LDATS2130

Introduction to bayesian statistics

Project report of Group F

Authors :

Mathias Dah Fienon, noma: 04452100

&

Lavinia Myo Lemegne, noma: 16562200

Academic year: 2022-2023

We are working under the following assumptions:

- the evolution of the numbers $y(t)$ of cancer cells detected by the technician over time is a poisson distribution $y(t_j) \sim$Pois$(\mu(t_j))$ with

$$\mu(t_j) = \beta_0 \exp\left(\frac{\beta_1}{\beta_2}(1 - e^{-\beta_2 t_j})\right)$$

where $\beta_k > 0$ $(k = 0, 1, 2)$ and the later is reparametrized as

$$\alpha_0 \exp(-\alpha_1 e^{-\alpha_2 t})$$

with $\alpha_k > 0$ $(k = 0, 1, 2)$

Question 1 :

(a) Examination of $\mu(t)$ and its relative change

Let's define $k(t) = (1 - e^{-\beta_2 t})$

$$k(t) = (1 - e^{-\beta_2 t}) \implies \mu(t) = \beta_0 \exp\left(\frac{\beta_1}{\beta_2} k(t)\right)$$

$$\frac{d\mu(t)}{dt} = \frac{\partial \mu(t)}{\partial k(t)} \frac{\partial k(t)}{\partial t} = \beta_0 \frac{\beta_1}{\beta_2} \exp\left(\frac{\beta_1}{\beta_2} k(t)\right) \beta_2 e^{-\beta_2 t}$$

$$\implies \frac{1}{\mu(t)} \frac{d\mu(t)}{dt} = \beta_1 e^{-\beta_2 t}$$

- **Parameters interpretation**

Table 1: values of μ and its relative change over time at $t = 0$ and $t \to \infty$

	$t = 0$	$t \to \infty$
$\mu(t)$	β_0	$\alpha_0 = \beta_0 \exp\left(\frac{\beta_1}{\beta_2}\right)$
$\frac{1}{\mu(t)} \frac{d\mu(t)}{dt}$	β_1	0

- β_0 is actually the expected value of the number of cancer cells in a given experiment at beginning of the latest.
- β_1 can be seen as the relative change of the expected mean of numbers of cancer cells in the experiment and $e^{-\beta_2}$ is the relative variation observed in this relative changes of the mean by time.

(b) Link between α_k and β_k

Given the available information,

$$\begin{aligned}
\mu(t_j) &= \beta_0 \exp\left(\frac{\beta_1}{\beta_2}(1-e^{-\beta_2 t_j})\right) \\
&= \beta_0 \exp\left(\frac{\beta_1}{\beta_2}\right)\exp\left(-\frac{\beta_1}{\beta_2}e^{-\beta_2 t}\right) = \alpha_0 \exp(-\alpha_1 e^{-\alpha_2 t})
\end{aligned}$$

$$\implies \alpha_0 = \beta_0 \exp\left(\frac{\beta_1}{\beta_2}\right), \alpha_1 = \frac{\beta_1}{\beta_2},\ \alpha_2 = \beta_2$$

Question 2

(a) Analytic form of likelihood function $\mathcal{L}(\boldsymbol{\alpha}/\mathcal{D}_i)$

By defining : $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)$, $y(t_j) = y_j$ the numbers of cancer cells detected by the technician at time t_j and $\mu(t_j) = \mu_j$ the parameter (the mean over time t_j) of the poisson distribution of y_j.

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}/\mathcal{D}_i) &= \prod_{j=1}^{J} p(y_j) = \prod_{j=1}^{J} \frac{e^{-\mu_j}\mu_j^{y_j}}{y_j!} = \frac{\prod_{j=1}^{J} e^{-\mu_j}\mu_j^{y_j}}{\prod_{j=1}^{J} y_j!} \\
&= \frac{e^{-\sum_{i=1}^{J}\mu_j}\prod_{j=1}^{J}\mu_j^{y_j}}{\prod_{j=1}^{J} y_j!}
\end{aligned}$$

(b) R function of log-likelihood

- **log-likelihood function** $\log(\mathcal{L}(\boldsymbol{\alpha}/\mathcal{D}_\rangle))$

$$\begin{aligned}
\log(\mathcal{L}(\boldsymbol{\alpha}/\mathcal{D}_\rangle)) &= \log(\prod_{j=1}^{J} p(y_j)) \\
&= \log(e^{-\sum_{i=1}^{J}\mu_j}\prod_{j=1}^{J}\mu_j^{y_j}) - \log(\prod_{j=1}^{J} y_j!) \\
&= \log(e^{-\sum_{i=1}^{J}\mu_j}) + \log(\prod_{j=1}^{J}\mu_j^{y_j}) - \log(\prod_{j=1}^{J} y_j!) \\
&= -\sum_{i=1}^{J}\mu_j \ + \sum_{i=1}^{J} y_j \log(\mu_j) \ - \sum_{i=1}^{J}\log(y_j!)
\end{aligned}$$

