Poisson CI's

Robert L. Wolpert
Department of Statistical Science
Duke University, Durham, NC, USA

1 Interval Estimates

Point estimates of unknown parameters θ governing the distribution of an observed quantity X are unsatisfying if they come with no measure of accuracy or precision. One approach to giving such a measure is to offer *interval* estimates for θ, rather than *point* estimates; upon observing X, we construct an interval $[a, b]$ which is very likely to contain θ, and which is very short. The approach to exactly how these are constructed and interpreted is different for inference in the Sampling Theory tradition, and in the Bayesian tradition. In these notes I'll present both approaches to estimating the mean of a Poisson random variable on the basis of a random sample of fixed size n.

1.1 Sampling Theory: Confidence Intervals

Let $\{X_j\}$ be n iid observations from the Poisson distribution with unknown mean θ, and let $0 < \gamma < 1$ be a number between zero and one. Denote by X the vector of all n components, and by $\mathcal{X} = \mathbb{Z}_+^n$ the space of its possible values, non-negative n-tuples of integers; also denote by $\Theta = \mathbb{R}_+$ the possible values of θ. A γ-*Confidence Interval* is a pair of functions $A: \mathcal{X} \to \mathbb{R}$ and $B: \mathcal{X} \to \mathbb{R}$ with the property that

$$(\forall \theta \in \Theta) \quad \mathsf{P}_\theta[A(X) < \theta < B(X)] \geq \gamma. \tag{1}$$

Notice that this probability is for each *fixed* θ; it is the endpoints of the interval (A, B) that are random in this calculation, not θ.

Let's try to find a γ-Confidence Interval that is

- *symmetric* in the sense that the two possible errors each have the same error bound,

$$\mathsf{P}_\theta[\theta \le A(X)] \le \frac{1-\gamma}{2}, \qquad \mathsf{P}_\theta[B(X) \le \theta] \le \frac{1-\gamma}{2} \tag{2}$$

- as short as possible, subject to the error bound.

Clearly each function A and B will be a monotonically increasing function of the sufficient statistic $S = \Sigma X_j \sim \mathsf{Po}(n\theta)$; let's write A_S and B_S for those functions, and consider A_S first. Before we do, remember that the arrival time T_k for the k'th event in a unit-rate Poisson process X_t has the $\mathsf{Ga}(k,1)$ distribution, and that $X_t \ge k$ if and only if $T_k \le t$ (at least k fish by time t if and only if the k'th fish arrives before time t)— hence, in `R`, the CDF functions for Gamma and Poisson are related for all $k \in \mathbb{N}$ and $t > 0$ by

$$1 - \mathtt{ppois}(\mathtt{k}-1, \mathtt{t}) = \mathtt{pgamma}(\mathtt{t}, \mathtt{k}, 1).$$

Also recall the Gamma quantile function in `R`, an inverse for the CDF function, which satisfies $\mathtt{p} = \mathtt{pgamma}(\mathtt{t}, \mathtt{k}, \mathtt{b})$ if and only if $\mathtt{t} = \mathtt{qgamma}(\mathtt{p}, \mathtt{k}, \mathtt{b})$, and that if $Y \sim \mathsf{Ga}(\alpha, \beta)$ and $b > 0$ then $Y/b \sim \mathsf{Ga}(\alpha, \beta\, b)$, so

$$\mathtt{pgamma}(\mathtt{b}\,\theta, \alpha, 1) = \mathtt{pgamma}(\theta, \alpha, \mathtt{b}).$$

Fix any positive integer k. To achieve Equation (2) for $\theta \le A_k$, we need:

$$\begin{aligned} \frac{1-\gamma}{2} &\ge \mathsf{P}_\theta[\theta \le A(X)] \\ &\ge \mathsf{P}_\theta[S \ge k] \\ &= 1 - \mathtt{ppois}(\mathtt{k}-1, \mathtt{n}\theta) \\ &= \mathtt{pgamma}(\mathtt{n}\theta, \mathtt{k}, 1) = \mathtt{pgamma}(\theta, \mathtt{k}, \mathtt{n}), \quad \text{i.e.} \\ \mathtt{qgamma}(\frac{1-\gamma}{2}, \mathtt{k}, \mathtt{n}) &\ge \theta \end{aligned}$$

for each $\theta \le A_k$. Evidently this happens if and only if:

$$\mathtt{qgamma}(\frac{1-\gamma}{2}, \mathtt{k}, \mathtt{n}) \ge A_k. \tag{3}$$

Similarly, for nonnegative k and $B_k \le \theta$, Equation (2) requires:

$$\begin{aligned}
\frac{1-\gamma}{2} &\ge \mathsf{P}_\theta[B(X) \le \theta] \\
&\ge \mathsf{P}_\theta[S \le k] \\
&= \texttt{ppois}(\texttt{k}, \texttt{n}\theta) \\
&= 1 - \texttt{pgamma}(\texttt{n}\theta, \texttt{k}+\texttt{1}, \texttt{1}), \quad \text{i.e.} \\
\frac{1+\gamma}{2} &\le \texttt{pgamma}(\theta, \texttt{k}+\texttt{1}, \texttt{n}), \quad \text{i.e.} \\
\texttt{qgamma}(\frac{1+\gamma}{2}, \texttt{k}+\texttt{1}, \texttt{n}) &\le \theta
\end{aligned}$$

for each $\theta \ge B_k$. This happens if and only if:

$$\texttt{qgamma}(\frac{1+\gamma}{2}, \texttt{k}+\texttt{1}, \texttt{n}) \le B_k. \tag{4}$$

