Inference in a Poisson model

Mark Andrews

Psychology, Nottingham Trent University

mark.andrews@ntu.ac.uk

@xmjandrews, @priorexposure

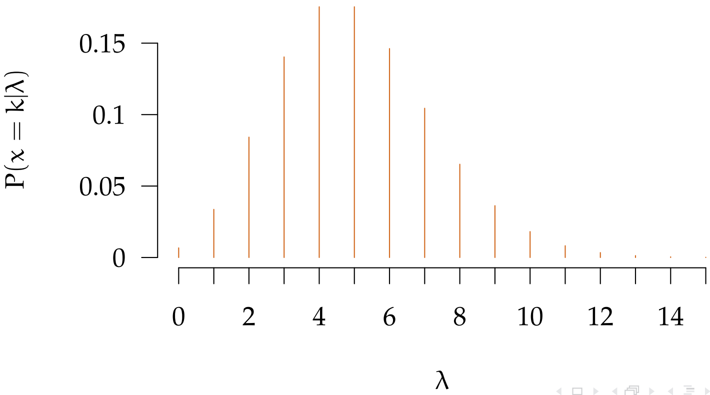
https://github.com/lawsofthought/psbayes

December, 2018

Inference in Poisson models

- The Poisson distribution can be used to model the rare occurrences in fixed intervals.
- If x is a Poisson random variable with rate λ, then

$$P(x = k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}.$$



Inference in Poisson models

- Let's say that the number of emails I get per hour, in $n = 10$ hours, is
$$D = 6, 8, 12, 7, 11, 14, 10, 12, 11, 16.$$
- We can assume that these frequencies are generated according to a Poisson process with rate λ per hour.
- Given this assumption and this observed data, what is the probable value of λ?
- In other words, what is
$$P(\lambda|D)?$$

Likelihood of a Poisson model

- Given a known value for λ, the probability of $x_1, x_2 \cdots x_n$ is

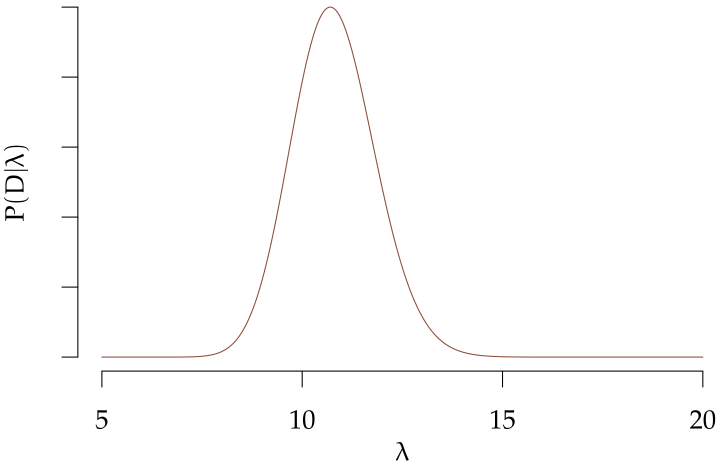
$$\begin{aligned} P(x_1, x_2 \cdots x_n | \lambda) &= \prod_{i=1}^{n} P(x_i|\lambda), \\ &= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}, \\ &\propto e^{-n\lambda} \prod_{i=1}^{n} \lambda^{x_i} = e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}. \end{aligned}$$

When $D = x_1, x_2 \cdots x_n = 6, 8 \cdots 16$, the likelihood of λ is

$$P(D|\lambda) = e^{-n\lambda} \lambda^{S},$$

where $S = \sum_{i=1}^{n} x_i = 107$.

Likelihood of a Poisson model

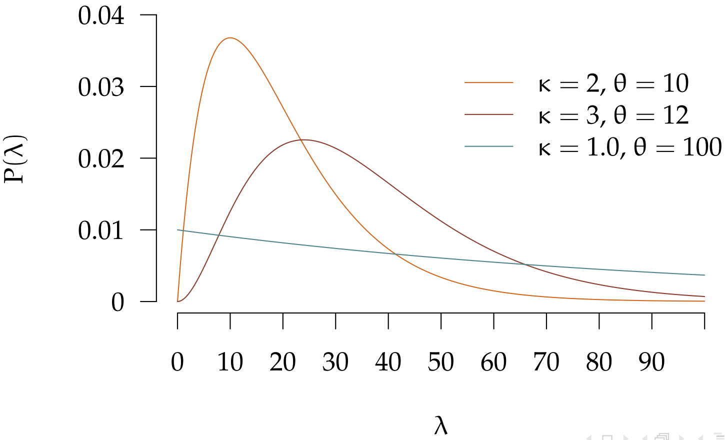


The likelihood of λ given the sufficient statistics $S = 107$ and $n = 10$.

