# Confidence & Credible Interval Estimates

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

## 1 Introduction

Point estimates of unknown parameters $\theta \in \Theta$ governing the distribution of an observed quantity $X \in \mathcal{X}$ are unsatisfying if they come with no measure of accuracy or precision. One approach to giving such a measure is to offer *set*-valued or, for one-dimensional $\Theta \subset \mathbb{R}$, *interval*-valued estimates for $\theta$, rather than *point* estimates. Upon observing $X = x$, we construct an interval $I(x)$ which is very likely to contain $\theta$, 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 integer-valued random variables with monotone CDFs.

### 1.1 Confidence Intervals and Credible Intervals

A $\gamma$-*Confidence Interval* is a random interval $I(X) \subset \Theta$ with the property that $\mathsf{P}_\theta[\theta \in I(X)] \geq \gamma$, *i.e.*, that it will contain $\theta$ with probability at least $\gamma$ no matter what $\theta$ might be. Such an interval may be specified by giving the end-points, a pair of functions $A : \mathcal{X} \to \mathbb{R}$ and $B : \mathcal{X} \to \mathbb{R}$ with the property that

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

Notice that this probability is for each *fixed* $\theta$; it is the endpoints of the interval $I(X) = (A, B)$ that are random in this calculation, not $\theta$, in the sampling-based Frequentist paradigm.

A $\gamma$-*Credible Interval* is an interval $I \subset \Theta$ with the property that $\mathsf{P}[\theta \in I \mid X] \geq \gamma$, *i.e.*, that the posterior probability that it contains $\theta$ is at least $\gamma$. A family of intervals $\{I(x) : x \in \mathcal{X}\}$ for all possible outcomes $x$ may be specified by giving the end-points as functions $A : \mathcal{X} \to \mathbb{R}$ and $B : \mathcal{X} \to \mathbb{R}$ with the property that

$$\mathsf{P}[A(x) < \theta < B(x) \mid X = x] \geq \gamma. \tag{2}$$

This expression is a *conditional* or *posterior* probability that the random variable $\theta$ will lie in the interval $[A, B]$, *given* the observed value $x$ of the random variable $X$. It is the parameter $\theta$ that is random in the Bayesian paradigm, not the endpoints of the interval $I(x) = (A, B)$.

## 1.2 Pivotal Quantities

Confidence intervals for many parametric distributions can be found using "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 $\alpha$ 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 $\mu$ is unknown but $\sigma^2$ is known, then $(X - \mu) \sim \mathsf{No}(0, \sigma^2)$ is pivotal and, for samples of size $n$, $\sqrt{n}(\bar{X}_n - \mu) \sim \mathsf{No}(0, \sigma^2)$ is pivotal.
- $\mathsf{No}(\mu, \sigma^2)$: If $\mu$ is known but $\sigma^2$ 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 $\mu$ and $\sigma^2$ are *both* unknown then for samples of size $n$, $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathsf{No}(0, 1)$ and $\sum(X_i - \bar{X}_n)^2/\sigma^2 \sim \chi^2_{n-1}$ are both pivotal. This is the key example below.
- $\mathsf{Un}(0, \theta)$: If $X \sim \mathsf{Un}(0, \theta)$ with $\theta$ 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.

## 1.3 Confidence Intervals from Pivotal Quantities

Pivotal quantities allow us to construct sampling-theory "confidence intervals" for uncertain parameters. For the simplest example, let's consider the problem of finding a $\gamma = 90\%$ Confidence Interval for the unknown rate parameter $\lambda$ from a single observation $X \sim \mathsf{Ex}(\lambda)$ from the exponential distribution. Since $\lambda X \sim \mathsf{Ex}(1)$ is pivotal, we have

$$\mathsf{P}_\lambda\big[a \le \lambda X \le b\big] = e^{-a} - e^{-b} = 0.95 - 0.05 = 0.90$$

if $a = -\log(0.95) = 0.513$ and $b = -\log(0.05) = 2.996$, so for every fixed $\lambda > 0$

$$0.90 = \mathsf{P}_\lambda\Big[\frac{0.513}{X} \le \lambda \le \frac{2.996}{X}\Big] \tag{3a}$$

will be a symmetric Confidence Interval for $\lambda$ for a single observation $X$. Note it is the *endpoints* of the interval that are random (through their dependence on $X$) in this sampling-distribution based approach to inference, while the parameter $\lambda$ is fixed. That's why the word "confidence" is used for these intervals, and not "probability." In Section (4) when we consider Bayesian interval estimates this will switch— the *observed* (and so fixed) values of $X$ will be used, while the parameter will be regarded as a random variable.

With a larger sample-size something similar can be done. For example, since $2n\lambda\bar{X}_n \sim \chi^2_{2n}$ is pivotal for iid $\{X_j\} \sim \mathsf{Ex}(\lambda)$, and since the 5% and 95% quantiles of the $\chi^2_{10}$ distribution are 3.940 and 18.307, a sample of size $n = 5$ from the $\mathsf{Ex}(\lambda)$ distribution satisfies

$$0.90 = \mathsf{P}_\lambda[3.940 \le 2n\lambda\bar{X}_n \le 18.307] = \mathsf{P}_\lambda\left[\frac{0.3940}{\bar{X}_5} \le \lambda \le \frac{1.8307}{\bar{X}_5}\right], \tag{3b}$$

so $[0.39/\bar{X}_5,\ \ 1.83/\bar{X}_5]$ is a 90% confidence interval for $\lambda$ based on a sample of size $n = 5$. Intervals (3a) and (3b) are both of the same form, straddling the MLE $\hat{\lambda} = 1/\bar{X}_n$, but (3b) is narrower because its sample-size is larger. This will be a recurring theme— larger samples will lead more precise estimates (*i.e.*, narrower intervals) at the cost of more sampling.

