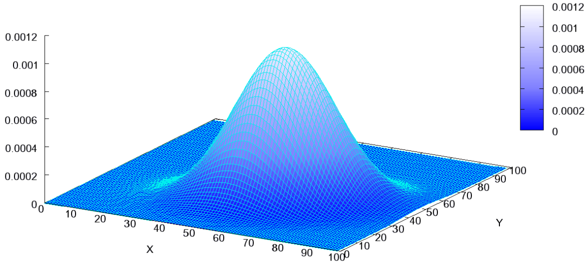
Modelos Gaussianas



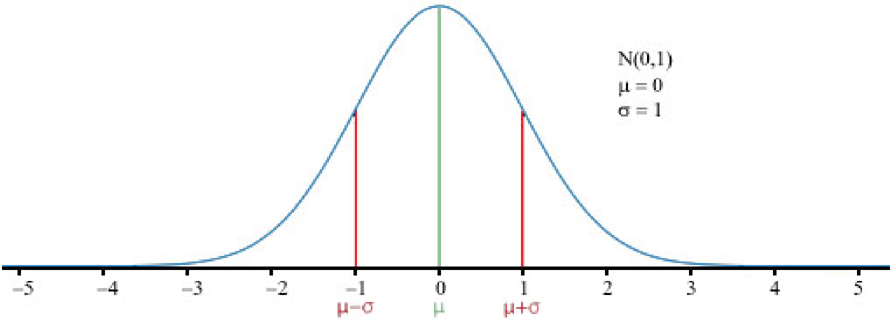
Mike Wiper

Departamento de Estadística

Universidad Carlos III de Madrid

Grado en Estadística y Empresa

Objetivo



Mostrar como implementar la inferencia bayesiana cuando los datos son normales o Gaussianos.

La verosimilitud Gaussiana

Supongamos que observamos datos $y_1, ..., y_n$ de una distribución Gaussiana: Normal $\left(\mu, \sigma^2\right)$. Luego, la función de verosimilitud es:

$$
\begin{aligned}
f(\text{datos}|\mu,\sigma^2) &= \prod_{i=1}^{n} f(y_i|\mu,\sigma^2) \\
&= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(y_i-\mu)^2\right) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}-\mu)^2\right)
\end{aligned}
$$

La verosimilitud Gaussiana

Ahora, resolvemos el cuadrado y tenemos:

$$
\begin{aligned}
f(\text{datos}|\mu, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (y_i - \bar{y})^2 + 2\sum_{i=1}^n (y_i - \bar{y})(\bar{y} - \mu) + \right.\right. \\
&\quad \left.\left. \sum_{i=1}^n (\bar{y} - \mu)^2\right]\right) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right]\right) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]\right)
\end{aligned}
$$

donde la cuasi-varianza, $s^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2$.

Estimación máximo-verosimil

Es fácil ver que el EMV de μ es $\hat{\mu} = \bar{y}$ (tanto si σ^2 es conocido como desconocido).

Cuando μ es desconocido, el EMV de σ^2 es la varianza muestral:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

El EMV de σ^2 cuando μ es desconocido es sesgado. Un estimador insesgado es la cuasi-varianza.

Estimación por intervalos

Se tiene que $\bar{Y}|\mu, \sigma^2 \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$.

En el caso de que σ^2 sea desconocido, recordamos que $\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sim \chi^2_{n-1}$ y luego, el intervalo para μ es:

$$\bar{y} \pm t_{n-1}\left(1 - \frac{\alpha}{2}\right)\frac{s}{\sqrt{n}},$$

dónde $t_a(p)$ es el $p-$ cuantíl de la distribución t de Student con a grados de libertad.

▸ ST

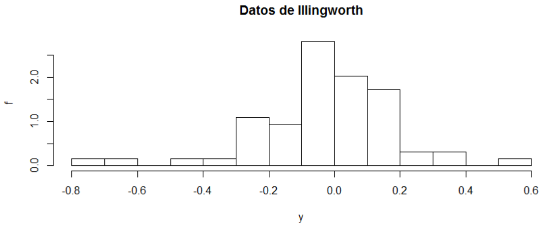
Cuando μ es desconocido, un intervalo para σ^2 es:

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)}, \quad \frac{(n-1)s^2}{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)}\right),$$

dónde $\chi^2_a(p)$ es el $p-$ cuantíl de una distribución ji-cuadrado con a grados de libertad.