Conjugate prior for the Poisson model

- The Gamma distribution with shape κ and scale θ is a conjugate prior for the Poisson model:

$$\text{Gamma}(\lambda|\kappa,\theta) = \frac{\lambda^{\kappa-1}e^{-\lambda/\theta}}{\theta^{\kappa}\Gamma(\kappa)}.$$



Posterior distribution

- With the Poisson likelihood, the Gamma prior leads to

$$
\begin{aligned}
P(\lambda|D,\kappa,\theta) &\propto e^{-n\lambda}\lambda^{S} \times \frac{\lambda^{\kappa-1}e^{-\lambda/\theta}}{\theta^{\kappa}\Gamma(\kappa)}, \\
&\propto e^{-\lambda\left(n+\frac{1}{\theta}\right)}\lambda^{S+\kappa-1}.
\end{aligned}
$$

Given that

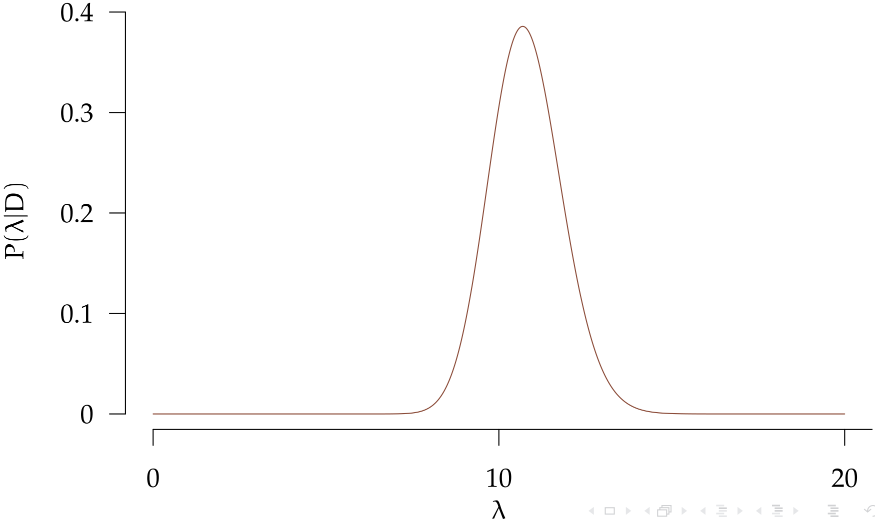
$$
\int e^{-\lambda\left(n+\frac{1}{\theta}\right)}\lambda^{S+\kappa-1}\, d\lambda = \left(n+\tfrac{1}{\theta}\right)^{-(S+\kappa)}\Gamma(S+\kappa),
$$

we have

$$
P(\lambda|D,\kappa,\theta) = \mathrm{Gamma}(\lambda|S+\kappa,(n+\tfrac{1}{\theta})^{-1}).
$$

Posterior distribution

- With prior hyper-parameters of $\kappa = 1.0$, $\theta = 100.0$, and sufficient statistics of $S = 107$ and $n = 10$, we have Gamma distribution with shape 108 and scale ≈ 0.1



Summarizing the posterior

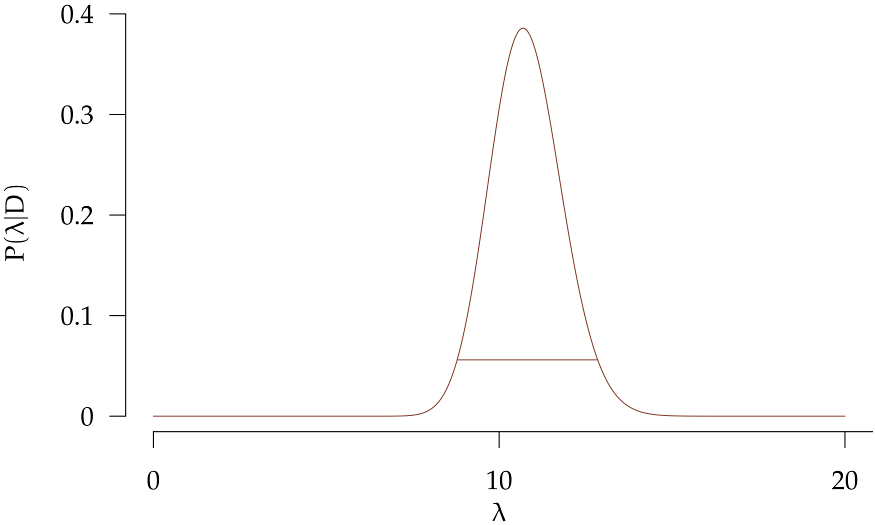
- The mean, variance and modes of any Gamma distribution with shape κ and scale θ are as follows:

$$\langle\lambda\rangle = \kappa\theta,$$
$$V(\lambda) = \kappa\theta^2,$$
$$\text{mode}(\lambda) = (\kappa - 1)\theta.$$

