Tema 3: Modelos de Inferencia para la distribución Normal

Introducción

En este tema se estudian métodos de inferencia para el modelo de la distribución normal. La importancia de este modelo se refleja en la teoría de errores, o en la aproximación, a través de la normal, de otras muchas distribuciones como resultado del Teorema Central del Límite.

NOTA: La función de densidad de una v.a. normal con parámetros μ y σ^2 ($X \sim N(\mu,\ \sigma)$) es

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

donde $-\infty < x < \infty$, siendo $E(X) = \mu$ y $Var(X) = \sigma^2$.

Modelo normal con varianza conocida

En el modelo básico se supone σ conocida, pero se desconoce μ y se desea hacer inferencias sobre μ.

Se tiene una serie de n observaciones y se formula el siguiente modelo

$$X_1, X_2, \ldots, X_n|\mu \sim N\left(\mu, \sigma_1^2\right)$$

(con σ_1 conocida) y por razones de flexibilidad y conveniencia matemática, se formula como distribución a priori de μ

$$\mu \sim N\left(\mu_0, \sigma_0^2\right).$$

Dados los datos $x_1, x_2, \ldots, x_n$, se procede del modo habitual y se calcula la distribución a posteriori como

$$
\begin{aligned}
&\pi\left(\mu \mid x_1, x_2, \ldots, x_n\right)=\frac{f\left(x_1, x_2, \ldots, x_n ; \mu\right)}{f\left(x_1, x_2, \ldots, x_n\right)}=\frac{f\left(x_1, x_2, \ldots, x_n \mid \mu\right) \pi(\mu)}{f\left(x_1, x_2, \ldots, x_n\right)} \propto \\
&\propto f\left(x_1, x_2, \ldots x_n \mid \mu\right) \pi(\mu) \propto\left(\prod_{i=1}^n \exp \left\{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma_1}\right)^2\right\}\right) \cdot \exp \left\{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right\} \propto \\
&\propto \exp \left\{-\frac{1}{2} \sum_{i=1}^n\left(\frac{x_i-\mu}{\sigma_1}\right)^2\right\} \cdot \exp \left\{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right\} \propto \\
&\propto \exp \left\{-\frac{1}{2}\left[\sum_{i=1}^n \frac{1}{\sigma_1^2}\left(x_i^2-2 x_i \mu+\mu^2\right)+\frac{1}{\sigma_0^2}\left(\mu^2-2 \mu \mu_0+\mu_0^2\right)\right]\right\} \propto \\
&\propto \exp \left\{-\frac{1}{2}\left[\sum_{i=1}^n \frac{1}{\sigma_1^2}\left(-2 x_i \mu+\mu^2\right)+\frac{1}{\sigma_0^2}\left(\mu^2-2 \mu \mu_0\right)\right]\right\} \propto \\
&\propto \exp \left\{-\frac{1}{2}\left(\mu^2\left(\frac{n}{\sigma_1^2}+\frac{1}{\sigma_0^2}\right)-2 \mu\left(\frac{\sum_{i=1}^n x_i}{\sigma_1^2}+\frac{\mu_0}{\sigma_0^2}\right)\right)\right\} \propto \\
&\propto \exp \left\{-\frac{1}{2}\left(\frac{n}{\sigma_1^2}+\frac{1}{\sigma_0^2}\right)\left(\mu^2-2 \mu\left(\frac{\frac{n}{\sigma_1^2} \bar{x}+\frac{1}{\sigma_0^2} \mu_0}{\frac{n}{\sigma_1^2}+\frac{1}{\sigma_0^2}}\right)\right)\right\},
\end{aligned}
$$

NOTA:

Si una v.a. Z se distribuye $N(m, s^2)$ entonces su función de densidad se puede escribir como

$$
f(z) \propto \exp\left[-\frac{1}{2}\cdot\frac{1}{s^2}(z-m)^2\right] \propto \exp\left[-\frac{1}{2}\cdot\frac{1}{s^2}\left(z^2-2zm\right)\right]
$$

Por tanto, observando la expresión anterior, resulta que la distribución a posteriori es una normal con media y varianza

$$
\begin{aligned}
a &= \frac{\frac{n}{\sigma_1^2}\bar{x}+\frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma_1^2}+\frac{1}{\sigma_0^2}}, \\
b^2 &= \frac{1}{\frac{n}{\sigma_1^2}+\frac{1}{\sigma_0^2}},
\end{aligned}
$$

esto es, $\mu|\mathbf{x} \sim N\left(a, b^2\right).$

Como consecuencia de todo lo anterior se verifica lo siguiente:

Teorema 1 *La distribución normal es una distribución* ***conjugada*** *respecto de la verosimilitud normal, con media desconocida y varianza conocida.*

Inferencia

Estimación puntual

En este caso, coinciden la media, la moda y la mediana a posteriori, así, como estimador puntual se puede usar la media a posteriori que es, observando la expresión anterior,

$$
\begin{aligned}
E\left[\mu|\mathbf{x}\right] &= \frac{\frac{n}{\sigma_1^2}\overline{x}+\frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma_1^2}+\frac{1}{\sigma_0^2}} = \\
&= \frac{\frac{n}{\sigma_1^2}}{\frac{n}{\sigma_1^2}+\frac{1}{\sigma_0^2}}\overline{x}+\frac{\frac{1}{\sigma_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}\mu_0,
\end{aligned}
$$

es, por tanto, una media ponderada entre la media muestral y la media a priori.

Estimación por intervalos

Sea $z_{\frac{\alpha}{2}}$ el valor, que para la normal $N\left(0,1\right)$, verifica que

$$P\left(|z|\leq z_{\frac{\alpha}{2}}\right)=1-\alpha,$$

por lo tanto, se toma como intervalo de probabilidad $1-\alpha$

$$\left[a-z_{\frac{\alpha}{2}}b,\ a+z_{\frac{\alpha}{2}}b\right],$$

donde

$$a=\frac{\frac{n}{\sigma_1^2}\overline{x}+\frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma_1^2}+\frac{1}{\sigma_0^2}}$$

y

$$b=\left(\frac{n}{\sigma_1^2}+\frac{1}{\sigma_0^2}\right)^{-\frac{1}{2}}.$$

Contraste de hipótesis

Hipótesis nulas no puntuales

Calculamos la probabilidad a posteriori de las hipótesis y nos quedamos con la que dé mayor probabilidad. Supongamos, por ejemplo, que hay dos hipótesis y éstas son

$$
\begin{aligned}
H_0 &: \mu\geq c \\
H_1 &: \mu< c
\end{aligned}
$$

Se tiene que, calculando las probabilidades a posteriori de cada una de las hipótesis,

$$
\begin{aligned}
P\left(H_0|\text{datos}\right) &= P\left(\mu\geq c|\text{datos}\right)=\int_c^\infty f\left(\mu|\text{datos}\right)d\mu=1-\Phi\left(\frac{c-a}{b}\right) \\
P\left(H_1|\text{datos}\right) &= \Phi\left(\frac{c-a}{b}\right).
\end{aligned}
$$

donde Φ es la función de distribución de una $N(0,1)$.

Así, se dice que los datos apoyan H_0 si

$$1 - \Phi\left(\frac{c-a}{b}\right) \geq \Phi\left(\frac{c-a}{b}\right)$$

o, equivalentemente,

$$\Phi\left(\frac{c-a}{b}\right) \leq \frac{1}{2}$$

y viceversa.

