Confidence & Credible Interval Estimates

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

1 Introduction

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 means of the Normal and Exponential distributions, using "pivotal quantities," and of Poisson random variables, using detailed features of the distribution, on the basis of a random sample of fixed size n.

1.1 Pivotal Quantities

A *pivotal quantity* is a function of the data *and* the parameters (so it's not a *statistic*) whose probability distribution does not depend on any uncertain parameter values. Some examples:

- $\mathsf{Ex}(\lambda)$**:** If $X \sim \mathsf{Ex}(\lambda)$ then $\lambda X \sim \mathsf{Ex}(1)$ is pivotal and, for samples of size n, $\lambda \bar{X}_n \sim \mathsf{Ga}(n, n)$ and $2n\lambda\bar{X}_n \sim \chi^2_{2n}$ are pivotal.
- $\mathsf{Ga}(\alpha, \lambda)$**:** If $X \sim \mathsf{Ga}(\alpha, \lambda)$ with α known then $\lambda X \sim \mathsf{Ga}(\alpha, 1)$ is pivotal and, for samples of size n, $\lambda\bar{X}_n \sim \mathsf{Ga}(\alpha n, n)$ and $2n\lambda\bar{X}_n \sim \chi^2_{2\alpha n}$ are pivotal.
- $\mathsf{No}(\mu, \sigma^2)$: If μ is unknown but σ^2 is known, then $(X-\mu)/\sigma \sim \mathsf{No}(0, 1)$ is pivotal and, for samples of size n, $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathsf{No}(0, 1)$ is pivotal.

- $\mathsf{No}(\mu, \sigma^2)$: If μ is known but σ^2 is unknown, then $(X-\mu)/\sigma \sim \mathsf{No}(0,1)$ is pivotal and, for samples of size n, $\sum(X_i-\mu)^2/\sigma^2 \sim \chi^2_n$ is pivotal.
- $\mathsf{No}(\mu, \sigma^2)$: If μ and σ^2 are *both* unknown then for samples of size n, both $\sum(X_i-\bar{X}_n)^2/\sigma^2 \sim \chi^2_{n-1}$ and $\sqrt{n}(\bar{X}_n-\mu)/\sigma \sim \mathsf{No}(0,1)$ are pivotal. This is the key example below.
- $\mathsf{Un}(0,\theta)$: If $X \sim \mathsf{Un}(0,\theta)$ with θ unknown then $(X/\theta) \sim \mathsf{Un}(0,1)$ is pivotal and, for samples of size n, $\max(X_i)/\theta \sim \mathsf{Be}(n,1)$ is pivotal. Find a sufficient pair of pivotal quantities for $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{Un}(\alpha,\beta)$.
- $\mathsf{We}(\alpha,\beta)$: If $X \sim \mathsf{We}(\alpha,\beta)$ has a Weibull distribution then $\beta X^\alpha \sim \mathsf{Ex}(1)$ is pivotal.

Pivotal quantites allow us to construct sampling-theory "confidence intervals" for uncertain parameters. For example, since $2n\lambda\bar{X}_n \sim \chi^2_{2n}$ for iid $\{X_j\} \sim \mathsf{Ex}(\lambda)$, and since the 5% and 95% quantiles of the χ^2_{10} distribution are 3.940299 and 18.307038, a sample of size $n=5$ from the $\mathsf{Ex}(\lambda)$ distribution satisfies

$$\begin{aligned} 0.90 &= \Pr[3.940299 < 2n\lambda\bar{X}_n < 18.307038] \\ &= \Pr\left[\frac{0.3940299}{\bar{X}_5} < \lambda < \frac{1.8307038}{\bar{X}_5}\right], \end{aligned}$$

so $[0.39/\bar{X}_5,\ \ 1.83/\bar{X}_5]$ is a 90% confidence interval for λ.

Most discrete distributions don't have (exact) pivotal quantities, but the central limit theorem usually leads to approximate confidence intervals for most distributions for large samples. Exact intervals are available for many distributions, with a little more work, even for small samples; see Section (4) for a construction of exact confidence and credible intervals for the Poisson distribution. Ask me if you're interested in more details.

