Poisson Likelihood

Suppose that a r.v. X obeys a Poisson distribution $\mathcal{P}(\lambda)$, characterized by the following Probability Mass Function (PMF)

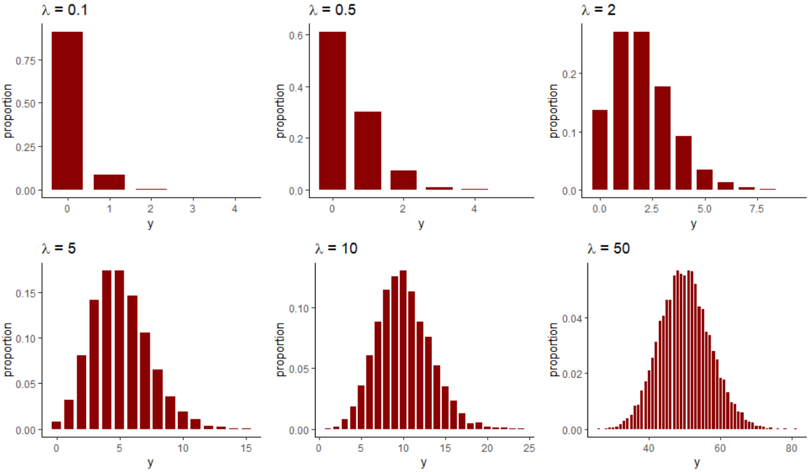
$$f_\lambda(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

If we observe a sample $x_1, ..., x_n$, then the likelihood is just the product of the individual PMF

$$\begin{aligned}\pi(x_1, ..., x_n \mid \lambda) &= \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n (x_i!)} \\ &= \frac{\lambda^{n\bar{x}} e^{-n\lambda}}{\prod_{i=1}^n (x_i!)}\end{aligned}$$

where the model parameter λ is the count of some event of interest and $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

Some examples of Poisson samples



Gamma prior

A convenient choice to model the uncertainty about λ is a Gamma distribution as prior, since the support of such a distribution is the interval $[0, \infty[$. A Gamma prior with shape parameter α and scale parameter β has the following form

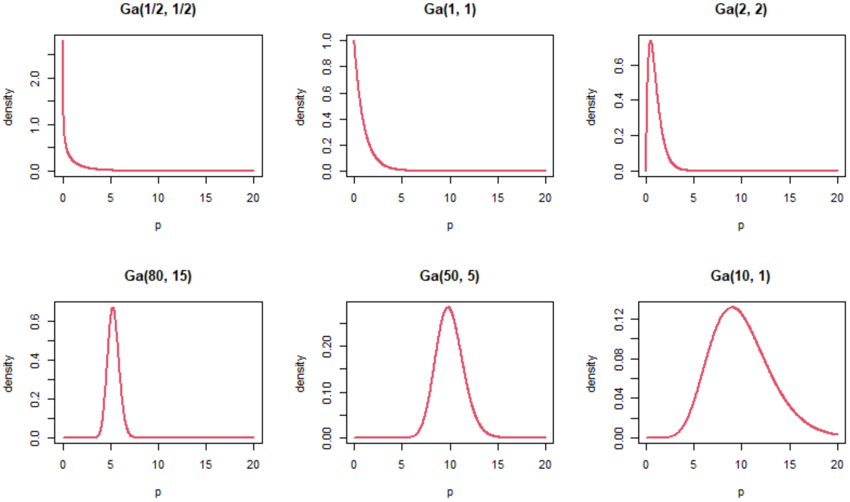
$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta}$$

where $\Gamma(\alpha)$ is the Gamma function, defined as $\int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha - 1)!$.

Let $X \sim \mathcal{G}a(\alpha, \beta)$. Then $E[X] = \alpha/\beta$ and $var(X) = \alpha/\beta^2$.

It is a convenient and flexible choice since the Gamma distribution can take a wide variety of shapes.

