`rstanarm` - Exercise 4

Bayesian Inference - Lab Sessions

Marika D'Agostini
marika.dagostini2@unibo.it

University of Bologna

November-December 2023

Exercise 4: Poisson Regression

Poisson regression model in a dose-response study.

- A dataset about the mutagenicity on salmonella is available.
- Three plates ($j = 1, 2, 3$) are processed at each dose (x_i; $i = 1, ..., 6$) of quinoline (liquid organic compound) and the number of revertant colonies of Salmonella were counted y_{ij}.

Exercise: compare a simple Poisson regression model and a Poisson model with random effects in terms of goodness of fit. Suppose that we want to study the effect of `quinoline` both at the measurement scale and at a logarithmic scale of the type $log(\texttt{quinoline} + 10)$. Start from a $\mathcal{N}(0, 10)$ prior for the β parameters and define a $\mathcal{N}(0, c)$ weakly informative prior.

Remarks:

- The Poisson model with random effects is usually assumed in order to take into account the eventual presence of overdispersion or underdispersion.
- The Poisson distribution has only one parameter ($\mathbb{E}[Y] = \mathbb{V}[Y] = \lambda$), and in fact the variance increases with the mean.

Simple Poisson Model

$\rightarrow$ First, a simple **Poisson regression model** is considered.

Its Bayesian formulation is:

$$
\begin{aligned}
y_i|\mu_i &\sim Poisson(\mu_i) \\
\log(\mu_i)|\boldsymbol{\beta} &= \beta_0 + \beta_1 log(x_i + 10) + \beta_2 x_i \\
&= \beta_0 + \beta_1 log(\texttt{quinoline}_i + 10) + \beta_2 \texttt{quinoline}_i \\
&= \beta_0 + \beta_1 \texttt{log\_quinoline}_i + \beta_2 \texttt{quinoline}_i, \quad i = 1, ..., n. \\
\beta_k &\sim \mathcal{N}(0, c), \;\; k = 0, 1, 2;
\end{aligned}
$$

```r
data4 <- read.csv("Data_Ex_4.csv")

mod_ex4a <- stan_glm(formula =
                  colonies~quinoline+log_quinoline,
        data = data4,
        family = "poisson",
        prior = normal(0,10, autoscale=T),
        prior_intercept = normal(0,10, autoscale=T))

prior_summary(mod_ex4a)
```

```
------
Intercept (after predictors centered)
 ~ normal(location = 0, scale = 10)

Coefficients
  Specified prior:
    ~ normal(location = [0,0], scale = [10,10])
  Adjusted prior:
    ~ normal(location = [0,0], scale = [0.027,6.091])
------
```

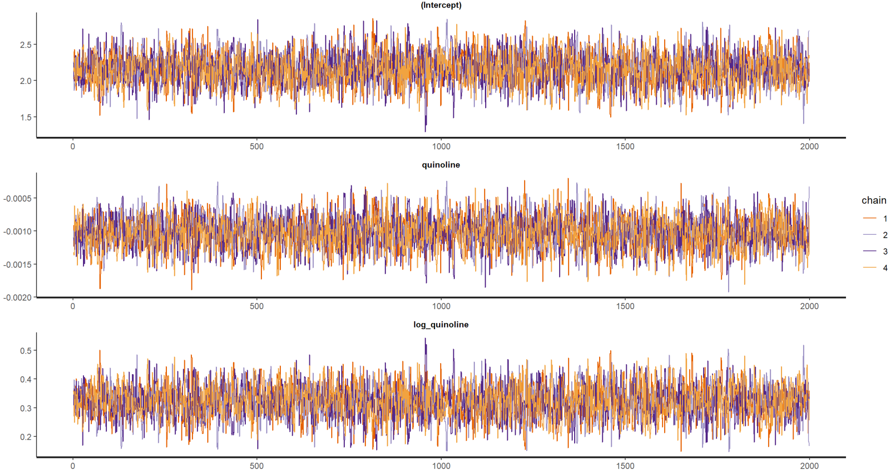
```
summary(mod_ex4a)
```

```
MCMC diagnostics
              mcse Rhat n_eff
(Intercept)   0.0  1.0  1711
quinoline     0.0  1.0  1669
log_quinoline 0.0  1.0  1629
mean_PPD      0.0  1.0  2750
log-posterior 0.0  1.0  1713
```