2 Confidence Intervals for a Normal Mean

First let's verify the claim above that, for a sample of size n from the $\mathsf{No}(\mu,\sigma^2)$ distribution, the pivotal quantities $\sum(X_i-\bar{X}_n)^2/\sigma^2 \sim \chi^2_{n-1}$ and $\sqrt{n}(\bar{X}_n-\mu)/\sigma \ \sim \mathsf{No}(0,1)$ have the distributions I claimed for them— and, moreover, that they are independent.

Let $\mathbf{x} = \{X_1,\dots,X_n\} \overset{\text{iid}}{\sim} \mathsf{No}(\mu,\sigma^2)$ be a simple random sample from a normal distribution mean μ and variance σ^2. The log likelihood function for

$\theta = (\mu, \sigma^2)$ is

$$\log f(\mathbf{x} \mid \theta) = \log \left\{ (2\pi\sigma^2)^{-n/2} \, e^{-\sum (x_i-\mu)^2/2\sigma^2} \right\}$$
$$= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \bar{x}_n)^2 - \frac{n(\bar{x}_n - \mu)^2}{2\sigma^2}$$

so the MLEs are

$$\hat{\mu} = \bar{x}_n = \frac{1}{n} \sum x_i \qquad \text{and} \qquad \hat{\sigma}^2 = s_n^2/n = \frac{1}{n} \sum (x_i - \bar{x}_n)^2.$$

First we turn to discovering the probability distributions of these estimators, so we can make sampling-based interval estimates for μ and σ^2.

Since the $\{X_i\}$ are independent, their sum has a $\mathsf{No}(n\mu,\, n\sigma^2)$ distribution and

$$\hat{\mu} = \bar{x}_n \sim \mathsf{No}(\mu, \sigma^2/n).$$

Since the *co*variance between $\bar{x}_n$ and each component of $(\mathbf{x} - \bar{x}_n)$ is zero, and since they're all jointly Gaussian, $\bar{x}_n$ must be *independent* of $(\mathbf{x} - \bar{x}_n)$ and hence of any function of $(\mathbf{x} - \bar{x}_n)$, including $s_n^2 = \sum (x_i - \bar{x}_n)^2$. Now we can use moment generating functions to discover the distribution of s_n^2.

Since $Z^2 \sim \chi_1^2 = \mathsf{Ga}(1/2, 1/2)$ for a standard normal $Z \sim \mathsf{No}(0,1)$, we have

$$n(\bar{x}_n - \mu)^2/\sigma^2 \sim \mathsf{Ga}(1/2, 1/2) \quad \text{and} \quad \sum (x_i - \mu)^2/\sigma^2 \sim \mathsf{Ga}(n/2, 1/2)$$

so

$$n(\bar{x}_n - \mu)^2 \sim \mathsf{Ga}(1/2, 1/2\sigma^2) \quad \text{and} \quad \sum (x_i - \mu)^2 \sim \mathsf{Ga}(n/2, 1/2\sigma^2).$$

By completing the square we have

$$\sum (x_i - \mu)^2 = \sum (x_i - \bar{x}_n)^2 + n(\bar{x}_n - \mu)^2$$

as the sum of two *independent* terms. Recall (or compute) that the Gamma $\mathsf{Ga}(\alpha, \beta)$ MGF is $(1 - t/\beta)^{-\alpha}$ for $t < \beta$, and that the MGF for the sum of independent random variables is the product of the individual MGFs, so

$$\begin{aligned}
\mathsf{E}\exp\left\{t\sum(x_i-\mu)^2\right\} &= (1-2\sigma^2 t)^{-n/2} \\
&= \mathsf{E}\exp\left\{t\sum(x_i-\bar{x}_n)^2\right\}\mathsf{E}\exp\left\{tn(\bar{x}_n-\mu)^2\right\} \\
&= \mathsf{E}\exp\left\{t\sum(x_i-\bar{x}_n)^2\right\}(1-2\sigma^2 t)^{-1/2}, \text{ so} \\
\mathsf{E}\exp\left\{t\sum(x_i-\bar{x}_n)^2\right\} &= (1-2\sigma^2 t)^{-(n-1)/2} \text{ by dividing. Thus} \\
s_n^2 \equiv \sum(x_i-\bar{x}_n)^2 &\sim \mathsf{Ga}\left(\frac{n-1}{2},\frac{1}{2\sigma^2}\right) \text{ and so} \\
\frac{s_n^2}{\sigma^2} &\sim \chi^2_{n-1} \text{ is independent of } \frac{\bar{x}_n-\mu}{\sqrt{\sigma^2/n}} \sim \mathsf{No}(0,1).
\end{aligned}$$

