4. Gaussian models

Box

Tiao

Objective

Introduce Bayesian inference for one and two sample problems with normally distributed data, illustrating the differences and similarities between Bayesian solutions and classical solutions.

Recommended reading

- Box, G.E. and Tiao, G.C. (1992). *Bayesian inference in statistical analysis*, Chapter 2.

- Lee, P.M. (2004). *Bayesian Statistics: An Introduction*, Chapter 2.

- Wiper, M.P., Girón, F.J. and Pewsey, A. (2008). Objective Bayesian inference for the half-normal and half-t distributions. *Communications in Statistics: Theory and Methods*, **37**, 3165–3185.

  http://docubib.uc3m.es/WORKINGPAPERS/WS/ws054709.pdf

Introduction: one sample inference problems

Initially, we shall consider the problem of estimating the mean, μ, and/or variance, σ^2, of a single normal population, $X|\mu,\sigma^2 \sim \mathcal{N}\left(\mu,\sigma^2\right)$, with density

$$f\left(x|\mu,\sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

from a Bayesian perspective.

We shall consider three different scenarios:

a) inference for μ when σ^2 is known,

b) inference for σ^2 when μ is known,

c) joint inference for μ,σ^2.

and will assume throughout that we observe a sample, $\mathbf{x}$, of size n.

Inference for μ when σ is known

In order to develop a conjugate prior, we must first represent the normal distribution in exponential family form. We have:

$$f(x|\mu) = \underbrace{\exp\left(-\frac{\mu^2}{2\sigma^2}\right)}_{C(\mu)} \underbrace{\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)}_{h(x)} \exp\left(\underbrace{\frac{\mu}{\sigma^2}}_{\phi(\mu)} \underbrace{x}_{s(x)}\right)$$

which is a one-parameter exponential family representation as in Definition 3.

A conjugate prior distribution

Given this exponential family representation, from Theorem 9, we can derive the form of a conjugate prior distribution as

$$\begin{aligned} p(\mu) &\propto C(\mu)^a \exp(\phi(\mu) b) \\ &\propto \exp\left(-\frac{\mu^2}{2\sigma^2}\right)^a \exp\left(\frac{\mu}{\sigma^2} b\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\left[a\mu^2 - 2b\mu\right]\right) \\ &\propto \exp\left(-\frac{a}{2\sigma^2}\left[\mu - \frac{b}{a}\right]^2\right) \end{aligned}$$

which is also a normal distribution, $\mu \sim \mathcal{N}\left(m, \tau^2\right)$, where $m = \frac{b}{a}$ and $\tau^2 = \frac{\sigma^2}{a}$.

Theorem 16

Let $X|\mu \sim \mathcal{N}\left(\mu, \sigma^2\right)$ with prior distribution $\mu \sim \mathcal{N}\left(m, \tau^2\right)$. Then the predictive distribution of X is

$$X \sim \mathcal{N}\left(m, \sigma^2 + \tau^2\right).$$

Given a sample $\mathbf{x} = (x_1, \ldots, x_n)$, the posterior distribution is $\mu|\mathbf{x} \sim \mathcal{N}\left(m^\star, \tau^{\star 2}\right)$ where

$$\begin{aligned} m^\star &= \frac{\frac{1}{\tau^2} m + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \\ \tau^{\star 2} &= \left(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\right)^{-1}. \end{aligned}$$

Derivation of the predictive distribution

Proof We can write $X = \mu + \epsilon$ where $\mu \sim \mathcal{N}\left(m, \tau^2\right)$ and $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$ are independent. Thus, the marginal distribution of X is

$$X \sim \mathcal{N}\left(m + 0, \sigma^2 + \tau^2\right)$$

and the result follows. ∎

We can derive the posterior distribution either directly via Bayes theorem, or alternatively via Theorem 9. In order to apply Bayes theorem, it is first useful to give a general expression for the normal likelihood function.

The normal likelihood function

Theorem 17

If a sample, $\mathbf{x}$, of size n is taken from the normal distribution $X|\mu,\sigma^2 \sim \mathcal{N}\left(\mu,\sigma^2\right)$, then the likelihood function is

$$l(\mu,\sigma|\mathbf{x}) = (2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\mu-\bar{x})^2\right]\right)$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$.

Proof

$$
\begin{aligned}
l(\mu, \sigma | \mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right) \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \mu)^2\right) \\
&= (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\mu - \bar{x})^2\right]\right)
\end{aligned}
$$

and the result follows. ∎

Now the posterior distribution can be derived from Bayes theorem in the usual way.

Proof

$$
\begin{aligned}
p(\mu|\mathbf{x}) &\propto p(\mu)l(\mu|\mathbf{x}) \\
&\propto \exp\left(-\frac{1}{2\tau^2}(\mu-m)^2\right)\sigma^{-n}\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2+n(\mu-\bar{x})^2\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\left(\frac{\mu-m}{\tau}\right)^2+n\left(\frac{\mu-\bar{x}}{\sigma}\right)^2\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)\mu^2-2\left(\frac{m}{\tau^2}+\frac{n\bar{x}}{\sigma^2}\right)\mu\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)\left[\mu-\frac{\frac{m}{\tau^2}+\frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\right]^2\right)
\end{aligned}
$$

which is a normal density; $\mu|\mathbf{x} \sim \mathcal{N}\left(\frac{\frac{m}{\tau^2}+\frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}, \left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)^{-1}\right)$. ∎

Derivation via Theorem 9

Proof In deriving the prior distribution, we found that $m = b/a$ and $\tau^2 = \sigma^2/a$ where $f(\mu) \propto C(\mu)^a \exp(\phi(\mu)b)$ is the conjugate representation of the prior in Theorem 9. Therefore, $a = \sigma^2/\tau^2$ and $b = m\sigma^2/\tau^2$. Also, we have $s(x) = x$ in our exponential family representation, which implies that, a posteriori,

$$\begin{aligned} a^\star &= a + n = \frac{\sigma^2}{\tau^2} + n = \sigma^2 \left(\frac{1}{\tau^2} + \frac{n}{\sigma^2} \right) \\ b^\star &= b + \sum_{i=1}^{n} s(x_i) = m\frac{\sigma^2}{\tau^2} + n\bar{x} = \sigma^2 \left(\frac{1}{\tau^2} m + \frac{n}{\sigma^2}\bar{x} \right), \end{aligned}$$

so that the posterior distribution is given by $\mu|\mathbf{x} \sim \mathcal{N}\left(m^\star, \tau^{\star 2}\right)$ where

$$m^\star = \frac{b^\star}{a^\star} = \frac{\frac{m}{\tau^2} + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \qquad \tau^{\star 2} = \frac{\sigma^2}{a^\star} = \left(\frac{1}{\tau^2} + \frac{n}{\sigma^2} \right)^{-1}.$$

∎

How to remember the updating formulae

There is a simple way of remembering the normal updating formulae. Note firstly that the maximum likelihood estimator for μ is

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Now, the posterior *precision* is given by $\frac{1}{\tau^{\star 2}} = \frac{1}{\tau^2} + \frac{n}{\sigma^2}$ where $\frac{1}{\tau^2}$ is the prior precision and $\frac{n}{\sigma^2}$ is the precision of the estimator $\bar{X}$ so,

$$\textit{posterior precision} = \textit{prior precision} + \textit{precision of MLE.}$$

Also, the posterior mean is simply a weighted average, $m^\star = wm + (1-w)\bar{x}$, where $w = \frac{1/\tau^2}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}$ is proportional to the prior precision.

$$\textit{posterior mean} = \frac{\textit{prior precision} \times \textit{prior mean} + \textit{precision of MLE} \times \textit{MLE}}{\textit{prior precision} + \textit{precision of MLE}}.$$

Interpretation

Write $\tau^2 = \sigma^2/\alpha$ so that the prior distribution is $\mu \sim \mathcal{N}\left(m, \frac{\sigma^2}{\alpha}\right)$ and the posterior distribution is

$$\mu|\mathbf{x} \sim \mathcal{N}\left(\frac{\alpha m + n\bar{x}}{\alpha + n}, \frac{\sigma^2}{\alpha + n}\right).$$

Thus we can interpret the information in the prior as being equivalent to the information in a sample of size $\alpha = \frac{\sigma^2}{\tau^2}$ with sample mean m.

Limiting results and comparison with classical results

Suppose that we let $\alpha \to 0$ (or equivalently, $\tau^2 \to \infty$). Then the posterior distribution tends to

$$\mu|\mathbf{x} \sim \mathcal{N}\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

and the posterior mean is equal to the MLE and, for example, a posterior 95% *credible interval*

$$\mu \pm 1.96\frac{\sigma}{\sqrt{n}}$$

is (numerically) equal to the classical 95% confidence interval.

The limiting (improper) prior distribution in this case is uniform,

$$p(\mu) \propto 1.$$

Inference for σ when μ is known

In order to find a conjugate density for σ (or σ^2), we can again represent the normal density in one-dimensional exponential family form as

$$f(x|\sigma) = \underbrace{\frac{1}{\sigma}}_{C(\sigma)} \underbrace{\frac{1}{\sqrt{2\pi}}}_{h(x)} \exp\left(\underbrace{-\frac{1}{2\sigma^2}}_{\phi(\sigma)} \underbrace{(x-\mu)^2}_{s(x)} \right).$$

This implies that a conjugate prior for σ takes the form

$$p(\sigma) \propto \sigma^{-c} \exp\left(-\frac{d}{2\sigma^2}\right)$$

for some values of c, d. This is not a well known distributional form, although we can derive a simple prior distribution using a transformation.

A conjugate prior for the precision ϕ

It is much easier to work in terms of the precision $\phi = \frac{1}{\sigma^2}$. In this case, we have $X|\phi \sim \mathcal{N}\left(\mu, \frac{1}{\phi}\right)$ and

$$f(x|\phi) = \sqrt{\frac{\phi}{2\pi}} \exp\left(-\frac{\phi}{2}(x-\mu)^2\right)$$

which is already in exponential family form and implies that a conjugate prior density for ϕ is

$$p(\phi) \propto \left(\sqrt{\phi}\right)^c \exp\left(-\frac{\phi}{2}d\right)$$

for some c, d, which is a gamma density, $\phi \sim \mathcal{G}\left(\frac{a}{2}, \frac{b}{2}\right)$, where $a = c+2$ and $b = d$.