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

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

so the MLEs are

$$\hat{\mu} = \bar{x}_n = \frac{1}{n}\sum x_i \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{n}S \qquad \text{where } S := \sum(x_i - \bar{x}_n)^2. \tag{4}$$

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

Since the $\{X_i\}$ are independent, their sum $\sum X_i$ 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 $S = \sum(x_i - \bar{x}_n)^2$. Now we can use moment generating functions to discover the distribution of $S$. Since $Z^2 \sim \chi^2_1 = \mathsf{Ga}(1/2, 1/2)$ 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 $\mathsf{Ga}(\alpha, \lambda)$ MGF is $(1 - t/\lambda)^{-\alpha}$ for $t < \lambda$, and that the MGF for the sum of independent random variables is the product of the individual MGFs, so

$$\begin{aligned}
\mathsf{E} \exp\left\{ t \sum (x_i - \mu)^2 \right\} &= (1 - 2\sigma^2 t)^{-n/2} \\
&= \mathsf{E} \exp\left\{ t \sum (x_i - \bar{x}_n)^2 \right\} \mathsf{E} \exp\left\{ tn(\bar{x}_n - \mu)^2 \right\} \\
&= \mathsf{E} \exp\left\{ t \sum (x_i - \bar{x}_n)^2 \right\} (1 - 2\sigma^2 t)^{-1/2}, \text{ so} \\
\mathsf{E} \exp\left\{ t \sum (x_i - \bar{x}_n)^2 \right\} &= (1 - 2\sigma^2 t)^{-(n-1)/2} \text{ by dividing. Thus} \\
S := \sum (x_i - \bar{x}_n)^2 &\sim \mathsf{Ga}\left( \frac{n-1}{2}, \frac{1}{2\sigma^2} \right)
\end{aligned}$$

and so $S/\sigma^2 \sim \chi^2_{n-1}$ is independent of $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathsf{No}(0, 1)$, as claimed.

## 2.1 Confidence Intervals for Mean $\mu$ when Variance $\sigma_0^2$ is Known

The pivotal quantity

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

has a standard $\mathsf{No}(0, 1)$ normal distribution, with CDF $\Phi(z)$. If $\sigma^2 = \sigma_0^2$ were known, then for any $0 < \gamma < 1$ and for $z^*$ such that $\Phi(z^*) = (1 + \gamma)/2$ we could write

$$\begin{aligned}
\gamma &= \mathsf{P}_\mu \left[ -z^* \le \frac{\bar{x}_n - \mu}{\sqrt{\sigma_0^2/n}} \le z^* \right] \\
&= \mathsf{P}_\mu \left[ \bar{x}_n - z^* \sigma_0/\sqrt{n} \le \mu \le \bar{x}_n + z^* \sigma_0/\sqrt{n} \right] \qquad (5)
\end{aligned}$$

to find a confidence interval $[\bar{x}_n - z^* \sigma_0/\sqrt{n},\ \bar{x}_n + z^* \sigma_0/\sqrt{n}]$ for $\mu$ by replacing $\bar{x}_n$ with its observed value.

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

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

obtained by replacing "$\sigma^2$" by its estimate "$\hat{\sigma}^2$" in the definition of $Z$ *does* have a probability distribution that does not depend on $\mu$ or $\sigma^2$, and so *is* pivotal, its distribution 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 $\mu$ *or* if the estimate $\hat{\sigma}^2$ for the variance is too small.

Following William S. Gosset (a statistician studying field data from barley cultivation for the Guinness Brewery in Dublin, Ireland) as adapted by Ronald A. Fisher (an English theoretical statistician), we consider the (slightly re-scaled, by Fisher) pivotal quantity:

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

for independent $Z \sim \mathsf{No}(0,1)$ and $Y \sim \chi^2_\nu$ with $\nu = (n-1)$ degrees of freedom. Now we turn to finding the density function for $t$, so we can find confidence intervals for $\mu$ in Section (2.3).

## 2.2 The $t$ pdf

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

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

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 Mean $\mu$ when Variance $\sigma^2$ is Unknown

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

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

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

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

where $s_n^2 = \hat{\sigma}_n^2 n/(n-1)$ is the usual unbiased estimator $s_n^2 := \frac{1}{n-1}\sum(x_i - \bar{x}_n)^2$ of $\sigma^2$. Once again we have a random interval that will contain $\mu$ with specified probability $\gamma$— and can replace the sufficient statistics $\bar{x}_n$, $\hat{\sigma}_n^2$ with their observed values to get a confidence interval. In this sampling theory approach the unknown $\mu$ and $\sigma^2$ are kept fixed and only the data $\mathbf{x}$ are treated as random; that's what the subscript "$\mu$" on $\mathsf{P}$ was intended to suggest.