2.1 Confidence Intervals for μ when the Variance is Known

The pivotal quantity $Z=(\bar{x}_n-\mu)/\sqrt{\sigma^2/n}$ has a standard $\mathsf{No}(0,1)$ normal distribution, with CDF $\Phi(z)$. If σ^2 were known, then for any $0<\gamma<1$ and for z^* such that $\Phi(z^*)=(1+\gamma)/2$ we could write

$$\begin{aligned}
\gamma &= \mathsf{P}_\mu\left[-z^*\le\frac{\bar{x}_n-\mu}{\sqrt{\sigma^2/n}}\le z^*\right] \\
&= \mathsf{P}_\mu\left[\bar{x}_n-z^*\sigma/\sqrt{n}\le\mu\le\bar{x}_n+z^*\sigma/\sqrt{n}\right]
\end{aligned}$$

and find a "confidence interval" $[\bar{x}_n-z^*\sigma/\sqrt{n},\ \bar{x}_n+z^*\sigma/\sqrt{n}]$ for μ by replacing $\bar{x}_n$ with its observed value. Note this is a *random interval*; from a sampling theory perspective, μ is fixed so once we replace "$\bar{x}_n$" with its value, the interval is no longer random and we can no longer make probability statements about it. That's why the word "confidence" is used for these intervals, and not "probability."

If σ^2 is *not* known, and must be estimated from the data, then we have a bit of a problem— because although the analogous quantity

$$\frac{\bar{x}_n-\mu}{\sqrt{\hat{\sigma}^2/n}}$$

does have a probability distribution that does not depend on μ or σ^2, and so *is* pivotal, its disribution is *not* Normal. Heuristically, this quantity has "fatter tails" than the normal density function, because it can be far from zero if *either* $\bar{x}_n$ is far from its mean μ *or* if the estimate $\hat{\sigma}^2$ for the variance is too small.

Following William S. Gosset (a statistician working for the Guinness Brewery in Dublin, Ireland) as adapted by Ronald A. Fisher (an English theoretical statistician), we consider the (slightly rescaled, by Fisher) pivotal quantity:

$$\begin{aligned} t &:= \frac{\bar{x}_n - \mu}{\sqrt{\hat{\sigma}^2/(n-1)}} = \frac{(\bar{x}_n - \mu)\sqrt{n}}{\sqrt{s_n^2/(n-1)}} \\ &= \frac{\frac{\bar{x}_n-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{s_n^2}{\sigma^2(n-1)}}} \qquad = \frac{Z}{\sqrt{Y/\nu}}, \end{aligned}$$

for independent $Z \sim \mathsf{No}(0,1)$ and $Y \sim \chi^2_\nu$ with $\nu = n-1$. Now we turn to finding the density function for t.

2.2 The t pdf

Note $X \equiv Z^2/2 \sim \mathsf{Ga}(1/2, 1)$ and $U \equiv Y/2 \sim \mathsf{Ga}(\nu/2, 1)$. Since Z has a symmetric distribution about zero, so does t and its pdf will satisfy $f_\nu(t) = f_\nu(-t)$. For $t > 0$,

$$\begin{aligned} \mathsf{P}\left[\frac{Z}{\sqrt{Y/\nu}} > t\right] &= \frac{1}{2}\mathsf{P}\left[\frac{Z^2}{Y/\nu} > t^2\right] = \frac{1}{2}\mathsf{P}\left[\frac{Z^2}{2} > \frac{Y}{2}\,\frac{t^2}{\nu}\right] \\ &= \frac{1}{2}\int_0^\infty \left\{\int_{ut^2/\nu}^\infty \frac{x^{-1/2}}{\Gamma(1/2)} e^{-x}\, dx\right\} \frac{u^{\nu/2-1}}{\Gamma(\nu/2)} e^{-u}\, du \end{aligned}$$

