EPG3343 - Seminario de Estadística III
Clase 9

Jonathan Acosta

Pontificia Universidad Católica de Chile

Segundo Semestre, 2022

Esquema

Esquema

- La estimación bayesiana se diferencia de los métodos clásicos de estimación puntual, al tratar el parámetro θ en un modelo de probabilidad como una variable aleatoria.

- La estimación bayesiana se diferencia de los métodos clásicos de estimación puntual, al tratar el parámetro θ en un modelo de probabilidad como una variable aleatoria.

- El valor de θ en un modelo paramétrico $f(x;\theta)$ tiene su propia fdp. denotada por $\pi(\theta;\lambda)$.

- La estimación bayesiana se diferencia de los métodos clásicos de estimación puntual, al tratar el parámetro θ en un modelo de probabilidad como una variable aleatoria.

- El valor de θ en un modelo paramétrico $f(x;\theta)$ tiene su propia fdp. denotada por $\pi(\theta;\lambda)$.

- $\pi(\theta;\lambda)$ se denomina distribución a priori y λ se conoce como hiperparámetro de la distribución a priori.

Estimación Puntual

- La distribución a priori está basada en los posibles valores de $\theta \in \Theta$ y refleja la incertidumbre sobre θ antes de que se hayan recopilado los datos.

Estimación Puntual

- La distribución a priori está basada en los posibles valores de $\theta \in \Theta$ y refleja la incertidumbre sobre θ antes de que se hayan recopilado los datos.

- La elección de la priori puede hacerse subjetiva u objetivamente y puede basarse en información previa sobre θ, conveniencia matemática o la falta total de información previa sobre θ.

Estimación Puntual

- La distribución a priori está basada en los posibles valores de $\theta \in \Theta$ y refleja la incertidumbre sobre θ antes de que se hayan recopilado los datos.
- La elección de la priori puede hacerse subjetiva u objetivamente y puede basarse en información previa sobre θ, conveniencia matemática o la falta total de información previa sobre θ.
- Las prioris subjetivas contienen información importante sobre θ, mientras que las prioris objetivas o prioris no informativas tienden a enfatizar la información contenida en la función de verosimilitud.

Estimación Puntual

- La distribución a priori está basada en los posibles valores de $\theta \in \Theta$ y refleja la incertidumbre sobre θ antes de que se hayan recopilado los datos.
- La elección de la priori puede hacerse subjetiva u objetivamente y puede basarse en información previa sobre θ, conveniencia matemática o la falta total de información previa sobre θ.
- Las prioris subjetivas contienen información importante sobre θ, mientras que las prioris objetivas o prioris no informativas tienden a enfatizar la información contenida en la función de verosimilitud.
- La información sobre θ contenida en la distribución a priori y la información sobre θ contenida en una muestra aleatoria del modelo de probabilidad $f(x \mid \theta)$ se utilizan para estimar θ.

Estimación Puntual

- Cuando θ es una variable aleatoria continua, el teorema de Bayes establece que la fdp de la distribución posterior es

$$f(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}, \theta; \lambda)}{f(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta; \lambda)}{\int_{\Theta} f(\boldsymbol{x}|\theta)\pi(\theta; \lambda)d\theta}.$$

Estimación Puntual

- Cuando θ es una variable aleatoria continua, el teorema de Bayes establece que la fdp de la distribución posterior es

$$f(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}, \theta; \lambda)}{f(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta; \lambda)}{\int_{\Theta} f(\boldsymbol{x}|\theta)\pi(\theta; \lambda) d\theta}.$$

- Cuando θ es una variable aleatoria discreta, el teorema de Bayes establece que la fmp de la distribución posterior es

$$f(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}, \theta; \lambda)}{f(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta; \lambda)}{\sum_{\theta \in \Theta} f(\boldsymbol{x}|\theta)\pi(\theta; \lambda)}.$$

Estimación Puntual

- Cuando θ es una variable aleatoria continua, el teorema de Bayes establece que la fdp de la distribución posterior es

$$f(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}, \theta; \lambda)}{f(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta; \lambda)}{\int_{\Theta} f(\boldsymbol{x}|\theta)\pi(\theta; \lambda)d\theta}.$$

- Cuando θ es una variable aleatoria discreta, el teorema de Bayes establece que la fmp de la distribución posterior es

$$f(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}, \theta; \lambda)}{f(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta; \lambda)}{\sum_{\theta \in \Theta} f(\boldsymbol{x}|\theta)\pi(\theta; \lambda)}.$$

- Note que en ambos casos, $f(\theta|\boldsymbol{x}) \propto \mathrm{L}(\theta)\pi(\theta; \lambda)$.