Note that using the change of variables formula on the prior for σ, we could also show that the implied prior for ϕ has the same structure.

Inference for ϕ when μ is known

Theorem 18
Let $X|\phi \sim \mathcal{N}\left(\mu, \frac{1}{\phi}\right)$ and suppose that $\phi \sim \mathcal{G}\left(\frac{a}{2}, \frac{b}{2}\right)$ for some a, b. Then, the marginal distribution of X is

$$\frac{X-\mu}{\sqrt{b/a}} \sim \mathcal{T}_a$$

and given a sample $\mathbf{x} = (x_1, \ldots, x_n)$, the posterior distribution of ϕ is

$$\phi|\mathbf{x} \sim \mathcal{G}\left(\frac{a+n}{2}, \frac{b+(n-1)s^2+n(\mu-\bar{x})^2}{2}\right).$$

Proof of the predictive distribution

Proof Define $Z = \sqrt{\phi}(X - \mu) \sim \mathcal{N}(0, 1)$ and $Y = b\phi \sim \chi^2_a$ and therefore, from standard distribution theory we know that $\frac{Z}{\sqrt{Y/a}} \sim \mathcal{T}_a$, is a Student's t distributed random variable with a degrees of freedom and transforming back, we have

$$\frac{Z}{\sqrt{Y/a}} = \frac{X - \mu}{\sqrt{b/a}}$$

which proves the predictive distribution formula. ■

In order to demonstrate the formula for the posterior distribution, it is first convenient to give the formula for the reparameterized normal likelihood function.

The reparameterized normal likelihood function

Theorem 19
If a sample, $\mathbf{x}$, of size n is taken from the normal distribution $X| \sim \mathcal{N}\left(\mu, \frac{1}{\phi}\right)$, then the likelihood function is

$$l(\mu, \phi|\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}\left[(n-1)s^2 + n(\mu - \bar{x})^2\right]\right)$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Proof This follows immediately by substituting $\phi = 1/\sigma^2$ in Theorem 17. ∎

Given this form for the normal likelihood function, the posterior distribution in Theorem 18 can be derived immediately from Bayes Theorem by combining prior and likelihood.

Interpretation

In this case, we can interpret the information in the prior as equivalent to the information in a sample $\mathbf{y}$ of size a where the sufficient statistic $\sum_{i=1}^{n}(y_i - \mu)^2 = b$.

Limiting results

Letting $a, b \to 0$, we have the limiting improper prior $f(\phi) \propto \frac{1}{\phi}$ and the limiting posterior distribution

$$\phi|\mathbf{x} \sim \mathcal{G}\left(\frac{n}{2}, \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2}\right) \quad \text{so that} \quad E[\phi|\mathbf{x}] = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right)^{-1}$$

which is the classical MLE of ϕ. Note however that

$$E\left[\frac{1}{\phi}\middle|\,\mathbf{x}\right] = E\left[\sigma^2|\mathbf{x}\right] = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n-2}$$

which is not the MLE of σ^2.

Interval estimation for σ (or σ^2)

Theorem 20

Let $\phi \sim \mathcal{G}\left(\frac{a}{2}, \frac{b}{2}\right)$ and define $\sigma = 1/\sqrt{\phi}$. Then the constant, c, such that $p = P(\sigma < c) = P(\sigma^2 < c^2)$, for given p, is $c = \sqrt{\frac{b}{\chi^2_a(1-p)}}$.

Proof

$$
\begin{aligned}
p &= P(\sigma < c) = P\left(\sigma^2 < c^2\right) = P\left(\phi > \frac{1}{c^2}\right) \\
&= P\left(b\phi > \frac{b}{c^2}\right) = P\left(Y > \frac{b}{c^2}\right) \quad \text{where } Y \sim \chi^2_a \\
\frac{b}{c^2} &= \chi^2_a(1-p) \\
c &= \sqrt{\frac{b}{\chi^2_a(1-p)}}.
\end{aligned}
$$

■

Limiting case posterior interval estimation of σ and σ^2

Theorem 21

Let $X|\phi \sim \mathcal{N}(\mu, 1/\phi)$ with (improper) prior distribution $f(\phi) \propto 1/\phi$. Then

$$\phi|\mathbf{x} \sim \mathcal{G}\left(\frac{n}{2}, \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}\right).$$

Moreover, let $\sigma = 1/\sqrt{\phi}$. Then a $100(1-p)\%$ credible interval for σ is

$$\sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi_n^2\left(1-\frac{p}{2}\right)}} < \sigma < \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi_n^2\left(\frac{p}{2}\right)}}$$

and a $100(1-p)\%$ credible interval for σ^2 is

$$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi_n^2\left(1-\frac{p}{2}\right)} < \sigma^2 < \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi_n^2\left(\frac{p}{2}\right)},$$

which are equal to the classical $100(1-p)\%$ confidence intervals for σ and σ^2.

Proof We have already given the posterior density for ϕ and the credible intervals follow immediately from Theorem 20 by setting $a = n$ and $b = \sum_{i=1}^{n}(x_i - \mu)^2$. ∎

Joint inference for μ and ϕ

When both parameters are unknown, it is again convenient to model in terms of the precision, ϕ, instead of the variance. Representing the normal density in exponential family form, we have

$$f(x|\mu,\phi) = \underbrace{\sqrt{\phi}\exp\left(-\frac{\phi\mu^2}{2}\right)}_{C(\mu,\phi)} \underbrace{\frac{1}{\sqrt{2\pi}}}_{h(x)} \exp\left(\underbrace{\phi\mu}_{\phi_1(\mu,\phi)} \underbrace{x}_{s_1(x)} \underbrace{-\frac{\phi}{2}}_{\phi_2(\mu,\phi)} \underbrace{x^2}_{s_2(x)} \right)$$

so that a conjugate prior for μ, ϕ can be found as

$$\begin{aligned} p(\mu,\phi) &\propto \left(\sqrt{\phi}\exp\left(-\frac{\phi\mu^2}{2}\right)\right)^a \exp\left(b_1\phi\mu - b_2\frac{\phi}{2}\right) \quad \text{for some } a, b_1, b_2 \\ &\propto \phi^{\frac{a}{2}}\exp\left(-\frac{\phi}{2}\left[a\mu^2 - 2b_1\mu + b_2\right]\right) \propto \phi^{\frac{a}{2}}\exp\left(-\frac{\phi}{2}\left[d + a(\mu-c)^2\right]\right) \end{aligned}$$

where $c = b_1/a$ and $d = b_2 - b_1^2/a$. This is a *normal-gamma distribution*.

Definition 7

Let $\mathbf{Y} = (Y_1, Y_2)$ and suppose that $Y_1|Y_2 \sim \mathcal{N}\left(\theta, \frac{1}{\lambda Y_2}\right)$ and $Y_2 \sim \mathcal{G}(\delta, \kappa)$. Then the distribution of $\mathbf{Y}$ is called a normal-gamma distribution, $\mathbf{Y} \sim \mathcal{NG}(\theta, \lambda, \delta, \kappa)$. The normal-gamma density function is

$$f(\mathbf{y}) = \frac{\kappa^\delta}{\Gamma(\delta)}\sqrt{\frac{\lambda}{2\pi}} y_2^{\frac{2\delta-1}{2}} \exp\left(-\frac{y_2}{2}\left[2\kappa + \lambda(y_1-\theta)^2\right]\right) \quad \text{for } -\infty < y_1 < \infty, y_2 > 0.$$

The following theorem gives the properties of the marginal distribution of Y_1.

Theorem 22

If $\mathbf{Y} = (Y_1, Y_2) \sim \mathcal{NG}(\theta, \lambda, \delta, \kappa)$, then $\frac{Y_1-\theta}{\sqrt{\kappa/(\lambda\delta)}} \sim \mathcal{T}_{2\delta}$. Furthermore, $E[Y_1] = \theta$ if $\delta > \frac{1}{2}$ and $V[Y_1] = \frac{\kappa}{\lambda(\delta-1)}$ if $\delta > 1$.

Proof

$$
\begin{aligned}
f(y_1) &= \int f(y_1, y_2)\, dy_2 \\
&= \frac{\kappa^\delta}{\Gamma(\delta)} \sqrt{\frac{\lambda}{2\pi}} \int_0^\infty y_2^{\frac{2\delta-1}{2}} \exp\left(-\frac{y_2}{2}\left[2\kappa + \lambda(y_1-\theta)^2\right]\right) dy_2 \\
&\propto \int_0^\infty y_2^{\frac{2\delta+1}{2}-1} \exp\left(-\frac{y_2}{2}\left[2\kappa + \lambda(y_1-\theta)^2\right]\right) dy_2 \\
&\propto \left(2\kappa + \lambda(y_1-\theta)^2\right)^{\frac{2\delta+1}{2}} \propto \left(1 + \frac{1}{2\delta}\left(\frac{y_1-\theta}{\sqrt{\kappa/(\lambda\delta)}}\right)^2\right)^{\frac{2\delta+1}{2}}
\end{aligned}
$$

and defining $T = \frac{Y_1-\theta}{\sqrt{\kappa/(\lambda\delta)}}$, we see that $f(t) \propto \left(1 + \frac{1}{2\delta}t^2\right)^{\frac{2\delta+1}{2}}$ which is the nucleus of a Student's t density with 2δ degrees of freedom. The mean and variance formulae follow from writing $Y_1 = \theta + \sqrt{\kappa/(\lambda\delta)}T$ and noting that $E[T] = 0$ for $2\delta > 1$ and $V[T] = \frac{2\delta}{2\delta-2} = \frac{\delta}{\delta-1}$ for $2\delta > 2$. ∎