- Thus in our case, we have

$$\langle\lambda\rangle = 108.0,$$
$$V(\lambda) = 1.08, \quad \text{sd}(\theta) = 1.04,$$
$$\text{mode}(\lambda) = 10.69.$$

The 0.95 HPD *interval*



The 0.95 HPD interval is from 8.78 to 12.86.

Posterior predictive distribution

- Given that the number of emails every hour were

$$D = 6, 8, 12, 7, 11, 14, 10, 12, 11, 16,$$

  what do we predict will be number of emails in the next hour?
- This is given by the posterior predictive distribution:

$$\begin{aligned}
P(x_{\text{next}} = k | D, \kappa, \theta) &= \int P(x_{\text{next}} = k | \lambda) P(\lambda | D, \kappa, \theta) d\lambda, \\
&= \int \frac{e^{-k}\lambda^k}{k!} \times \frac{(n + \frac{1}{\theta})^{S+\kappa}}{\Gamma(S+\kappa)} e^{-\lambda(n+\frac{1}{\theta})}\lambda^{S+\kappa-1}, \\
&= \frac{\Gamma(S+\kappa+k)}{\Gamma(S+\kappa)\Gamma(k+1)} q^k (1-q)^{S+\kappa}, \\
&= \text{NegativeBinomial}(k | S+\kappa, q),
\end{aligned}$$

  with $q = \left(n + \frac{1}{\theta} + 1\right)^{-1}$.

Posterior predictive distribution

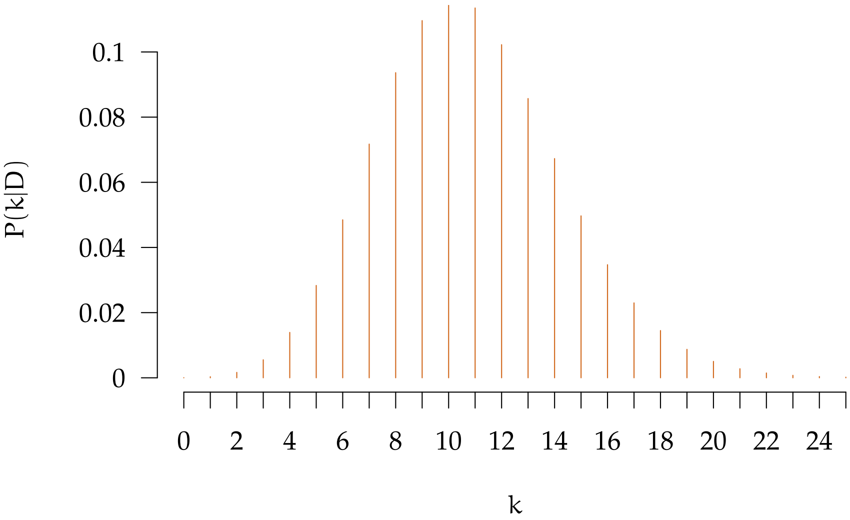
- The negative binomial distribution gives the number of “successes” until a predefined number r of “failures” have occurred, with the probability of a success being q.
- The mean and variance of the negative binomial are

$$\langle k \rangle = \frac{qr}{1-q}, \quad V(k) = \frac{qr}{(1-q)^2}.$$

- In our case, $r = S + \kappa = 108$ and $q = \left(n + \frac{1}{\theta} + 1\right)^{-1} = \left(10 + \frac{1}{100} + 1\right)^{-1} = 0.091.$
- The mean and variance of the negative binomial are

$$\langle k \rangle = \frac{S+\kappa}{n+\frac{1}{\theta}} = 10.79, \quad V(k) = \frac{S+\kappa}{n+\frac{1}{\theta}}(n + \frac{1}{\theta} + 1) = 11.87.$$

Posterior predictive distribution



The posterior predictive distribution $P(x_{next} = k|D, \kappa, \theta)$ is a negative binomial distribution with $r = 108$, $q = 0.091$.