Estimación Puntual

Ejemplo: Sea $X_1, \ldots, X_n \overset{\text{iid}}{\sim} b(1, \theta)$. Suponga que $\theta \sim U(0,1)$ (priori no informativa). La función de verosimilitud para la muestra observada es

$$\mathrm{L}(\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{t}(1-\theta)^{n-t}, \quad \text{donde} \quad t = \sum_{i=1}^{n} x_i.$$

Ejemplo: Sea $X_1, \ldots, X_n \overset{\text{iid}}{\sim} b(1, \theta)$. Suponga que $\theta \sim U(0,1)$ (priori no informativa). La función de verosimilitud para la muestra observada es

$$\mathrm{L}(\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^t(1-\theta)^{n-t}, \quad \text{donde} \quad t = \sum_{i=1}^{n} x_i.$$

Ahora, dado que la distribución a priori es continua, la distribución posterior es

$$\begin{aligned} f(\theta|\boldsymbol{x}) &= \frac{\mathrm{L}(\theta)\pi(\theta;\lambda)}{\displaystyle\int_{\Theta} \mathrm{L}(\theta)\pi(\theta;\lambda)d\theta} = \frac{\theta^t(1-\theta)^{n-t}}{\displaystyle\int_0^1 \theta^t(1-\theta)^{n-t}d\theta} \\ &= \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)}\theta^t(1-\theta)^{n-t}, \end{aligned}$$

Ejemplo: Sea $X_1, \ldots, X_n \overset{\text{iid}}{\sim} b(1, \theta)$. Suponga que $\theta \sim U(0,1)$ (priori no informativa). La función de verosimilitud para la muestra observada es

$$\mathrm{L}(\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^t (1-\theta)^{n-t}, \quad \text{donde} \quad t = \sum_{i=1}^{n} x_i.$$

Ahora, dado que la distribución a priori es continua, la distribución posterior es

$$\begin{aligned} f(\theta|\boldsymbol{x}) &= \frac{\mathrm{L}(\theta)\pi(\theta;\lambda)}{\int_{\Theta} \mathrm{L}(\theta)\pi(\theta;\lambda)d\theta} = \frac{\theta^t(1-\theta)^{n-t}}{\int_0^1 \theta^t(1-\theta)^{n-t}d\theta} \\ &= \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)}\theta^t(1-\theta)^{n-t}, \end{aligned}$$

es decir, $\theta \mid \boldsymbol{x} \sim \mathrm{Beta}(t+1, n-t+1)$.

Estimación Puntual

- Note que sin datos muéstrales, la media de θ basada en la distribución a priori $\mathrm{E}(\theta) = 0{,}5$.

Estimación Puntual

- Note que sin datos muéstrales, la media de θ basada en la distribución a priori $\mathrm{E}(\theta) = 0{,}5$.

- Sin embargo, dada la información contenida en una muestra observada, la media a posterior es

$$\mathrm{E}(\theta|\boldsymbol{x}) = \frac{\sum_{i=1}^{n} x_i + 2}{n+2} = \frac{n\overline{x} + 1}{n+2} = \frac{n}{n+2}\overline{x} + \frac{2}{n+2} \cdot 0{,}5$$

Estimación Puntual

- Note que sin datos muéstrales, la media de θ basada en la distribución a priori $\mathrm{E}(\theta) = 0{,}5$.

- Sin embargo, dada la información contenida en una muestra observada, la media a posterior es

$$\mathrm{E}(\theta|\boldsymbol{x}) = \frac{\sum_{i=1}^{n} x_i + 2}{n+2} = \frac{n\overline{x} + 1}{n+2} = \frac{n}{n+2}\overline{x} + \frac{2}{n+2} \cdot 0{,}5$$

- Por ejemplo, para una muestra de $n = 25$ con $t = 21$, la media a posteriori es $\mathrm{E}(\theta|\boldsymbol{x}) = 0{,}815$. Mientras que $\overline{X} = 0{,}84$.

Estimación Puntual

- Note que sin datos muéstrales, la media de θ basada en la distribución a priori $\mathrm{E}(\theta) = 0{,}5$.
- Sin embargo, dada la información contenida en una muestra observada, la media a posterior es

$$\mathrm{E}(\theta|\boldsymbol{x}) = \frac{\sum_{i=1}^{n} x_i + 2}{n+2} = \frac{n\overline{x} + 1}{n+2} = \frac{n}{n+2}\overline{x} + \frac{2}{n+2} \cdot 0{,}5$$

- Por ejemplo, para una muestra de $n = 25$ con $t = 21$, la media a posteriori es $\mathrm{E}(\theta|\boldsymbol{x}) = 0{,}815$. Mientras que $\overline{X} = 0{,}84$.
- Si tenemos la distribución a posteriori completa, ¿por qué fijarnos solo en la media?

Estimación Puntual

- A continuación responderemos la pregunta: ¿por qué fijarnos solo en la media a posterior?.

Estimación Puntual

- A continuación responderemos la pregunta: ¿por qué fijarnos solo en la media a posterior?.
- Todas las inferencias bayesianas sobre θ se basan en la distribución posterior.

Estimación Puntual

- A continuación responderemos la pregunta: ¿por qué fijarnos solo en la media a posterior?.
- Todas las inferencias bayesianas sobre θ se basan en la distribución posterior.
- Para cuantificar que tan cerca se encuentra un estimador de θ debemos definir una función de pérdida y ponderar sobre todos los valores posible del parámetro.

Definición (Función de Pérdida)

Para un estimador T de θ, una función de pérdida $\mathscr{L}(T;\theta)$ es una función de valor real no negativa con tal que $\mathscr{L}(\theta;\theta) = 0$

Estimación Puntual

Definición (Función de Pérdida)

Para un estimador T de θ, una función de pérdida $\mathscr{L}(T;\theta)$ es una función de valor real no negativa con tal que $\mathscr{L}(\theta;\theta) = 0$

Algunos ejemplos de funciones de pérdidas son,

i) Pérdida cuadrática, $\mathscr{L}_2(T;\theta) = (T-\theta)^2$,

ii) Pérdida absoluta, $\mathscr{L}_1(T;\theta) = |T-\theta|$,

iii) Norma L_p, $\mathscr{L}_p(T;\theta) = |T-\theta|^p$, para $p > 0$.