Ejemplo

Famosos datos utilizados para demostrar empíricamente la teoría de la relatividad.



$$n = 64, \quad \bar{y} = -0{,}015 \quad s^2 = 0{,}042$$

Intervalos de 95 % de confianza para μ, σ^2 son $(-0{,}066, 0{,}036)$ y $(0{,}0310{,}062)$ respectivamente.

Inferencia bayesiana

¿Existe una distribución a priori conjugada (para μ y/o σ^2) en el caso de datos Gaussianos?

Miramos la función de verosimilitud otra vez:

$$l(\mu, \sigma^2|\text{datos}) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right).$$

En términos de μ, ¿parece como una distribución conocida?

Inferencia bayesiana

¿Existe una distribución a priori conjugada (para μ y/o σ^2) en el caso de datos Gaussianos?

Miramos la función de verosimilitud otra vez:

$$l(\mu, \sigma^2|\text{datos}) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right).$$

En términos de μ, ¿parece como una distribución conocida?

En términos de σ^2, ¿parece como una distribución conocida?

Reescribiendo la verosimilitud

Sea $\phi = 1/\sigma^2$ la precisión del modelo. Luego, escribimos la verosimilitud en términos de ϕ:

$$l(\mu, \phi|\text{datos}) \propto \phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right).$$

Ahora, en términos de ϕ, ¿parece como una distribución conocida?

Una distribución a priori conjugada: la distribución normal-gamma

Consideramos la situación donde tanto μ como ϕ son desconocidos. Una distribución a priori natural es suponer que:

$$\begin{aligned} \phi &\sim \text{Gamma}\left(\frac{a}{2}, \frac{b}{2}\right) \\ \mu|\phi &\sim \text{Normal}\left(m, \frac{1}{c\phi}\right) \end{aligned}$$

Esta distribución se llama la distribución normal-gamma con parámetros m, c, a, b.

Propiedades de la distribución normal-gamma

La distribución conjunta de μ, ϕ es:

$$f(\mu, \phi) = \sqrt{\frac{c}{2\pi}} \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \phi^{\frac{a+1}{2}-1} \exp\left(-\frac{\phi}{2}\left[b + c(\mu - m)^2\right]\right)$$

La distribución marginal de μ es:

$$\begin{aligned} f(\mu) &= \int_0^\infty f(\mu, \phi)\, d\phi \\ &= \sqrt{\frac{c}{2\pi}} \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \int_0^\infty \phi^{\frac{a+1}{2}-1} \exp\left(-\frac{\phi}{2}\left[b + c(\mu - m)^2\right]\right) d\phi \\ &= \sqrt{\frac{c}{2\pi}} \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \frac{\Gamma\left(\frac{a+1}{2}\right)}{\left(\frac{1}{2}\left[b + c(\mu - m)^2\right]\right)^{\frac{a+1}{2}}} \end{aligned}$$

Propiedades de la distribución normal-gamma

$$
\begin{aligned}
f(\mu) &= \sqrt{\frac{c}{b\pi}}\frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)}\left(1+\frac{c}{b}(\mu-m)^2\right)^{-\frac{a+1}{2}} \\
&= \sqrt{\frac{ac}{b}}\frac{\Gamma\left(\frac{a+1}{2}\right)}{\sqrt{a\pi}\Gamma\left(\frac{a}{2}\right)}\left(1+\frac{1}{a}\left(\frac{\mu-m}{\sqrt{b/(ac)}}\right)^2\right)^{-\frac{a+1}{2}}
\end{aligned}
$$

La distribución marginal de μ es una distribución t de Student, escalada y no centrada:

$$\frac{\mu-m}{\sqrt{b/(ac)}} \sim t_a.$$

La distribución a posteriori

Se puede demostrar que dada la muestra, $y_1, \ldots, y_n$, la distribución a posteriori es también normal gamma con parámetros m^*, c^*, a^*, b^* donde:

$$
\begin{aligned}
c^* &= c + n \\
m^* &= \frac{cm + n\bar{y}}{c + n} \\
a^* &= a + n \\
b^* &= b + (n-1)s^2 + \frac{cn}{c+n}(m - \bar{y})^2
\end{aligned}
$$