Taking the negative derivative wrt t, and noting $\Gamma(1/2) = \sqrt{\pi}$,

$$\begin{aligned} f_\nu(t) &= \frac{1}{2\sqrt{\pi}} \int_0^\infty \left\{\frac{2ut}{\nu}(u\,t^2/\nu)^{-1/2} e^{-ut^2/\nu}\, \frac{u^{\nu/2-1}}{\Gamma(\nu/2)} e^{-u}\right\} du \\ &= \frac{1}{\sqrt{\pi\nu}\;\Gamma(\nu/2)} \int_0^\infty u^{\frac{\nu+1}{2}-1}\; e^{-u(1+t^2/\nu)}\, du \\ &= \left[\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\;\Gamma(\nu/2)}\right] \frac{1}{(1+t^2/\nu)^{(\nu+1)/2}}, \end{aligned}$$

the Student t density function. It's not important to remember or be able to reproduce the derivation, or to remember the normalizing constant— but it's useful to know a few things about the t density function:

- $f_\nu(t) \propto (1+t^2/\nu)^{-(\nu+1)/2}$ is symmetric and bell-shaped, but falls off to zero as $t \to \pm\infty$ more slowly than the Normal density ($f_\nu(t) \asymp |t|^{-\nu-1}$ while $\phi(z) \asymp e^{-z^2/2}$).

- For one degree of freedom $\nu = 1$, the t_1 is identical to the standard Cauchy distribution $f_1(t) = \pi^{-1}/(1+t^2)$, with undefined mean and infinite variance.

- As $\nu \to \infty$, $f_\nu(t) \to \phi(t)$ converges to the standard Normal $\mathsf{No}(0,1)$ density function.

2.3 Confidence Intervals for μ when σ^2 is Unknown

With the t distribution (and hence its CDF $F_\nu(t)$) now known, for any random sample $\mathbf{x} = \{X_1, \dots, X_n\} \overset{\text{iid}}{\sim} \mathsf{No}(\mu, \sigma^2)$ from the normal distribution we can set $\nu \equiv n-1$ and compute sufficient statistics

$$\bar{x}_n = \frac{1}{n}\sum x_i \qquad \hat{\sigma}_n^2 = \frac{1}{n}s_n^2 = \frac{1}{n}\sum (x_i - \bar{x}_n)^2$$

and, for any $0 < \gamma < 1$, find t^* such that $F_\nu(t^*) = (1+\gamma)/2$, then compute

$$\begin{aligned}\gamma &= \mathsf{P}_\mu\left[-t^* \le \frac{\bar{x}_n - \mu}{\sqrt{\hat{\sigma}_n^2/\nu}} \le t^*\right] \\ &= \mathsf{P}_\mu\left[\bar{x}_n - t^*\hat{\sigma}/\sqrt{\nu} \le \mu \le \bar{x}_n + t^*\hat{\sigma}/\sqrt{\nu}\right]\end{aligned}$$

Once again we have a random interval that will contain μ with specified probability γ— and can replace the sufficient statistics $\bar{x}_n$, $\hat{\sigma}^2$ with their observed values to get a confidence interval. In this sampling theory approach the unknown μ and σ^2 are kept fixed and only the data $\mathbf{x}$ are treated as random; that's what the subscript "μ" on P was intended to suggest.

For example, with just $n = 2$ observations the t distribution will have only $\nu = 1$ degrees of freedom, so it coincides with the Cauchy distribution with CDF

$$F_1(t) = \int_{-\infty}^{t} \frac{1/\pi}{1+x^2}\,dx = \frac{1}{2} + \frac{1}{\pi}\arctan(t).$$

For a confidence interval with $\gamma = 0.95$ we need $t^* = \tan[\pi(0.975 - 0.5)] = \tan(0.475\pi) = 12.70620$ (*much* larger than $z^* = 1.96$); the interval is

$$0.95 = \mathsf{P}_\mu[\bar{x} - 12.7\hat{\sigma} \le \mu \le \bar{x} + 12.7\hat{\sigma}]$$

with $\bar{x} = (x_1 + x_2)/2$ and $\hat{\sigma} = |x_1 - x_2|/2$.

When σ^2 is known it is better to use its known value than to estimate it, because (on average) the interval estimates will be shorter. Of course some things we "know" turn out to be false... or, as Will Rogers put it,

It isn't what we don't know that gives us trouble,
it's what we know that ain't so.