Some examples of the Gamma family



Gamma posterior (1/2)

The posterior distribution for λ if our data $x_1, ..., x_n$ is modeled with a Poisson likelihood and a Gamma prior is chosen for λ will also have the functional form of a Gamma r.v. Using the Bayes theorem, we have that

$$
\begin{aligned}
\pi(\lambda \mid x_1, ..., x_n) &= \frac{\pi(x_1, ..., x_n \mid \lambda)\pi(\lambda)}{\pi(x_1, ..., x_n)} \\
&= \frac{\lambda^{n\bar{x}} e^{-n\lambda}}{\prod_{i=1}^{n}(x_i!)} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta} \\
&= \underbrace{\frac{1}{\prod_{i=1}^{n}(x_i!)} \frac{\beta^\alpha}{\Gamma(\alpha)}}_{\text{do NOT depend on } \lambda} \lambda^{n\bar{x}} e^{-n\lambda} \lambda^{\alpha-1} e^{-\lambda\beta} \\
&\propto \lambda^{\alpha+n\bar{x}-1} e^{-(n+\beta)\lambda}
\end{aligned}
$$

Gamma posterior (2/2)

So we find that $\pi(\lambda \mid x_1, ..., x_n) \propto Gamma(\alpha + n\bar{x}, n + \beta)$. The posterior mean and variance are given by

$$E[\lambda] = \frac{\alpha + n\bar{x}}{n + \beta} \qquad var(\lambda) = \frac{\alpha + n\bar{x}}{(n + \beta)^2}$$

where $n\bar{x} = \sum_{i=1}^{n} x_i$ is the sum of the counts and n is the sample size. Let's take a few examples and plot the likelihood, a possible prior and the posterior, all at once in R.

General remark on Bayesian inference

Unlike in the traditional Frequentist framework, the Bayesian approach views parameters as random variables rather than fixed, unknown quantities. Given a Poisson sample $x_1, ..., x_n$ and a Poisson parameter λ, from the Bayes theorem, we can write

$$\pi(\lambda \mid x_1, ..., x_n) = \frac{\pi(x_1, ..., x_n \mid \lambda)\ \pi(\lambda)}{\pi(x_1, ..., x_n)}$$

Adopting the 'proportional' notation, the constant term in the denominator is dropped so that the above expression is rewritten as $\pi(\lambda \mid x_1, ..., x_n) \propto \pi(x_1, ..., x_n \mid \lambda)\ \pi(\lambda)$

When conjugate models are used (as in the case of a Poisson-Gamma model), the posterior distribution can be identified and closed-form quantities of interest like a mean, a variance or quantiles can be computed. Most of the time in practice, the posterior distribution is intractable so that it is necessary to resort to MCMC techniques.

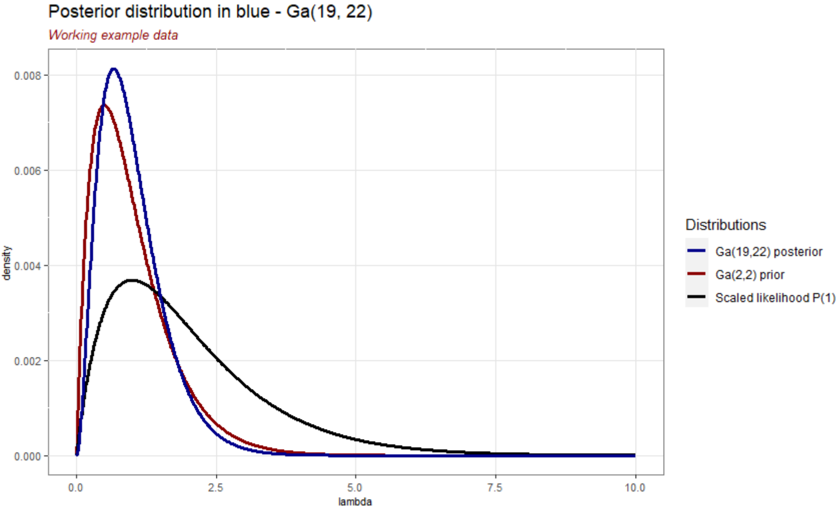
Working example

Suppose that we record the number of a specific bacteria present in 20 water samples taken in the Mekong Delta (Vietnam) so that we have the following data at hand:

$$x_i = 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 2, 0, 0, 5, 2, 0, 0, 2, 0, 1$$

So assuming a Poisson likelihood with parameter $\lambda = 1$, namely a $\mathcal{P}(1)$ likelihood for the data and using a $\mathcal{G}a(2, 2)$ prior, with mean $2/2 = 1$ and variance $2/2^2 = 0.5$, what is the posterior mean and the 95% credible intreval for the model parameter?