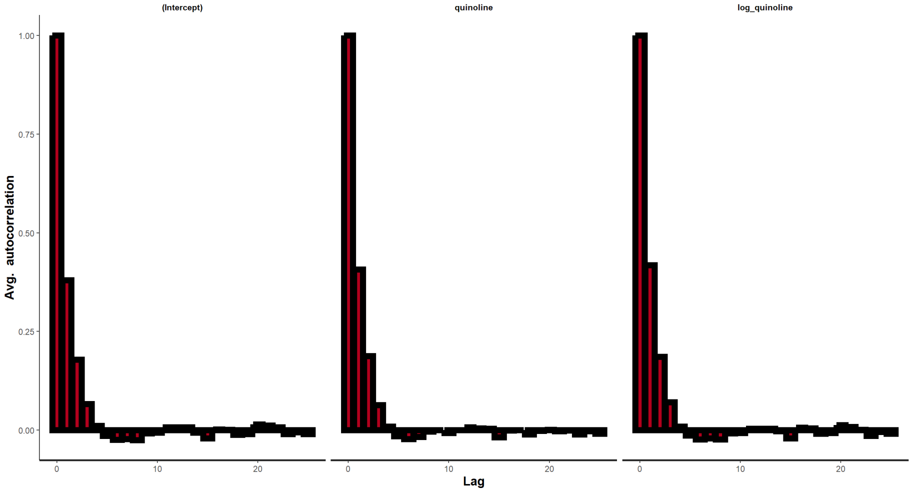
```
mod_ex4a<-update(mod_ex4a, iter=4000)
summary(mod_ex4a)
```

```
MCMC diagnostics
              mcse Rhat n_eff
(Intercept)   0.0  1.0  3545
quinoline     0.0  1.0  3442
log_quinoline 0.0  1.0  3395
mean_PPD      0.0  1.0  5403
log-posterior 0.0  1.0  3085
```

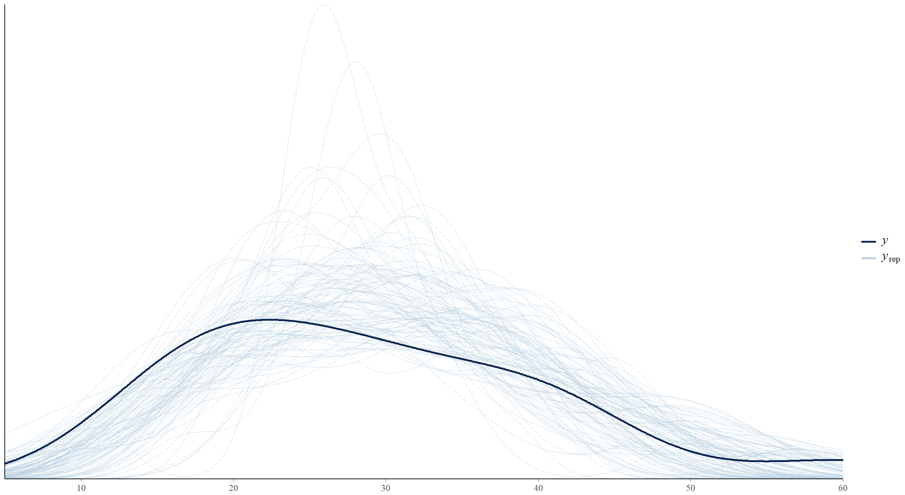
```
stan_trace(mod_ex4a, nrow=3, ncol=1)
```