3 Bayesian Credible Intervals for a Normal Mean

How would we make inference about μ for the normal distribution using Bayesian methods?

3.1 Unknown mean μ, known precision $\tau = \sigma^{-2}$

When only the mean μ is uncertain but the *precision* $\tau \equiv 1/\sigma^2$ is known, the normal likelihood function

$$f(\mathbf{x} \mid \mu) = (\tau/2\pi)^{n/2}\, e^{-\frac{\tau}{2}\sum(x_i-\bar{x}_n)^2-\frac{\tau}{2}n(\bar{x}_n-\mu)^2} \propto e^{-\frac{\tau_0}{2}(\mu-\mu_0)^2}$$

is proportional to a normal density in μ with mean $\mu_0 = \bar{x}_n$ and precision $\tau_0 = n\tau$, so $\pi_\mu(\mu) \sim \mathsf{No}(\mu_0, \tau_0^{-1})$ is a conjugate prior distribution. The posterior in this case is again normal $\mu \mid \mathbf{x} \sim \mathsf{No}(\mu_1, \tau_1^{-1})$ with updated "hyper-parameters"

$$\mu_1 = \frac{\tau_0\mu_0 + n\tau\bar{x}_n}{\tau_0 + n\tau}, \qquad \tau_1 = \tau_0 + n\tau.$$

Posterior credible intervals are available of any size $0 < \gamma < 1$; using the critical value z^* such that $\Phi(z^*) = (1+\gamma)/2$, we have

$$\gamma = \mathsf{P}\Big[\mu_1 - z^*/\sqrt{\tau_1} < \mu < \mu_1 + z^*/\sqrt{\tau_1} \quad \Big|\ \mathbf{x}\Big].$$

The limit as $\mu_0 \to 0$ and $\tau_0 \to 0$ leads to the improper uniform prior $\pi_\mu(\mu) \propto 1$ with posterior $\mu \mid \mathbf{x} \sim \mathsf{No}(\mu_1 = \bar{x}_n, \tau_1^{-1} = \sigma^2/n)$, with a Bayesian credible interval $[\bar{x}_n - z^*\sigma/\sqrt{n},\ x_n + z^*\sigma/\sqrt{n}]$ identical to the sampling theory confidence interval of Section (2.1), but with a different interpretation: here $\gamma = \mathsf{P}[\bar{x}_n - z^*\sigma/\sqrt{n} < \mu < x_n + z^*\sigma/\sqrt{n} \mid \mathbf{x}]$ (with μ random, $\bar{x}_n$ fixed), while for the sampling-theory interval $\gamma = \mathsf{P}_\mu[\bar{x}_n - z^*\sigma/\sqrt{n} < \mu < x_n + z^*\sigma/\sqrt{n}]$ (with $\bar{x}_n$ random for fixed but unknown μ, τ).

3.2 Known mean μ, unknown precision $\tau = \sigma^{-2}$

When μ is known and only the precision τ is uncertain,

$$f(\mathbf{x} \mid \tau) \propto \tau^{\alpha_0 - 1} e^{-\beta_0 \tau}$$

is proportional to a gamma density in τ with shape $\alpha_0 = 1 + n/2$ and rate $\beta_0 = \Sigma(x_i - \mu)^2$, so $\pi_\tau(\tau) \sim \mathsf{Ga}(\alpha_0, \beta_0)$ is conjugate for τ. The posterior distribution is $\tau \mid \mathbf{x} \sim \mathsf{Ga}(\alpha_1, \beta_1)$ with updated hyper-parameters

$$\alpha_1 = \alpha_0 + \frac{n}{2}, \qquad \beta_1 = \beta_0 + \frac{1}{2}\sum(x_i - \mu)^2.$$

There's no point in giving interval estimates for μ (since it's known), but we can use the fact that $2\beta_1\tau \sim \chi^2_\nu$ with $\nu = 2\alpha_1$ degrees of freedom to generate intervals for τ or σ^2. For numbers $0 < \gamma_1 < \gamma_2 < 1$ find quantiles $0 < c_1 < c_2 < \infty$ of the χ^1_ν distribution that satisfy $\mathsf{P}[\chi^2_\nu \le c_j] = \gamma_j$; then

$$\gamma_2 - \gamma_1 = \mathsf{P}\Big[\frac{c_1}{2\beta_1} < \tau < \frac{c_2}{2\beta_1} \quad \Big| \, \mathbf{x}\Big] = \mathsf{P}\Big[\frac{2\beta_1}{c_2} < \sigma^2 < \frac{2\beta_1}{c_1} \quad \Big| \, \mathbf{x}\Big]$$