Conjugate inference for μ, ϕ given a normal-gamma prior

Theorem 23

Let $X|\mu,\phi \sim \mathcal{N}\left(\mu, \frac{1}{\phi}\right)$ and $\mu,\phi \sim \mathcal{NG}\left(m, \alpha, \frac{a}{2}, \frac{b}{2}\right)$ so that $\mu|\phi \sim \mathcal{N}\left(m, \frac{1}{\alpha\phi}\right)$ and $\phi \sim \mathcal{G}\left(\frac{a}{2}, \frac{b}{2}\right)$. Then the marginal distribution of X is

$$\frac{X - m}{\sqrt{(\alpha+1)b/(\alpha a)}} \sim \mathcal{T}_a.$$

Given a sample, $\mathbf{x}$, the posterior distribution of μ, ϕ is

$$\begin{aligned}
\mu, \phi &\sim \mathcal{NG}\left(m^\star, \alpha^\star, \frac{a^\star}{2}, \frac{b^\star}{2}\right) \quad \text{where} \\
m^\star &= \frac{\alpha m + n\bar{x}}{\alpha + n} \\
\alpha^\star &= \alpha + n \\
a^\star &= a + n \\
b^\star &= b + (n-1)s^2 + \frac{\alpha n}{\alpha + n}(m - \bar{x})^2.
\end{aligned}$$

Proof Note first that

$$X|\phi \sim \mathcal{N}\left(m, \left(1+\frac{1}{\alpha}\right)\frac{1}{\phi}\right).$$

Therefore, X and ϕ have a normal gamma distribution.

$$X, \phi \sim \mathcal{NG}\left(m, \left(1+\frac{1}{\alpha}\right)^{-1}, a/2, b/2\right)$$

and the marginal distribution of X follows immediately from Theorem 22 setting $(\theta, \lambda, \delta, \kappa) = \left(m, \left(1+\frac{1}{\alpha}\right)^{-1}, a/2, b/2\right)$. In order to calculate the posterior distribution, we can use Bayes theorem:

$$
\begin{aligned}
p(\mu,\phi|\mathbf{x}) &\propto p(\mu,\phi)l(\mu,\phi|\mathbf{x}) \\
&\propto \phi^{\frac{a-1}{2}}\exp\left(-\frac{\phi}{2}\left[b+\alpha(\mu-m)^2\right]\right)\phi^{\frac{n}{2}}\exp\left(-\frac{\phi}{2}\left[(n-1)s^2+n(\mu-\bar{x})^2\right]\right) \\
&\propto \phi^{\frac{a+n-1}{2}}\exp\left(-\frac{\phi}{2}\left[b+(n-1)s^2+(\alpha+n)\mu^2-2(\alpha m+n\bar{x})\mu+\alpha m^2+n\bar{x}^2\right]\right) \\
&\propto \phi^{\frac{a+n-1}{2}}\exp\left(-\frac{\phi}{2}\left[b+(n-1)s^2+(\alpha+n)\left(\mu-\frac{\alpha m+n\bar{x}}{\alpha+n}\right)^2+\alpha m^2+n\bar{x}^2-\right.\right. \\
&\qquad \left.\left.-\frac{(\alpha m+n\bar{x})^2}{\alpha+n}\right]\right) \\
&\propto \phi^{\frac{a+n-1}{2}}\exp\left(-\frac{\phi}{2}\left[b+(n-1)s^2+\frac{\alpha n}{\alpha+n}(m-\bar{x})^2+\right.\right. \\
&\qquad \left.\left.(\alpha+n)\left(\mu-\frac{\alpha m+n\bar{x}}{\alpha+n}\right)^2\right]\right)
\end{aligned}
$$

which is the nucleus of the given normal-gamma density. ∎

Results with the limiting prior

Suppose that we use the improper prior $p(\mu, \phi) \propto \frac{1}{\phi}$. Then,

$$\begin{aligned} p(\mu, \phi|\mathbf{x}) &\propto \frac{1}{\phi}\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}\left[(n-1)s^2 + n(\mu - \bar{x})^2\right]\right) \\ &\propto \phi^{\frac{(n-1)-1}{2}} \exp\left(-\frac{\phi}{2}\left[(n-1)s^2 + n(\mu - \bar{x})^2\right]\right) \quad \text{so that} \\ \mu, \phi|\mathbf{x} &\sim \mathcal{NG}\left(\bar{x}, n, \frac{n-1}{2}, \frac{(n-1)s^2}{2}\right) \end{aligned}$$

Now, from Theorem 22, we have that

$$\frac{\mu - \bar{x}}{\sqrt{(n-1)s^2/(n(n-1))}} = \frac{\mu - \bar{x}}{s/\sqrt{n}} \sim \mathcal{T}_{n-1}.$$

Therefore, for example, a 100(1-p)% posterior credible interval for μ is

$$\bar{x} \pm \frac{s}{\sqrt{n}} \mathcal{T}_{n-1}\left(1 - \frac{p}{2}\right)$$

which is equal to the classical confidence interval.

Semi-conjugate inference via Gibbs sampling

The fully conjugate, normal-gamma prior distribution is slightly unnatural from the point of view of real prior choice, in that the distribution of μ depends on the unknown model precision ϕ. An alternative is to assume that μ and ϕ have *independent* prior distributions, say

$$\mu \sim \mathcal{N}\left(m, \frac{1}{\psi}\right) \quad \phi \sim \mathcal{G}\left(\frac{a}{2}, \frac{b}{2}\right)$$

where m, ψ, a, b are all known.

In this case, the joint posterior distribution given a sample, $\mathbf{x}$, is

$$p(\mu, \phi|\mathbf{x}) \propto \phi^{\frac{a+n}{2}-1} \exp\left(-\frac{1}{2}\left[\left(b + (n-1)s^2\right)\phi + n\phi(\mu - \bar{x})^2 + \psi(\mu - m)^2\right]\right).$$

The marginal distributions of μ and ϕ are

$$\begin{aligned} p(\mu|\mathbf{x}) &\propto \exp\left(-\frac{\psi}{2}(\mu-m)^2\right)\left(b+(n-1)s^2+n(\mu-\bar{x})^2\right)^{-\frac{a+n}{2}} \\ p(\phi|\mathbf{x}) &\propto \frac{\phi^{\frac{a+n}{2}-1}}{\sqrt{n\phi+\psi}}\exp\left(-\frac{\phi}{2}\left[b+(n-1)s^2+\frac{n\psi}{n\phi+\psi}(m-\bar{x})^2\right]\right). \end{aligned}$$

Neither the joint posterior density or the marginal posterior densities has a standard form. One possible solution is to use numerical integration techniques to estimate the integration constants, moments etc. of these densities.

Another possibility is to use a Monte Carlo approach. In this case, direct Monte Carlo sampling cannot easily be applied. However, an indirect approach is available.

Note firstly that the *conditional* posterior densities of $\mu|\phi,\mathbf{x}$ and of $\phi|\mu,\mathbf{x}$ are available in closed form. We have

$$\begin{aligned}\mu|\phi,\mathbf{x} &\sim \mathcal{N}\left(\frac{n\phi\bar{x}+\psi m}{n\phi+\psi},\frac{1}{n\phi+\psi}\right)\\ \phi|\mu,\mathbf{x} &\sim \mathcal{G}\left(\frac{a+n}{2},\frac{b+(n-1)s^2+n(\mu-\bar{x})^2}{2}\right)\end{aligned}$$

It is straightforward to sample from these distributions. Therefore, we can set up a *Gibbs sampler* to give an approximate Monte Carlo sample from the joint posterior distribution.

Aside: The (two variable) Gibbs sampler

Assume that we wish to sample from the density of $\mathbf{X} = (X_1, X_2)$. Suppose that the conditional densities, $p(\cdot|X_2)$ and $p(\cdot|X_1)$ are known. Then the (two variable) Gibbs sampler (Geman and Geman 1984) is a scheme for iteratively sampling these conditional densities as follows.

1. $t = 0$. Fix an initial value $x_1^{(0)}$.
2. Sample $x_2^{(t)} \sim p\left(x_2|x_1^{(t-1)}\right)$.
3. Sample $x_1^{(t)} \sim p\left(x_1|x_2^{(t)}\right)$.
4. Set $t = t + 1$.
5. Go to 2.

As t increases, it can be demonstrated that the sampled values approximate a Monte Carlo sample from the joint density, $p(\mathbf{X})$ and can be used in the same way as any other Monte Carlo sample.

The Gibbs sampler is an example of a Markov chain Monte Carlo (MCMC) algorithm and will be discussed in full in chapter 6.

Applying the Gibbs sampler to sampling $p(\mu, \phi|\mathbf{x})$

In our case, we can implement the Gibbs sampler as follows.

1. $t = 0$. Fix an initial value $\mu^{(0)}$.

2. Generate $\phi^{(t)} \sim \mathcal{G}\left(\frac{a+n}{2}. \frac{b+(n-1)s^2+n\left(\mu^{(t-1)}-\bar{x}\right)^2}{2}\right)$.