Working example: posterior



Posterior quantities obtained from direct sampling

```r
# Posterior mean, posterior variance and 95% Credible Interval including the sample median
set.seed(2023)
data1 = rpois(n = n, lambda = lambda1)
alpha_posterior = round(alpha1 + n*mean(data1), 2) # 19
beta_posterior = n + beta1 # 22

pmean = alpha_posterior / beta_posterior
pmean
# [1] 0.8636364

pvariance = alpha_posterior / beta_posterior^2
pvariance
# [1] 0.0392562

# 95% Credible Interval obtained by direct sampling (simulation)
set.seed(2023)
round(quantile(rgamma(n = 10^8, alpha_posterior, beta_posterior), probs = c(0.025, 0.5, 0.975)),4)
#   2.5%    50%  97.5%
# 0.5200 0.8486 1.2931

# Posterior mean obtained from direct sampling
set.seed(2023)
mean(rgamma(n = 10^8, alpha_posterior, beta_posterior))
# [1] 0.8928863
```

Jeffreys' prior: definition

Let us first give the general deffinition of the Jeffreys' prior $p_J(\theta)$:

$$p_J(\theta) = \sqrt{\mathcal{I}_n(\theta)}$$

where $\mathcal{I}_n(\theta)$ is the Fisher information of the sample. The Fisher information is defined as follows:

$$\mathcal{I}_n(\theta) = -E_\theta \left[\frac{\partial^2 ln \mathcal{L}(\theta \mid \mathbf{x})}{\partial \theta^2} \right]$$

or equivalently

$$\mathcal{I}_n(\theta) = var_\theta \left(\frac{\partial ln \mathcal{L}(\theta \mid \mathbf{x})}{\partial \theta} \right) = E_\theta \left[\left(\frac{\partial ln \mathcal{L}(\theta \mid \mathbf{x})}{\partial \theta} \right)^2 \right]$$

The first derivative of the log-likelihood function with respect to the model parameter $\frac{\partial ln \mathcal{L}(\theta|\mathbf{x})}{\partial \theta}$ is sometimes referred to as the score function.

Likelihood and derived functions for a Poisson model

Likelihood and log-likelihood, score function and second derivative of the log-likelihood function for a Poisson model:

$$\mathcal{L}(\lambda \mid \mathbf{x}) = \prod_{i=1}^{n} f_\lambda(x_i) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

$$l(p \mid \mathbf{x}) = ln\big(\mathcal{L}(\lambda \mid \mathbf{x})\big) = \sum_{i=1}^{n} x_i \; ln(\lambda) + -n\lambda - \sum_{i=1}^{n} log(x_i!)$$

$$\frac{\partial ln\mathcal{L}(\lambda \mid \mathbf{x})}{\partial\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

$$\frac{\partial^2 ln\mathcal{L}(\lambda \mid \mathbf{x})}{\partial\lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2}$$

Fisher information and Jeffrey's prior for a Poisson model (1/2)

So we want

$$\mathcal{I}_n(\lambda) = -E\left[\frac{\partial^2 ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial\lambda^2}\right]$$

and we note that

$$E\left[\sum_{i=1}^{n} x_i\right] = E[n\overline{x}] = nE[\overline{x}] = n\lambda$$

Fisher information and Jeffrey's prior for a Poisson model (2/2)

Thus we have that

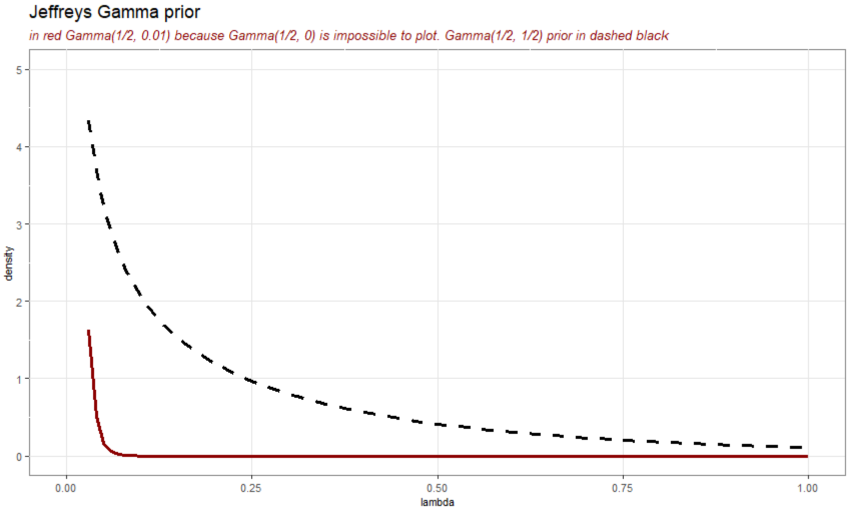
$$\begin{aligned}
\mathcal{I}_n(\lambda) &= \frac{E\left[\sum_{i=1}^{n} x_i\right]}{\lambda^2} \\
&= \frac{n\lambda}{\lambda^2} \\
&= \frac{n}{\lambda} \\
&\propto \lambda^{-1}
\end{aligned}$$