Estimación Puntual

Definición (Función de Pérdida)

Para un estimador T de θ, una función de pérdida $\mathscr{L}(T;\theta)$ es una función de valor real no negativa con tal que $\mathscr{L}(\theta;\theta) = 0$

Algunos ejemplos de funciones de pérdidas son,

i) Pérdida cuadrática, $\mathscr{L}_2(T;\theta) = (T-\theta)^2$,

ii) Pérdida absoluta, $\mathscr{L}_1(T;\theta) = |T-\theta|$,

iii) Norma L_p, $\mathscr{L}_p(T;\theta) = |T-\theta|^p$, para $p > 0$.

El valor esperado de la función de pérdida, tomado como función de θ, es llamado función de riesgo.

Estimación Puntual

Definición (Función de Riesgo)

Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una muestra aleatoria. Para un estimador $T(\boldsymbol{X})$ de θ y una función de pérdida $\mathscr{L}(T;\theta)$, la función de riesgo asociada con T y $\mathscr{L}$ es el valor esperado de la función de pérdida con respecto a $f(\boldsymbol{x} \mid \theta)$; la cual denotamos por $R(T;\theta) = \mathsf{E}\{\mathscr{L}(T;\theta)|\theta\}$.

Estimación Puntual

Definición (Función de Riesgo)

Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una muestra aleatoria. Para un estimador $T(\boldsymbol{X})$ de θ y una función de pérdida $\mathscr{L}(T;\theta)$, la función de riesgo asociada con T y $\mathscr{L}$ es el valor esperado de la función de pérdida con respecto a $f(\boldsymbol{x} \mid \theta)$; la cual denotamos por $R(T;\theta) = \mathsf{E}\{\mathscr{L}(T;\theta)|\theta\}$.

- Cuando la función de pérdida es la cuadrática, la función de riesgo se denomina error cuadrático medio (MSE) del estimador.

Definición (Función de Riesgo)

Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una muestra aleatoria. Para un estimador $T(\boldsymbol{X})$ de θ y una función de pérdida $\mathscr{L}(T;\theta)$, la función de riesgo asociada con T y $\mathscr{L}$ es el valor esperado de la función de pérdida con respecto a $f(\boldsymbol{x} \mid \theta)$; la cual denotamos por $R(T;\theta) = \mathsf{E}\{\mathscr{L}(T;\theta)|\theta\}$.

- Cuando la función de pérdida es la cuadrática, la función de riesgo se denomina error cuadrático medio (MSE) del estimador.
- Cuando la función de pérdida es la pérdida absoluta, entonces la función de riesgo se conoce como la desviación media absoluta (MAD) del estimador.

Estimación Puntual

- En el enfoque clásico, a menudo se elige un estimador que minimice el valor esperado de la función de pérdida, denominado riesgo del estimador.

Estimación Puntual

- En el enfoque clásico, a menudo se elige un estimador que minimice el valor esperado de la función de pérdida, denominado riesgo del estimador.
- En el enfoque bayesiano, se eligen estimadores que minimicen el valor esperado de la función de riesgo, llamado riesgo de Bayes.

Estimación Puntual

- En el enfoque clásico, a menudo se elige un estimador que minimice el valor esperado de la función de pérdida, denominado riesgo del estimador.

- En el enfoque bayesiano, se eligen estimadores que minimicen el valor esperado de la función de riesgo, llamado riesgo de Bayes.

Definición (Riesgo de Bayes)

Si $X_1, \ldots, X_n$ es una muestra de va iid con fdp común $f(x|\theta)$, el riesgo de Bayes de un estimador T asociado con una función de pérdida $\mathscr{L}$ y una distribución a priori $\pi(\theta; \lambda)$ es $\mathsf{E}_\theta\{R(T; \theta)\}$.

Estimación Puntual

Definición (Estimador de Bayes)

Sea $\mathscr{L}$ una función de pérdida y $\pi(\theta; \lambda)$ una distribución a priori para θ. Un estimador $T^\star$ de θ con mínimo riesgo de Bayes se denomina estimador bayesiano de θ.

Estimación Puntual

Definición (Estimador de Bayes)

Sea $\mathscr{L}$ una función de pérdida y $\pi(\theta;\lambda)$ una distribución a priori para θ. Un estimador $T^{\star}$ de θ con mínimo riesgo de Bayes se denomina estimador bayesiano de θ.

Así, $T^{\star}$ es el estimador bayesiano de θ cuando $\mathrm{E}_{\theta}\{R(T^{\star};\theta)\} \leq \mathrm{E}_{\theta}\{R(T;\theta)\}$ para cualquier otro estimador T de θ.

Definición (Estimador de Bayes)

Sea $\mathscr{L}$ una función de pérdida y $\pi(\theta; \lambda)$ una distribución a priori para θ. Un estimador $T^\star$ de θ con mínimo riesgo de Bayes se denomina estimador bayesiano de θ.