3. Generate $\mu^{(t)} \sim \mathcal{N}\left(\frac{n\phi^{(t)}\bar{x}+\psi m}{n\phi^{(t)}+\psi}, \frac{1}{n\phi^{(t)}+\psi}\right)$

4. Set $t = t + 1$.

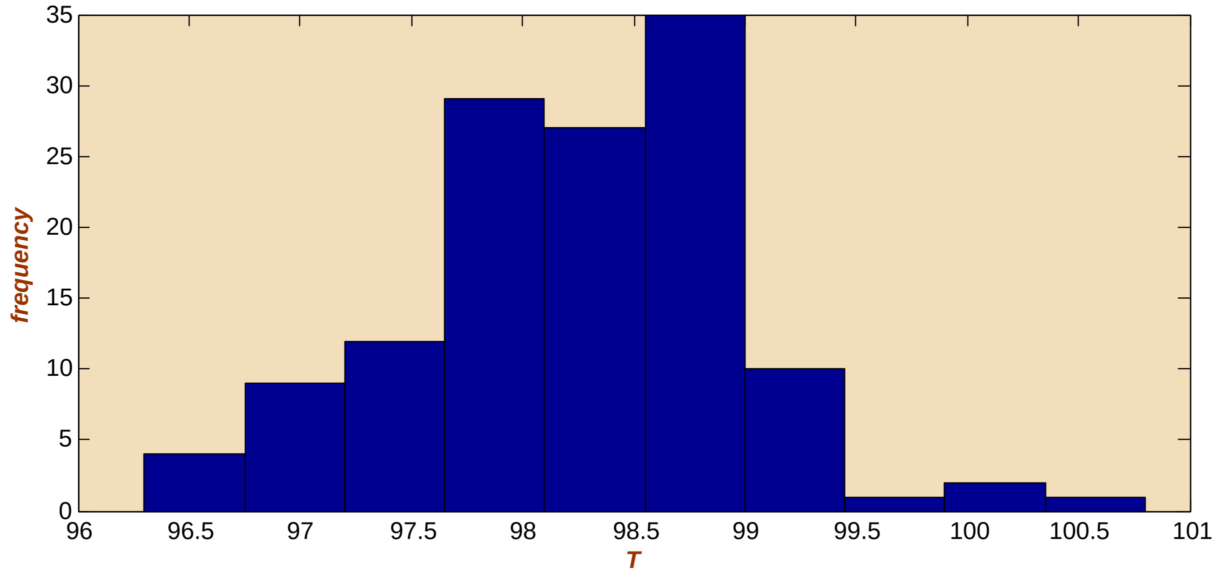
5. Go to 2.

In order to estimate the marginal density of μ, for example, we can recall that $p(\mu|\mathbf{x}) = \int p(\mu|\phi, \mathbf{x})p(\phi|\mathbf{x}) \approx \frac{1}{T}\sum_{t=1}^{T} p\left(\mu|\phi^{(t)}, \mathbf{x}\right)$.

Example 20

The normal core body temperature, T, of a healthy adult is supposed to be 98.6 degrees fahrenheit or 37 degrees celsius on average. As temperature can very around the mean, depending on given conditions, a normal model for temperatures, say $T|\mu,\phi \sim \mathcal{N}(\mu, 1/\phi)$, has been proposed.

Mackowiak et al (1992) measured the core body temperatures of 130 individuals and a histogram of the results is given below.



The sample mean temperature is $\bar{x} = 98.2492$ fahrenheit with sample standard deviation $s = 0.7332$. Thus, a classical 95% confidence interval for the true mean temperature is

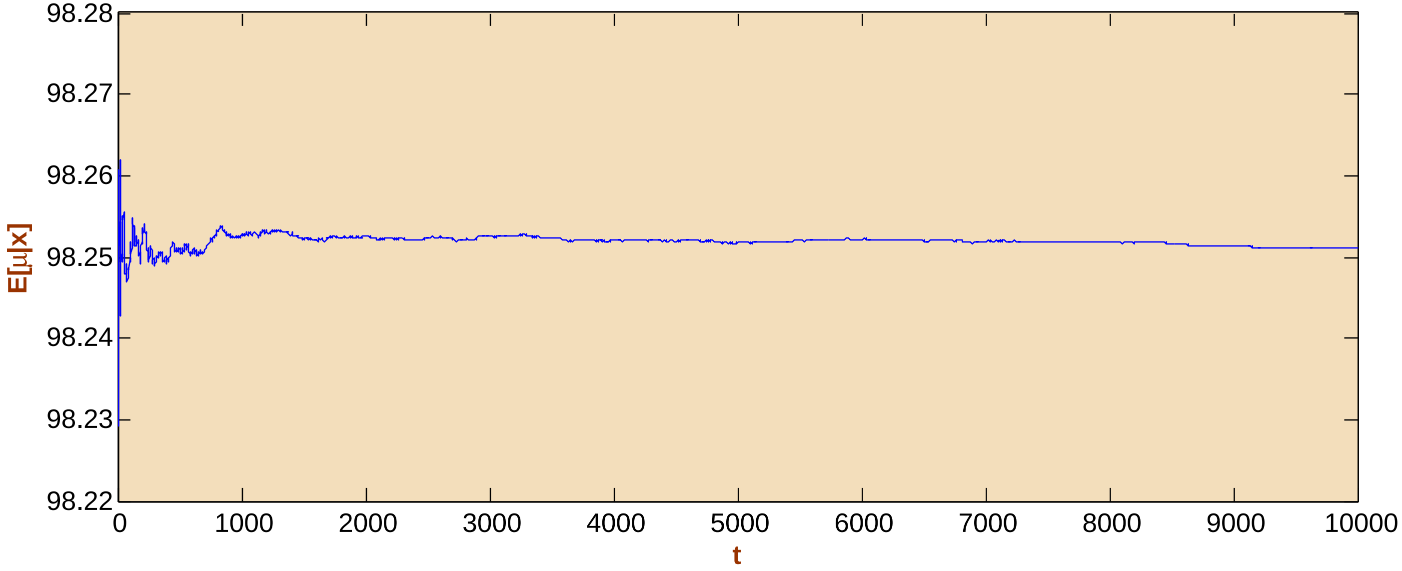
$$98.2492 \pm 1.96 \times 0.7332/\sqrt{130} = (98.1232, 98.3752)$$

and the hypothesis that the true temperature is 98.6 degrees is clearly rejected.

From a Bayesian viewpoint, we will analyze using two different prior distributions. Firstly, we assume the limiting prior $p(\mu, \phi) \propto \frac{1}{\phi}$ and secondly, we assume independent priors with relatively large variances, that is $\mu \sim \mathcal{N}(98.6, 1)$ and $\phi \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2}\right)$.

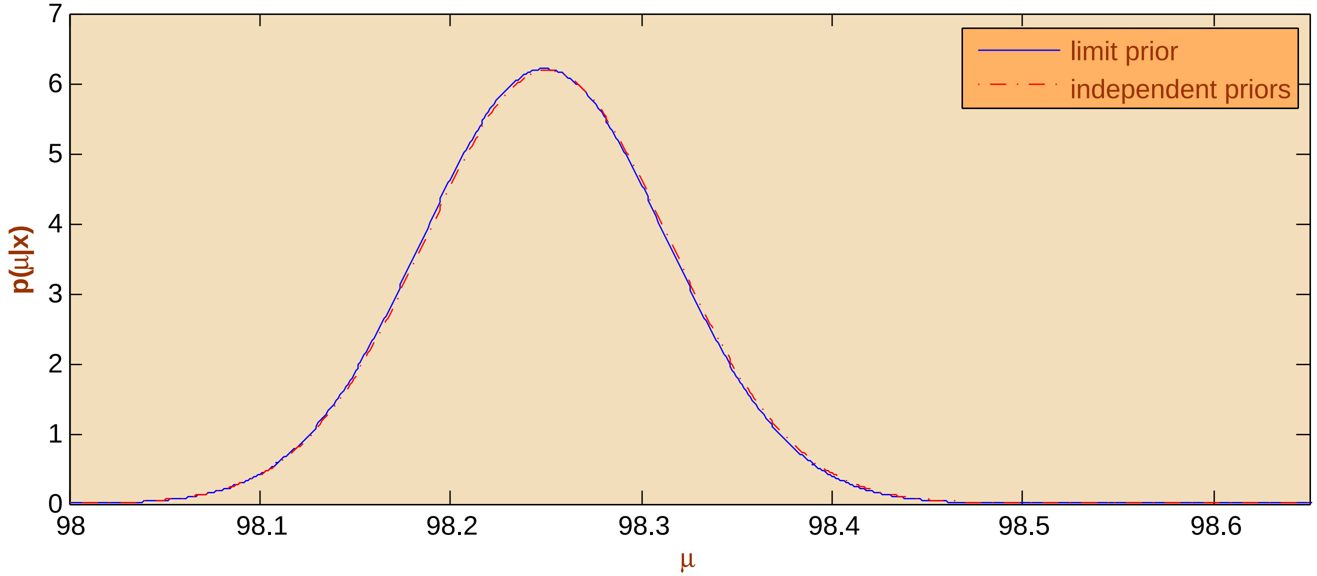
Given the limiting prior, we know that the posterior density of μ is $\frac{\mu - 98.2492}{0.7332/\sqrt{130}} \sim \mathcal{T}_{129}$ and a 95% posterior credible interval for μ coincides with the classical confidence interval.

Given the independent priors, the Gibbs sampler was run for 10000 iterations, starting at $\mu^{(0)} = 98.2492$. The following diagram shows a plot of the estimated posterior mean of μ versus the number of iterations.



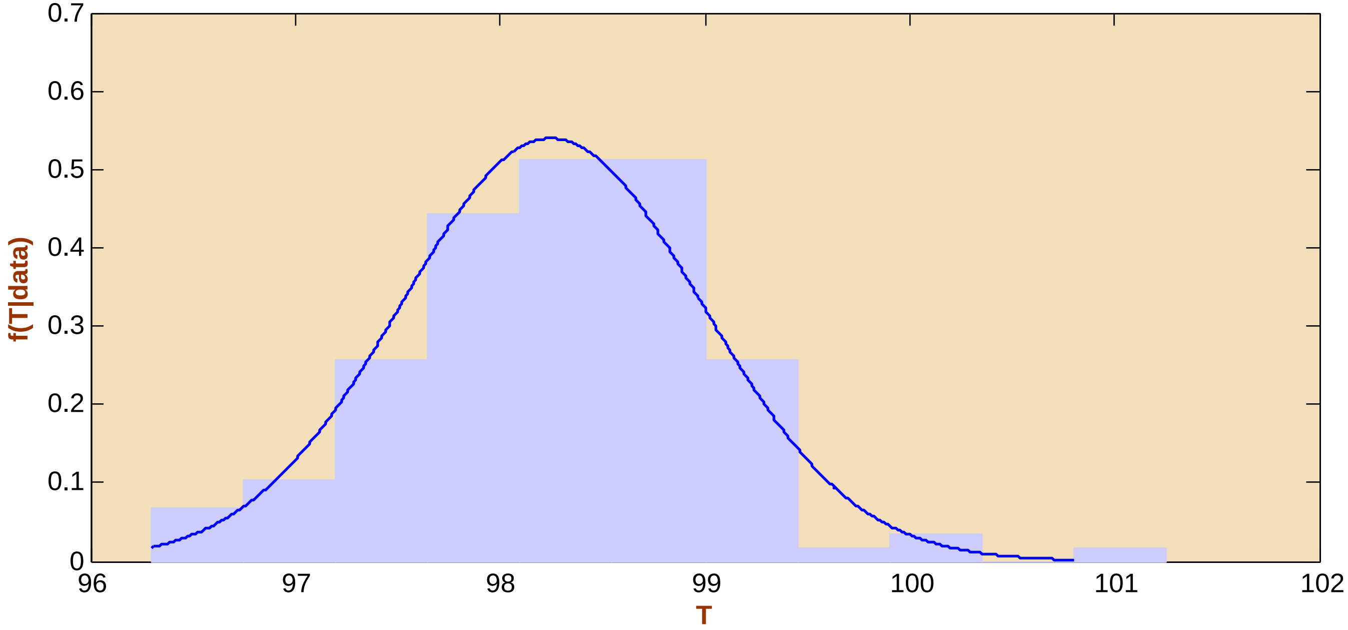
The estimated mean appears to have converged after 2000 iterations or less. The estimated posterior mean of μ was $E[\mu|\mathbf{x}] \approx 98.2512$.