Hipótesis nulas de una cola

Supongamos, por ejemplo, la hipótesis de una cola

$$H_0 : \mu = c$$

$$H_1 : \mu \neq c$$

Se calcula un intervalo de probabilidad $1-\alpha$. Se dice, entonces, que los datos apoyan H_0 si el valor crítico c está dentro del intervalo. Específicamente, se dice que los datos apoyan H_0 si

$$c \in \left[a - z_{\frac{\alpha}{2}} b, a + z_{\frac{\alpha}{2}} b\right],$$

donde

$$a = \frac{\frac{n}{\sigma_1^2}\overline{x} + \frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}}$$

$$b = \left(\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}\right)^{-\frac{1}{2}}$$

o, alternativamente, si

$$\left|\frac{a-c}{b}\right| \leq z_{\frac{\alpha}{2}}.$$

Predicción

Se tiene que calcular la distribución de $X_{n+1}|\mathbf{x}$. En general, se tiene que

$$f\left(x_{n+1}|\mathbf{x}\right) = \int_{-\infty}^{\infty} f\left(x_{n+1}, \mu|\mathbf{x}\right) d\mu = \int_{-\infty}^{\infty} f\left(x_{n+1}|\mu, \mathbf{x}\right) f\left(\mu|\mathbf{x}\right) d\mu =$$
$$= \int_{-\infty}^{\infty} f\left(x_{n+1}|\mu\right) f\left(\mu|\mathbf{x}\right) d\mu.$$

En este caso, su esperanza es

$$E\left(X_{n+1}|\mathbf{x}\right) = \int_{-\infty}^{\infty} x_{n+1} f\left(x_{n+1}|\mathbf{x}\right) dx_{n+1} =$$
$$= \int_{-\infty}^{\infty} x_{n+1} \int_{-\infty}^{\infty} f\left(x_{n+1}|\mu\right) f\left(\mu|\mathbf{x}\right) d\mu dx_{n+1} =$$
$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x_{n+1} f\left(x_{n+1}|\mu\right) dx_{n+1}\right] f\left(\mu|\mathbf{x}\right) d\mu =$$
$$= \int_{-\infty}^{\infty} \mu f\left(\mu|\mathbf{x}\right) d\mu = a.$$

y de modo análogo se calcula su varianza.

$$Var\left(X_{n+1}|\mathbf{x}\right) = \sigma_1^2 + b^2.$$

Se puede demostrar, escribiendo $X_{n+1} = (X_{n+1} - \mu) + \mu$ y observando que ambos términos son variables independientes:

$$\begin{aligned}(X_{n+1} - \mu) &\sim N\left(0, \sigma_1^2\right) \\ \mu|\mathbf{x} &\sim N\left(a, b^2\right)\end{aligned}$$

de modo que

$$X_{n+1} \sim N\left(a, \left(\sigma_1^2 + b^2\right)\right).$$

donde

$$\begin{aligned}a &= \frac{\frac{n}{\sigma_1^2}\overline{x} + \frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}} \\ b^2 &= \left(\frac{n}{\sigma_1^2} + \frac{1}{\sigma_0^2}\right)^{-1}.\end{aligned}$$

Ejemplo

Supongamos que nos proponen comparar los métodos de producción de cierta sustancia. Se realizan 16 mediciones y se calcula la diferencia de rendimiento entre ambos métodos (el del primero menos el del segundo).

3,7	−6,7	−10,5	−6,1	−17,6	2,3	−7,9	−8,9
−4,5	−7,7	−9,4	−10,4	−10,9	−9,3	−16,7	−7,2

Suponemos que

$$X_i \sim N\left(\mu,\ \sigma_1^2 = 25\right),$$

donde la distribución a priori de μ es

$$\mu \sim N\left(\mu_0 = 0,\ \sigma_0^2 = 100\right).$$

Se puede observar que el intervalo de probabilidad 0.95 a priori es $(-20, 20)$, que incluye todas las observaciones.

Como tenemos $\overline{x} = -7{,}99$, $n = 16$, aplicando los resultados anteriores, se tiene

$$\begin{aligned}a &= \frac{\frac{16}{25}(-7{,}99) + \frac{1}{100}0}{\frac{16}{25} + \frac{1}{100}} = -7{,}87 \\ b^2 &= \frac{1}{\frac{16}{25} + \frac{1}{100}} = 1{,}54.\end{aligned}$$

Estimador puntual: el estimador puntual de la diferencia entre ambas medidas es la media a posteriori, esto es, $a = -7{,}87$.
Estimación por intervalos: el intervalo de probabilidad 0.95 es

$$\begin{aligned} &\left[a - z_{\frac{\alpha}{2}} b, a + z_{\frac{\alpha}{2}} b\right] = \\ &= \left[-7{,}87 - 1{,}96 \cdot \sqrt{1{,}54}; -7{,}87 + 1{,}96 \cdot \sqrt{1{,}54}\right] = \\ &= [-10{,}288; -5{,}452] \end{aligned}$$

Contraste de hipótesis: el contraste de hipótesis que se considera es:

$$H_0 : \mu \geq 0$$
$$H_1 : \mu < 0,$$

luego

$$P\left(H_0|\text{datos}\right) = P\left(\mu \geq 0\right) = P\left(\frac{\mu + 7{,}87}{\sqrt{1{,}54}} \geq \frac{7{,}87}{\sqrt{1{,}54}}\right) = P\left(Z \geq 6{,}34\right) \approx 0,$$

por tanto, se rechaza la hipótesis nula, esto es, el rendimiento del segundo es mayor.
Predicción:

$$X_{17} \sim N\left(-7{,}87;\ \left(1{,}54 + 5^2\right)\right) \equiv N\left(-7{,}87; 26{,}54\right).$$

Asignación de la distribución a priori

Se considera ahora la asignación de la distribución a priori de μ, $\mu \sim N\left(\mu_0,\ \sigma_0^2\right)$. Se necesitan dos juicios para fijar los dos parámetros. Habitualmente, se emplea un valor típico como media (que es igual a la moda) y un valor extremo con su probabilidad asociada, para asignar la desviación típica. Usualmente se hace alguna comprobación adicional para evaluar la consistencia de las asignaciones.

Ejemplo

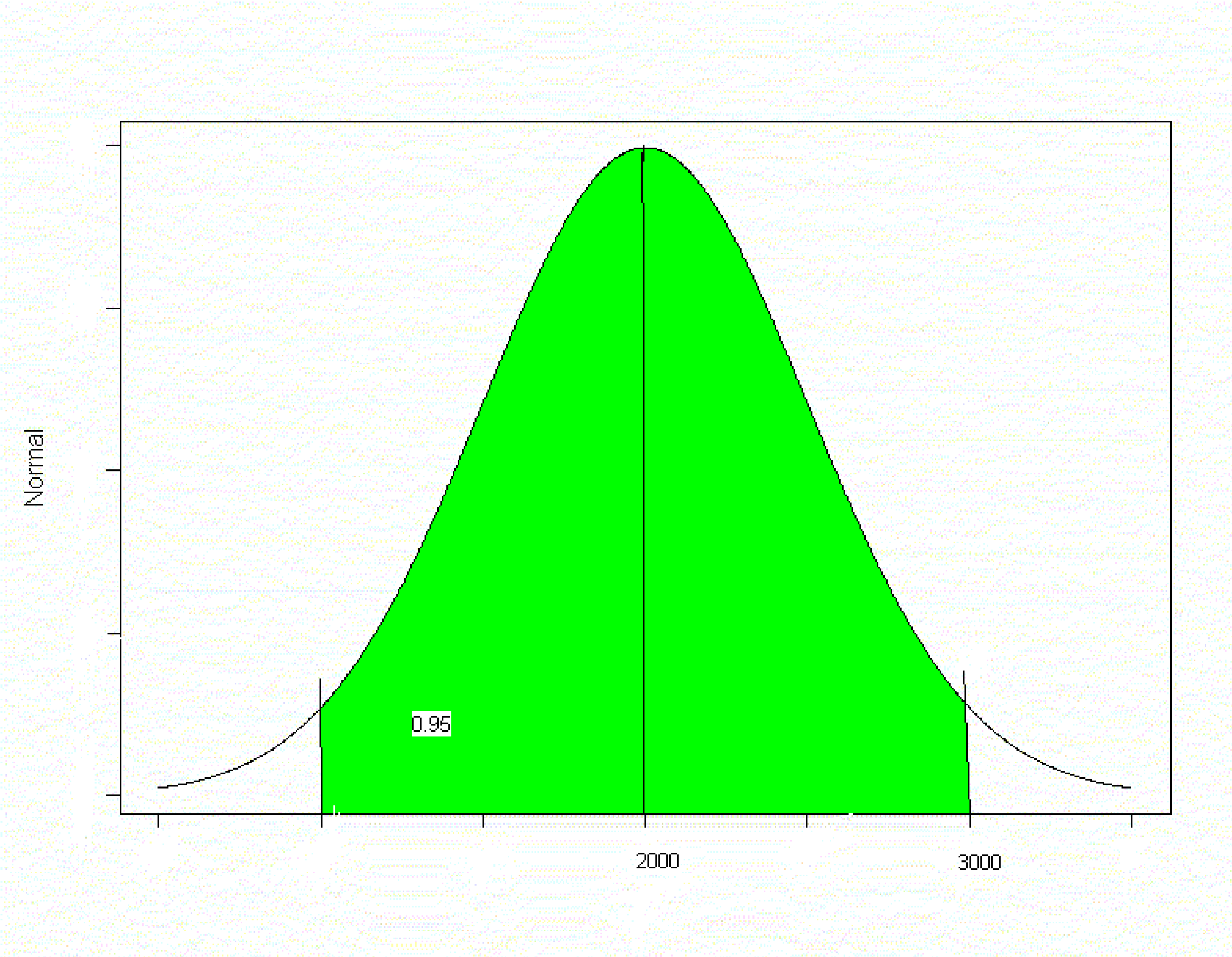
Se considera el volumen de agua que entra en un embalse durante el mes de Enero. Se consulta a un experto. Este dice que el valor típico es 2000 y nunca suele entrar más de 3000.