Así, $T^\star$ es el estimador bayesiano de θ cuando $\mathrm{E}_\theta\{R(T^\star; \theta)\} \leq \mathrm{E}_\theta\{R(T; \theta)\}$ para cualquier otro estimador T de θ.

Teorema

Sea $X_1, \ldots, X_n$ una muestra aleatoria con fdp conjunta $f(\boldsymbol{x}|\theta)$, $\pi(\theta; \lambda)$ una distribución a priori para θ, y $\mathscr{L}$ una función de pérdida. El estimador de Bayes de θ es el estimador $T^\star$ que minimiza

$$\int_\Theta \mathscr{L}(T; \theta) f(\theta|\boldsymbol{x}) \pi(\theta; \lambda) d\theta.$$

A esta cantidad se le denomina riesgo de Bayes a posterior.

Dem. Tenemos que

$$
\begin{aligned}
\mathrm{E}_{\theta}\{R(T;\theta)\} &= \int_{\Theta}\int_{\mathcal{X}} \mathscr{L}(T;\theta) f(\boldsymbol{x}|\theta)\pi(\theta;\lambda) d\boldsymbol{x} d\theta \\
&= \int_{\mathcal{X}} \left[\int_{\Theta} \mathscr{L}(T;\theta) f(\boldsymbol{x}|\theta)\pi(\theta;\lambda) d\theta\right] d\boldsymbol{x} \\
&= \int_{\mathcal{X}} \left[\int_{\Theta} \mathscr{L}(T;\theta) f(\boldsymbol{x}|\theta)\pi(\theta;\lambda) \frac{f(\boldsymbol{x})}{f(\boldsymbol{x})} d\theta\right] d\boldsymbol{x}
\end{aligned}
$$

$$
= \int_{\mathcal{X}} \left[\int_{\Theta} \mathscr{L}(T;\theta) f(\theta|\boldsymbol{x}) d\theta\right] f(\boldsymbol{x}) d\boldsymbol{x}.
$$

Ya que $\mathscr{L}$ es no negativa, se tiene que $\mathrm{E}_{\theta}\{R(T;\theta)\}$ es mínimo cuando la integral $\int_{\Theta} \mathscr{L}(T;\theta) f(\theta|\boldsymbol{x}) d\theta$ sea mínima, para cada $\boldsymbol{x}$.

Teorema

Si $X_1, \ldots, X_n$ es una muestra de va iid con fdp común $f(x|\theta)$, $\pi(\theta;\lambda)$ es una distribución a priori para θ, y $\mathscr{L}(\hat{\theta};\theta)$ es la pérdida cuadrática, entonces la media de la distribución posterior, $\mathsf{E}(\theta|\boldsymbol{x})$, es el estimador de Bayesiano de θ.

Estimación Puntual

Teorema

Si $X_1, \ldots, X_n$ es una muestra de va iid con fdp común $f(x|\theta)$, $\pi(\theta; \lambda)$ es una distribución a priori para θ, y $\mathscr{L}(\hat{\theta}; \theta)$ es la pérdida cuadrática, entonces la media de la distribución posterior, $\mathsf{E}(\theta|\boldsymbol{x})$, es el estimador de Bayesiano de θ.

Teorema

Si $X_1, \ldots, X_n$ es una muestra de variables aleatorias iid con fdp común $f(x|\theta)$, $\pi(\theta; \lambda)$ es una distribución a priori para θ, y $\mathscr{L}(\hat{\theta}; \theta)$ es la función de pérdida absoluta, entonces la mediana de la distribución posterior es el estimador de Bayes de θ.

Estimación Puntual

Ejemplo: Sabemos que si $X_i|\theta \sim P(\theta)$, y la distribución a priori es $\text{Exp}(\lambda)$, la distribución posterior de $\theta|\boldsymbol{x}$ es Gamma $\left(\sum_{i=1}^{n} x_i + 1, \frac{1}{n+\frac{1}{\lambda}}\right)$.

Estimación Puntual

Ejemplo: Sabemos que si $X_i|\theta \sim P(\theta)$, y la distribución a priori es $\text{Exp}(\lambda)$, la distribución posterior de $\theta|\boldsymbol{x}$ es Gamma $\left(\sum_{i=1}^n x_i + 1, \frac{1}{n+\frac{1}{\lambda}}\right)$. En el caso de pérdida cuadrática,

$$\widehat{\theta} = \mathbb{E}(\theta|\boldsymbol{x}) = \frac{\lambda n}{\lambda n + 1}\overline{x} + \frac{1}{\lambda n + 1}\lambda.$$

Mientras que para la pérdida absoluta, no hay una fórmula cerrada.

Estimación Puntual

Ejemplo: Sabemos que si $X_i|\theta \sim P(\theta)$, y la distribución a priori es $\mathrm{Exp}(\lambda)$, la distribución posterior de $\theta|\boldsymbol{x}$ es Gamma $\left(\sum_{i=1}^{n} x_i + 1, \frac{1}{n+\frac{1}{\lambda}}\right)$. En el caso de pérdida cuadrática,

$$\widehat{\theta} = \mathbb{E}(\theta|\boldsymbol{x}) = \frac{\lambda n}{\lambda n + 1}\overline{x} + \frac{1}{\lambda n + 1}\lambda.$$

Mientras que para la pérdida absoluta, no hay una fórmula cerrada.

