

Poisson Likelihood

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

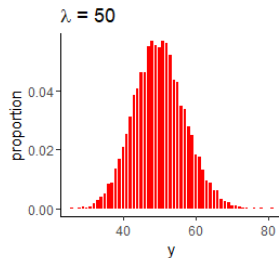
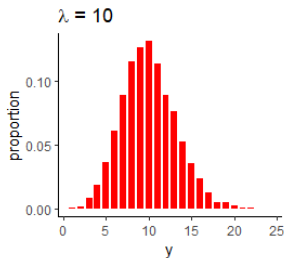
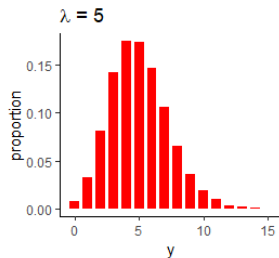
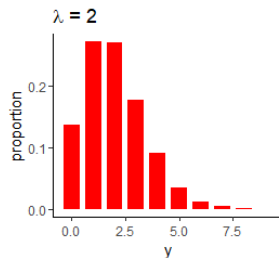
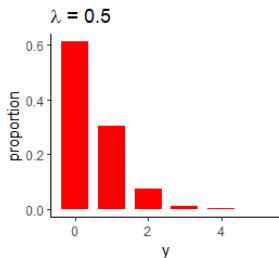
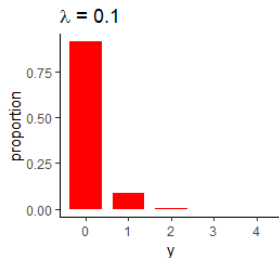
$$f_{\lambda}(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

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

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

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

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

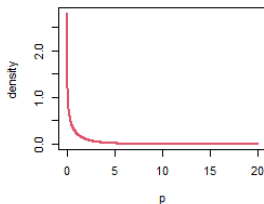
where $\Gamma(\alpha)$ is the Gamma function, defined as $\int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha - 1)!$.

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

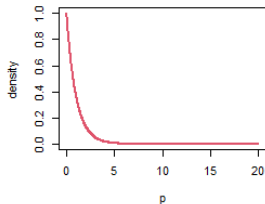
It is a convenient and flexible choice since the Gamma distribution can take a wide variety of shapes.

Some examples of the Gamma family

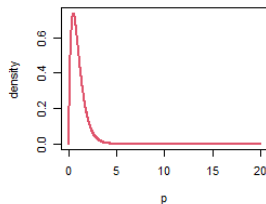
Ga(1/2, 1/2)



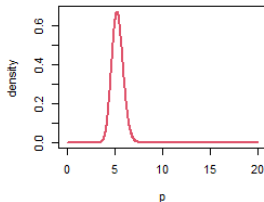
Ga(1, 1)



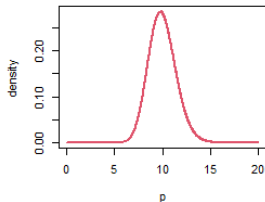
Ga(2, 2)



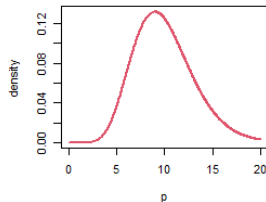
Ga(80, 15)



Ga(50, 5)



Ga(10, 1)



Gamma posterior (1/2)

The posterior distribution for λ if our data x_1, \dots, 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

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

Gamma posterior (2/2)

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

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

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, \dots, x_n and a Poisson parameter λ , from the Bayes theorem, we can write

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

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

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.