For $0 < \gamma < 1$ the symmetric case of $\gamma_1 = (1-\gamma)/2$, $\gamma_2 = (1+\gamma)/2$ is not the shortest possible interval of size $\gamma = \gamma_2 - \gamma_1$, because the χ^2 density isn't symmetric. The shortest choice is called the "HPD" or "highest posterior density" interval because the $\mathsf{Ga}(\alpha_1, \beta_1)$ density function takes equal values at c_1, c_2 and is higher inside $[c_1, c_2]$ than outside. Typically HPDs are found by a numerical search.

In the limit as $\alpha_0 \to 0$ and $\beta_0 \to 0$, we have the improper prior $\pi_\tau(\tau) \propto \tau^{-1}$ with posterior $\tau \mid \mathbf{x} \sim \mathsf{Ga}(n/2, \Sigma(x_i - \mu)^2/2)$ proportional to a χ^2_n distribution, and the Bayesian credible interval coincides with a sampling theory confidence interval for σ^2 (again, with the conditioning reversed).

3.3 Both mean μ *and* precision $\tau = \sigma^{-2}$ Unknown

When *both* parameters are uncertain, there is no conjugate family with μ, τ independent under both prior and posterior— but there is a four-parameter family, the "normal-gamma" distributions, that *is* conjugate (see §7.6 in (3/e) or §8.6 in (4/e) of the text). It is usually expressed in conditional form:

$$\tau \sim \mathsf{Ga}(\alpha_0, \beta_0), \qquad \mu \mid \tau \sim \mathsf{No}\big(\mu_0, [\lambda_0\tau]^{-1}\big)$$

with the prior precision for μ proportional to the data precision τ. Its density function is seldom needed, but easy to write down:

$$\pi(\mu, \tau \mid \alpha_0, \beta_0, \mu_0, \lambda_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\sqrt{\frac{\lambda_0}{2\pi}}\;\tau^{\alpha_0 - 1/2}\, e^{-\tau[\beta_0 + \lambda_0(\mu-\mu_0)^2/2]} \qquad (1)$$

The posterior distribution is again of the same form, with updated hyper-parameters that depend on n and the sufficient statistics $\bar{x}_n$ and $\hat{\sigma}_n^2$:

$$\alpha_1 = \alpha_0 + \frac{n}{2} \qquad \beta_1 = \beta_0 + \frac{n}{2}\left[\hat{\sigma}_n^2 + \frac{\lambda_0(\bar{x}_n - \mu_0)^2}{\lambda_0 + n}\right]$$
$$\mu_1 = \frac{\lambda_0\mu_0 + n\bar{x}_n}{\lambda_0 + n} \qquad \lambda_1 = \lambda_0 + n$$

The conventional "non-informative" or "vague" improper prior distribution for a location-scale family like this is $\pi(\mu, \tau) = \tau^{-1}$, invariant under changes in both location $\mathbf{x} \rightsquigarrow \mathbf{x} + a$ and scale $\mathbf{x} \rightsquigarrow c\mathbf{x}$; from Equation (1) we see this can be achieved (apart from the irrelevant normalizing constant) by taking $\alpha_0 = -1/2$ and $\beta_0 = \lambda_0 = 0$, with μ_0 arbitrary. In this limiting case we find posterior distributions of $\tau \sim \mathsf{Ga}(\nu/2, s_n^2/2)$ and $\mu \mid \tau \sim \mathsf{No}(\bar{x}_n, n\tau)$ with $\nu = n - 1$, so

$$\frac{\mu - \bar{x}_n}{\sqrt{\hat{\sigma}_n^2/\nu}} \sim t_\nu$$

and the Bayesian posterior credible interval

$$\gamma = \mathsf{P}\left[\bar{x}_n - t^*\frac{\hat{\sigma}_n}{\sqrt{n-1}} < \mu < \bar{x}_n + t^*\frac{\hat{\sigma}_n}{\sqrt{n-1}} \quad \Big| \; \mathbf{x}\right]$$

coincides with the sampling-theory confidence interval

$$\gamma = \mathsf{P}\left[\bar{x}_n - t^*\frac{\hat{\sigma}_n}{\sqrt{n-1}} < \mu < \bar{x}_n + t^*\frac{\hat{\sigma}_n}{\sqrt{n-1}} \quad \Big| \; \mu, \tau\right]$$

of Section (2.3), with a different interpretation.