The shortest interval subject to the two constraints of Equations (3, 4) is:

$$A_k = \texttt{qgamma}(\frac{1-\gamma}{2}, \texttt{k}, \texttt{n}) \qquad B_k = \texttt{qgamma}(\frac{1+\gamma}{2}, \texttt{k}+\texttt{1}, \texttt{n}). \tag{5}$$

1.2 Bayesian Credible Intervals

A conjugate Bayesian analysis for iid Poisson data $\{X_j\} \overset{\text{iid}}{\sim} \mathsf{Po}(\theta)$ begins with the selection of parmeters $\alpha > 0$, $\beta > 0$ for a $\mathsf{Ga}(\alpha, \beta)$ prior density

$$\pi(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$$

and calculation of the likelihood function

$$\begin{aligned}
f(x \mid \theta) &= \prod_{j=1}^{n} \left[\frac{\theta^{x_j}}{x_j!} e^{-\theta} \right] \\
&\propto \theta^S e^{-n\theta},
\end{aligned}$$

where again $S = \sum_{j=1}^n X_j$. The posterior distribution is

$$\begin{aligned}
\pi(\theta \mid x) &\propto \theta^{\alpha+S-1} e^{-(\beta+n)\theta} \\
&\sim \mathsf{Ga}(\alpha+S,\ \beta+n).
\end{aligned}$$

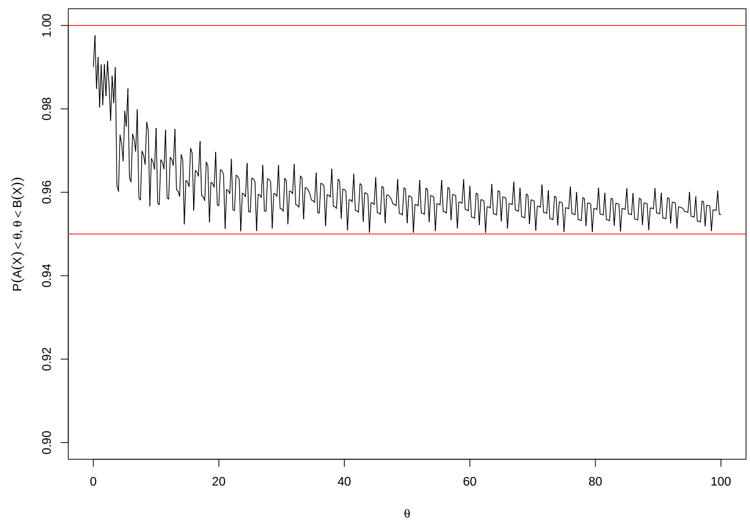

Figure 1: Exact coverage probability for 95% Poisson Confidence Intervals

Thus a symmetric γ posterior ("credible") interval for θ can be given by

$$\gamma = \mathsf{P}[a(X) < \theta < b(X) \mid X] \tag{6}$$

where $a(X) = a_S$ and $b(X) = b_S$ with

$$a_k = \texttt{qgamma}(\frac{1-\gamma}{2}, \alpha+\texttt{k}, \beta+\texttt{n}) \quad b_k = \texttt{qgamma}(\frac{1+\gamma}{2}, \alpha+\texttt{k}, \beta+\texttt{n}). \tag{7}$$

1.3 Comparison

The two probability statements in Equations (1, 6) are different— in Equation (1) the value of θ is fixed while X (and hence the sufficient statistic S) is random. Because S has a discrete distribution it is not possible to achieve exact equality for all θ; the probability $\mathsf{P}_\theta[A(X) < \theta < B(X)]$ (as a function of θ) jumps at each of the points $\{A_k, B_k\}$ (see Figure (1)). Instead we guarantee a minumum probability of γ ($\gamma = 0.95$ in Figure (1)) that θ will be captured by the interval. In Equation (6), however, X (and hence S) are fixed, and we consider θ to be random; it has a continuous distribution, and it is possible to achieve exact equality.

The formulas for the interval endpoints given in Equations (5, 7) are similar— if we take $\beta = 0$ and $\alpha = \frac{1}{2}$ they will be as close as possible to each other.

Note that this corresponds to an improper $\mathsf{Ga}(\frac{1}{2}, 0)$ prior distribution

$$\pi(\theta) \propto \theta^{-1/2} \mathbf{1}_{\{\theta > 0\}},$$

but the *posterior* distribution $\pi(\theta \mid X) \sim \mathsf{Ga}(S+\frac{1}{2}, n)$ is proper for any $X \in \mathcal{X}$. For any α and β, all the intervals have the same asymptotic behavior for large n; by the central limit theorem,

$$A(X), a(X) \rightsquigarrow \bar{X} - z_\gamma \sqrt{\bar{X}/n}, \qquad B(X), b(X) \rightsquigarrow \bar{X} + z_\gamma \sqrt{\bar{X}/n}$$

where $\Phi(z_\gamma) = (1 + \gamma)/2$, so $\gamma = \Phi(z_\gamma) - \Phi(-z_\gamma)$.