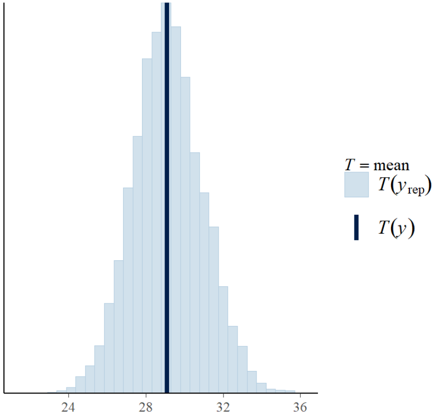
stan_ac(mod_ex4a)

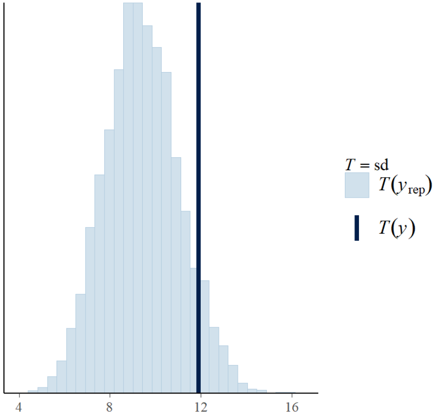


```r
# Posterior checks
y_tilde4a <- posterior_predict(mod_ex4a)
ppc_dens_overlay(y = data4$colonies,
                 yrep = y_tilde4a[1100:1200,])
```

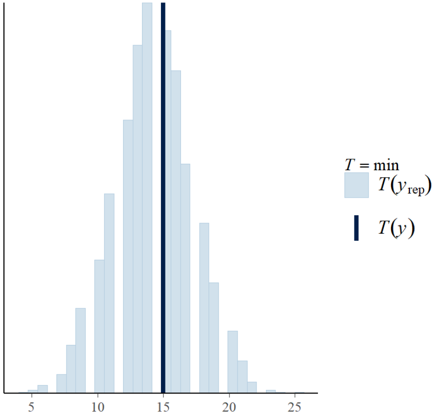


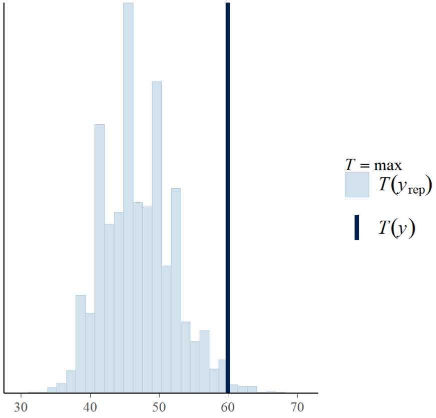
```
ppc_stat(y = data4$colonies,
                yrep = y_tilde4a, stat = "mean")
ppc_stat(y = data4$colonies,
                yrep = y_tilde4a, stat = "sd")
```





```
ppc_stat(y = data4$colonies,
                yrep = y_tilde4a, stat = "min")
ppc_stat(y = data4$colonies,
                yrep = y_tilde4a, stat = "max")
```





Poisson Regression Model with random effects

$\rightarrow$ To take into account the eventual presence of overdispersion, the term λ_j, that is *plate-specific* is included in the linear predictor.

Likelihood:

$$y_{ij}|\mu_{ij} \sim Poisson(\mu_{ij})$$
$$\log(\mu_{ij})|\boldsymbol{\beta}, \lambda_j = \beta_0 + \beta_1 \log(x_{ij} + 10) + \beta_2 x_{ij} + \lambda_j$$

Priors:

$$\lambda_j|\sigma^2_\lambda \sim \mathcal{N}(0, \sigma^2_\lambda),\ \ j = 1, 2, 3;$$
$$\beta_k \sim \mathcal{N}(0, c),\ \ k = 0, 1, 2.$$