Si $X_1, \ldots, X_{25}$ es una muestra aleatoria con $\sum_{i=1}^{25} x_i = 52$ y hiperparámetro $\lambda = 5$, entonces $\theta|\boldsymbol{x} \sim \mathrm{Gamma}(53, 0{,}04)$. En este caso, la mediana de la distribución posterior es

$$> \text{ qgamma } (0{,}5, 53, \text{ scale } = 0{,}04)$$
$$[1] \quad 2{,}11$$

Para propósitos comparativos, el EMV de θ para una distribución Poisson es $\overline{X}$, que en este es $\overline{x} = 2{,}08$. Mientras que el estimador de Bayes bajo pérdida cuadrática es $\widehat{\theta} = 2{,}12$

Estimación Puntual

- La inferencia bayesiana permite la especificación probabilística de creencias previas a través de una distribución a priori.

Estimación Puntual

- La inferencia bayesiana permite la especificación probabilística de creencias previas a través de una distribución a priori.
- A menudo es útil y está justificado restringir el rango de posibles distribuciones a priori a una familia específica con uno o dos parámetros.

Estimación Puntual

- La inferencia bayesiana permite la especificación probabilística de creencias previas a través de una distribución a priori.
- A menudo es útil y está justificado restringir el rango de posibles distribuciones a priori a una familia específica con uno o dos parámetros.
- La elección de esta familia puede basarse en el tipo de función de verosimilitud encontrada.

Estimación Puntual

- La inferencia bayesiana permite la especificación probabilística de creencias previas a través de una distribución a priori.
- A menudo es útil y está justificado restringir el rango de posibles distribuciones a priori a una familia específica con uno o dos parámetros.
- La elección de esta familia puede basarse en el tipo de función de verosimilitud encontrada.

Definición (Distribución a priori conjugada)

Sea $\mathsf{L}(\theta) = f(x|\theta)$ una función de verosimilitud basada en la observación $X = x$. Una clase $\mathcal{G}$ de distribuciones se denomina conjugada con respecto a $\mathsf{L}(\theta)$ si la distribución posterior $f(\theta|x)$ está en $\mathcal{G}$ para todo x, siempre que la distribución a priori $\pi(\theta;\lambda)$ esté en $\mathcal{G}$.

Estimación Puntual

Algunas distribuciones a priori conjugadas para diferentes funciones de verosimilitud, se muestran a continuación:

Verosimilitud	Priori conjugada	Posteriori
$X \mid \pi \sim \mathrm{b}(n, \pi)$	$\pi \sim \mathrm{Beta}(\alpha, \beta)$	$\pi \mid x \sim \mathrm{Beta}(\alpha + x, \beta + n - x)$
$X \mid \pi \sim \mathrm{Geo}(\pi)$	$\pi \sim \mathrm{Beta}(\alpha, \beta)$	$\pi \mid x \sim \mathrm{Beta}(\alpha + 1, \beta + x - 1)$
$X \mid \lambda \sim \mathrm{Po}(\lambda)$	$\lambda \sim \mathrm{G}(\alpha, \beta)$	$\lambda \mid x \sim \mathrm{G}(\alpha + x, \beta + e)$
$X \mid \lambda \sim \mathrm{Exp}(\lambda)$	$\lambda \sim \mathrm{G}(\alpha, \beta)$	$\lambda \mid x \sim \mathrm{G}(\alpha + 1, \beta + x)$
$X \mid \mu \sim \mathrm{N}\left(\mu, \sigma^2 \text{ conocida }\right)$	$\mu \sim \mathrm{N}\left(v, \tau^2\right)$	ver Ec. (1.1) abajo
$X \mid \sigma^2 \sim \mathrm{N}\left(\mu \text{ conocida, } \sigma^2\right)$	$\sigma^2 \sim \mathrm{IG}(\alpha, \beta)$	$\sigma^2 \mid x \sim \mathrm{IG}\left(\alpha + \frac{1}{2}, \beta + \frac{1}{2}(x - \mu)^2\right)$

$$\mu | x \sim \mathrm{N}\left(\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}\left(\frac{x}{\sigma^2} + \frac{v}{\tau^2}\right), \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}\right) \tag{1}$$

Estimación Puntual

Algunas distribuciones a priori conjugadas para diferentes funciones de verosimilitud, se muestran a continuación:

Verosimilitud	Priori conjugada	Posteriori
$X \mid \pi \sim \mathrm{b}(n, \pi)$	$\pi \sim \mathrm{Beta}(\alpha, \beta)$	$\pi \mid x \sim \mathrm{Beta}(\alpha + x, \beta + n - x)$
$X \mid \pi \sim \mathrm{Geo}(\pi)$	$\pi \sim \mathrm{Beta}(\alpha, \beta)$	$\pi \mid x \sim \mathrm{Beta}(\alpha + 1, \beta + x - 1)$
$X \mid \lambda \sim \mathrm{Po}(\lambda)$	$\lambda \sim \mathrm{G}(\alpha, \beta)$	$\lambda \mid x \sim \mathrm{G}(\alpha + x, \beta + e)$
$X \mid \lambda \sim \mathrm{Exp}(\lambda)$	$\lambda \sim \mathrm{G}(\alpha, \beta)$	$\lambda \mid x \sim \mathrm{G}(\alpha + 1, \beta + x)$
$X \mid \mu \sim \mathrm{N}\left(\mu, \sigma^2 \text{ conocida }\right)$	$\mu \sim \mathrm{N}\left(v, \tau^2\right)$	ver Ec. (1.1) abajo
$X \mid \sigma^2 \sim \mathrm{N}\left(\mu \text{ conocida}, \sigma^2\right)$	$\sigma^2 \sim \mathrm{IG}(\alpha, \beta)$	$\sigma^2 \mid x \sim \mathrm{IG}\left(\alpha + \frac{1}{2}, \beta + \frac{1}{2}(x - \mu)^2\right)$

$$\mu | x \sim \mathrm{N}\left(\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \left(\frac{x}{\sigma^2} + \frac{v}{\tau^2}\right), \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}\right) \tag{1}$$