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

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

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

$$0.95 = \mathsf{P}_\mu[\bar{x} - 12.7\hat{\sigma}_2 \le \mu \le \bar{x} + 12.7\hat{\sigma}_2]$$

with $\bar{x} = (x_1 + x_2)/2$ and $\hat{\sigma}_2 = |x_1 - x_2|/2$.

When $\sigma^2$ 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 Confidence Intervals for a Normal Variance

Estimating variances arises less frequently than estimating means does, but the methods are illuminating and the problem does arise on occasion.

### 3.1 Confidence Intervals for Variance $\sigma^2$, when Mean $\mu_0$ is Known

If $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{No}(\mu_0, \sigma^2)$ then

$$\hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \mu_0)^2 \sim \mathsf{Ga}\big(n/2, n/2\sigma^2\big)$$

has a Gamma distribution with rate proportional to $\sigma^{-2}$, so the re-scaled

$$n\hat{\sigma}^2/\sigma^2 \sim \mathsf{Ga}\big(n/2, 1/2\big) = \chi^2_n$$

is a pivotal quantity. We can construct an exact $100\gamma\%$ Confidence Interval for $\sigma^2$ from

$$\gamma = \mathsf{P}\left[a \le \frac{\sum (X_i - \mu_0)^2}{\sigma^2} \le b\right]$$

for any $0 \le a < b < \infty$ with $\gamma = \texttt{pgamma}(\texttt{b}, \texttt{n}/2, 1/2) - \texttt{pgamma}(\texttt{a}, \texttt{n}/2, 1/2)$ or, equivalently, $\gamma = \texttt{pchisq}(\texttt{b}, \texttt{n}) - \texttt{pchisq}(\texttt{a}, \texttt{n})$. For a symmetric interval, take $a = \texttt{qgamma}(\frac{1-\gamma}{2}, \texttt{n}/2, 1/2)$, $b = \texttt{qgamma}(\frac{1+\gamma}{2}, \texttt{n}/2, 1/2)$, to get

$$\gamma = \mathsf{P}\left[\frac{\sum (X_i - \mu_0)^2}{b} \le \sigma^2 \le \frac{\sum (X_i - \mu_0)^2}{a}\right],$$

a symmetric $100\gamma\%$ confidence interval for $\sigma^2$, with known mean $\mu_0$.

### 3.2 Confidence Intervals for Variance $\sigma^2$, when Mean $\mu$ is Unknown

In the more realistic case that $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{No}(\mu, \sigma^2)$ with both $\mu$ and $\sigma^2$ unknown, we find that

$$\hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \bar{X}_n)^2 \sim \mathsf{Ga}\big(\frac{n-1}{2}, \frac{n}{2\sigma^2}\big)$$

again has a Gamma distribution with rate proportional to $\sigma^{-2}$, so again the re-scaled

$$n\hat{\sigma}^2/\sigma^2 \sim \mathsf{Ga}\big(\nu/2, 1/2\big) = \chi^2_\nu$$

is a pivotal quantity with a $\chi^2_\nu$ distribution, now with one fewer degrees of freedom $\nu = (n-1)$. This again leads to an exact $100\gamma\%$ Confidence Interval for $\sigma^2$ from

$$\gamma = \mathsf{P}\left[a \le \frac{\sum (X_i - \bar{X}_n)^2}{\sigma^2} \le b\right]$$

for any $0 \le a < b < \infty$ with $\gamma = \texttt{pgamma}(\texttt{b}, \nu/2, 1/2) - \texttt{pgamma}(\texttt{a}, \nu/2, 1/2)$ or, equivalently, $\gamma = \texttt{pchisq}(\texttt{b}, \nu) - \texttt{pchisq}(\texttt{a}, \nu)$ for $\nu = (n-1)$. For a symmetric interval, take

$$\begin{aligned} a &= \texttt{qgamma}(\tfrac{1-\gamma}{2}, \tfrac{\nu}{2}, 1/2) & b &= \texttt{qgamma}(\tfrac{1+\gamma}{2}, \tfrac{\nu}{2}, 1/2) \\ &= \texttt{qchisq}(\tfrac{1-\gamma}{2}, \nu) & &= \texttt{qchisq}(\tfrac{1+\gamma}{2}, \nu) \end{aligned}$$

to get the $100\gamma\%$ confidence interval for $\sigma^2$, with unknown mean $\mu$:

$$\gamma = \mathsf{P}\left[\frac{\sum (X_i - \bar{X}_n)^2}{b} \le \sigma^2 \le \frac{\sum (X_i - \bar{X}_n)^2}{a}\right].$$

# 4 Bayesian Credible Intervals for a Normal Mean

How would we make inference about $\mu$ for the normal distribution using Bayesian methods?