Hyperprior:

$$\sigma_\lambda \sim \pi(\sigma_\lambda).$$

```
mod_ex4b <- stan_glmer(formula =
         colonies~quinoline+log_quinoline+(1|plate),
        data = data4,
        family = "poisson",
        prior = normal(0,10, autoscale=T),
        prior_intercept = normal(0,10, autoscale=T))

prior_summary(mod_ex4b)
```

```
Priors for model 'mod_ex4b'
------
Intercept (after predictors centered)
 ~ normal(location = 0, scale = 10)

Coefficients
  Specified prior:
    ~ normal(location = [0,0], scale = [10,10])
  Adjusted prior:
    ~ normal(location = [0,0], scale = [0.027,6.091])

Covariance
 ~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)
------
```

```
summary(mod_ex4b)
```

```
MCMC diagnostics
                                      mcse Rhat n_eff
(Intercept)                           0.0  1.0  1312
quinoline                             0.0  1.0  1819
log_quinoline                         0.0  1.0  1769
b[(Intercept) plate:A]                0.0  1.0  1072
b[(Intercept) plate:B]                0.0  1.0  1059
b[(Intercept) plate:C]                0.0  1.0  1067
Sigma[plate:(Intercept),(Intercept)]  0.0  1.0  1031
mean_PPD                              0.0  1.0  4138
```

```
mod_ex4b <- update(mod_ex4b, iter=8000)
summary(mod_ex4b)
```

```
MCMC diagnostics
                                     mcse Rhat n_eff
(Intercept)                          0.0  1.0   4092
quinoline                            0.0  1.0   8308
log_quinoline                        0.0  1.0   8350
b[(Intercept) plate:A]               0.0  1.0   3723
b[(Intercept) plate:B]               0.0  1.0   3604
b[(Intercept) plate:C]               0.0  1.0   3609
Sigma[plate:(Intercept),(Intercept)] 0.0  1.0   2906
mean_PPD                             0.0  1.0  16445
```

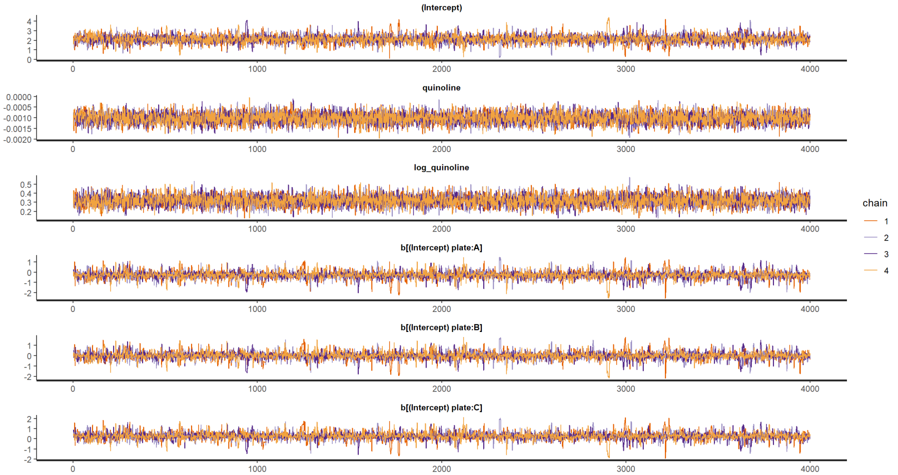
```
Warning messages:
1: There were 18 divergent transitions after warmup. See
https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
to find out why this is a problem and how to eliminate them.
2: Examine the pairs() plot to diagnose sampling problems
```

- Classical warning when we work with the Poisson regression with random effects.

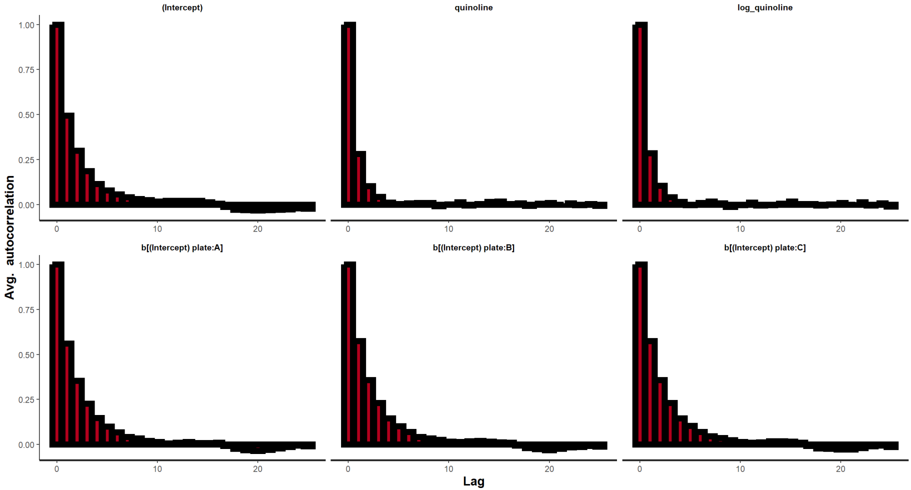
```
model <- update(model, adapt_delta=.99)
```