Si bien matemáticamente esto es muy conveniente, si una familia conjugada es una opción razonable para un problema particular, es una cuestión que debe dejarse al experimentador.

Estimación Puntual

Una alternativa para no sesgar los resultados con la elección de la distribución a priori, es utilizar una priori no informativa.

Estimación Puntual

Una alternativa para no sesgar los resultados con la elección de la distribución a priori, es utilizar una priori no informativa.

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

Estimación Puntual

Una alternativa para no sesgar los resultados con la elección de la distribución a priori, es utilizar una priori no informativa.

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

Definición (Priori de Jeffreys)

La distribución a priori de Jeffreys es una distribución proporcial a la raíz de la información de Fisher, es decir

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

Estimación Puntual

Una alternativa para no sesgar los resultados con la elección de la distribución a priori, es utilizar una priori no informativa.

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

Definición (Priori de Jeffreys)

La distribución a priori de Jeffreys es una distribución proporcial a la raíz de la información de Fisher, es decir

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

- La distribución a priori de Jeffreys es invariante en el sentido de que la inferencia no depende de la escala elegida para el parámetro.

Estimación Puntual

Una alternativa para no sesgar los resultados con la elección de la distribución a priori, es utilizar una priori no informativa.

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

Definición (Priori de Jeffreys)

La distribución a priori de Jeffreys es una distribución proporcial a la raíz de la información de Fisher, es decir

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

- 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.

Métodos Numéricos para Estimadores Bayesianos

- Sabemos que los estimadores Bayesianos dependen de la distribución a priori y de la distribución condicional de los datos.

Métodos Numéricos para Estimadores Bayesianos

- Sabemos que los estimadores Bayesianos dependen de la distribución a priori y de la distribución condicional de los datos.

- Siempre se tiene,
$$f(\theta|\boldsymbol{x}) \propto \mathrm{L}(\theta)\pi(\theta;\lambda).$$

Métodos Numéricos para Estimadores Bayesianos

- Sabemos que los estimadores Bayesianos dependen de la distribución a priori y de la distribución condicional de los datos.

- Siempre se tiene,
$$f(\theta|\boldsymbol{x}) \propto \mathrm{L}(\theta)\pi(\theta;\lambda).$$

- Para el caso de pérdida cuadrática el estimador de $h(\theta)$ es
$$\mathbb{E}[h(\theta)|\boldsymbol{x}] = \frac{\int_\Theta h(\theta) f(\boldsymbol{x}|\theta)\pi(\theta;\lambda)d\theta}{\int_\Theta f(\boldsymbol{x}|\theta)\pi(\theta;\lambda)d\theta}.$$

Métodos Numéricos para Estimadores Bayesianos

- Sabemos que los estimadores Bayesianos dependen de la distribución a priori y de la distribución condicional de los datos.

- Siempre se tiene,
$$f(\theta|\boldsymbol{x}) \propto \mathrm{L}(\theta)\pi(\theta;\lambda).$$

- Para el caso de pérdida cuadrática el estimador de $h(\theta)$ es
$$\mathbb{E}[h(\theta)|\boldsymbol{x}] = \frac{\int_\Theta h(\theta)f(\boldsymbol{x}|\theta)\pi(\theta;\lambda)d\theta}{\int_\Theta f(\boldsymbol{x}|\theta)\pi(\theta;\lambda)d\theta}.$$

- Como se puede apreciar la dificultad en los métodos bayesianos es el cálculo de una integral.

Métodos Numéricos para Estimadores Bayesianos

- Sabemos que los estimadores Bayesianos dependen de la distribución a priori y de la distribución condicional de los datos.

- Siempre se tiene,
$$f(\theta|\boldsymbol{x}) \propto \mathrm{L}(\theta)\pi(\theta;\lambda).$$

- Para el caso de pérdida cuadrática el estimador de $h(\theta)$ es
$$\mathbb{E}[h(\theta)|\boldsymbol{x}] = \frac{\int_\Theta h(\theta) f(\boldsymbol{x}|\theta)\pi(\theta;\lambda)d\theta}{\int_\Theta f(\boldsymbol{x}|\theta)\pi(\theta;\lambda)d\theta}.$$

- Como se puede apreciar la dificultad en los métodos bayesianos es el cálculo de una integral.

- A continuación veremos algunas alternativas para estimar de forma numérica dichas integrales.

Métodos Numéricos para Estimadores Bayesianos

- Sabemos que los estimadores Bayesianos dependen de la distribución a priori y de la distribución condicional de los datos.