More generally, for *any* positive $\alpha_*, \beta_*, \mu_*, \lambda_*$, the marginal distribution for μ is that of a shifted (or "non-central") and scaled t_ν distribution with $\nu = 2\alpha_*$ degrees of freedom; specifically,

$$\frac{\mu - \mu_*}{\sqrt{\beta_*/\alpha_*\lambda_*}} \sim t_\nu, \qquad \nu = 2\alpha_*$$

One way to see that is to begin with the relations

$$\tau \sim \mathsf{Ga}(\alpha_*, \beta_*), \qquad \mu \mid \tau \sim \mathsf{No}\Big(\mu_*, \frac{1}{\lambda_*\tau}\Big)$$

and, after scaling and centering, find

$$Z \equiv (\mu - \mu_*)\sqrt{\lambda_*\tau} \sim \mathsf{No}(0,1) \quad \perp\!\!\!\perp \quad Y \equiv 2\beta_*\tau \sim \mathsf{Ga}(\alpha_*, 1/2) = \chi^2_\nu$$

for $\nu \equiv 2\alpha_*$ and hence

$$\frac{Z}{\sqrt{Y/\nu}} = \frac{\mu - \mu_*}{\sqrt{\beta_*/\alpha_*\lambda_*}} \sim t_\nu.$$

With the normal-gamma prior, and with t^* chosen so that $F_\nu(t^*) = (1+\gamma)/2$, a Bayesian posterior credible interval is:

$$\gamma = \mathsf{P}\left[\mu_1 - t^*\sqrt{\beta_1/\alpha_1\lambda_1} \le \mu \le \mu_1 + t^*\sqrt{\beta_1/\alpha_1\lambda_1} \quad \Big| \; \mathbf{x}\right],$$

an interval that *is* a meaningful probability statement even after $\bar{x}_n$ and $\hat{\sigma}_n^2$ (and hence α_1, β_1, μ_1, λ_1) are replaced with their observed values from the data.

4 Interval Estimates for Poisson Means

Let $\mathbf{x} = \{X_j\}$ be a sample of 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 $\mathcal{X} = \mathbb{N}_0^n$ the space of all possible values of $\mathbf{x}$, n-tuples of non-negative integers; also denote by $\Theta = \mathbb{R}_+$ the possible values of θ.

4.1 Sampling Theory: Confidence Intervals

A γ-*Confidence Interval* is a random interval $I(\mathbf{x})$ with the property that $\mathsf{P}_\theta[\theta \in I(\mathbf{x})] \ge \gamma$ for every fixed value of $\gamma > 0$, *i.e.*, that will contain θ with probability at least γ no matter what θ might be. We can specify an interval by giving its end-points, a pair of functions $A: \mathcal{X} \to \mathbb{R}$ and $B: \mathcal{X} \to \mathbb{R}$ with the property that

$$(\forall \theta \in \Theta) \quad \mathsf{P}_\theta[A(\mathbf{x}) < \theta < B(\mathbf{x})] \ge \gamma. \tag{2}$$

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,

$$\mathsf{P}_\theta[\theta \le A(\mathbf{x})] \le \frac{1-\gamma}{2}, \qquad \mathsf{P}_\theta[B(\mathbf{x}) \le \theta] \le \frac{1-\gamma}{2} \tag{3}$$

- 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 $X_t \geq k$ if and only if $T_k \leq t$ (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