The following diagram shows the posterior densities of μ given the limit and independent priors.



It can be seen that the posterior probability that μ is bigger than 98.6 fahrenheit is virtually zero.

The final diagram shows the predictive density of T given the observed data and the independent prior distributions.



The fit of the normal distribution is reasonable although it might be improved by incorporating other covariates, e.g. sex or heart rate.

A simple MATLAB code for executing the Gibbs sampler is as below.

```
n=length(temp); xbar=mean(temp); s=std(temp); % Summary statistics.
a=1; b=1; m=98.6; psi=1;                      % Priors for phi and mu.
iters=10000;mu=zeros(1,iters);phi=zeros(1,iters);                    % Sample
murange=98+[0:1000]/1000;pmu=zeros(1,length(murange));pmuuninf=pmu % initialization
xrange=min(temp)+[0:1000]*(max(temp)-min(temp))/1000; fx=zeros(1,length(xrange));

% Start of Gibbs sampler
mu(1)=xbar; phi(1)=gamrnd((a+n)/2,2/(b+(n-1)*s*s+n*((mu(1)-xbar)^2)));
for t=2:iters
    mu(t)=normrnd((psi*m+n*phi(t-1)*xbar)/(psi+n*phi(t-1)),1/sqrt(psi+n*phi(t-1)));
    phi(t)=gamrnd((a+n)/2,2/(b+(n-1)*s*s+n*((mu(t)-xbar)^2)));
    pmu=pmu+normpdf(murange,(psi*m+n*phi(t)*xbar)/(psi+n*phi(t)),1/sqrt(psi+n*phi(t)));
    fx=fx+normpdf(xrange,mu(t),1/sqrt(phi(t)));
end
pmu=pmu/iters; fx=fx/iters;  %Normalization.

pmuuninf=normpdf(murange,xbar,s/sqrt(n)); %Posterior density given 1/phi prior.
% Plots
figure; plot([1:iters],cumsum(mu)./[1:iters]); figure;
plot(murange,pmuuninf); hold on; plot(murange,pmu,'-.r'); hold off;
figure; histadjusted(temp,10); hold on; plot(xrange,fx); hold off;
```

Two sample problems

Various two sample problems have been studied.

1. Paired data samples: difference in means.

2. Unpaired samples:

   (a) Difference of two population means: variances known,
   (b) Difference of two population means: variances unknown but equal,
   (c) Difference of two population means: unknown variances,
   (d) Ratio of two population variances.

We shall consider Bayesian inference for each situation in turn.

Paired data

As in classical inference, by considering the differences of each pair, this reduces to a single normal sample inference problem.

Unpaired data: known variances

Assume that we have $X|\mu_X \sim \mathcal{N}\left(\mu_X, \sigma_X^2\right)$ and $Y \sim \mathcal{N}\left(\mu_Y, \sigma_Y^2\right)$ where σ_X^2 and σ_Y^2 are known. Suppose also that we use independent, uniform priors for μ_X and μ_Y. Then, given samples $\mathbf{x}$ and $\mathbf{y}$ of sizes n_X and n_Y respectively, we know that the posterior distributions are independent normal

$$\mu_X|\mathbf{x} \sim \mathcal{N}\left(\bar{x}, \frac{\sigma_X^2}{n_X}\right) \quad \mu_Y|\mathbf{y} \sim \mathcal{N}\left(\bar{y}, \frac{\sigma_Y^2}{n_Y}\right).$$

Therefore, if we define $\delta = \mu_X - \mu_Y$ to be the difference in the two population means, then the posterior distribution of δ is

$$\delta|\mathbf{x}, \mathbf{y} \sim \mathcal{N}\left(\bar{x} - \bar{y}, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right).$$

In this case, a 95% credible interval for δ is

$$\bar{x} - \bar{y} \pm 1.96\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

which is equal to the corresponding classical confidence interval.

It is, of course, straightforward to generalize this result to the case when proper priors for μ_X and μ_Y are used.

Unpaired data: unknown, equal variances

Suppose now that $X|\mu_X,\phi \sim \mathcal{N}\left(\mu_X, \frac{1}{\phi}\right)$ and $Y|\mu_Y,\phi \sim \mathcal{N}\left(\mu_Y, \frac{1}{\phi}\right)$ have common, unknown precision ϕ and that we observe samples $\mathbf{x}$ and $\mathbf{y}$ of sizes n_X and n_Y respectively. Consider the joint prior distribution

$$p(\mu_X, \mu_Y, \phi) \propto \frac{1}{\phi}.$$

Then, the joint posterior distribution is

$$\begin{aligned} p(\mu_X, \mu_Y, \phi|\mathbf{x}, \mathbf{y}) \quad \propto \quad & \phi^{\frac{n_X+n_Y}{2}-1} \exp\left(-\frac{\phi}{2}\left[(n_X-1)s_X^2 + (n_Y-1)s_Y^2 + \right.\right. \\ & \left.\left. n_X(\mu_X-\bar{x})^2 + n_Y(\mu_Y-\bar{y})^2\right]\right). \end{aligned}$$

Given this joint distribution, we can see that

$$\mu_X|\phi,\mathbf{x},\mathbf{y} \sim \mathcal{N}\left(\bar{x}, \frac{1}{n_X\phi}\right) \quad \mu_Y|\phi,\mathbf{x},\mathbf{y} \sim \mathcal{N}\left(\bar{y}, \frac{1}{n_Y\phi}\right).$$

Integrating out μ_X and μ_Y successively from the joint distribution, we have

$$\phi|\mathbf{x} \sim \mathcal{G}\left(\frac{n_X + n_Y - 2}{2}, \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{2}\right).$$

Also, from the previous slide, if $\delta = \mu_X - \mu_Y$, then

$$\delta|\phi, \mathbf{x}, \mathbf{y} \sim \mathcal{N}\left(\bar{x} - \bar{y}, \frac{1}{\phi}\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)\right).$$

Thus, the joint posterior distribution of δ and ϕ is normal-gamma

$$\delta, \phi|\mathbf{x}, \mathbf{y} \sim \mathcal{NG}\left(\bar{x} - \bar{y}, \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)^{-1}, \frac{n_X + n_Y - 2}{2}, \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{2}\right).$$

Therefore, the marginal posterior distribution of δ is Student's t:

$$\frac{\delta - (\bar{x} - \bar{y})}{\sqrt{\frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)^{-1}(n_X + n_y - 2)}}} = \frac{\delta - (\bar{x} - \bar{y})}{\sqrt{\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)\frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_y - 2}}} \sim \mathcal{T}_{n_X + n_Y - 2}.$$

Now, a 95% credible interval for δ is given by

$$\bar{x} - \bar{y} \pm \mathcal{T}_{n_X + n_y - 2}(0.975)\sqrt{\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)} s_c$$

where s_c^2 is the (classical) combined variance estimator

$$s_c^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_y - 2}.$$

This is equal to the usual, classical confidence interval.

The Behrens-Fisher problem

Now assume that $X|\mu_X, \phi_X \sim \mathcal{N}\left(\mu_X, \frac{1}{\phi_X}\right)$ and $Y|\mu_Y, \phi_Y \sim \mathcal{N}\left(\mu_Y, \frac{1}{\phi_Y}\right)$ where all parameters are unknown. Given the (improper) prior distributions

$$p(\mu_X, \phi_X) \propto \frac{1}{\phi_X} \quad \text{and} \quad p(\mu_Y, \phi_Y) \propto \frac{1}{\phi_Y}$$

then we know that the marginal posterior distributions of μ_X and μ_Y are independent, shifted, scaled t distributions:

$$\frac{\mu_X - \bar{x}}{s_X/\sqrt{n_X}} \sim \mathcal{T}_{n_X-1} \quad \frac{\mu_Y - \bar{y}}{s_Y/\sqrt{n_Y}} \sim \mathcal{T}_{n_Y-1}.$$

Thus, if $\delta = \mu_X - \mu_Y$, we have

$$\delta = T_{n_X-1}\frac{s_X}{\sqrt{n_X}} - T_{n_Y-1}\frac{s_Y}{\sqrt{n_Y}} - (\bar{x} - \bar{y})$$

where T_d represents a t distributed variable with d degrees of freedom.

We could try to derive the density formula of δ directly from this expression. However, it is more straightforward to work in terms of a transformed variable.

Theorem 24
Let $\delta' = \frac{\delta-(\bar{x}-\bar{y})}{\sqrt{s_X^2/n_X+s_Y^2/n_Y}}$. Then we can write

$$\delta' = T_{n_X-1}\sin\omega - T_{n_Y-1}\cos\omega$$

where $\omega = \tan^{-1}\frac{s_X/\sqrt{n_X}}{s_Y/\sqrt{n_Y}}$.