And we conclude that

$$p_J(\lambda) = \sqrt{\mathcal{I}(\lambda)} \propto \lambda^{-1/2}$$

which is the Gamma distribution $\mathcal{G}a(1/2, 0)$ (improper prior)

Gamma Jeffreys prior



Back to the working example (1/2)

Getting back to our water samples data

$$x_i = 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 2, 0, 0, 5, 2, 0, 0, 2, 0, 1$$

Assuming a Poisson likelihood for the data and using weakly informative prior close to the Jeffreys prior, namely a $Gamma(1/2, 1/2)$, what is the posterior mean and the 95% credible intreval for the model parameter ?

Back to the working example (2/2)

From the Bayes Theorem, we know that

$$p(\lambda \mid x) = \frac{p(x \mid \theta)\ p(\theta)}{p(x)} \propto p(x \mid \theta)\ p(\theta)$$

In our case, we have that

$$\begin{aligned} p(\lambda \mid x) &\propto \underbrace{\lambda^{\sum_{i=1}^{n} x_i} e^{-10\lambda}}_{\text{Poisson Likelihood}}\ \underbrace{p(\lambda)}_{\text{Prior}} \\ &\propto \lambda^{17} e^{-20\lambda}\ \lambda^{-1/2} e^{-1/2\lambda} \\ &\propto \lambda^{-16.5}\ e^{-19.5\lambda} \end{aligned}$$

and we recognize the functional form of a Gamma density, that is $Gamma(\alpha = 17.5, \beta = 19.5)$. The posterior mean is given by $\alpha/\beta = 0.8974$.

Working example: in conclusion

So the theoretical posterior mean is given by

$$E[\lambda] = \frac{\alpha + n\bar{x}}{n + \beta} = \frac{2 + 20 * 0.85}{20 + 2} = 19/22 = 0.8636364$$

By direct sampling, using 108 number of simulations, the posterior sample mean is 0.8636725.

By direct sampling, a 95% Credible Interval is given by

$$[0.5200, 1.2931]$$

So, combining modeling and simulations, we are now able to generalize and infer to the whole population of bacteria in the Mekong Delta those values from a sample of size 20.

Poisson-Gamma Gibbs sampling: example

Example of a multi-stage Gibbs sampler for reliability analysis. We want to set up a Bayesian analysis to eventually draw from the posterior distribution of the number of failures for a given time of observation of nuclear plant pumps based on some initial data.

Keywords: Reliability, Poisson process, Gibbs sampling, Poisson-Gamma model, Empirical Bayes

Pump data

The data: number of failures and times of observation of 10 pumps in a nuclear plant water system (Source: Gaver and O'Muircheartaigh, 1987)

	Pump	Failures	Time
1	1	5.00	94.32
2	2	1.00	15.72
3	3	5.00	62.88
4	4	14.00	125.76
5	5	3.00	5.24
6	6	19.00	31.44
7	7	1.00	1.05
8	8	1.00	1.05
9	9	4.00	2.10
10	10	22.00	10.48

Model and assumptions

The failure of the ith pump follow a Poisson process with parameter λ_i, for $i = 1, 2, ..., 10$. For an observed time t_i the number of failures p_i is thus a Poisson $\mathcal{P}(\lambda_i t_i)$ r.v.

Likelihood of failure

$$y_i \sim \mathcal{P}(\lambda_i t_i)$$

Prior on λ_i and prior on β

$$\lambda_i \sim \mathcal{G}a(\alpha, \beta)$$

$$\beta \sim \mathcal{G}a(\gamma, \delta)$$

with $\alpha = 1.8$, $\gamma = 0.01$ and $\delta = 1$ (see. Gaver and O'Muircheartaigh 1987 for a motivation of these numerical values)