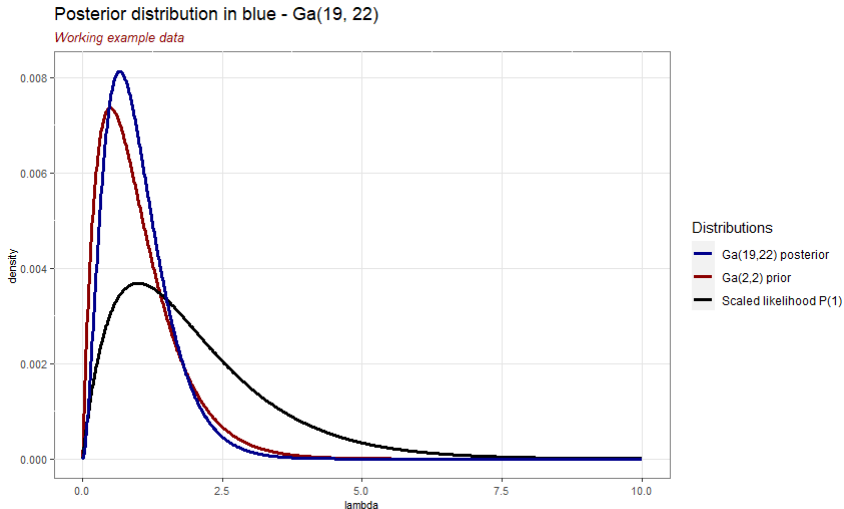
Working example

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 $Gamma(2, 2)$ prior, with mean $2/2 = 1$ and variance $2/2^2 = 0.5$, what is the posterior mean and the 95% credible interval for the model parameter?

Working example: posterior



Posterior quantities obtained from direct sampling

```
1 # Posterior mean, posterior variance and 95% Credible Interval including the
  sample median
2 set.seed(2023)
3 data1 = rpois(n = n, lambda = lambda1)
4 alpha_posterior = round(alpha1 + n*mean(data1), 2) # 19
5 beta_posterior = n + beta1 # 22
6
7 pmean = alpha_posterior / beta_posterior
8 pmean
9 # [1] 0.8636364
10
11 pvariance = alpha_posterior / beta_posterior^2
12 pvariance
13 # [1] 0.0392562
14
15 # 95% Credible Interval obtained by direct sampling (simulation)
16 set.seed(2023)
17 round(quantile(rgamma(n = 10^8, alpha_posterior, beta_posterior), probs = c
  (0.025, 0.5, 0.975)),4)
18 #      2.5%      50%     97.5%
19 # 0.5200 0.8486 1.2931
20
21 # Posterior mean obtained from direct sampling
22 set.seed(2023)
23 mean(rgamma(n = 10^8, alpha_posterior, beta_posterior))
24 # [1] 0.8928863
```

Jeffreys' prior: definition

Let us first give the general definition 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

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

The first derivative of the log-likelihood function with respect to the model parameter $\frac{\partial \ln \mathcal{L}(\theta \mid \mathbf{x})}{\partial \theta}$ 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:

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

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

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

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

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

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

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

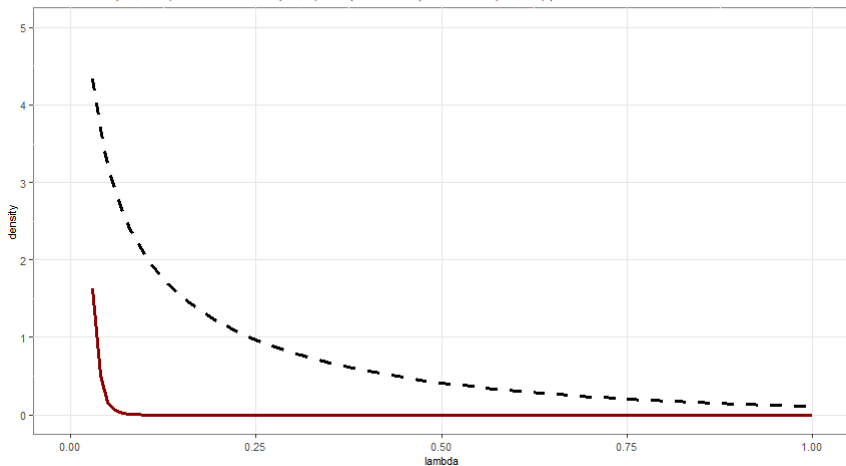
$$p_J(\lambda) = \sqrt{\mathcal{I}(\lambda)} \propto \lambda^{-1/2}$$

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

Gamma Jeffreys prior

Jeffreys Gamma prior

in red Gamma(1/2, 0.01) because Gamma(1/2, 0) is impossible to plot. Gamma(1/2, 1/2) prior in dashed black



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 interval 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

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

and we recognize the functional form of a Gamma density, that is $\text{Gamma}(\alpha = 17.5, \beta = 20.5)$. The posterior mean is given by $\alpha/\beta = 0.8537$.

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.

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](https://github.com/JRigh/Poisson-Gamma-example-in-R/blob/main/Poisson-Gamma)