- `adapt_delta` is the target average proposal acceptance probability
- In general you should not need to change `adapt_delta` unless you see a warning message about divergent transitions, in which case you can increase `adapt_delta` from the default to a value closer to 1 (e.g. from 0.95 to 0.99, or from 0.99 to 0.999, etc).
- The step size used by the numerical integrator is a function of `adapt_delta` in that increasing `adapt_delta` will result in a smaller step size and fewer divergences.
- Increasing `adapt_delta` will typically result in a slower sampler, but it will always lead to a more robust sampler.

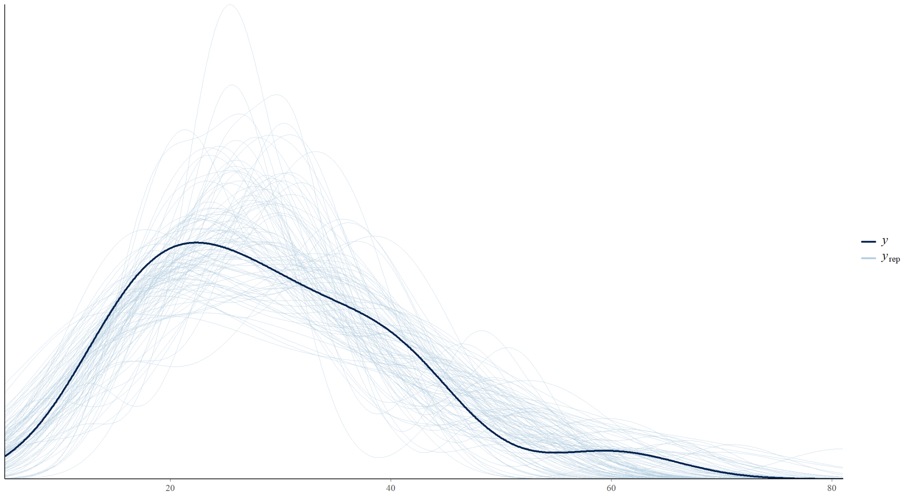
stan_trace(mod_ex4b, nrow=6, ncol=1)



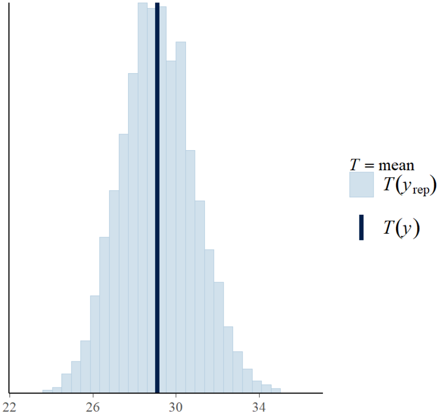
stan_ac(mod_ex4b)

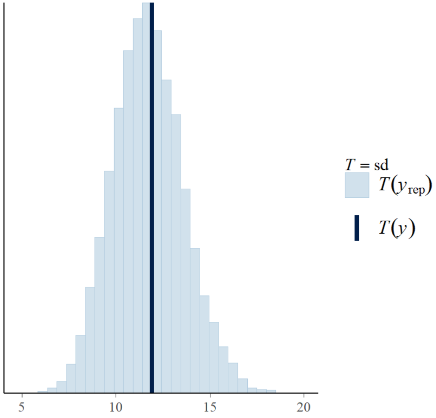


```r
# Posterior checks
y_tilde4b <- posterior_predict(mod_ex4b)
ppc_dens_overlay(y = data4$colonies,
                 yrep = y_tilde4b[1100:1200,])
```