Hago $\mu_0 = 2000$ y

$$1{,}96\sigma_0 = 3000 - 2000$$

de modo que $\sigma_0 \approx 500$, esto es,

$$\mu \sim N\left(2000,\ \sigma_0 = 500\right).$$



Como recomendación general, cuanto mayor sea la incertidumbre (o menor sea la información disponible), σ_0 deberá ser mayor de modo que la densidad pasa a ser casi plana en la región de interés.

Distribuciones de Referencia o *no informativas*

A veces no se quiere introducir información en la distribución a priori, porque

(i) No se sabe *nada* previamente sobre el problema,

(ii) Se quiere ser *objetivo*.

En estas situaciones se tienen que elegir distribuciones a priori *no informativas*. Se pueden considerar diferentes soluciones.

El principio de la *razón insuficiente*:

Este principio (Bayes 1763, Laplace 1814), que es muy intuitivo, dice que si no hay información para diferenciar entre valores diferentes de θ, se debe dar la misma probabilidad a todos los valores. Esto implica considerar una distribución a priori uniforme para θ.

Si el soporte de θ es infinito, la distribución a priori será impropia (no integra 1) resulta ser,

$$\pi(\theta) \propto 1.$$

La crítica más importante es que la distribución uniforme no es invariante en caso de transformaciones, es decir, frente a cambios de unidades de medida, por ejemplo.

Distribuciones a priori de Jeffreys

Jeffreys introdujo una distribución a priori con la propiedad de invarianza.

Sea θ unidimensional, la distribución a priori de Jeffreys es

$$\pi(\theta) \propto \sqrt{I(\theta)}$$

donde

$$I(\theta) = \mathrm{Var}_X \left[\frac{\partial}{\partial \theta} \log f(X|\theta) \right] = \mathrm{E}_X \left[\left(\frac{\partial}{\partial \theta} \log f(X|\theta) \right)^2 \right] = -\mathrm{E}_X \left[\frac{\partial^2}{\partial \theta^2} \log f(X|\theta) \right]$$

es la información esperada de Fisher.

Si $\phi = \phi(\theta)$ y se elige $\pi(\theta) \propto \sqrt{I(\theta)}$, entonces $\pi(\phi) \propto \sqrt{I(\phi)}$.

La distribución a priori de Jeffreys es invariante en el sentido de que la inferencia no depende de la escala elegida para el parámetro. El problema es que no funciona adecuadamente para parámetros multivariantes.

En el caso de la normal se busca una distribución a priori sobre el parámetro μ; si se aplica el principio de la razón insuficiente (que coincide en este caso con el método de Jeffreys) la distribución impropia inicial de μ es

$$\pi(\mu) \propto 1, \quad \mu \in \mathbb{R}$$

de manera que la metodología clásica es equivalente a la bayesiana.

Si $\pi(\mu) \propto 1$ entonces

$$\begin{aligned}
\pi(\mu|\mathbf{x}) &\propto \pi(\mu) f(x_1, \ldots, x_n \mid \mu) \propto \\
&\propto 1 \cdot \exp \left[-\frac{1}{2\sigma_1^2} \sum_{i=1}^n (x_i - \mu)^2 \right] \\
&\propto \exp \left[-\frac{1}{2\sigma_1^2} \left(n\mu^2 - 2n\overline{x}\mu \right) \right] \\
&\propto \exp \left[-\frac{1}{2\left(\frac{\sigma_1^2}{n} \right)} (\mu - \overline{x})^2 \right],
\end{aligned}$$

resultando,

$$\mu|\mathbf{x} \sim N\left(\overline{x}, \frac{\sigma_1^2}{n} \right)$$

que no depende de la información a priori sobre μ.

Este caso equivale a considerar que $\sigma_0 \longrightarrow \infty$, ya que

$$\begin{aligned}
&\pi(\mu|\mathbf{x}) \propto f(x_1, x_2, \ldots x_n|\mu)\, \pi(\mu) \propto \\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma_1}\right)^2\right\} \cdot \exp\left\{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right\} \underset{\sigma_0\longrightarrow\infty}{\propto} \\
&\underset{\sigma_0\longrightarrow\infty}{\propto} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma_1}\right)^2\right\} \propto \exp\left[-\frac{1}{2\left(\frac{\sigma_1^2}{n}\right)}(\mu-\overline{x})^2\right]
\end{aligned}$$

Así pues, en el caso no informativo se tiene que

1. **Estimación puntual**: $\overline{x}$
2. **Estimación por intervalos**: $\left[\overline{x} - z_{\frac{\alpha}{2}}\frac{\sigma_1}{\sqrt{n}}; \overline{x} + z_{\frac{\alpha}{2}}\frac{\sigma_1}{\sqrt{n}}\right]$
3. **Contraste de hipótesis**:

   a) Caso puntual $H_0 : \mu = 0$ frente a $H_1 : \mu \neq 0$

   Los datos apoyan la hipótesis $H_0 : \mu = 0$ si
   $$0 \in \left[\overline{x} - z_{\frac{\alpha}{2}}\frac{\sigma_1}{\sqrt{n}}; \overline{x} + z_{\frac{\alpha}{2}}\frac{\sigma_1}{\sqrt{n}}\right],$$
   ó alternativamente si
   $$\left|\frac{\sqrt{n}\overline{x}}{\sigma_1}\right| \leq z_{\frac{\alpha}{2}}$$

   b) Caso $H_0 : \mu \leq 0$ frente a $H_1 : \mu > 0$.

   Los datos apoyan la hipótesis $H_0 : \mu \leq 0$ si
   $$P(\mu \leq 0|\text{ datos}) > P(\mu > 0|\text{ datos}),$$
   esto es, si
   $$\Phi\left(-\frac{\bar{x}}{\sigma_1/\sqrt{n}}\right) > \frac{1}{2}.$$

4. **Predicción**:
   $$X_{n+1}|\text{datos} \sim N\left(\overline{x},\ \sigma_1^2\left(\frac{1}{n}+1\right)\right).$$

Ejemplo

Se consideran 40 generadores nuevos y se mide su potencia

18	18	19	17	25	18	18	17	18	18	22	20	20	17	20	18	18	20	15	17
18	17	19	17	17	17	19	18	18	18	16	17	18	18	18	17	17	22	16	20

Suponemos que $\sigma_1 = 1{,}8$.

Si se asume una distribución a priori no informativa, como $\overline{x} = 18{,}25$, se tiene

$$\mu|\text{ datos} \sim N\left(18{,}25,\ \sigma^2 = \frac{1{,}8^2}{40}\right).$$

Se quiere calcular, además, la probabilidad de que el siguiente aparato tenga una potencia mayor que 20. Se tiene que

$$X_{41}|\text{datos} \sim N\left(18{,}25,\ \sigma^2 = 1{,}8^2\left(\frac{1}{40}+1\right)\right).$$

Se tiene que

$$P\left(X_{41} > 20|\text{datos}\right) = P\left(Z > \frac{20-18{,}25}{1{,}821}\right) = P\left(Z \geq 0{,}96\right) = 0{,}17.$$

Programa de Bugs:

```
rm(list=ls(all=TRUE))
ruta <- "C:/dondesea"
setwd(ruta)

#.................................................................
# PROGRAMA de BUGS
#.................................................................
cat("
# Modelo normal, varianza conocida
model
{
# tao = 1/varianza
tao <- 1/sigma2

for (i in 1:n)
{
x[i] ~ dnorm(mu,tao)
}

# A priori con varianza grande:
mu ~ dnorm(0,0.001)
}
",
file="modelo.txt")

#.................................................................

x = c(-1.10635822, 0.56352639, -1.62101846, 0.06205707, 0.50183464,
0.45905694, -1.00045360, -0.58795638, 1.01602187, -0.26987089,
0.18354493, 1.64605637, -0.96384666, 0.53842310, -1.11685831,
0.75908479, 1.10442473, -1.71124673, -0.42677894, 0.68031412)
n = length(x)
sigma2 = 1
datos = list("x","sigma2","n")
parametros = "mu"
# iniciales = function(){list(mu=rnorm(1))}

#.................................................................
```