▸ Demostración

Ejemplo

Usamos una distribución a priori normal-gamma con $m = 0$, $c = 1$, $a = b = 1$. Luego,

$$E[\mu|\text{datos}] = -0{,}014 \quad E[\sigma^2|\text{datos}] = 0{,}058$$

y intervalos de 95 % de credibilidad para μ, σ^2 son $(-0{,}073, 0{,}044)$ y $(0{,}041, 0{,}082)$ respectivamente

Predicción

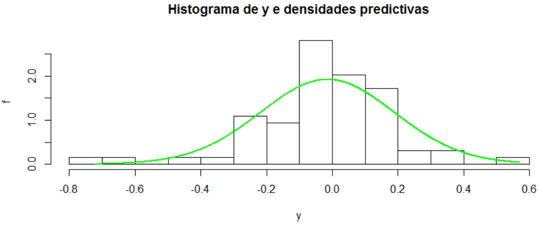
Supongamos que queremos predecir un nuevo dato: $Y|\mu, \phi \sim \text{Normal}\left(\mu, \frac{1}{\phi}\right)$.

Luego, tenemos $Y = \mu + \epsilon$ donde $\mu|\phi, \text{datos} \sim \text{Normal}\left(m^*, \frac{1}{c^*\phi}\right)$, $\epsilon|\phi \sim \text{Normal}\left(0, \frac{1}{\phi}\right)$, $\phi|\text{datos} \sim \text{Gamma}\left(\frac{a^*}{2}, \frac{b^*}{2}\right)$.

Luego, $Y|\phi, \text{datos} \sim \text{Normal}\left(m^*, \frac{1}{c^{**}\phi}\right)$ donde $c^{**} = \frac{c^*}{c^*+1}$ y la distribución conjunta de Y, ϕ es normal gamma.

Entonces, la distribución marginal de Y es otra t de Student trasladada y escalada.

Ejemplo



La distribución predictiva se ajusta bastante bien a los datos.

Una distribución a priori "no-informativa"

Dejando $c, m, a, b \to 0$ proporciona la distribución a priori impropia:

$$f(\mu, \phi) \propto \frac{1}{\phi}.$$

En este caso, la distribución a posteriori es

$$\begin{aligned} f(\mu, \phi|\text{datos}) &\propto \frac{1}{\phi}\phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}\left[(n-1)s^2 + n(\mu-\bar{y})^2\right]\right) \\ &\propto \phi^{\frac{(n-1)+1}{2}-1} \exp\left(-\frac{\phi}{2}\left[(n-1)s^2 + n(\mu-\bar{y})^2\right]\right) \end{aligned}$$

que es una distribución normal-gamma:

$$\begin{aligned} \mu|\phi, \text{datos} &\sim \text{Normal}\left(\bar{y}, \frac{1}{n\phi}\right) \\ \phi|\text{datos} &\sim \text{Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right) \end{aligned}$$

Estimación de μ

Claramente, la media a posteriori de μ es $E[\mu|\text{datos}] = \bar{y} = \hat{\mu}$.

Además, recordando la fórmula para la distribución marginal de μ, tenemos:

$$\frac{\mu - \bar{y}}{s/\sqrt{n}} \sim t_{n-1}$$

y luego, un intervalo bayesiano de credibilidad para μ coincide con el intervalo frecuentista.

Estimación de ϕ y σ^2

La media a posteriori de ϕ es $E[\phi|\text{datos}] = \frac{1}{s^2}$, correspondiendo al inverso del estimador insesgado frecuentista de σ^2.

No obstante, la media a posteriori de σ^2 es

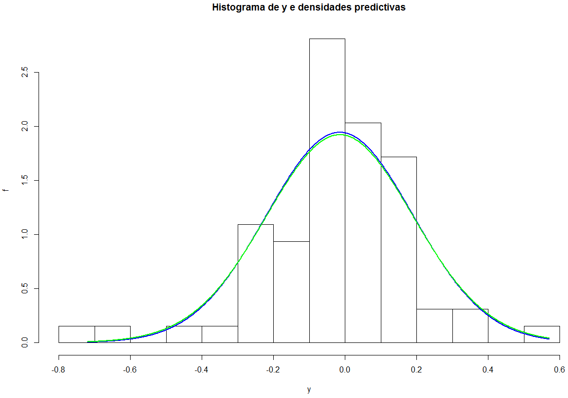
$$E[\sigma^2|\text{datos}] = E\left[\frac{1}{\phi}|\text{datos}\right] = \frac{n-1}{n-3}s^2$$

que es distinto al EMV.

