Methods and Applications." *Applied Stochastic Models in Business and Industry* 28 (2). Wiley Online Library: 104–16.

Winkelmann, Rainer. 1995. "Duration Dependence and Dispersion in Count-Data Models." *Journal of Business & Economic Statistics* 13 (4). Taylor & Francis Group: 467–74.

———. 2008. *Econometric Analysis of Count Data.* Springer Science & Business Media.

Yildirim, Ilker. 2012. "Bayesian Inference: Metropolis-Hastings Sampling." *Dept. of Brain and Cognitive Sciences, Univ. of Rochester, Rochester, NY.*