## 4.1 Unknown mean $\mu$, known precision $\tau := \sigma^{-2}$

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

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

is proportional to a normal density for the parameter $\mu$ with mean $\bar{x}_n$ and precision $n\tau$. With the improper uniform Jeffreys' Rule or Reference prior for this problem, $\pi_J(\mu) \propto 1$, the posterior distribution would be

$$\pi_J(\mu \mid \mathbf{x}) \sim \mathsf{No}\big(\bar{x}_n,\ (n\tau)^{-1}\big)$$

leading to posterior $100\gamma\%$ intervals of the form

$$\gamma = \mathsf{P}_J\Big[\,\bar{x}_n - z^*/\sqrt{n\tau} < \mu < \bar{x}_n + z^*/\sqrt{n\tau} \quad \Big|\ \mathbf{x}\,\Big]$$

where $z^* = \texttt{qnorm}(\frac{1+\gamma}{2})$ is the normal quantile such that $\Phi(z^*) = \frac{1+\gamma}{2}$, identical to the frequentist Confidence Interval of (5)

More generally, for any $\mu_0 \in \mathbb{R}$ and any $\tau_0 \geq 0$, the prior distribution $\pi_0(\mu) \sim \mathsf{No}(\mu_0, \tau_0^{-1})$ is conjugate in the sense that the posterior distribution is again normal, in this case $\mu \mid \mathbf{x} \sim \mathsf{No}(\mu_1, \tau_1^{-1})$ with updated "hyper-parameters"

$$\mu_1 = \frac{\tau_0\mu_0 + n\tau\bar{x}_n}{\tau_0 + n\tau}, \qquad \tau_1 = \tau_0 + n\tau. \tag{6}$$

The posterior precision is the sum of the prior precision $\tau_0$ and the data precision $n\tau$, while the posterior mean is the precision-weighted average of the prior mean $\mu_0$ and the data mean $\bar{x}_n$. The Reference example above was the special case of $\tau_0 = 0$ and arbitrary $\mu_0$, leading to $\mu_1 = \bar{x}_n$ and $\tau_1 = n\tau$.

Posterior credible intervals are available of any size $0 < \gamma < 1$. Using the quantile $z^*$ such that $\Phi(z^*) = (1+\gamma)/2$, from (6) 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 $\pi_0(\mu) \propto 1$ 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 $\mu$ 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 $\mu$).

### 4.2 Known mean $\mu$, unknown precision $\tau = \sigma^{-2}$

When $\mu$ is known and only the precision $\tau$ is uncertain,

$$f(\mathbf{x} \mid \tau) \propto \tau^{n/2}\ e^{-\tau \sum (x_i-\mu)^2/2}$$

is proportional to a gamma density in $\tau$, so a gamma prior $\pi_\tau(\tau) \sim \mathsf{Ga}(\alpha_0, \lambda_0)$ is conjugate for $\tau$. The posterior distribution is $\tau \mid \mathbf{x} \sim \mathsf{Ga}(\alpha_1, \lambda_1)$ with updated hyper-parameters

$$\alpha_1 = \alpha_0 + \frac{n}{2}, \qquad \lambda_1 = \lambda_0 + \frac{1}{2}\sum (x_i - \mu)^2.$$

There's no point in giving interval estimates for $\mu$ (since it's known), but we can use the fact that $2\lambda_1\tau \sim \chi^2_\nu$ with $\nu = 2\alpha_1$ degrees of freedom to generate interval estimates for $\tau$ or $\sigma^2$. For numbers $0 < \gamma_1 < \gamma_2 < 1$ find quantiles $0 < c_1 < c_2 < \infty$ of the $\chi^1_\nu$ distribution that satisfy $\mathsf{P}[\chi^2_\nu \le c_j] = \gamma_j$; then

$$\gamma_2 - \gamma_1 = \mathsf{P}\Big[\ \frac{c_1}{2\lambda_1} < \tau < \frac{c_2}{2\lambda_1}\quad \Big|\ \mathbf{x}\ \Big] = \mathsf{P}\Big[\ \frac{2\lambda_1}{c_2} < \sigma^2 < \frac{2\lambda_1}{c_1}\quad \Big|\ \mathbf{x}\ \Big]$$

For $0 < \gamma < 1$ the symmetric case of $\gamma_1 = \frac{1-\gamma}{2}$, $\gamma_2 = \frac{1+\gamma}{2}$ is not the shortest possible interval of size $\gamma = \gamma_2 - \gamma_1$, because the $\chi^2$ density isn't symmetric. The shortest choice is called the "HPD" or "highest posterior density" interval because the $\mathsf{Ga}(\alpha_1, \lambda_1)$ 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 $\pi_\tau(\tau) \propto \tau^{-1}$ for the precision, with posterior $\tau \mid \mathbf{x} \sim \mathsf{Ga}(n/2, \Sigma(x_i - \mu)^2/2)$ proportional to a $\chi^2_n$ distribution, and the Bayesian credible interval coincides with a sampling theory confidence interval for $\sigma^2$ (again, with the conditioning reversed).

### 4.3 Both mean $\mu$ *and* precision $\tau = \sigma^{-2}$ Unknown

When *both* parameters are uncertain, there is no conjugate family with $\mu, \tau$ independent under both prior and posterior— but there is a four-parameter family, the "normal-gamma" distributions, that *is* conjugate. It is usually expressed in conditional form:

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

for parameter vector $\theta_0 = (\alpha_0, \beta_0, \mu_0, \lambda_0)$, with prior precision $\lambda_0\tau$ for $\mu$ proportional to the data precision $\tau$. Its density function on $\mathbb{R} \times \mathbb{R}_+$ is seldom needed, but is easily found:

$$\pi(\mu, \tau \mid \alpha_0, \beta_0, \mu_0, \lambda_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\sqrt{\frac{\lambda_0}{2\pi}}\ \tau^{\alpha_0 - 1/2}\ e^{-\tau[\beta_0 + \lambda_0(\mu-\mu_0)^2/2]} \tag{7}$$