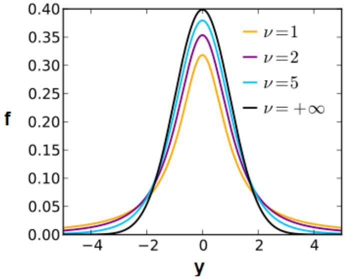
El intervalo de credibilidad para σ^2 sí coincide con el intervalo frecuentista.

Predicción y ejemplo

En el caso de la predicción frecuentista, lo normal sería utilizar estimadores “plug-in"de μ, σ^2, y luego, la distribución predictiva, frecuentista sería Normal $\left(\bar{y}, s^2\right)$. La distribución predictiva bayesiana es una t de Student y muestra alguna diferencia con la predicción frecuentista.



Apéndice: la distribución t de Student



Una variable continua, Y tiene una distribución t de Student con $\nu > 0$ grados de libertad si:

$$f(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{2\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad \forall y \in \mathbb{R}.$$

Se tiene $E[Y] = 0$ para $\nu > 1$ y $V[Y] = \frac{\nu}{\nu-2}$ para $\nu > 2$.

Resumen y siguiente sesión

$$\hat{\nu} \approx \frac{(g_1 + g_2)^2}{g_1^2/(n_1 - 1) + g_2^2/(n_2 - 1)} \quad \text{where } g_i = s_i^2/n_i.$$

En esta sección hemos comparado la inferencia frecuentista y la inferencia bayesiana para una población normal.

En la siguiente sesión estudiamos problemas de dos muestrar y introduciremos como usar el método Monte Carlo para implementar la inferencia bayesiana en uno de ellos, que es difícil de resolver con métodos básicos frecuentistas.

Apéndice: demostración de la distribución a posteriori

Por el teorema de Bayes:

$$\begin{aligned}
f(\mu,\phi|\text{datos}) &\propto f(\text{datos}|\mu,\phi)f(\mu,\phi) \\
&\propto \phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}\left[(n-1)s^2 + n(\mu-\bar{y})^2\right]\right) \times \\
&\quad \sqrt{\frac{c}{2\pi}}\frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)}\phi^{\frac{a+1}{2}-1}\exp\left(-\frac{\phi}{2}\left[b+c(\mu-m)^2\right]\right) \\
&\propto \phi^{\frac{a+n+1}{2}-1}\exp\left(-\frac{\phi}{2}\left[b+(n-1)s^2+c(\mu-m)^2+n(\bar{y}-\mu)^2\right]\right) \\
&\propto \phi^{\frac{a+n+1}{2}-1}\exp\left(-\frac{\phi}{2}\left[b+(n-1)s^2+(c+n)\mu^2-\right.\right. \\
&\quad \left.\left. 2\mu(cm+n\bar{y})+cm^2+n\bar{y}^2\right]\right)
\end{aligned}$$

$$
\begin{aligned}
f(\mu,\phi|\text{datos}) &\propto \phi^{\frac{a+n+1}{2}-1}\exp\left(-\frac{\phi}{2}\left[b+(n-1)s^2+(c+n)\left(\mu-\frac{cm+n\bar{y}}{c+n}\right)^2\right.\right. \\
&\qquad \left.\left.+cm^2+n\bar{y}^2-\frac{(cm+n\bar{y})^2}{c+n}\right]\right) \\
&\propto \phi^{\frac{a+n+1}{2}-1}\exp\left(-\frac{\phi}{2}\left[b+(n-1)s^2+\frac{cn}{c+n}(m-\bar{y})^2+\right.\right. \\
&\qquad \left.\left.\left(\mu-\frac{cm+n\bar{y}}{c+n}\right)^2\right]\right) \\
&\propto \phi^{\frac{a^*+1}{2}-1}\exp\left(-\frac{\phi}{2}\left[b^*+c^*\left(\mu-m^*\right)^2\right]\right)
\end{aligned}
$$

que es el núcleo de la distribución normal-gamma con parámetros a^*, b^*, c^*, m^*. ☺