Poisson Likelihood

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

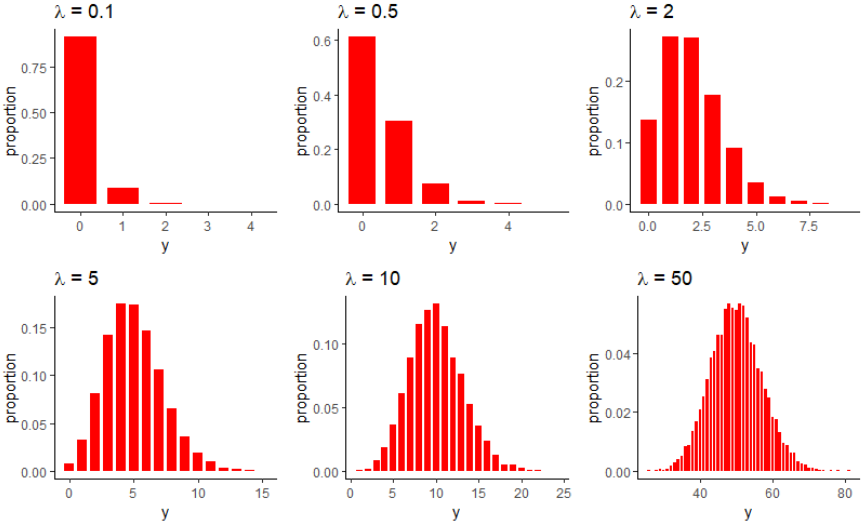
$$f_\lambda(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

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

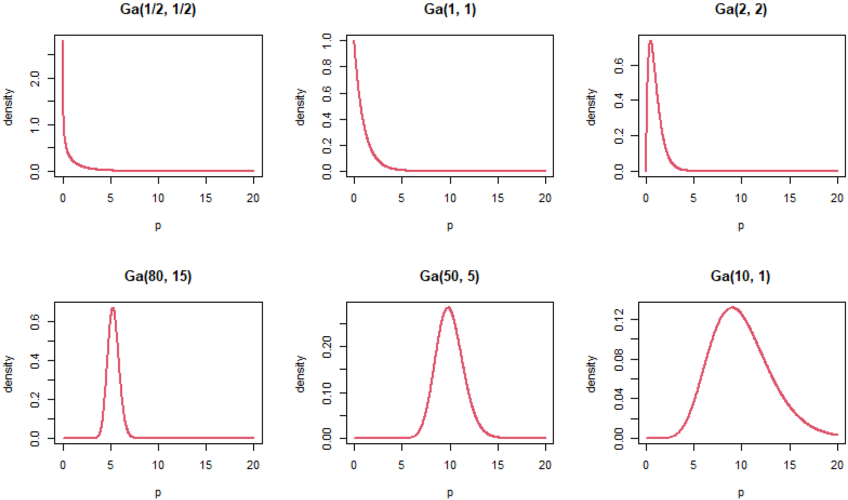
$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta}$$

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

Let $X \sim Gamma(\alpha, \beta)$. 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

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

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.

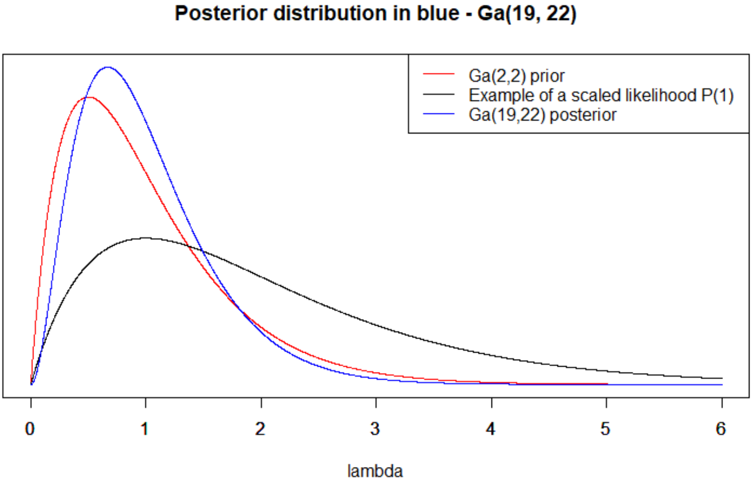
Working example

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 $Gamma(8, 8)$ 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
```

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 10^8 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.

Further reading and code

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