The posterior distribution is again of the same form, with updated hyper-parameters that depend on the sample size $n$ and the sufficient statistics $\bar{x}_n$ and $\hat{\sigma}^2_n$ (see (4)):

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

$$\mu_1 = \mu_0 + \frac{n(\bar{x}_n - \mu_0)}{\lambda_0 + n} \qquad \lambda_1 = \lambda_0 + n$$

It will be proper so long as $\alpha_0 > -1$, $\beta_0 \geq 0$, and $\lambda_0 > -1$. 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 Eqn (7) 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 $\mu_0$ arbitrary. In this limiting case we find posterior distributions of $\tau \sim \mathsf{Ga}(\nu/2, S/2)$ 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* $\theta_* = (\alpha_*, \beta_*, \mu_*, \lambda_*)$, the marginal distribution for $\tau$ is $\mathsf{Ga}(\alpha_*, \beta_*)$ and the marginal distribution for $\mu$ is that of a shifted (or "non-central") and scaled $t_\nu$ 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

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

for $\nu := 2\alpha_*$ 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 $\mathsf{P}[t_\nu \leq t^*] = \frac{1+\gamma}{2}$, a Bayesian posterior credible interval for the mean $\mu$ is:

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

an interval that *is* a meaningful probability statement even after $\bar{x}_n$ and $\hat{\sigma}_n^2$ (and hence $\alpha_1$, $\beta_1$, $\mu_1$, $\lambda_1$) are replaced with their observed values from the data.

# 5 Confidence Intervals for Distributions with Monotone CDFs

Many statistical models feature a one-dimensional sufficient statistic $T$ whose CDF

$$F_\theta(t) := \mathsf{P}_\theta[T(\mathbf{x}) \le t]$$

is a monotonic function of a one-dimensional parameter $\theta \in \Theta \subset \mathbb{R}$, for each fixed value of $t \in \mathbb{R}$. For these we can find Confidence Intervals $[A(\mathbf{x}), B(\mathbf{x})]$ with endpoints of the form $A(\mathbf{x}) = a(T(\mathbf{x}))$ and $B(\mathbf{x}) = b(T(\mathbf{x}))$ as follows.

## 5.1 CIs for Continuous Distributions with Monotone CDFs

If $T(X)$ has a continuous probability distribution with a CDF $F_\theta(t)$ that is monotone *decreasing* in $\theta$ for each $t$, so larger values of $\theta$ are associated with larger values of $T(\mathbf{x})$, then set

$$a(t) := \sup\left\{\theta:\ F_\theta(t) \ge \tfrac{1+\gamma}{2}\right\} \qquad b(t) := \inf\left\{\theta:\ F_\theta(t) \le \tfrac{1-\gamma}{2}\right\} \tag{8a}$$

Conversely, if $F_\theta(t)$ that is monotone *increasing* in $\theta$ for each $t$, set

$$a(t) := \sup\left\{\theta:\ F_\theta(t) \le \tfrac{1-\gamma}{2}\right\} \qquad b(t) := \inf\left\{\theta:\ F_\theta(t) \ge \tfrac{1+\gamma}{2}\right\} \tag{8b}$$

and, in either case,

$$A(\mathbf{x}) := a\big(T(\mathbf{x})\big) \qquad B(\mathbf{x}) := b\big(T(\mathbf{x})\big).$$

For most continuous distributions these infima and suprema $a(t)$, $b(t)$ will be the unique values of $\theta$ such that $F_\theta(t) = \frac{1\pm\gamma}{2}$. Now the interval $[A(\mathbf{x}), B(\mathbf{x})]$ will be an exact $100\gamma\%$ symmetric confidence interval, *i.e.*, will satisfy

$$(\forall\theta \in \Theta) \quad \mathsf{P}_\theta\left[A(\mathbf{x}) < \theta < B(\mathbf{x})\right] = \gamma.$$

For example, if $X \sim \mathsf{Ex}(\theta)$ has the exponential distribution, then $T(X) := X$ has CDF $F_\theta(t) = \mathsf{P}_\theta(X \le t) = [1 - \exp(-\theta t)]_+$, a monotone increasing function of $\theta > 0$ for any fixed $t > 0$. The functions in (8) above are $a(t) = -\log(\frac{1+\gamma}{2})/t$ and $b(t) = -\log(\frac{1-\gamma}{2})/t$, leading for $\gamma = 0.90$ to the interval $[A(X) = 0.0513/X,\ B(X) = 2.996/X]$.

If $\{X_1 \dots X_n\} \overset{\text{iid}}{\sim} \mathsf{Un}(0,\theta)$ then $T(X) := \max\{X_i\}$ has CDF $F_\theta(t) = (t/\theta)_+^n \wedge 1$, decreasing in $\theta > t$ for any $t > 0$, so $a(t) = \left(\frac{1+\gamma}{2}\right)^{-1/n}$ and $b(t) = \left(\frac{1-\gamma}{2}\right)^{-1/n}$. For $\gamma = 0.90$ and $n = 4$ this leads to the interval $[A(X) = 1.013/T,\ B(X) = 2.115/T]$ for $\theta$.

