A short introduction to Bayesian statistics, part III

D. Joyce, Clark University

Apr 2009

5 The Poisson process

A Poisson process is the continuous version of a Bernoulli process. In a Bernoulli process, time is discrete, and at each time unit there is a certain probability p that success occurs, the same probability at any given time, and the events at one time instant are independent of the events at other time instants.

In a Poisson process, time is continuous, and there is a certain rate λ of events occurring per unit time that is the same for any time interval, and events occur independently of each other. Whereas in a Bernoulli process at most one event occurs in a unit time interval, in a Poisson process any non-negative whole number of events can occur in unit time.

As in a Bernoulli process, you can ask various questions about a Poisson process, and the answers will have various distributions. If you ask how many events occur in an interval of length t, then the answer will have a Poisson distribution, $\text{Poisson}(\lambda t)$. It's probability mass function is

$$f(x) = \frac{1}{x!}(\lambda t)^x e^{-\lambda t} \qquad \text{for } x = 0, 1, \ldots.$$

If you ask how long until the first event occurs, then the answer will have an exponential distribution, $\text{Exponential}(\lambda)$, with probability density function

$$f(x) = \lambda e^{-\lambda x} \qquad \text{for } x \in [0, \infty).$$

If you ask how long until the r^{th} event, then the answer will have a gamma distribution $\text{Gamma}(\lambda, r)$. There are a couple different ways that gamma distributions are parametrized—either in terms of λ and r as done here, or in terms of α and β. The connection is $\alpha = r$, and $\beta = 1/\lambda$, which is the expected time to the first event in a Poisson process. The probability density function for a gamma distribution is

$$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}$$

for $x \in [0, \infty)$. The mean of a gamma distribution is $\alpha\beta = r/\lambda$ while its variance is $\alpha\beta^2 = r/\lambda^2$.

Our job is to get information about this parameter λ. Using the Bayesian approach, we have a prior density function $f(\lambda)$ on λ. Suppose over a time interval of length t we observe k events. The posterior density function is proportional to a conditional probability times the prior density function

$$f(\lambda \,|\, k) \propto P(k \,|\, \lambda)\, f(\lambda).$$

Now, k and t are constants, so

$$\begin{aligned} P(k \,|\, \lambda) &= P(k \text{ successes in time } t \,|\, \lambda) \\ &= \frac{1}{k!}(\lambda t)^k e^{-\lambda t} \\ &\propto \lambda^k e^{-\lambda t} \end{aligned}$$

Therefore, we have the following proportionality relating the posterior density function to the prior density function

$$f(\lambda \,|\, k) \propto \lambda^k e^{-\lambda t}\, f(\lambda).$$

Finding a family of conjugate priors. Again, we have the problem of deciding on what the prior density functions $f(\lambda)$ should be. Let's take one that seems to be natural and see what family of distributions it leads to. We know λ is some positive value, so we need a distribution on $(0, \infty)$. The exponential distributions are common distributions defined on $(0, \infty)$, so let's take the simplest one, with density

$$f(\lambda) = e^{-\lambda}$$

for $\lambda \geq 0$. Then

$$f(\lambda \,|\, k) \propto \lambda^k e^{-\lambda t} \, e^{-\lambda} = \lambda^k e^{-\lambda(t+1)}.$$

That makes the posterior distribution $f(\lambda \,|\, k)$ a gamma distribution $\textsc{Gamma}(\lambda, r) = \textsc{Gamma}(t + 1, k + 1)$ distribution since a $\textsc{Gamma}(\lambda, r)$ distribution has the density function

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} \propto x^{r-1} e^{-\lambda x}.$$

We have a little notational problem right now since we're using the symbol λ in two ways. First, it's the parameter to the Poisson process with a distribution; second, it's one of the two parameters of that distribution. From now on, I'll decorate the second use with subscripts somehow.

In this paragraph we have found that if λ had a prior distribution which was exponential, which in fact is a special case of a gamma distribution $\textsc{Gamma}(1, 1)$, then the posterior distribution was also a gamma distribution $\textsc{Gamma}(t + 1, k + 1)$.

More generally, the prior distribution can be any gamma distribution $\textsc{Gamma}(\lambda_0, r_0)$. Then if k successes are observed in time t, the posterior distribution will also be a gamma distribution, namely, $\textsc{Gamma}(\lambda_0 + t, r_0 + k)$. Essentially, the first coordinate keeps track of the total elapsed time while the second keeps track of the number of events.

Thus, a family of conjugate priors for the Poisson process parameter λ is the family of gamma distributions.

Selecting the prior distribution. How do you choose the right prior out of the family $\textsc{Gamma}(\lambda_0, r_0)$, that is, what do you choose for λ_0 and r_0?

One possibility is that you have a prior notion for the mean μ and variance σ^2 of λ. The mean for a $\textsc{Gamma}(\lambda_0, r_0)$ distribution is $\mu = r_0/\lambda_0$ and its variance is $\sigma^2 = r_0/\lambda_0^2$. These two equations can be solved for r_0 and λ_0 to give

$$r_0 = \mu^2/\sigma^2 \quad \text{ and } \lambda_0 = \mu/\sigma^2.$$

But what if you don't have any prior information? What's a good know-nothing prior? That's like saying that we've had no successes in no time. That suggests taking $\textsc{Gamma}(0, 0)$ as the prior on λ. Now $\textsc{Gamma}(\lambda, r)$ describes a gamma distribution only when $\lambda > 0$ and $r > 0$, so $\textsc{Gamma}(0, 0)$ is only a formal symbol. Nonetheless, as soon as we make an observation of k events in time t, with k at least 1, we can use the rule developed above to update it to $\textsc{Gamma}(t, k)$ which is an actual distribution.

A point estimator for λ. As mentioned above, the mean of a distribution on a parameter is a commonly taken as a point estimator for that parameter. Let the prior distribution for λ be $\textsc{Gamma}(\lambda_0, r_0)$. Then the prior estimator for λ is $\mu_\lambda = \dfrac{r_0}{\lambda_0}$. After an observation $\mathbf{x}$ with k events in time t, the posterior distribution will be $\textsc{Gamma}(\lambda_0 + t, r_0 + k)$, so the posterior estimator for λ is $\mu_{\lambda|\mathbf{x}} = \dfrac{r_0 + k}{\lambda_0 + t}$. If we took the prior to be the no-nothing prior of $\textsc{Gamma}(0, 0)$, that implies that posterior estimator for λ is just k/t, the rate of observed occurrences.