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

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

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

Fix any positive integer k. To achieve Equation (3) for $\theta \leq A_k$, we need:

$$\begin{aligned} \frac{1-\gamma}{2} &\geq \mathsf{P}_\theta[\theta \leq A(\mathbf{x})] \\ &\geq \mathsf{P}_\theta[S \geq k] \\ &= 1 - \texttt{ppois}(\texttt{k} - 1, \texttt{n}\theta) \\ &= \texttt{pgamma}(\texttt{n}\theta, \texttt{k}, 1) = \texttt{pgamma}(\theta, \texttt{k}, \texttt{n}), \quad \text{i.e.} \end{aligned}$$

$$\texttt{qgamma}(\frac{1-\gamma}{2}, \texttt{k}, \texttt{n}) \geq \theta$$

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

$$\texttt{qgamma}(\frac{1-\gamma}{2}, \texttt{k}, \texttt{n}) \geq A_k. \tag{4}$$

Similarly, for nonnegative k and $B_k \leq \theta$, Equation (3) requires:

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

$$\texttt{qgamma}(\frac{1+\gamma}{2}, \texttt{k} + 1, \texttt{n}) \leq \theta$$

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

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

The shortest interval subject to the two constraints of Equations (4, 5) is:

$$A_k = \texttt{qgamma}(\frac{1-\gamma}{2}, \texttt{k}, \texttt{n}) \qquad B_k = \texttt{qgamma}(\frac{1+\gamma}{2}, \texttt{k}+\texttt{1}, \texttt{n}). \tag{6}$$

4.2 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 hyperparameters $\alpha > 0$, $\beta > 0$ 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

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

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

$$\gamma = \mathsf{P}_\pi[a(\mathbf{x}) < \theta < b(\mathbf{x}) \mid \mathbf{x}] \tag{7}$$

where $a(\mathbf{x}) = a_S$ and $b(\mathbf{x}) = b_S$ with

$$a_k = \texttt{qgamma}(\frac{1-\gamma}{2}, \alpha+\texttt{k}, \beta+\texttt{n}) \quad b_k = \texttt{qgamma}(\frac{1+\gamma}{2}, \alpha+\texttt{k}, \beta+\texttt{n}). \tag{8}$$

4.3 Comparison

The two probability statements in Equations (2, 7) are different— in Equation (2) the value of θ is fixed while $\mathbf{x}$ (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 $\mathsf{P}_\theta[A(\mathbf{x}) < \theta < B(\mathbf{x})]$ (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 (7), however, $\mathbf{x}$ (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.

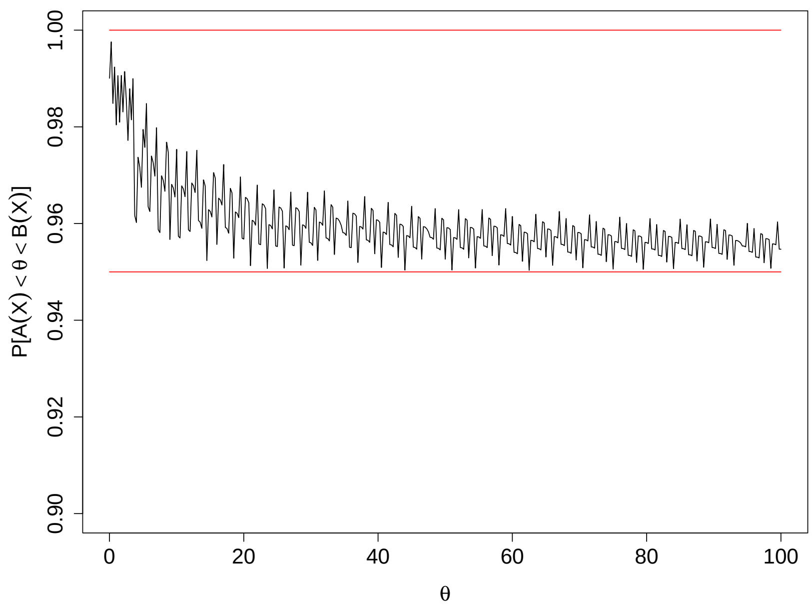

Figure 1: Exact coverage probability for 95% Poisson Confidence Intervals

The formulas for the interval endpoints given in Equations (6, 8) 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 $\pi(\theta \mid \mathbf{x}) \sim \mathsf{Ga}(S + \frac{1}{2}, n)$ is proper for any $\mathbf{x} \in \mathcal{X}$. For any α and β, all the intervals have the same asymptotic behavior for large n; by the central limit theorem,

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

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