## 5.2 CIs for Discrete Distributions with Monotone CDFs

For discrete distributions it will be impossible to attain precisely probability $\gamma$ except possibly for a discrete set of $\{\gamma_i\} \subset (0,1)$, but we can find conservative intervals whose probability of containing $\theta$ is *at least* $\gamma$ for any $\gamma \in (0,1)$.

Let $T$ be an integer-valued sufficient statistic with CDF $F_\theta(t)$ that decreases monotonically in $\theta$ for each fixed $t$. For any sequence $\cdots < c_0 < c_1 < c_2 < \cdots$ spanning $\Theta$ and any integer $k \in \mathbb{Z}$, then

$$c_k < \theta < c_{k+1} \quad \Longrightarrow \quad \mathsf{P}_\theta\left[c_{T(\mathbf{x})} < \theta\right] = \mathsf{P}_\theta\left[c_{T(\mathbf{x})} \le \theta\right] = \mathsf{P}_\theta\left[T(\mathbf{x}) \le k\right] = F_\theta(k)$$

To achieve the upper bound $\mathsf{P}_\theta\left[\theta < a\big(T(\mathbf{x})\big)\right] \le \frac{1-\gamma}{2}$ for the left endpoint of the interval, *i.e.*, $\mathsf{P}_\theta\left[a\big(T(\mathbf{x})\big) \le \theta\right] \ge \frac{1+\gamma}{2}$, evidently entails $F_\theta(k) \ge \frac{1+\gamma}{2}$ for each $k$ and each $\theta \in \big(a(k),\ a(k+1)\big)$. Since $F_\theta(t)$ is decreasing in $\theta$, the inequality will hold for all $\theta$ in the interval if and only if $F_\theta(k) \ge \frac{1+\gamma}{2}$ for $\theta = a(k+1)$. Thus the largest permissible left endpoint is $A(\mathbf{x}) = a\big(T(\mathbf{x})\big)$ for

$$a(t) := \sup\left\{\theta :\ \mathsf{P}_\theta[T(\mathbf{x}) \le (t-1)] \ge \tfrac{1+\gamma}{2}\right\} \tag{9a}$$

Similarly the requirement for the right endpoint that $\mathsf{P}_\theta\left[b\big(T(\mathbf{x})\big) \le \theta\right] \le \frac{1-\gamma}{2}$ entails $F_\theta(k) \le \frac{1-\gamma}{2}$ for each $k$ and each $\theta \in \big(b(k),\ b(k+1)\big)$ or, by monotonicity, simply that $F_\theta(k) \le \frac{1+\gamma}{2}$ for $\theta = b(k)$. The smallest permissible right endpoint is $B(\mathbf{x}) = b\big(T(\mathbf{x})\big)$ for

$$b(t) := \inf\left\{\theta :\ \mathsf{P}_\theta[T(\mathbf{x}) \le t] \le \tfrac{1-\gamma}{2}\right\} \tag{9b}$$

With these choices the interval with endpoints $A(\mathbf{x}) := a\big(T(\mathbf{x})\big)$ and $B(\mathbf{x}) := b\big(T(\mathbf{x})\big)$ will satisfy

$$(\forall\theta \in \Theta) \quad \mathsf{P}_\theta\left[A(\mathbf{x}) < \theta < B(\mathbf{x})\right] \ge \gamma. \tag{10}$$

For example, the statistic $T(\mathbf{x}) = \sum X_i$ is sufficient for iid Bernoulli random variables $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{Bi}(1,\theta)$ with a CDF $F_\theta(t) = \texttt{pbinom(t,n,theta)}$ that decreases in $\theta$ for each fixed $t$. For $T = 7$ successes in a sample of size $n = 10$, an exact 90% confidence interval for $\theta$ would be $[A, B]$, where

$$\begin{aligned} a(7) &= \sup\{\theta : \texttt{pbinom(6,10,theta)} \ge \texttt{0.95}\} = 0.3933 \\ b(7) &= \inf\{\theta : \texttt{pbinom(7,10,theta)} \le \texttt{0.05}\} \ = 0.9128 \end{aligned}$$

found using the `R` code

```
th <- seq(0,1,,10001);
a  <- max(th[pbinom(6,10,th) >= 0.95]);
b  <- min(th[pbinom(7,10,th) <= 0.05]);
```

Or, using the identity `pbinom(x,n,p) = 1-pbeta(p,x+1,n-x)` relating binomial and beta CDFs, the simpler and more precise

$$a(x) \leftarrow \texttt{qbeta}(\tfrac{1-\gamma}{2}, \texttt{x}, \texttt{n}-\texttt{x}+\texttt{1}) \qquad b(x) \leftarrow \texttt{qbeta}(\tfrac{1+\gamma}{2}, \texttt{x}+\texttt{1}, \texttt{n}-\texttt{x})$$

If the CDF $F_\theta(t)$ of an integer-valued statistic $T(\mathbf{x})$ is monotonically *increasing* in $\theta$ for each fixed $t$, then applying (9) to the statistic $-T(\mathbf{x})$ whose CDF decreases in $\theta$ leads to an interval of the form $A(\mathbf{x}) := a\big(T(\mathbf{x})\big)$ and $B(\mathbf{x}) := b\big(T(\mathbf{x})\big)$ where

$$a(t) := \sup\left\{\theta :\ \mathsf{P}_\theta[T(\mathbf{x}) \le t] \le \tfrac{1-\gamma}{2}\right\} \tag{11a}$$

$$b(t) := \inf\left\{\theta :\ \mathsf{P}_\theta[T(\mathbf{x}) \le t-1] \ge \tfrac{1+\gamma}{2}\right\} \tag{11b}$$