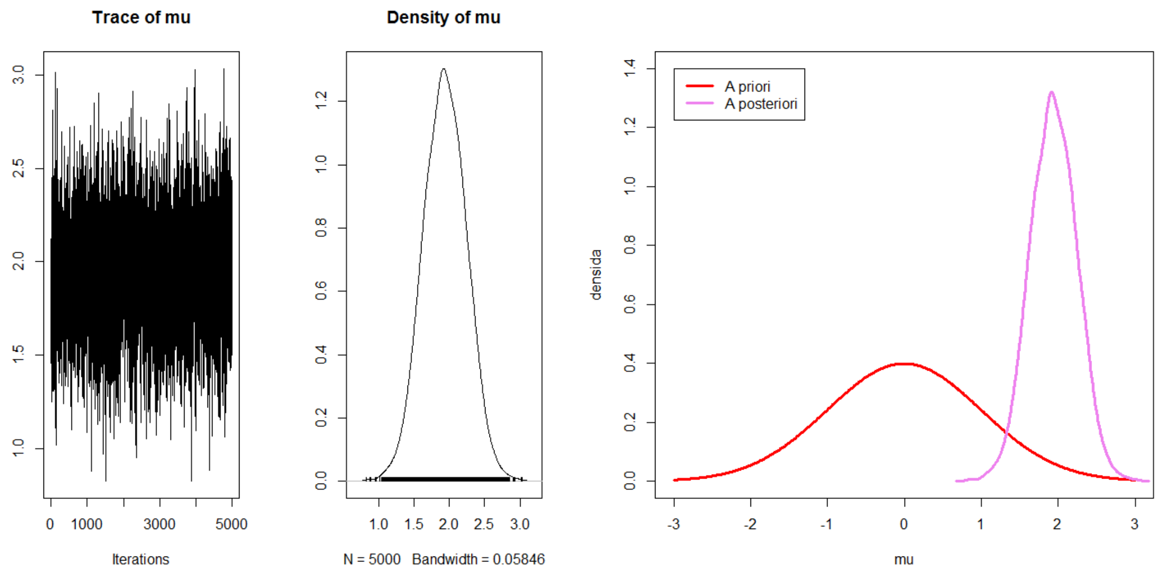
```
# Programa con BRugs
library(BRugs)
modelo <- BRugsFit(data=datos, inits=NULL,
nBurnin = 10000, nIter = 20000,
para=parametros, modelFile="modelo.txt",
numChains=3,working.directory=ruta)

print(modelo)

samplesStats("*")

samplesDensity("*", mfrow=c(1,1), col=4, main="Parametro Media",
xlab=expression(mu))

#.........................................

# Programa Original OpenBugs
library(R2OpenBUGS)

modelo <- bugs(data=datos,inits=NULL,
parameters.to.save=parametros,
model.file="modelo.txt",
n.chains=3, n.iter=20000, n.burnin=10000,
working.directory=ruta,
clearWD=TRUE, debug=TRUE)

print(modelo)

file.remove("modelo.txt")
```

Se pueden considerar también los resultados obtenidos con la librería `MCMCpack` de `R`:

```
y = c(2.65, 1.80, 2.29, 2.11, 2.27, 2.61, 2.49, 0.96, 1.72, 2.40)

library(MCMCpack)
posterior = MCnormalnormal(y, 1, 0, 1, 5000)
summary(posterior)

X11()
plot(posterior)

grid = seq(-3, 3, 0.01)
X11()
plot(grid, dnorm(grid, 0, 1), type="l", col="red", lwd=3,
ylim=c(0,1.4), xlab="mu", ylab="densida")
lines(density(posterior), col="violet", lwd=3)
legend(-3, 1.4, c("A priori", "A posteriori"), lwd=3,
col=c("red", "violet"))
```

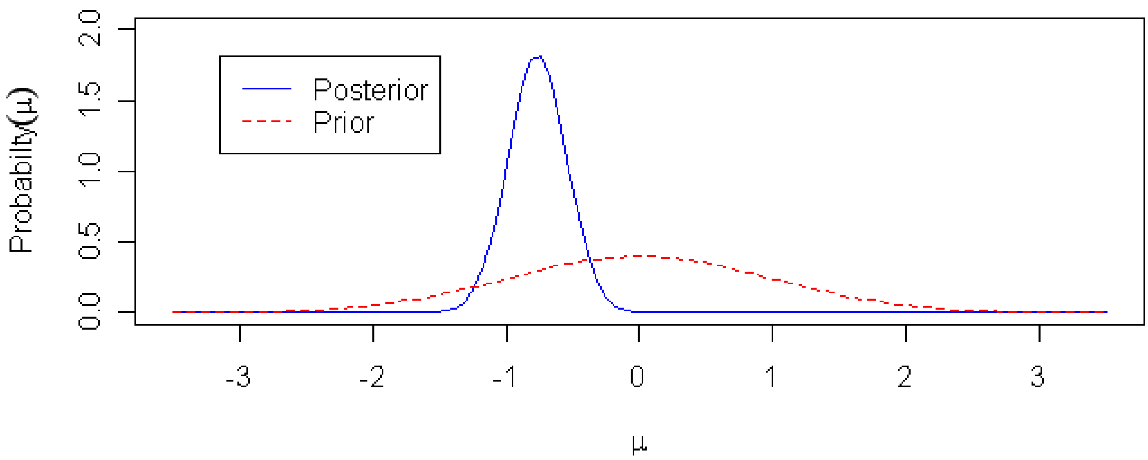
Se obtiene

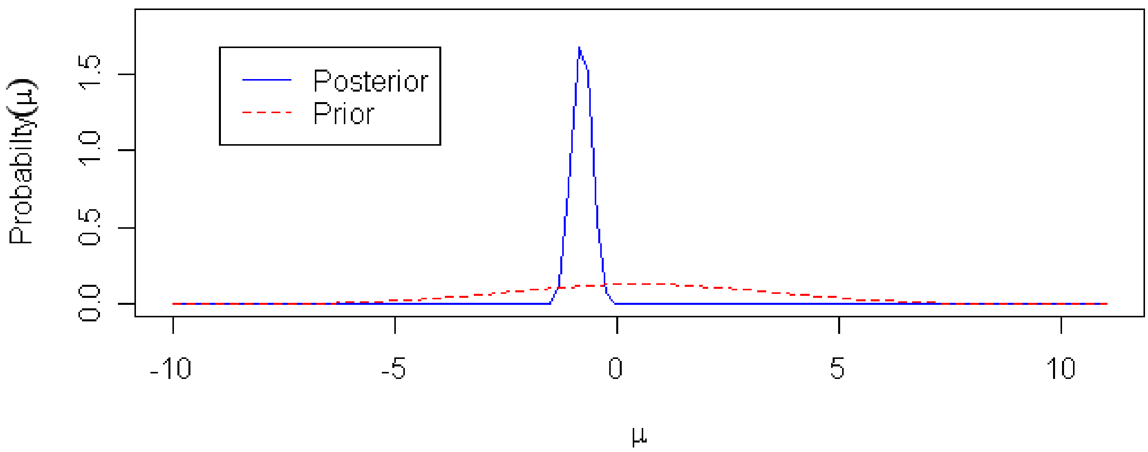
```
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:
          Mean             SD       Naive SE Time-series SE
      1.938369       0.302926       0.004284       0.004319

2. Quantiles for each variable:

 2.5%   25%   50%   75% 97.5%
1.349 1.730 1.937 2.144 2.525
```



Se pueden considerar también los resultados obtenidos con la librería `Boldstad` de `R`:

```
library(Bolstad)
# Generas una muestra de 20 observaciones de una N(-0.5,1)
x <- rnorm(20,-0.5,1)
par(mfrow=c(2,1))
# Determinas la densidad a posteriori con con una N(0,1)
# a priori sobre mu
normnp(x,1)

# Determinas la densidad a posteriori con con una N(0.5,3)
# a priori sobre mu
normnp(x,1,0.5,3)
```