- Siempre se tiene,
$$f(\theta|\boldsymbol{x}) \propto \mathrm{L}(\theta)\pi(\theta;\lambda).$$

- Para el caso de pérdida cuadrática el estimador de $h(\theta)$ es
$$\mathbb{E}[h(\theta)|\boldsymbol{x}] = \frac{\int_\Theta h(\theta)f(\boldsymbol{x}|\theta)\pi(\theta;\lambda)d\theta}{\int_\Theta f(\boldsymbol{x}|\theta)\pi(\theta;\lambda)d\theta}.$$

- Como se puede apreciar la dificultad en los métodos bayesianos es el cálculo de una integral.

- A continuación veremos algunas alternativas para estimar de forma numérica dichas integrales.

- Estas se basan principalmente en simular números aleatorios de la densidad posterior.

Metropolis-Hasting

- El objetivo del algoritmo Metropolis-Hasting es obtener numeros aleatorios desde una densidad de probabilidad.

Metropolis-Hasting

- El objetivo del algoritmo Metropolis-Hasting es obtener numeros aleatorios desde una densidad de probabilidad.
- En este sentido, el algoritmo de Metropolis-Hasting, se utiliza en los métodos de estimación bayesiano, simulando números aleatorios de la distribución posterior.

Metropolis-Hasting

- El objetivo del algoritmo Metropolis-Hasting es obtener numeros aleatorios desde una densidad de probabilidad.
- En este sentido, el algoritmo de Metropolis-Hasting, se utiliza en los métodos de estimación bayesiano, simulando números aleatorios de la distribución posterior.
- Metropolis-Hasting utiliza cadena de Markov (MCMC) en conjunto de algoritmo de aceptación-rechazo.

Metropolis-Hasting

- El objetivo del algoritmo Metropolis-Hasting es obtener numeros aleatorios desde una densidad de probabilidad.
- En este sentido, el algoritmo de Metropolis-Hasting, se utiliza en los métodos de estimación bayesiano, simulando números aleatorios de la distribución posterior.
- Metropolis-Hasting utiliza cadena de Markov (MCMC) en conjunto de algoritmo de aceptación-rechazo.
- La idea principal de los algoritmos de aceptación-rechazo, se basa en definir un criterio para aceptar o rechazar si un número aleatorio simulado de otra densidad es factible que sea de la densidad objetivo.

- Suponga que se quiere simular desde $f(\theta \mid x)$ sin necesariamente conocer la constante de integración.

Metropolis-Hasting

Aceptación - Rechazo

- Suponga que se quiere simular desde $f(\theta \mid x)$ sin necesariamente conocer la constante de integración.

- El primer paso en el algoritmo de aceptación-rechazo, consiste en encontrar una densidad $p(\theta)$ tal que:
  - Sea fácil de simular,
  - La densidad p se asemeja a la densidad posterior de interés f en términos de localización y dispersión,
  - Para todo θ, se tiene que $f(\theta \mid x) \leq cp(\theta)$.

- Suponga que se quiere simular desde $f(\theta \mid x)$ sin necesariamente conocer la constante de integración.

- El primer paso en el algoritmo de aceptación-rechazo, consiste en encontrar una densidad $p(\theta)$ tal que:
  - Sea fácil de simular,
  - La densidad p se asemeja a la densidad posterior de interés f en términos de localización y dispersión,
  - Para todo θ, se tiene que $f(\theta \mid x) \leq cp(\theta)$.

- Luego de definir la fdp p, el algoritmo de aceptación-rechazo sigue como:
  1. Simular independientemente $\theta \sim p(\theta)$ y $U \sim U(0, 1)$.
  2. Si $cUp(\theta) \leq f(\theta \mid x)$, entonces acepte θ como una extracción de la densidad posterior f; en caso contrario, rechace θ.
  3. Continuar los pasos 1 y 2 hasta que se haya recogido un número suficiente número de θ aceptados.

Metropolis-Hasting

- La estrategia de muestreo MCMC establece una cadena de Markov irreducible y aperiódica donde la distribución estacionaria es igual a la distribución posterior de interés.

Metropolis-Hasting

- La estrategia de muestreo MCMC establece una cadena de Markov irreducible y aperiódica donde la distribución estacionaria es igual a la distribución posterior de interés.
- Un algoritmo de Metropolis-Hastings comienza con un valor inicial θ^0 y genera una secuencia θ^t dado el valor de θ^{t-1} a través de una regla.

Metropolis-Hasting

- La estrategia de muestreo MCMC establece una cadena de Markov irreducible y aperiódica donde la distribución estacionaria es igual a la distribución posterior de interés.
- Un algoritmo de Metropolis-Hastings comienza con un valor inicial θ^0 y genera una secuencia θ^t dado el valor de θ^{t-1} a través de una regla.
- Esta regla consiste en simular un valor candidato θ^* de una densidad de propuesta p, y además del cálculo de una probabilidad de aceptación α, que indica la probabilidad de que el valor candidato sea aceptado como el siguiente valor de la secuencia.