For example, for any fixed $x \in \mathbb{Z}_+$ the CDF $F_\theta(x) = 1-(1-\theta)^{x+1}$ for a single observation $X \sim \mathsf{Ge}(\theta)$ from the geometric distribution increases monotonically from $F_0(x) = 0$ to $F_1(x) = 1$ as $\theta$ increases from zero to one, so a symmetric 90% interval for $\theta$ from the single observation $X = 10$ would be

$$\big[A(10) = 1-(0.95)^{1/11} = 0.0047, \qquad B(10) = 1-(0.05)^{1/10} = 0.2589\big]$$

### 5.3 Confidence Intervals for the Poisson Distribution

Let $\mathbf{x} = \{X_1, \cdots, X_n\} \overset{\text{iid}}{\sim} \mathsf{Po}(\theta)$ be a sample of size $n$ from the Poisson distribution. The natural sufficient statistic $S(\mathbf{x}) = \sum X_i$ has a $\mathsf{Po}(n\theta)$ distribution whose CDF $F_\theta(s)$ decreases with $\theta$ for each fixed $s \in \mathbb{Z}_+$, so by (9) a $100\gamma\%$ Confidence Interval would be $[A(\mathbf{x}),\ B(\mathbf{x})]$ with $A(\mathbf{x}) = a(S)$ and $B(\mathbf{x}) = b(S)$ given by $a(s) = \sup\left\{\theta:\ \mathtt{ppois(s-1,n*}\theta\mathtt{)} \ge \frac{1+\gamma}{2}\right\}$ and $b(s) = \inf\left\{\theta:\ \mathtt{ppois(s,n*}\theta\mathtt{)} \le \frac{1-\gamma}{2}\right\}$.

These bounds can be found without a numerical search, by exploiting the relation between the Poisson and Gamma distributions. Recall 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 the identity

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

Using this, we can express the Poisson confidence interval limits as

$$\begin{aligned}
a(s) &= \sup\left\{\theta:\ \mathtt{ppois(s-1,n*}\theta\mathtt{)} \ge \tfrac{1+\gamma}{2}\right\} & b(s) &= \inf\left\{\theta:\ \mathtt{ppois(s,n*}\theta\mathtt{)} \le \tfrac{1-\gamma}{2}\right\}\\
&= \sup\left\{\theta:\ \mathtt{pgamma(n*}\theta\mathtt{,s)} \le \tfrac{1-\gamma}{2}\right\} & &= \inf\left\{\theta:\ \mathtt{pgamma(n*}\theta\mathtt{,s+1)} \ge \tfrac{1+\gamma}{2}\right\}\\
&= \mathtt{qgamma(}\tfrac{1-\gamma}{2}\mathtt{,s)/n} & &= \mathtt{qgamma(}\tfrac{1+\gamma}{2}\mathtt{,s+1)/n}\\
&= \mathtt{qgamma(}\tfrac{1-\gamma}{2}\mathtt{,s,n)} & &= \mathtt{qgamma(}\tfrac{1+\gamma}{2}\mathtt{,s+1,n)} \qquad (12)
\end{aligned}$$

Here we have used the Gamma quantile function `qgamma()` in R, an inverse for the CDF function `pgamma()`, and its optional rate parameter. Since a re-scaled random variable $Y \sim \mathsf{Ga}(\alpha,\lambda)$ satisfies $\lambda Y \sim \mathsf{Ga}(\alpha, 1)$ with unit rate (the default for `pgamma()` and `qgamma()`), these are related by

$$\mathtt{p = pgamma(}\lambda\,\mathtt{q,}\alpha\mathtt{) = pgamma(q,}\alpha,\lambda\mathtt{)} \quad\Leftrightarrow\quad \mathtt{q = qgamma(p,}\alpha\mathtt{)}/\lambda = \mathtt{qgamma(p,}\alpha,\lambda\mathtt{)}.$$

### 5.4 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 Poisson likelihood function

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

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 $\gamma$ posterior ("credible") interval for $\theta$ can be given by

$$\mathsf{P}_\pi[A(\mathbf{x}) < \theta < B(\mathbf{x}) \mid \mathbf{x}] = \gamma \tag{13}$$

where $A(\mathbf{x}) = a(S)$ and $B(\mathbf{x}) = b(S)$ with

$$a(s) := \texttt{qgamma}(\tfrac{1-\gamma}{2}, \alpha{+}\texttt{s}, \beta{+}\texttt{n}) \quad b(s) := \texttt{qgamma}(\tfrac{1+\gamma}{2}, \alpha{+}\texttt{s}, \beta{+}\texttt{n}). \tag{14}$$

## 5.5 Comparison

The two probability statements in Eqns (10, 13) look similar but they mean different things— in Eqn (10) the value of $\theta$ is fixed while $\mathbf{x}$ (and hence the sufficient statistic $S$) is random, while Eqn (13) expresses a conditional probability given $\mathbf{x}$ and hence $A(\mathbf{x})$ and $B(\mathbf{x})$ are fixed, while $\theta$ is random. Because $S$ has a discrete distribution it is not possible to achieve exact equality for all $\theta$; the coverage probability $\mathsf{P}_\theta[A(\mathbf{x}) < \theta < B(\mathbf{x})]$ as a function of $\theta$ jumps at each of the points $\{A_s, B_s\}$ (see Figure (1)). Instead we guarantee a minimum probability of $\gamma$ ($\gamma = 0.95$ in Figure (1)) that $\theta$ will be captured by the interval. In Eqn (2), however, $\mathbf{x}$ (and hence $S$) are fixed, and we consider $\theta$ to be random; it has a continuous distribution, and it is possible to achieve exact equality.