Modelo con varianza desconocida y media conocida

Supongamos una v.a. distribuida como una normal con μ_0 *conocida* y σ^2 desconocida, esto es,

$$X \sim N\left(\mu_0, \sigma^2\right)$$

Se obtiene una serie de n observaciones y se formula el siguiente modelo

$$X_1, X_2, \ldots, X_n | \sigma^2 \sim N\left(\mu_0, \sigma^2\right)$$

Para determinar la distribución a priori no informativa de σ, se puede aplicar la regla de Jeffreys,

$$f(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_0)^2}{2\sigma^2}\right)$$

$$\begin{aligned}
\log f(x|\sigma) &= -\log\sqrt{2\pi} - \log\sigma - \frac{(x-\mu_0)^2}{2\sigma^2} \\
\frac{\partial}{\partial\sigma}\log f(x|\sigma) &= -\frac{1}{\sigma} + \frac{2(x-\mu_0)^2}{2\sigma^3} \\
\frac{\partial^2}{\partial(\sigma)^2}\log f(x|\sigma) &= \frac{1}{\sigma^2} - \frac{3(x-\mu_0)^2}{\sigma^4}
\end{aligned}$$

$$\begin{aligned}
I(\sigma) &= -E_X\left[\frac{\partial^2}{\partial(\sigma)^2}\log f(x|\sigma)\right] = -\frac{1}{\sigma^2} + \frac{3E[x-\mu_0]^2}{\sigma^4} = -\frac{1}{\sigma^2} + \frac{3\sigma^2}{\sigma^4} \\
&= \frac{2}{\sigma^2},
\end{aligned}$$

por tanto,

$$\pi(\sigma) \propto \sqrt{I(\sigma)} \propto \frac{1}{\sigma}$$

y, aplicando el cambio de variable, se puede tomar como distribución inicial de referencia sobre σ^2,

$$\pi\left(\sigma^2\right) \propto \frac{1}{\sigma^2}.$$

En este caso, la verosimilitud de una m.a.s. de tamaño n se puede reexpresar como

$$\begin{aligned}
f\left(x_1, \ldots, x_n \mid \sigma^2\right) &\propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_0}{\sigma}\right)^2\right\} \\
&\propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{nS^2}{2}\frac{1}{\sigma^2}\right]
\end{aligned}$$

donde

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2,$$

de modo que la distribución a posteriori de σ^2 será:

$$\pi\left(\sigma^2 \mid x_1,\ldots,x_n\right) \propto \pi\left(\sigma^2\right)\cdot f\left(x_1,\ldots,x_n \mid \sigma^2\right) \propto$$
$$\propto \left(\sigma^2\right)^{-\frac{n}{2}-1}\exp\left[-\frac{nS^2}{2}\frac{1}{\sigma^2}\right],$$

$$\sigma^2 \mid \mathbf{x} \sim \chi^{-2}_{(n,nS^2)}.$$

es una χ^2-inversa con n grados de libertad con parámetro de escala igual a nS^2.

Tenemos, por tanto, que

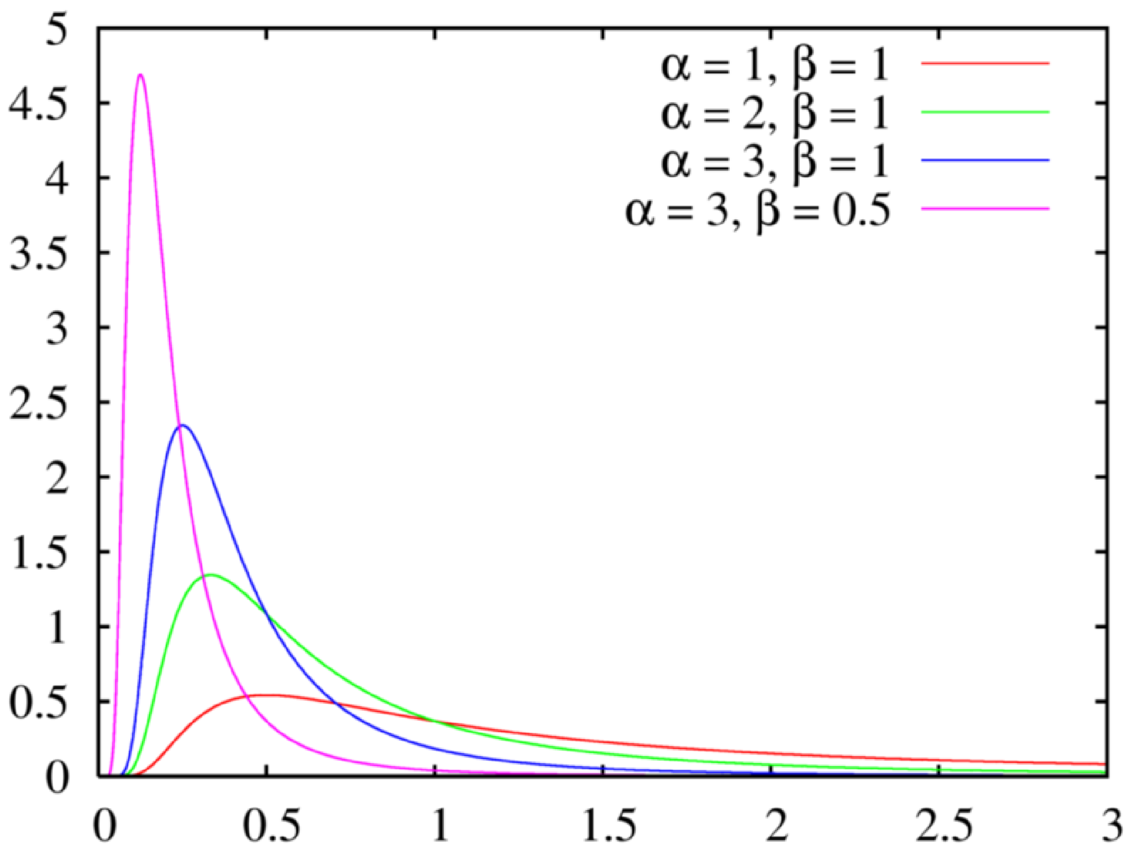
$$\frac{nS^2}{\sigma^2} \mid \mathbf{x} \sim \chi^2_n$$

Observación: En Inferencia clásica se deduce que una cantidad **pivotal** para σ^2 es

$$\frac{nS^2}{\sigma^2} \mid \sigma^2 \sim \chi^2_n$$

NOTAS:

Distribución gamma-inversa: Una v.a. $X \sim \mathrm{Inv}-\mathrm{Gam}(\alpha,\beta)$ si su función de densidad es



Si una v.a. X se distribuye como una distribución gamma, entonces $\dfrac{1}{X}$ se distribuye como una gamma inversa.

Distribución chi-cuadrado: Una v.a. X se distribuye como una chi cuadrado, $X \sim \chi^2_k$, si su función densidad es

$$f(x|k) = \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} \exp\left[-\frac{x}{2}\right]$$

es decir equivale a una densidad Gamma $\left(\frac{k}{2}, \frac{1}{2}\right)$.

Así, si una v.a. X se distribuye como una distribución chi cuadrado, entonces $Y = \frac{1}{X}$ se distribuye como una *chi cuadrado inversa*:

$$f(y|k) = \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma(\frac{k}{2})} y^{-\frac{k}{2}-1} \exp\left[-\frac{1}{2y}\right]$$

A su vez, la distribución chi cuadrado inversa escalada $\chi^{-2}_{(k,k\sigma^2)}$ tiene como función de densidad

$$f(x|k,\sigma) = \frac{\left(\frac{k\sigma^2}{2}\right)^{\frac{k}{2}}}{\Gamma(\frac{k}{2})} x^{-\frac{k}{2}-1} \exp\left[-\frac{k\sigma^2}{2x}\right]$$

Si una v.a. X se distribuye como una distribución chi cuadrado inversa escalada por $k\sigma^2$, entonces $Y = \dfrac{k\sigma^2}{X}$ se distribuye como una chi cuadrado.

Mira en:

```
http://en.wikipedia.org/wiki/Inverse-gamma_distribution

http://en.wikipedia.org/wiki/Inverse-chi-square_distribution

http://en.wikipedia.org/wiki/Scale-inverse-chi-square_distribution
```

Distribución inicial conjugada para σ^2

Tomamos como distribución inicial sobre σ^2 una χ^2-inversa escalada con parámetros conocidos $(\nu_0, \nu_0 S_0^2)$, es decir, alternativamente se puede decir que

$$\frac{\nu_0 S_0^2}{\sigma^2} \sim \chi^2_{\nu_0}$$

cuya función de densidad es:

$$\pi\left(\sigma^2\right) \propto \left(\sigma^2\right)^{-\frac{\nu_0}{2}-1} \exp\left[-\frac{\nu_0 S_0^2}{2}\frac{1}{\sigma^2}\right].$$