Metropolis-Hasting

- La estrategia de muestreo MCMC establece una cadena de Markov irreducible y aperiódica donde la distribución estacionaria es igual a la distribución posterior de interés.
- Un algoritmo de Metropolis-Hastings comienza con un valor inicial θ^0 y genera una secuencia θ^t dado el valor de θ^{t-1} a través de una regla.
- Esta regla consiste en simular un valor candidato θ^* de una densidad de propuesta p, y además del cálculo de una probabilidad de aceptación α, que indica la probabilidad de que el valor candidato sea aceptado como el siguiente valor de la secuencia.
- El algoritmo Metropolis-Hasting puede describirse como sigue:

Metropolis-Hasting

1. Simular $\theta^* \sim p(\theta^* \mid \theta^{t-1})$.
2. Calcular el radio

$$R = \frac{f(\theta^* \mid x)p(\theta^{t-1} \mid \theta^*)}{f(\theta^{t-1} \mid x)p(\theta^* \mid \theta^{t-1})}$$

3. Calcular la probabilidad de aceptación

$$\alpha = \text{mín}\{R, 1\}$$

4. Simular $U \sim U(0,1)$; luego comparar α y U, de modo que:

$$\text{Si} \;\; U \leq \alpha \;\Rightarrow\; \theta^t = \theta^* \;\; \text{si no} \;\; \theta^t = \theta^{t-1}$$

Métodos Numéricos para Estimadores Bayesianos

- Existen varios otros métodos numéricos, como Importance Sample o Gibbs Sample.

Métodos Numéricos para Estimadores Bayesianos

- Existen varios otros métodos numéricos, como Importance Sample o Gibbs Sample.

- Volviendo al objetivo de estimador de bayes, y asumiendo que $\theta^1, \theta^2, \ldots, \theta^n$ es una muestra aleatoria desde la distribución porterior $f(\theta \mid x)$, entonces, el estimador de bayes de θ es

$$\widehat{\theta}_B = \frac{1}{n}\sum_{i=1}^{n} \theta^i \qquad \text{bajo pérdida cuadrática}$$

$$\widehat{\theta}_B = \theta^{((n+1)/2)} \qquad \text{bajo pérdida absoluta}$$

Estimación Intervalar

Una región de credibilidad bayesiana es un estimador de intervalo basado en la distribución posterior.

Estimación Intervalar

Una región de credibilidad bayesiana es un estimador de intervalo basado en la distribución posterior.

Definición

Sea $X_1, \ldots, X_n$ una muestra aleatoria y $f(\theta|\boldsymbol{x})$ la distribución posterior de θ para una priori $\pi(\theta; \lambda)$. Un estimador de intervalo $[T_1, T_2]$ es un intervalo de credibilidad bayesiano del $100(1-\alpha)\,\%$ para θ si

$$P(T_1 \leq \theta \leq T_2|\boldsymbol{x}) = \int_{t_1}^{t_2} f(\theta|\boldsymbol{x}) d\theta = 1 - \alpha$$

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra de va iid $P(\theta)$ y $\theta \sim \text{Exp}(\lambda)$ la distribución a priori. La distribución posterior es

$$\theta|\boldsymbol{x} \sim \text{Gamma}\Big(\sum_{i=1}^{n} x_i + 1, \frac{1}{n + \frac{1}{\lambda}}\Big)$$

Estimación Intervalar

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra de va iid $P(\theta)$ y $\theta \sim \text{Exp}(\lambda)$ la distribución a priori. La distribución posterior es

$$\theta|\boldsymbol{x} \sim \text{Gamma}\Big(\sum_{i=1}^{n} x_i + 1, \frac{1}{n + \frac{1}{\lambda}}\Big)$$

Así un IC (de credibilidad) del $100(1-\alpha)\,\%$ para θ es $[t_{\alpha/2}, t_{1-\alpha/2}]$ donde $t_{\alpha/2}$ y $t_{1-\alpha/2}$ son los cuantíles $\alpha/2$ y $1-\alpha/2$ de la distribución Gamma $\left(\sum_{i=1}^{n} x_i + 1, \frac{1}{n+\frac{1}{\lambda}}\right)$.

Ejemplo: Sea $X_1, \dots, X_n$ una muestra de va iid $P(\theta)$ y $\theta \sim \mathrm{Exp}(\lambda)$ la distribución a priori. La distribución posterior es

$$\theta|\boldsymbol{x} \sim \mathrm{Gamma}\Big(\sum_{i=1}^{n} x_i + 1, \frac{1}{n + \frac{1}{\lambda}}\Big)$$

Así un IC (de credibilidad) del $100(1-\alpha)\,\%$ para θ es $[t_{\alpha/2}, t_{1-\alpha/2}]$ donde $t_{\alpha/2}$ y $t_{1-\alpha/2}$ son los cuantíles $\alpha/2$ y $1-\alpha/2$ de la distribución Gamma $\left(\sum_{i=1}^{n} x_i + 1, \frac{1}{n+\frac{1}{\lambda}}\right)$.

Por ejemplo, si, $n = 25, \sum_{i=1}^{25} x_i = 112$, y $\lambda = 2$, la distribución posterior es $\mathrm{Gamma}(113, 0{,}0392)$, y un IC del $95\,\%$ para θ es $[3{,}65, 5{,}28]$.

Bibliografia

Banerjee S., Carlin B., and Gelfand A. (2015) *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman Hall/CRC.

Cressie, N. (1993) *Statistics for Spatial Data*. New York: Wiley.

Matheron, G., (1965) Les variables régionalisés et leur estimation. Masson, Paris.

Shabenberger, O., Gotway, C. A. (2005) *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.

¿Alguna Consulta?