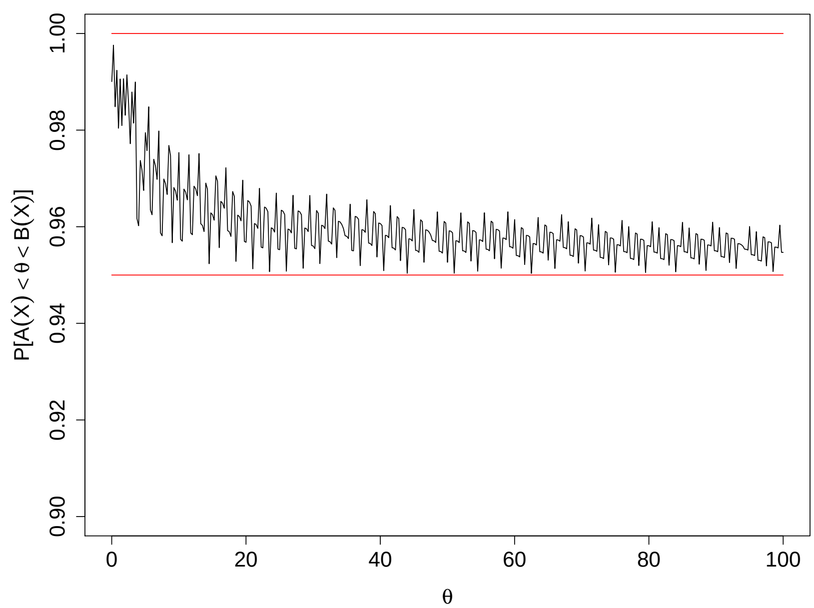


Figure 1: Exact coverage probability for 95% Poisson Confidence Intervals

The formulas for the interval endpoints given in Eqns (12, 14) are similar. If we take $\beta = 0$ and $\alpha = \frac{1}{2}$ they will be as close as possible to each other. This is the objective "Jeffreys' Rule" or "Reference" prior distribution for the Poisson, the $\mathsf{Ga}(\frac{1}{2}, 0)$ distribution with density

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

It is "improper" in the sense that $\int_\Theta \pi(\theta)\, d\theta = \infty$, 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 $\alpha \geq 0$ and $\beta \geq 0$, all the intervals have the same asymptotic behavior for large $n$: by the central limit theorem, in both cases

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

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

### 5.6 One-sided Intervals

For rare events one is often interested in *one*-sided confidence intervals of the form

$$(\forall \theta \in \Theta) \quad \mathsf{P}_\theta[0 \le \theta \le B(\mathbf{x})] \ge \gamma$$

or one-sided credible intervals

$$\mathsf{P}[0 \le \theta \le B(\mathbf{x}) \mid \mathbf{x}] \ge \gamma.$$

For example, if *zero* events have been observed in $n$ independent tries, how large might $\theta$ plausibly be? The solutions follow from the same reasoning that led to the symmetric two-sided intervals of Eqns (12, 14): $B(\mathbf{x}) = B_S$ and $B(\mathbf{x}) = b_S$ for $S := \sum_{j=1}^n X_j$, with

$$B_k = \mathtt{qgamma}(\gamma, \mathtt{k+1}, \mathtt{n}) \quad b_k = \mathtt{qgamma}(\gamma, \alpha{+}\mathtt{k}, \beta{+}\mathtt{n}). \tag{12',14'}$$

For example, the Reference Bayesian $\gamma = 90\%$ one-sided interval for $\theta$ upon observing $k = 0$ events in $n = 10$ tries would be $[0,\ 0.1353]$ with upper limit $b_0 = \mathtt{qgamma}(0.90, 0.50, 10)$, tighter and probably more useful than the two-sided interval $[0.0001966,\ 0.192073]$.

### 5.7 Poisson HPD Intervals

Sometimes interest focuses on the *shortest* interval $[a(\mathbf{x}),\ b(\mathbf{x})]$ with the posterior coverage probability $\mathsf{P}_\pi[\theta \in [a(\mathbf{x}),\ b(\mathbf{x})] \mid X] \ge \gamma$. In general there is no closed-form expression, but the solution can be found by setting $a_k = a(\xi^\star)$ and $b_k = b(\xi^\star)$ for the number

$$\xi^\star := \operatorname*{argmin}_{0 \le \xi \le 1-\gamma} [b(\xi) - a(\xi)]$$

for the functions

$$a(\xi) := \mathtt{qgamma}(\xi, \alpha{+}\mathtt{S}, \beta{+}\mathtt{n}) \quad b(\xi) := \mathtt{qgamma}(\xi + \gamma, \alpha{+}\mathtt{S}, \beta{+}\mathtt{n}),$$

which can be found with a one-dimensional numerical search.