Proof From the previous page, we have

$$\begin{aligned}\delta &= T_{n_X-1}\frac{s_X}{\sqrt{n_X}} - T_{n_Y-1}\frac{s_Y}{\sqrt{n_Y}} - (\bar{x}-\bar{y}) \\ \frac{\delta-(\bar{x}-\bar{y})}{\sqrt{s_X^2/n_X+s_Y^2/n_Y}} &= \frac{T_{n_X-1}\frac{s_X}{\sqrt{n_X}}}{\sqrt{s_X^2/n_X+s_Y^2/n_Y}} - \frac{T_{n_Y-1}\frac{s_Y}{\sqrt{n_Y}}}{\sqrt{s_X^2/n_X+s_Y^2/n_Y}}\end{aligned}$$

and the result follows by observing that the squares of the coefficients on the right hand side sum to 1. ∎

Definition 8

The distribution of

$$Y = T_{\nu_1} \sin \omega - T_{\nu_2} \cos \omega$$

where T_{ν_1} and T_{ν_2} are standard Student's t variables with ν_1 and ν_2 degrees of freedom, is the *Behrens-Fisher distribution*, (Behrens 1929) with parameters ν_1, ν_2 and ω. In this case, we write $Y \sim \mathcal{BF}(\nu_1, \nu_2, \omega)$.

Thus, from Theorem 24, we have

$$\delta' | \mathbf{x}, \mathbf{y} \sim \mathcal{BF}\left(n_X - 1, n_Y - 1 \tan^{-1} \frac{s_X/\sqrt{n_X}}{s_Y/\sqrt{n_Y}} \right).$$

Note that this Bayesian solution to the Behrens-Fisher problem corresponds to the *fiducial* solution and was first developed in this context by Fisher (1935).

Patil's approximation and other appoximate methods

Tables of the percentage points of the Behrens-Fisher distribution are available for given values of ω, see e.g. Kim and Cohen (1996). For other values, Patil's (1965) approximation may be applied.

If $Y \sim \mathcal{BF}(\nu_1, \nu_2, \omega)$ then we have $Y/a \approx \mathcal{T}_b$ where

$$
\begin{aligned}
b &= 4 + \left(\frac{\nu_1 \cos^2 \omega}{\nu_1 - 2} + \frac{\nu_2 \sin^2 \omega}{\nu_2 - 2} \right)^2 \left[\frac{\nu_1^2 \cos^4 \omega}{(\nu_1 - 2)^2 (\nu_1 - 4)} + \frac{\nu_2^2 \sin^4 \omega}{(\nu_2 - 2)^2 (\nu_2 - 4)} \right]^{-1} \\
a &= \frac{b}{b-2} \left[\frac{\nu_1 \cos^2 \omega}{\nu_1 - 2} + \frac{\nu_2 \sin^2 \omega}{\nu_2 - 2} \right]^{-1}
\end{aligned}
$$

Alternative approximation methods are given by Ghosh (1975) or Willink (2004).

A Monte Carlo approach

Another approach is to use Monte Carlo sampling to approximate a Behrens Fisher distribution as follows. Suppose that $Y \sim \mathcal{BF}(\nu_1, \nu_2, \omega)$. Then we can generate a sample from the Behrens Fisher distribution as follows:

1. Fix some large M.
2. For $i = 1, \ldots, M$
   (a) Generate $t_1^{(i)} \sim \mathcal{T}_{\nu_1}$ and $t_2^{(i)} \sim \mathcal{T}_{\nu_2}$.
   (b) Define $y^{(i)} = t_1^{(i)} \cos\omega - t_2^{(i)} \sin\omega$.

Then, approximate 95% credible intervals for Y could be estimated using the sample quantiles. If M is increased sufficiently, the accuracy of these estimates can be derived to any fixed precision.

The frequentist approach to the Behrens-Fisher problem

No equivalent solution exists in frequentist inference because in this case, the sampling distribution of δ' depends upon the ratio, ϕ_X/ϕ_Y, of the unknown precisions.

The usual frequentist approximation to this sampling distribution is:

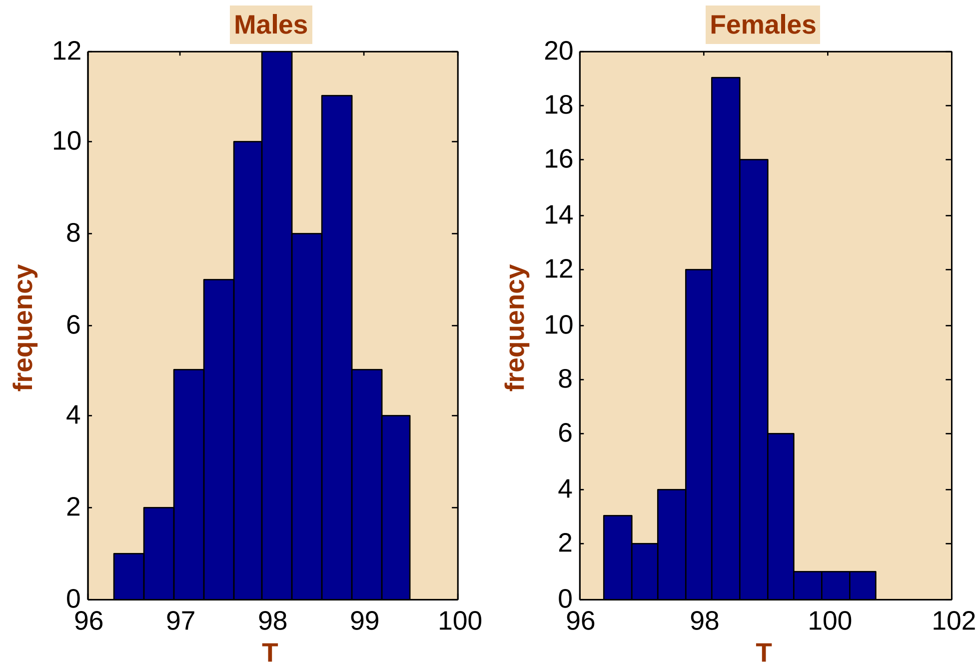
$$\delta' \approx \mathcal{T}_r \quad \text{where} \quad r = \frac{(s_X^2/n_X + s_Y^2/n_Y)^2}{\left(\frac{s_X^4/n_X^2}{n_X-1} + \frac{s_Y^4/n_Y^2}{n_Y-1}\right)}$$

However, we cannot know how good this approximation is for any given sample.

Another look at the normal body temperature example

Example 21

The sample contained 122 males and 22 females. The following two histograms illustrate the temperature distributions for both groups.



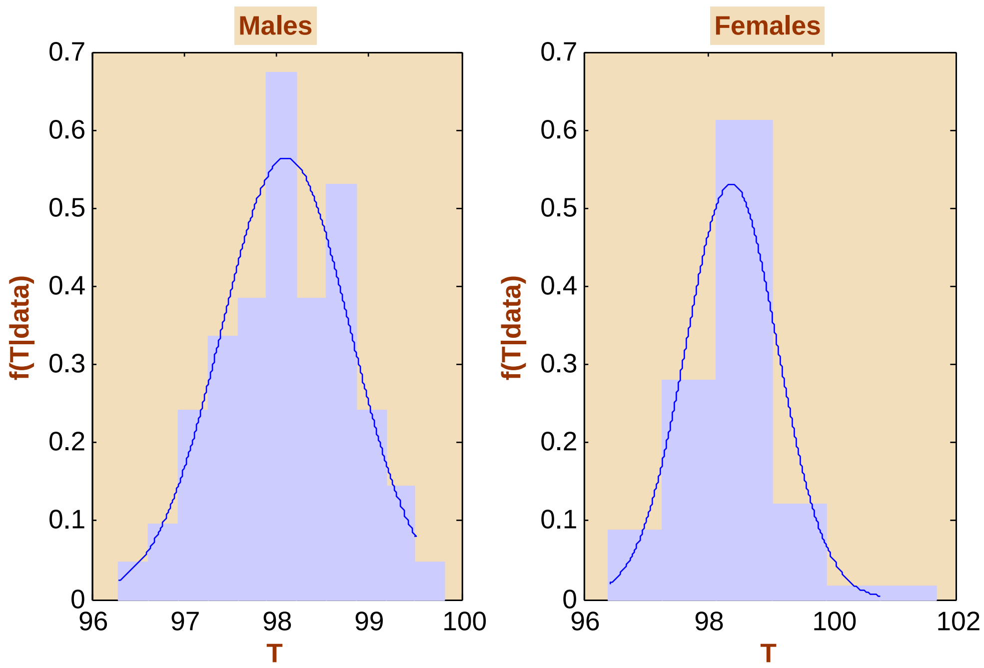
We may wish to question whether or not the mean body temperatures are the same in both groups.

The means (and standard deviations) for males and females are 98.1046 (0.6988) and 98.3938 (0.7435) respectively.

In this case, an approximate classical 95% confidence interval for the mean difference δ is given by $(-0.5396, -0.0388)$ and there is strong evidence that the mean body temperature is higher in females than in males.

Using a Bayesian approach with improper priors as outlined earlier, a 95% posterior credible interval for δ is given by $(-0.5359, -0.0426)$ which is slightly narrower than the classical interval.

Also, the fitted predictive posterior densities for both groups are given in the following diagram. The normal fits appear to have improved somewhat.



Comparing two population variances

Suppose as earlier that $X|\mu_X, \phi_X \sim \mathcal{N}\left(\mu_X, \frac{1}{\phi_X}\right)$ and $Y|\mu_Y, \phi_Y \sim \mathcal{N}\left(\mu_Y, \frac{1}{\phi_Y}\right)$ with the usual improper priors;

$$p(\mu_X, \phi_X) \propto \frac{1}{\phi_X} \quad p(\mu_Y, \phi_Y) \propto \frac{1}{\phi_Y}.$$

Then, from earlier, given samples of sizes n_X and n_Y, we know that

$$\phi_X|\mathbf{x} \sim \mathcal{G}\left(\frac{n_X - 1}{2}, \frac{(n_X - 1)s_X^2}{2}\right) \quad \phi_Y|\mathbf{y} \sim \mathcal{G}\left(\frac{n_Y - 1}{2}, \frac{(n_Y - 1)s_Y^2}{2}\right)$$

and therefore,

$$(n_X - 1)s_X^2\phi_X \sim \chi^2_{n_X-1} \quad (n_Y - 1)s_Y^2\phi_Y \sim \chi^2_{n_Y-1}.$$

Recalling that the ratio of two chi-squared distributions divided by their degrees of freedom is F distributed, we thus have

$$\frac{s_X^2}{s_Y^2}\frac{\phi_X}{\phi_Y} \sim \mathcal{F}_{n_Y-1}^{n_X-1}$$

and Bayesian and classical credible intervals will coincide.