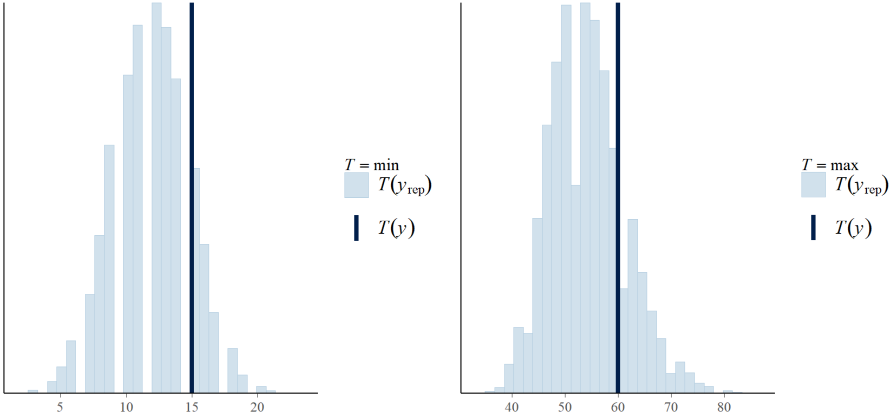


```
ppc_stat(y = data4$colonies,
                yrep = y_tilde4a, stat = "mean")
ppc_stat(y = data4$colonies,
                yrep = y_tilde4a, stat = "sd")
```





```
ppc_stat(y = data4$colonies,
                yrep = y_tilde4a, stat = "min")
ppc_stat(y = data4$colonies,
                yrep = y_tilde4a, stat = "max")
```



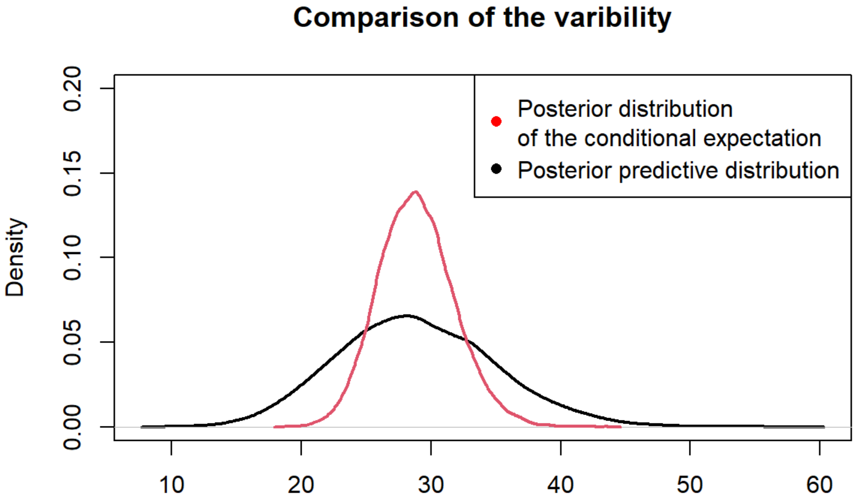
Suppose now that we want to use our model to perform inference on a new covariate pattern `quinoline = 500` and `plate = "A"`

```
# Generate the new dataset
data4_new <- data.frame(quinoline=500,
                log_quinoline = log(500+10),
                plate ="A")

# evaluation linear predictor
mu_new <- posterior_epred(mod_ex4b, newdata = data4_new)

# posterior predictive distribution
y_tilde_new <- posterior_predict(mod_ex4b,
                newdata = data4_new)
```

```r
#comparison
plot(density(y_tilde_new), ylim=c(0,0.2), lwd=2)
lines(density(mu_new), col="red", lwd=2)
```



```
mean(mu_new); sd(mu_new)
mean(y_tilde_new); sd(y_tilde_new)
```

```
> mean(mu_new);sd(mu_new)
[1] 28.85901
[1] 2.91659
> mean(y_tilde_new);sd(y_tilde_new)
[1] 28.85675
[1] 6.095023
```