- **R function defining the log-likelihood**

```
loglikelihood <- function(theta0, theta1, theta2, data = dt[[2]], day = dt[[1]]){

  sum_mu_j <- sum(exp(theta0)*exp(exp(-(exp(theta1)*exp(-day*exp(theta2))))))
  sum_yj_log_muj <- sum(data*
                         log(exp(theta0)*exp(exp(-(exp(theta1)
                                                  *exp(-day*exp(theta2)))))))
  sum_log_yj <- sum(factorial(log(data)))

  return(-sum_mu_j + sum_yj_log_muj - sum_log_yj)
}

loglikelihood(.5, .2, .3)
```

```
## [1] -364958
```

(c) R function of log-posterior $p(\boldsymbol{\theta}/\mathcal{D}_i)$

- **log-posterior function**

Let's define the prior distribution of parameters $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$

Under independence assumptions between θs and the large variances,

$$p(\boldsymbol{\theta}) = p(\theta_0, \theta_1, \theta_2,) \propto 1$$

Considering $\theta_k = \log(\alpha_k)$ we have

$$\mu(t_j) = e^{\theta_0} \exp(-e^{\theta_1} e^{-t_j e^{\theta_2}})$$

This boils down the posterior distribution of $\boldsymbol{\theta}$ to the following :

$$\begin{aligned}
p(\boldsymbol{\theta}|\mathcal{D}_\rangle) &= \mathcal{L}(\mathcal{D}_\rangle|\boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \\
&= \mathcal{L}(\mathcal{D}_\rangle|\boldsymbol{\theta}) \times 1 \\
&= \frac{e^{-\sum_{j=1}^{J} \mu_j} \prod_{j=1}^{J} \mu_j^{y_j}}{\prod_{j=1}^{J} y_j!} \\
&\propto e^{-\sum_{j=1}^{J} \mu_j} \prod_{j=1}^{J} \mu_j^{y_j} \\
&\propto \exp\left(-e^{\theta_0} \sum_{j=1}^{J} \exp(-e^{\theta_1} e^{-t_j e^{\theta_2}})\right) \prod_{j=1}^{J} \left(e^{\theta_0} \exp(-e^{\theta_1} e^{-t_j e^{\theta_2}})\right)^{y_j}
\end{aligned}$$

From the above, the log-posterior is :

$$
\begin{aligned}
\log(p(\boldsymbol{\theta}|\mathcal{D}_\rangle)) &= \log(\mathcal{L}(\mathcal{D}_\rangle|\boldsymbol{\theta})) + \log(p(\boldsymbol{\theta})) \\
&\propto \left(-e^{\theta_0}\sum_{j=1}^{J}\exp(-e^{\theta_1}e^{-t_je^{\theta_2}})\right) + \sum_{j=1}^{J} y_j \log\left(e^{\theta_0}\exp(-e^{\theta_1}e^{-t_je^{\theta_2}})\right) \\
&\propto \left(-e^{\theta_0}\sum_{j=1}^{J}\exp(-e^{\theta_1}e^{-t_je^{\theta_2}})\right) + \sum_{j=1}^{J} y_j \left(\theta_0 - e^{\theta_1}e^{-t_je^{\theta_2}}\right) \\
&\propto \left(-e^{\theta_0}\sum_{j=1}^{J}\exp(-e^{\theta_1}e^{-t_je^{\theta_2}})\right) + \theta_0\sum_{j=1}^{J} y_j \; - e^{\theta_1}\sum_{j=1}^{J} y_j e^{-t_je^{\theta_2}}
\end{aligned}
$$

- **R function of log-posterior**

```
logposterior <- function(theta0,theta1, theta2, data=dt[[2]], day=dt[[1]],variance=1){
  firstsum <- -exp(theta0) * sum(exp(-exp(theta1)*exp(-day*exp(theta2))))
  secondsum <- theta0*sum(data)
  thirdsum <- exp(theta1)*sum(data*exp(-day*exp(theta2)))
  return(firstsum+secondsum-thirdsum)
}

logposterior(.5, .2, .3)
```

```
## [1] 50634.28
```