La verosimilitud de la muestra observada, como función de σ^2, es proporcional a

$$f\left(x_1, \ldots, x_n \mid \sigma^2\right) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{nS^2}{2}\frac{1}{\sigma^2}\right]$$

donde

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu_0\right)^2.$$

Por tanto, la distribución a posteriori de σ^2, cuando tomamos como distribución inicial una χ^2 inversa escalada con parámetros $(\nu_0, \nu_0 S_0^2)$, es:

$$\begin{aligned}
\pi\left(\sigma^2 \mid x_1, \ldots, x_n\right) &\propto \pi\left(\sigma^2\right) f\left(x_1, \ldots, x_n \mid \sigma^2\right) \propto \\
&\propto \left(\sigma^2\right)^{-\frac{\nu_0}{2}-1} \exp\left[-\frac{\nu_0 S_0^2}{2}\frac{1}{\sigma^2}\right] \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{nS^2}{2}\frac{1}{\sigma^2}\right] \\
&\propto \left(\sigma^2\right)^{-\frac{(\nu_0+n)}{2}-1} \exp\left[-\frac{\left(\nu_0 S_0^2 + nS^2\right)}{2}\frac{1}{\sigma^2}\right].
\end{aligned}$$

de modo que

$$\sigma^2 \mid \mathbf{x} \sim \chi^{-2}_{\left(\nu_1, \nu_1 S_1^2\right)}$$

donde

$$\nu_1 = \nu_0 + n$$

$$\nu_1 S_1^2 = \nu_0 S_0^2 + nS^2,$$

siendo la distribución a posteriori de σ^2 condicionada por la observación de la muestra, una χ^2-inversa escalada con ν_1 grados de libertad y escalada por $\nu_1 S_1^2$, ó equivalentemente,

$$\frac{\nu_1 S_1^2}{\sigma^2} \mid \mathbf{x} \sim \chi^2_{\nu_1}$$

Modelo con varianza y media desconocidas

Supongamos una v.a. distribuida como una normal con parámetros μ y σ^2 desconocidos, esto es,

$$X \sim N\left(\mu, \sigma^2\right).$$

Se obtiene una serie de n observaciones y se formula el siguiente modelo

$$X_1, X_2, \ldots, X_n | \mu, \sigma^2 \sim N\left(\mu, \sigma^2\right)$$

La verosimilitud es, así,

$$\begin{aligned} L\left(\mu, \sigma^2; x_1, \ldots, x_n\right) &= f\left(x_1, \ldots, x_n \mid \mu, \sigma^2\right) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2\right] \end{aligned}$$

Distribución inicial conjunta no informativa para μ y σ^2

Si no tenemos ninguna información sobre los parámetros μ y σ^2 tomamos una distribución inicial conjunta no informativa sobre ellos, en la que suponemos que las distribuciones de cada parámetro son independientes, es decir,

$$\pi\left(\mu, \sigma^2\right) = \pi\left(\mu\right) \cdot \pi\left(\sigma^2\right)$$

(i) Como μ es un parámetro de localización, podemos tomar como distribución inicial sobre μ

$$\pi\left(\mu\right) \propto 1$$

(ii) Como σ es un parámetro de escala, podemos tomar como distribución inicial sobre σ

$$\pi\left(\sigma\right) \propto \frac{1}{\sigma},$$

que, aplicando el cambio de variable, resulta: $\pi\left(\sigma^2\right) \propto \frac{1}{\sigma^2}$.

Así, tenemos como distribución inicial conjunta sobre μ y σ^2 (que no coincide en realidad con la dada por la regla de Jeffreys),

$$\pi\left(\mu, \sigma^2\right) \propto \frac{1}{\sigma^2}$$

La distribución a posteriori conjunta de μ y σ^2, es

$$\begin{aligned} \pi\left(\mu, \sigma^2 \mid x_1, \ldots, x_n\right) &\propto \pi\left(\mu, \sigma^2\right) \cdot L\left(\mu, \sigma^2; x_1, \ldots, x_n\right) \\ &\propto \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2\right] \\ &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2\right]. \end{aligned}$$

Desarrollando el término $\sum_{i=1}^{n}(x_i-\mu)^2$ tenemos:

$$\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}\left((x_i-\overline{x})+(\overline{x}-\mu)\right)^2=$$

$$\sum_{i=1}^{n}\left[(x_i-\overline{x})^2+(\overline{x}-\mu)^2+2(\overline{x}-\mu)(x_i-\overline{x})\right]$$

$$=\sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\mu)^2+2(\overline{x}-\mu)\sum_{i=1}^{n}(x_i-\overline{x})$$

$$=(n-1)s^2+n(\overline{x}-\mu)^2$$

(ya que $2(\overline{x}-\mu)\sum_{i=1}^{n}(x_i-\overline{x})=0$) donde se define como s^2 a la cuasivarianza muestral.

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

Resulta así que

$$\pi\left(\mu,\sigma^2\mid x_1,\ldots,x_n\right)\propto\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1}\exp\left[-\frac{1}{2\sigma^2}\left[(n-1)s^2+n(\overline{x}-\mu)^2\right]\right].$$

y la expresión anterior se puede expresar de forma que,

$$\pi\left(\mu,\sigma^2\mid\overline{x},s^2\right)\propto\left(\frac{1}{\sigma^2}\right)^{\frac{n-1}{2}+1}\exp\left[-\frac{(n-1)s^2}{2}\frac{1}{\sigma^2}\right]\cdot\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}}\exp\left[-\frac{1}{2}\frac{(\mu-\overline{x})^2}{\sigma^2/n}\right]$$

$$\propto\pi\left(\sigma^2\mid s^2\right)\cdot\pi\left(\mu\mid\sigma^2,\overline{x}\right)$$

Se obtiene que la distribución a posteriori marginal de σ^2 condicionada por la muestra, a través del estadístico s^2, será:

$$\sigma^2\mid s^2\sim\chi^{-2}_{(n-1,(n-1)s^2)}$$

es decir, una χ^2-inversa con $n-1$ grados de libertad y escalada por $(n-1)s^2$, o alternativamente,

$$\frac{(n-1)s^2}{\sigma^2}\mid s^2\sim\chi^2_{n-1}.$$

La distribución a posteriori de $\mu\mid\sigma^2$, condicionada por la muestra a través del estadístico $\overline{x}$, será:

$$\mu\mid\sigma^2,\overline{x}\sim N\left(\overline{x},\frac{\sigma^2}{n}\right),$$

o alternativamente,

$$\frac{\mu-\overline{x}}{\sigma/\sqrt{n}}\mid\sigma,\overline{x}\sim N(0,1).$$

Observación: El resultado de Inferencia Clásica asegura la existencia de una cantidad pivotal para σ^2

$$\frac{(n-1)s^2}{\sigma^2}\mid\sigma^2\sim\chi^2_{n-1}$$

y para σ,μ

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\mid\sigma,\mu\sim N(0,1).$$

Distribución marginal de $\mu \mid x_1, \ldots, x_n$

La distribución marginal de $\mu \mid x_1, \ldots, x_n$ es

$$\begin{aligned}
\pi\left(\mu \mid x_1, \ldots, x_n\right) &\propto \int_0^{\infty} \pi\left(\mu, \sigma^2 \mid x_1, \ldots, x_n\right) d\sigma^2 \\
&\propto \int_0^{\infty} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} \exp\left[-\frac{1}{2\sigma^2}\left[(n-1)\, s^2 + n\left(\overline{x}-\mu\right)^2\right]\right] d\sigma^2.
\end{aligned}$$

Como sabemos que, a partir de la definición de la función de densidad de una gamma inversa,

$$\int_0^{\infty} e^{-a\left[\frac{1}{x}\right]} \cdot \left[\frac{1}{x}\right]^{p+1} dx = \frac{\Gamma\left(p\right)}{a^p},$$

denotando en la expresión anterior, como

$$a = \frac{(n-1)\, s^2 + n\left(\overline{x}-\mu\right)^2}{2}$$

$$p = \frac{n}{2}$$

resulta que

$$\begin{aligned}
\pi\left(\mu \mid x_1, \ldots, x_n\right) &\propto \Gamma\left(\frac{n}{2}\right) \cdot \left(\frac{(n-1)\, s^2 + n\left(\overline{x}-\mu\right)^2}{2}\right)^{-\frac{n}{2}} \\
&\propto \left[(n-1)\, s^2 + n\left(\mu-\overline{x}\right)^2\right]^{-\frac{n}{2}}
\end{aligned}$$

Esta distribución está relacionada con la distribución t de Student. Basta considerar el siguiente cambio de variable:

$$t = \frac{\mu-\overline{x}}{\frac{s}{\sqrt{n}}}$$

obteniéndose

$$\mu = \frac{s}{\sqrt{n}} t + \overline{x}$$

$$\begin{aligned}
\pi\left(t \mid \overline{x}, s\right) &\propto \pi\left(\mu\left(t\right)\right) \cdot \left|\frac{d\mu\left(t\right)}{dt}\right| \\
&\propto \left[(n-1)\, s^2 + n\left(t\frac{s}{\sqrt{n}} + \overline{x} - \overline{x}\right)^2\right]^{-\frac{n}{2}} \frac{s}{\sqrt{n}} \\
&\propto \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}
\end{aligned}$$

que es el núcleo de la función de densidad de una distribución t de Student con $n-1$ grados de libertad. Es decir, si μ es la variable aleatoria y la muestra es $\mathbf{x} = \left(x_1, \ldots, x_n\right)$,

$$\frac{\mu-\overline{x}}{\frac{s}{\sqrt{n}}} \mid \overline{x}, s \sim t_{n-1}$$