Application II: Bayesian inference for the half-normal distribution

We will follow the analysis of Wiper et al (2008).

The half-normal distribution

If $Z \sim \mathcal{N}(0,1)$ then $X = |Z|$ has a half-normal distribution, $X \sim \mathcal{HN}(0,1)$, with probability density function

$$\begin{aligned} f(x) &= 2\phi(x) \quad \text{for } x > 0\text{, where} \\ \phi(x) &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \quad \text{is the standard normal pdf.} \end{aligned}$$

The generalized half-normal distribution

In this case, we allow the location and scale parameters to vary. Thus, let $X = \xi + \eta|Z|$ where Z is a standard normal random variable as earlier. Then we say that X has a generalized half-normal distribution, $X|\xi, \eta \sim \mathcal{HN}(\xi, \eta)$, with density

$$f(x|\xi, \eta) = \frac{2}{\eta}\phi\left(\frac{x-\xi}{\eta}\right) \quad \text{for } x > \xi.$$

Applications

The half-normal distribution is a model for truncated data with applications in various areas:

- measuring body fat indices in athletes (Pewsey 2002, 2004).
- stochastic frontier modeling (Aigner et al 1977, Meeusen and van den Broeck, 1977).
- fluctuating asymmetry modeling (Palmer and Strobeck 1986).
- fibre buckling (Haberle 1991).
- blowfly dispersion (Dobzhansky and Wright 1947).

Classical inference for the half-normal distribution

Suppose we have a sample $\mathbf{x} = (x_1, \ldots, x_n)$ from $\mathcal{HN}(\xi, \eta)$ and that we wish to estimate ξ (and η). Now from Pewsey (2002,2004):

- The MLE for ξ is $\hat{\xi} = x_{(1)}$, the minimum of the data.
- Uncorrected and bias corrected estimators for η are

$$\hat{\eta} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - x_{(1)})^2} \qquad \hat{\eta}_{BC} = \sqrt{\frac{n}{n-1}}\hat{\eta}$$

- An (asymptotic) 95% confidence interval for ξ:

$$x_{(1)} + \log(\alpha)\hat{\eta}\Phi^{-1}\left(\frac{1}{2} + \frac{1}{2n}\right) < \xi < x_{(1)}.$$

Bayesian inference for the half-normal distribution

Inference is similar to the usual normal case. First we reparameterize by defining $\tau = \frac{1}{\eta^2}$ so that

$$f(x|\xi,\tau) = 2\sqrt{\tau}\phi\left(\sqrt{\tau}(x-\xi)\right).$$

Now we can define a generalized version of the normal-gamma prior distribution, that is the *right-truncated normal-gamma* distribution.

The right-truncated normal-gamma distribution

We say that, ξ, τ have a RTNG with parameters ξ_0, m, α, a, b,

$$\xi, \tau \sim \mathcal{RTNG}\left(\xi_0, m, \alpha, \frac{a}{2}, \frac{b}{2}\right) \quad \text{if}$$

$$p(\xi, \tau) = \frac{1}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)} \frac{(b/2)^{a/2}}{\Gamma(a/2)} \sqrt{\frac{\alpha}{2\pi}} \tau^{\frac{a+1}{2}-1} \exp\left\{-\frac{\tau}{2}\left[b + \alpha(\xi - m)^2\right]\right\}$$

for $\xi < \xi_0$ where $\Phi_d(\cdot)$ is the Student's t cdf with d degrees of freedom;

$$\Phi_d(z) = \int_{-\infty}^{z} \phi_d(y)\, dy \quad \text{and}$$

$$\phi_d(y) = \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \frac{1}{\sqrt{\pi d}} \left(1 + \frac{y^2}{d}\right)^{-\frac{d+1}{2}}$$

What is the motivation for this density?

$$p(\xi,\tau) = \frac{1}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)} \frac{(b/2)^{a/2}}{\Gamma(a/2)} \sqrt{\frac{\alpha}{2\pi}} \tau^{\frac{a+1}{2}-1} \exp\left\{-\frac{\tau}{2}\left[b + \alpha(\xi - m)^2\right]\right\}$$

$$p(\xi,\tau) = \frac{1}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)} \underbrace{\frac{(b/2)^{a/2}}{\Gamma(a/2)} \sqrt{\frac{\alpha}{2\pi}} \tau^{\frac{a+1}{2}-1} \exp\left\{-\frac{\tau}{2}\left[b + \alpha(\xi - m)^2\right]\right\}}_{\text{the usual normal gamma density}}$$

$$p(\xi,\tau) = \underbrace{\frac{1}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)}}_{\text{from the truncation}} \frac{(b/2)^{a/2}}{\Gamma(a/2)} \sqrt{\frac{\alpha}{2\pi}} \tau^{\frac{a+1}{2}-1} \exp\left\{-\frac{\tau}{2}\left[b + \alpha(\xi - m)^2\right]\right\}$$

The marginal density of ξ

$$
\begin{aligned}
p(\xi) &= \frac{1}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)} \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \sqrt{\frac{\alpha}{b\pi}} \left(1 + \frac{1}{a}\left(\frac{\xi - m}{\sqrt{b/(a\alpha)}}\right)^2\right)^{-\frac{a+1}{2}} \\
&= \sqrt{\frac{a\alpha}{b}} \frac{\phi_a\left(\frac{\xi - m}{\sqrt{b/(a\alpha)}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)} \quad \text{for } \xi < \xi_0
\end{aligned}
$$

Thus, $\xi' = \frac{\xi - m}{\sqrt{b/(a\alpha)}} \sim \mathcal{T}_a$, truncated onto the region $\xi' < \frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}$.

Moments of ξ

$$E[\xi] = m - \sqrt{\frac{b}{a\alpha}} \frac{a + \left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)^2}{a-1} \frac{\phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)} \quad \text{for } a > 1$$

$$V[\xi] = \frac{b}{a\alpha(a-2)} \left\{ a - \left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right) \left(a + \left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)^2\right) \frac{\phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)} \right\}$$

$$- \frac{b}{a\alpha} \left(\frac{a + \left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)^2}{a-1} \frac{\phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)} \right)^2 \quad \text{for } a > 2$$

We have explicit formulae for all moments of ξ.

The conditional and marginal densities of τ

The conditional density of τ given ξ is the same as in the usual normal-gamma model, that is

$$\tau|\xi \sim \mathcal{G}\left(\frac{a}{2}, \frac{b+\alpha(\xi-m)^2}{2}\right) \quad \text{for } \tau > 0.$$

The marginal distribution of τ is a *Gaussian modulated gamma distribution*, that is $\tau \sim \mathcal{GMG}(\sqrt{\alpha}(\xi_0 - m), a, b)$, with density

$$p(\tau) = \frac{\Phi\left(\sqrt{\alpha\tau}(\xi_0 - m)\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)} \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \tau^{\frac{a}{2}-1} e^{-\frac{b}{2}\tau}.$$

Moments of τ

$$E[\tau] = \frac{\Phi_{a+2}\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha(a+2))}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)} \frac{a}{b}$$

$$E\left[\tau^2\right] = \frac{\Phi_{a+4}\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha(a+4))}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)} \frac{a(a+2)}{b^2}$$

We can derive all positive and negative moments of τ and therefore η.

The RTNG distribution is conjugate to the truncated normal

From standard normal gamma distribution theory, it is clear that given a sample $\mathbf{x}$ of half-normal data, and a RTNG prior, $\xi, \tau \sim \mathcal{RTNG}\left(\xi_0, m, \alpha, \frac{a}{2}, \frac{b}{2}\right)$, then the posterior distribution is

$$\begin{aligned}
\xi, \tau|\mathbf{x} &\sim \mathcal{RTNG}\left(\xi_0^\star, m^\star, \alpha^\star, \frac{a^\star}{2}, \frac{b^\star}{2}\right) \quad \text{where} \\
\xi_0^\star &= \min\{\xi_0, \mathbf{x}\} \\
m^\star &= \frac{\alpha m + n\bar{x}}{\alpha + n} \\
\alpha^\star &= \alpha + n \\
a^\star &= a + n \\
b^\star &= b + (n-1)s^2 + \frac{\alpha n}{\alpha + n}(m - \bar{x})^2
\end{aligned}$$

The limiting case posterior

Suppose that we define the improper prior $p(\xi, \tau) \propto \frac{1}{\tau}$ for $-\infty < \xi < \infty$. Then, the posterior distribution is

$$\xi|\mathbf{x} \sim \mathcal{RTNG}\left(\min\{\mathbf{x}\}, \bar{x}, n, \frac{n-1}{2}, \frac{(n-1)s^2}{2}\right).$$

Also, the posterior mean in this case is:

$$E[\xi|\mathbf{x}] = \bar{x} - \frac{s}{\sqrt{n}} \frac{n + \left(\frac{\min\{\mathbf{x}\}-\bar{x}}{s/\sqrt{n}}\right)^2}{n-1} \frac{\phi_n\left(\frac{\min\{\mathbf{x}\}-\bar{x}}{s/\sqrt{n}}\right)}{\Phi_n\left(\frac{\min\{\mathbf{x}\}-\bar{x}}{s/\sqrt{n}}\right)} \neq \hat{\xi} = \min\{\mathbf{x}\}.$$

The classical and Bayesian estimator for ξ are different.