Joint and conditional distributions

Joint posterior distribution

$$
\begin{aligned}
&\pi(\lambda_1, ..., \lambda_{10}, \beta \mid t_1, ..., t_{10}, p_1, ..., p_{10}) \\
&\propto \prod_{i=1}^{10} \left((\lambda_i t_i)^{p_i} e^{-\lambda_i t_i} \lambda_i^{\alpha-1} e^{-\beta\lambda_i} \right) \beta^{10\alpha} \beta^{\gamma-1} e^{-\delta\beta} \\
&\propto \prod_{i=1}^{10} \left((\lambda_i)^{p_i+\alpha-1} e^{-(t_i+\beta)\lambda_i} e^{-\beta\lambda_i} \right) \beta^{10\alpha+\gamma-1} e^{-\delta\beta}
\end{aligned}
$$

and a natural decomposition of π in conditional distributions is

$$\lambda_i \mid \beta, t_i, p_i \sim \mathcal{G}a(p_i + \alpha, t_i + \beta)$$

$$\beta \mid \lambda_1, ..., \lambda_{10} \sim \mathcal{G}a\left(\gamma + 10\alpha, \delta \sum_{i=1}^{10} \lambda_i\right)$$

Multistage Gibbs sampler for Poisson-Gamma model

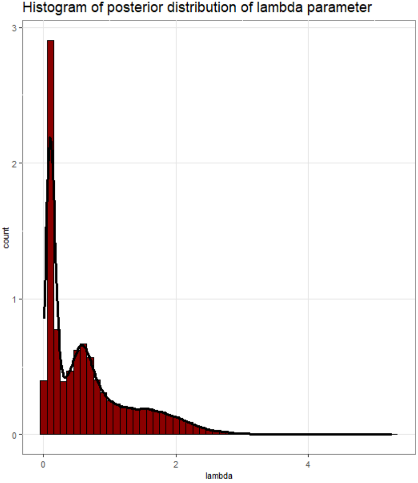
```
Gibbs_sampler_PG <- function(nsim, beta, alpha, gamma, delta, y, t, burnin) {

  X = matrix(0, nrow = nsim, ncol = length(y)+1) # empty matrix to record the simulated values
  X[1,1] = beta  # beta prior parameter
  X[1,c(2:(length(y)+1))] = rgamma(length(y), y + alpha, t + X[1,1]) # initial lambda

  for(i in 2:nsim) {

    X[i,c(2:(length(y)+1))] = rgamma(length(y), y + alpha, t + X[i-1,1]) # update lambda
    X[i,1] = rgamma(1, length(y) * alpha + gamma, delta + sum(X[i-1,c(2:(length(y)+1))])) # update beta
  }

  b <- burnin + 1 # record the burn in period (observations to be discarded)
  x <- X[b:nsim, ]

  return(list('lambda' = as.numeric(x[,c(2:(length(y)+1))]), 'beta' = x[,1] ))
}

# posterior
set.seed(2023)
posterior <- Gibbs_sampler_PG(nsim = 10000, beta = 1, alpha = 1.8, gamma = 0.01,
                              delta = 1, y = dataset[,2], t = dataset[,3], burnin = 1000)
```

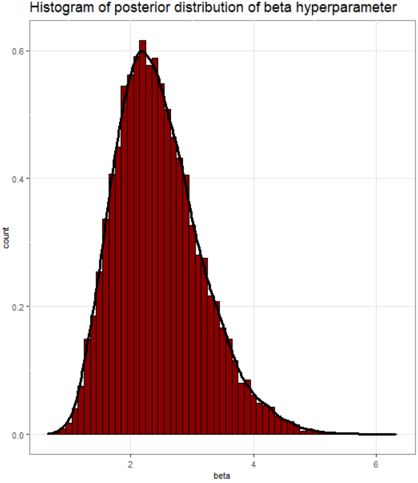
Posterior quantities obtained from Gibbs sampling

	lambda	beta
post mean	0.6510	2.4598
post sd	0.6534	0.7101
2.5%	0.0431	1.3196
50%	0.4626	2.3690
97.5%	2.2652	4.0806

We see that the distribution of the number of failures given time is bimodal and right skewed. We can now draw our conclusions not only based on point estimates but we have the whole distribution.

Posterior distributions





References and code

Donald P. Gaver I. G. O'Muircheartaigh (1987) Robust Empirical Bayes Analyses of Event Rates, Technometrics, 29:1, 1-15, DOI: 10.1080/00401706.1987.10488178

Robert, C. P., Casella, G. (1999). Monte Carlo statistical methods (Vol. 2). New York: Springer.

The R Project for Statistical Computing:
https://www.r-project.org/

Accessing the R code:
https://github.com/JRigh/Poisson-Gamma-example-in-R/blob/main/Poisson-Gamma