Un intervalo para μ de probabilidad $1-\alpha$ será, así,

$$P\left\{-t_{n-1,\frac{\alpha}{2}} \leq \frac{\mu-\overline{x}}{\frac{s}{\sqrt{n}}} \leq t_{n-1,\frac{\alpha}{2}} \mid \overline{x}, s\right\} = 1-\alpha$$

es decir,

$$I.C._{1-\alpha}(\mu) = \left[\overline{x} - t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \overline{x} + t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$$

Distribución inicial conjunta conjugada para μ y σ^2

Se puede definir una distribución conjugada para ambos parámetros, tomando como distribución inicial conjunta sobre μ y σ^2

$$\pi\left(\mu, \sigma^2\right) = \pi\left(\sigma^2\right) \cdot \pi\left(\mu \mid \sigma^2\right)$$

donde σ^2 sigue una distribución χ^2-inversa escalada con parámetros $(\nu_0, \nu_0 s_0^2)$,

$$\frac{\nu_0 s_0^2}{\sigma^2} \sim \chi^2_{\nu_0}$$

es decir,

$$\pi\left(\sigma^2\right) \propto \left(\sigma^2\right)^{-\frac{\nu_0}{2}-1} \exp\left[-\frac{\nu_0 s_0^2}{2}\frac{1}{\sigma^2}\right].$$

A su vez se toma para $\mu \mid \sigma^2$ una distribución normal con parámetros (μ_0, n_0),

$$\mu \mid \sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right) \Longrightarrow \frac{\mu-\mu_0}{\sigma/\sqrt{n_0}} \mid \sigma \sim N(0,1)$$

es decir,

$$\pi\left(\mu \mid \sigma^2\right) \propto \left(\sigma^2\right)^{-\frac{1}{2}} \exp\left[-\frac{n_0}{2\sigma^2}(\mu-\mu_0)^2\right].$$

Entonces distribución inicial conjunta sobre μ y σ^2 es

$$\begin{aligned}
\pi\left(\mu, \sigma^2\right) &= \pi\left(\sigma^2\right) \cdot \pi\left(\mu \mid \sigma^2\right) \\
&\propto \left(\sigma^2\right)^{-\frac{\nu_0}{2}-1} \exp\left[-\frac{\nu_0 s_0^2}{2}\frac{1}{\sigma^2}\right] \cdot \left(\sigma^2\right)^{-\frac{1}{2}} \exp\left[-\frac{n_0}{2\sigma^2}(\mu-\mu_0)^2\right] \\
&\propto \left(\sigma^2\right)^{-\frac{\nu_0+1}{2}-1} \exp\left[-\frac{1}{2\sigma^2}\left[\nu_0 s_0^2 + n_0(\mu-\mu_0)^2\right]\right].
\end{aligned}$$

Se puede definir como una distribución *normal gamma inversa* de parámetros $(\nu_0, s_0^2, \mu_0, n_0)$.

Por otro lado, la verosimilitud de la muestra observada será proporcional, como función de μ y σ^2, a

$$f\left(x_1, \ldots, x_n \mid \mu, \sigma^2\right) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\overline{x}-\mu)^2\right]\right]$$

con

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

También, se puede denotar como

$$Q_0(\mu) = \nu_0 s_0^2 + n_0 (\mu - \mu_0)^2 = n_0\mu^2 - 2n_0\mu_0\mu + \left(\nu_0 s_0^2 + n_0\mu_0^2\right).$$

Luego la distribución a posteriori conjunta de μ y σ^2, cuando tomamos como distribución a priori una *normal gamma inversa* de parámetros $(\nu_0 + 1, Q_0(\mu))$, será:

$$\begin{aligned}
&\pi\left(\mu, \sigma^2 \mid x_1, \ldots, x_n\right) \propto \pi\left(\mu, \sigma^2\right) f\left(x_1, \ldots, x_n \mid \mu, \sigma^2\right) \\
&\propto \left(\sigma^2\right)^{-\frac{\nu_0+1}{2}-1} \exp\left[-\frac{1}{2\sigma^2}\left[\nu_0 s_0^2 + n_0(\mu-\mu_0)^2\right]\right] \cdot \\
&\qquad \cdot \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\overline{x}-\mu)^2\right]\right] \\
&\propto \left(\sigma^2\right)^{-\frac{\nu_0+n+1}{2}-1} \exp\left[-\frac{1}{2\sigma^2}\left[\nu_0 s_0^2 + n_0(\mu-\mu_0)^2 + (n-1)s^2 + n(\overline{x}-\mu)^2\right]\right] \\
&\propto \left(\sigma^2\right)^{-\frac{\nu_0+n+1}{2}-1} \exp\left[-\frac{1}{2\sigma^2}\left[Q_0(\mu) + (n-1)s^2 + n(\overline{x}-\mu)^2\right]\right].
\end{aligned}$$

Desarrollando respecto de μ el corchete de la exponencial, resulta:

$$\begin{aligned}
Q_1(\mu) &= Q_0(\mu) + (n-1)s^2 + n(\overline{x}-\mu)^2 \\
&= (n_0+n)\mu^2 - 2(n_0\mu_0 + n\overline{x})\mu + \left(\nu_0 s_0^2 + n_0\mu_0^2 + (n-1)s^2 + n\overline{x}^2\right),
\end{aligned}$$

que tiene la misma forma que la distribución conjunta inicial sobre μ y σ^2: una *normal gamma inversa* de parámetros $(\nu_1, s_1^2, \mu_1, n_1)$, cuya expresión es

$$\begin{aligned}
\nu_1 &= \nu_0 + n \\
n_1 &= n_0 + n \\
\mu_1 &= \frac{n_0\mu_0 + n\overline{x}}{n_0+n} = \frac{n_0}{n_0+n}\mu_0 + \frac{n}{n_0+n}\overline{x}
\end{aligned}$$

$$\begin{aligned}
s_1^2 &= \frac{\nu_0 s_0^2 + n_0\mu_0^2 + (n-1)s^2 + n\overline{x}^2 - n_1\mu_1^2}{\nu_1} = \\
&= \frac{1}{\nu_0+n}\left[\nu_0 s_0^2 + n_0\mu_0^2 + (n-1)s^2 + n\overline{x}^2 - n_1\mu_1^2\right] \\
&= \frac{1}{\nu_0+n}\left[\nu_0 s_0^2 + (n-1)s^2 + \frac{n_0 n}{n_0+n}(\overline{x}-\mu_0)^2\right]
\end{aligned}$$

(i) La distribución a posteriori marginal de σ^2 condicionada por la muestra, $\sigma^2 \mid x_1, \ldots, x_n$, a través del estadístico s_1^2 (que es función de $\overline{x}$ y s^2) es

$$\sigma^2 \mid x_1, \ldots, x_n \sim \chi^{-2}_{\left(\nu_1, \nu_1 s_1^2\right)},$$

es decir, una χ^2-inversa con ν_1 grados de libertad y escalada por $\nu_1 s_1^2$:

$$\frac{\nu_1 s_1^2}{\sigma^2} \mid s_1^2 \sim \chi^2_{\nu_1}$$

(ii) La distribución a posteriori condicional de $\mu \mid \sigma^2$, condicionada por la muestra, será:

$$\mu \mid \sigma^2, \overline{x} \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right)$$

$$\frac{\mu - \mu_1}{\sigma/\sqrt{n_1}} \mid \sigma, \overline{x} \sim N(0,1)$$

Teorema. La distribución conjunta sobre μ y σ^2, normal gamma inversa de parámetros (ν, s^2, μ, n) es conjugada respecto de la verosimilitud proveniente de una distribución normal de parámetros μ y σ^2 desconocidos.

Por otro lado, como

$$\frac{\mu - \mu_1}{\sigma/\sqrt{n_1}} \mid \sigma, \overline{x} \sim N(0,1)$$

$$\frac{\nu_1 s_1^2}{\sigma^2} \mid s_1^2 \sim \chi^2_{\nu_1}$$

y se tiene por definición que

$$\frac{N(0,1)}{\sqrt{\frac{\chi^2_\nu}{\nu}}} \sim t_\nu,$$

resulta que

$$\frac{\frac{\mu - \mu_1}{\sigma/\sqrt{n_1}}}{\sqrt{\frac{\nu_1}{\nu_1}\frac{s_1^2}{\sigma^2}}} = \frac{\mu - \mu_1}{s_1/\sqrt{n_1}} \mid \overline{x}, s \sim t_{\nu_1}.$$

NOTA:

Cuando el tamaño muestral n es grande ($n \geq 50$) la distribución t de *Student* se aproxima a la normal. Así, en las aplicaciones básicas se puede hacer:

(i) Para n suficientemente grande ($n \geq 50$), se sustituye σ^2 por s^2 y se procede como en el modelo básico con μ desconocida y σ^2 fijada.