Comparison of the Bayesian posterior mean and the MLE

Bias

We simulated 1000 samples of various sizes from $\mathcal{HN}(0,1)$. The Table compares the estimated *biases* of the Bayesian posterior mean estimator of ξ with the MLE.

n	$\hat{\xi}$	$E[\xi\|\mathbf{X}]$
5	0.217	-0.063
10	0.118	-0.011
20	0.060	-0.003
50	0.025	-0.0005
100	0.012	-0.0002
1000	0.001	2×10^{-6}

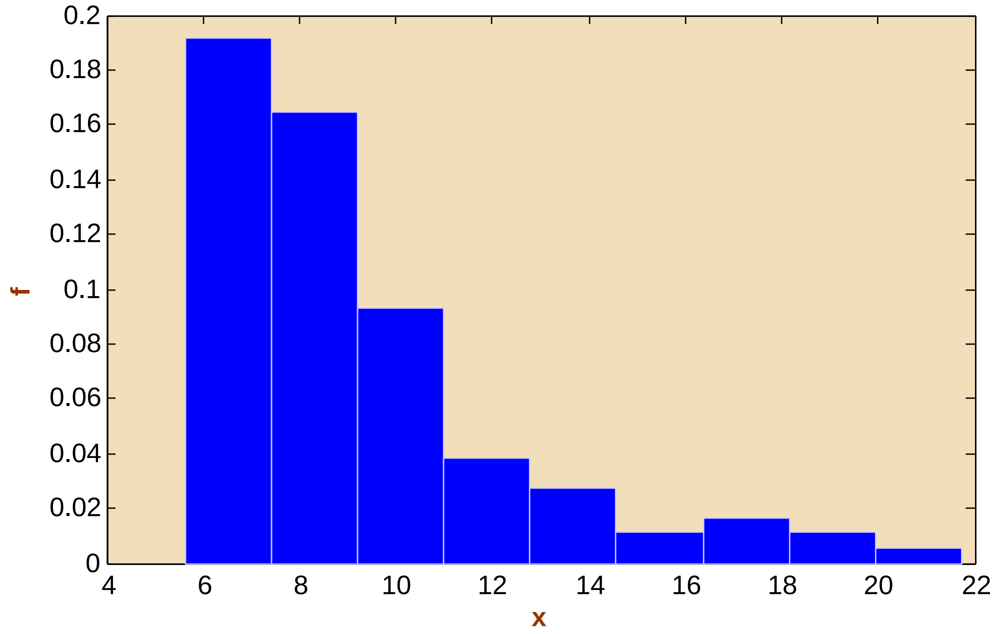
The bias of the Bayesian estimator is much lower than that of the MLE and compares with Pewsey's (2002,2004) unbiased estimator

Coverage of credible and confidence intervals

	Classical intervals				Bayesian intervals	
	Uncorrected		Bias Corrected			
n	coverage	length	coverage	length	coverage	length
5	0.880	0.585	0.904	0.655	0.950	0.910
10	0.916	0.332	0.927	0.350	0.945	0.398
20	0.937	0.177	0.943	0.181	0.951	0.192
50	0.944	0.073	0.946	0.074	0.949	0.076
100	0.948	0.037	0.948	0.037	0.950	0.038
1000	0.950	0.004	0.950	0.004	0.949	0.004

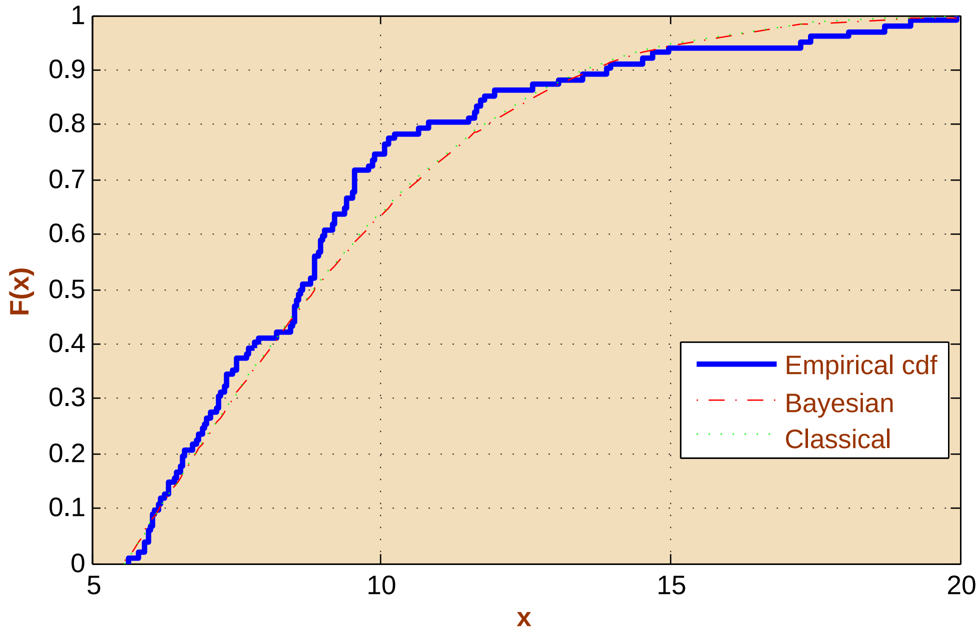
The athletes data example

Pewsey (2002,2004) analyzes data on the body fat indices of athletes.



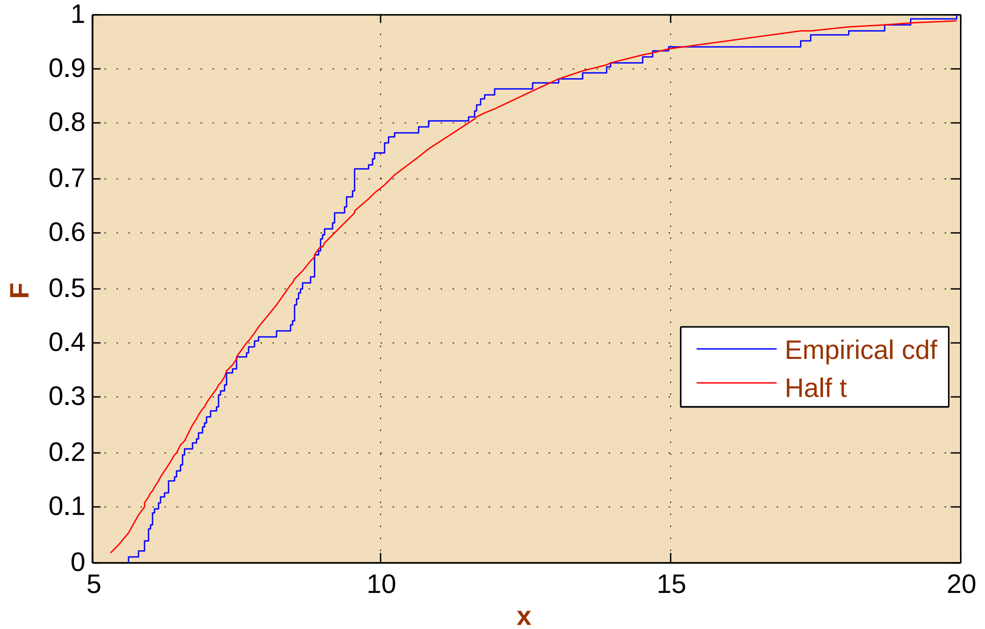
Maybe they can be well fitted by a half-normal model.

The predictive distribution function of X using Bayesian and classical methods



The fit doesn't look too good, but . . .

Extension: the half-t model



. . . results of fitting a half-t distribution model are much more promising. But that's another story.

References

Aigner, D.J., Lovell, C.A.K., Schmidt, P. (1977). Formulation and estimation of stochastic frontier production models. *Journal of Econometrics*, **6**, 21–37.

Behrens, W.-V. (1929). Ein Beitrag zur Fehlerberchnung bei Wenigen Beobachtungen, *Landw. Jb.*, **LXVIII**, 807–837.

Dobzhansky, T., Wright, S. (1947). Genetics of natural populations X. Dispersal rates in *drosophila pseudoobscura*. *Genetics*, **28**, 304–340.

Fisher, R.A. (1935). The Fiducial Argument in Statistical Inference, *Annals of Eugenics*, **VI**, 91–98.

Ghosh, B.K. (1975). A two-stage procedure for the Behrens-Fisher problem. *Journal of the American Statistical Association*, **70**, 457-462.

Haberle, J.G. (1991). Strength and failure mechanisms of unidirectional carbon fibre-reinforced plastics under axial compression. Unpublished Ph.D. thesis, Imperial College, London, U.K.

Kim, S.H. and Cohen, A.S. (1998). On the Behrens-Fisher problem: a review. *Journal of Educational and Behavioral Statistics*, **23**, 356–377.

Mackowiak, P.A., Wasserman, S.S. and Levine, M.M. (1992). A critical appraisal of 98.6 degrees F, the upper limit of the normal body temperature, and other legacies of Carl

Reinhold August Wunderlich. *The Journal of the American Medical Association*, **268**, 1578–1580.

Meeusen, W.J., van den Broeck, J. (1977). Efficiency estimation from Cobb Douglas production functions with composed error. *International Economic Review*, **8**, 435–444.

Palmer, A.R. and Strobeck, C. (1986). Fluctuating Asymmetry: Measurement, Analysis, Patterns. *Annual Review of Ecology and Systematics*, **17**, 391–421.

Patil, V.H. (1965). Approximation to the Behrens Fisher distributions. *Biometrika*, **52**, 267–71.

Pewsey, A. (2002). Large-sample inference for the general half-normal distribution. *Communications in Statistics – Theory and Methods*, **31**, 1045–1054.

Pewsey, A. (2004). Improved likelihood based inference for the general half-normal distribution. *Communications in Statistics – Theory and Methods*, **33**, 197–204.

Willink, R. (2004). Approximating the difference of two t-variables for all degrees of freedom using truncated variables. *Australian & New Zealand Journal of Statistics*, **46**, 495–504.

Wiper, M.P., Girón, F.J. and Pewsey, A. (2008). Objective Bayesian inference for the half-normal and half-t distributions. *Communications in Statistics: Theory and Methods*, **37**, 3165–3185.