(ii) Para $n < 50$, se estima σ^2 mediante $s^2\left(1 + \frac{20}{n^2}\right)^2$.

Comparación de dos medias

Se considera ahora el problema de comparación de dos medias suponiendo conocida la varianza. Las hipótesis que se asumen son las siguientes:

1. Las muestras son independientes

2. Las distribuciones a priori son independientes.

Como consecuencia de las hipótesis anteriores, las distribuciones a posteriori son también independientes (y normales).

En el caso de la normal se tiene que si

$$\begin{aligned} Y_1 &\sim N\left(m_1,\ \sigma_1^2\right), \\ Y_2 &\sim N\left(m_2,\ \sigma_2^2\right), \end{aligned}$$

y son independientes, entonces

$$Y_1 - Y_2 \sim N\left(m_1 - m_2,\ \sigma_1^2 + \sigma_2^2\right).$$

Se ilustra el procedimiento con un ejemplo.

Ejemplo

Comparamos las dos muestras de los generadores estudiados antes. Suponemos distribuciones a priori no informativas y varianzas desconocidas. Los datos obtenidos son los siguientes:

	n	$\overline{x}$	s
1	40	21.14	4.03
2	28	18.25	1.79

Se estiman los valores de σ_i^2 mediante $s_i^2\left(1+\dfrac{20}{n_i^2}\right)^2$:

$$\begin{aligned} \sigma_1^2 &= 16{,}65 \Longrightarrow \frac{\sigma_1}{\sqrt{n_1}} = 0{,}645 \\ \sigma_2^2 &= 3{,}37 \Longrightarrow \frac{\sigma_2}{\sqrt{n_2}} = 0{,}347 \end{aligned}$$

Tenemos entonces,

$$\begin{aligned} \mu_1|\text{datos} &\sim N\left(21{,}14,\ \sigma = 0{,}645\right) \\ \mu_2|\text{datos} &\sim N\left(18{,}25,\ \sigma = 0{,}347\right), \end{aligned}$$

siendo independientes.

Nos preguntamos si son diferentes ambos generadores.

Se tiene que, como $\sqrt{0{,}645^2 + 0{,}347^2} = 0{,}732$,

$$d = \mu_1 - \mu_2 | \text{datos} \sim N\left(2{,}89,\ \sigma = 0{,}732\right).$$

Se contrasta, entonces,

$$H_0 : d \geq 0$$
$$H_1 : d < 0.$$

$$\begin{aligned} P\left(d \geq 0 | \text{datos}\right) = P\left(\frac{d - 2{,}89}{0{,}732} \geq -3{,}94\right) = \\ = P\left(z \geq -3{,}94\right) = 0{,}9999 > P\left(d < 0 | \text{datos}\right), \end{aligned}$$

de modo que $P\left(d < 0 | \text{datos}\right) = 0{,}0001$.

La información disponible apoya que los generadores de tipo 1 son más potentes.

Contrastamos ahora,

$$H_0 : d = 0$$
$$H_1 : d \neq 0.$$

Calculamos un intervalo a posteriori de probabilidad 0.99, es decir, $z_{\frac{\alpha}{2}} = 2{,}58$

$$[2{,}89 - 2{,}58 \cdot 0{,}732; 2{,}89 + 2{,}58 \cdot 0{,}732] = [1{,}001;\ 4{,}779],$$

luego, como $0 \notin [1{,}001;\ 4{,}779]$, la información disponible no apoya la hipótesis de que los generadores son iguales.

Ejemplo

Se toman 28 observaciones de un generador:

25	20	25	18	24	26	24	24	24	24	18	18	17	17
21	15	13	12	25	19	18	24	25	18	25	24	24	25

Se quiere calcular $P(\mu \geq 20)$ y $P(X_{29} \geq 20)$.

Suponemos que $X_1, \ldots, X_{28} \sim N(\mu,\ \sigma)$ y asignamos una distribución no informativa para μ. Por tanto, $\mu|\text{datos} \sim N\left(\overline{x}, \frac{\sigma^2}{n}\right)$.

Sustituimos σ^2 por $s^2\left(1+\frac{20}{n^2}\right)^2$ y, como $\overline{x} = 21{,}14$ y $s^2 = 16{,}86$, se tiene que

$$\begin{aligned} \mu|\text{datos} &\sim N\left(21{,}14,\ \sigma^2 = \frac{16{,}86\cdot\left(1+\frac{20}{28^2}\right)^2}{28}\right) = \\ &\sim N\left(21{,}14,\ \sigma^2 = \frac{17{,}73}{28}\right) \\ &N\left(21{,}14,\ \sigma^2 = 0{,}63\right). \end{aligned}$$

Por tanto,

$$\begin{aligned} P(\mu \geq 20|\text{datos}) &= P\left(Z \geq \frac{20-21{,}14}{\sqrt{0{,}63}}\right) = \\ P(Z \geq -1{,}44) &= P(Z \leq 1{,}44) = 0{,}93. \end{aligned}$$

Además,

$$\begin{aligned} X_{29}|\text{datos} &\sim N\left(21{,}14,\ \sigma^2 = 17{,}73 + 0{,}63\right) = \\ &N\left(21{,}14,\ \sigma^2 = 18{,}36\right) \end{aligned}$$

y

$$P(X_{29} \geq 20|\text{datos}) = 0{,}61,$$

valor que se puede calcular usando la siguiente sentencia de `R`:

```
1 - pnorm(20, 21.14, sqrt(18.36))
```

Programa de Bugs:

```
rm(list=ls(all=TRUE))
ruta <- "C:/dondesea"
setwd(ruta)

#.................................................................
# PROGRAMA de BUGS
#.................................................................
cat("
# Modelo normal, varianza desconocida
model
{
    for (i in 1:n)
      {
         x[i] ~ dnorm(mu,tao);
      }
         mu ~ dnorm(0,0.001);
         tao ~ dgamma(0.01,0.01);
         sigma2 <- 1/tao;
}
",
file="modelo.txt")

#.................................................................

eso <-
list(x=c(-1.10635822,0.56352639,-1.62101846,0.06205707,0.50183464,
           0.45905694,-1.00045360,-0.58795638,1.01602187,-0.26987089,
           0.18354493,1.64605637,-0.96384666,0.53842310,-1.11685831,
           0.75908479,1.10442473,-1.71124673,-0.42677894,0.68031412))

attach(eso)
n <- length(x)

datos <- list ("x","n")
parametros <- c("mu", "sigma2")

# iniciales <- function() {list(mu=rnorm(1),tao=runif(1))}
```

```
# Programa con BRugs
library(BRugs)
modelo <- BRugsFit(data=datos, inits=NULL,
nBurnin = 10000, nIter = 20000,
para=parametros, modelFile="modelo.txt",
numChains=3,working.directory=ruta)

modelo
samplesStats("*")
samplesDensity("*", mfrow=c(2,1))

#.................................................................

library(R2OpenBUGS)
modelo <- bugs(data=datos,inits=NULL,
parameters.to.save=parametros,
model.file="modelo.txt",
n.chains=3, n.iter=20000, n.burnin=10000,
working.directory=ruta,
clearWD=TRUE, debug=TRUE)

print(modelo)

detach(eso)
file.remove("modelo.txt")
```

Programa de SAS:

```
OPTIONS nodate;
TITLE 'Ejemplo Normal-Normal Bayes';
ODS rtf file='/folders/myfolders/cosa.rtf' style=minimal
   startpage=no;

DATA cosas;
     id+1;
INPUT puntos @@;
DATALINES;
-1.10635822 0.56352639 -1.62101846 0.06205707 0.50183464
0.45905694 -1.00045360 -0.58795638 1.01602187 -0.26987089
0.18354493 1.64605637 -0.96384666 0.53842310 -1.11685831
0.75908479 1.10442473 -1.71124673 -0.42677894 0.68031412
;

PROC mcmc data=cosas nmc=10000 nbi=5000 outpost=sale DIC;

parms mu0 sigma20;

prior mu0 ~ normal(mean=0, sd=100);
prior sigma20 ~ igamma(shape=0.01, scale=0.01);

model puntos ~ normal(mean=mu0, sd=sqrt(sigma20));
RUN;

ODS rtf close;
```