Notas Curso de Estadística (Parte I)

Maikol Solís

Actualizado el 11 August, 2020

Índice general

Capítulo 1

Introducción

Capítulo 2

Inferencia estadística

Definición: Hacer afirmaciones probabilísticas respecto a (acerca de) cantidades desconocidas.

2.1. Ejemplo

***Pregunta**: ¿Será posible modelar cuánto dura un componente electrónico en fallar?

Solución: Podemos responder esta pregunta dividiéndola en dos partes:

1. **Modelo probabilístico:** Asuma que los tiempos de vida del componente son exponenciales (en años).
2. **Parámetro:** Sea $\theta > 0$ la tasa de fallo (unidades: 1/Tiempo(años)).

Es decir, tenemos un modelo (exponencial) y estamos decretando que su información estará concentrada en el parámetro θ.

Nota: El parámetro θ contiene la información del modelo, pero ¿Cómo obtenemos esa información

Muestra: Secuencia (sucesión) de variables aleatorias independientes $X_1, X_2, \ldots, X_n, \ldots$. Tomemos una muestra $X_1, X_2, \ldots, X_n, \ldots \overset{i.i.d}{\sim} \text{Exp}(\theta)$.

Objetivos

- Estimar $X_m, X_{m+1}, \ldots$ si se observa $X_1, X_{m-1}, \ldots$ (Predicción).

- Estimar θ usando información.

Datos: Realizaciones de variables aleatorias $X_1, \ldots, X_m$ pertenecientes a la muestra.

Estimación de θ

Dado que $\mathbb{E}(X) = \dfrac{1}{\theta}$ con $X \sim \text{Exp}(\theta)$, por la ley de grandes números se tiene que

$$\underbrace{\frac{1}{n}\sum_{i=1} X_i}_{\bar{X}_n} \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}(X) = \frac{1}{\theta}$$

por propiedad de convergencia en probabilidad.

Un posible candidato para estimar θ es $\dfrac{1}{\bar{X}_n}$, bajo el supuesto por Ley de Grandes Números que θ es una constante (frecuentista).

Realidad: θ no necesariamente es determinístico (factores externos, por la naturaleza del fenómeno).

Asumimos un modelo probabilístico para θ (tasa siempre positiva):

$$\theta \sim \Gamma(\alpha_0, \beta_0)$$

Luego, según estudios previos la tasa esperada es 0.5/año

$$\mathbb{E}(\theta) = \frac{1}{2} = \frac{\alpha_0}{\beta_0}.$$

Un primer indicio de que se podría establecer que $\alpha_0 = 1$ y de $\beta_0 = 2$.

2.2. Modelo estadístico

Vamos a definir como típicamente se define un modelo estadístico.

1. Variables aleatorias observables / hipotéticamente observables:

$$\underbrace{X_t}_{\text{Observable}} = \underbrace{Y_t}_{\text{Hip. observable}} + \underbrace{\epsilon}_{\text{Ruido}}$$

En otras palabras Y_t sería la el dato *"verdadero"* que pasó exactamente en el fenómeno analizado. Esta observación es afectada por muchos factores no observables (por ejemplo: errores de medición, cambio de las condiciones de la economía, etc.). La variable ϵ captura toda esa aleatoriedad que no es parte del fénomeno.

Claramente ni Y_t ni ϵ se pueden medir y la mejor representación del nuestro es fenómeno es a partir de X_t.

2. Distribución conjunta de una muestra de variables observables.

   Es decir cuál es el supuesto general que estoy usando para describir mis observaciones.

3. Parámetros que son hipotéticamente observables (desconocidos).

   ¿Cuál sería la mejor calibración de los componentes del modelo anterior de modo que mi modelo se ajuste a los datos?

4. (Opcional) Distribución conjunta de los parámetros.

   En el caso de Bayes, los parámetro dejan de ser simple valores puntuales y se convierten en distribuciones completas.

- **Inferencia estadística**: procedimiento que genera afirmaciones probabilísticas de un modelo estadístico.

Ejemplo de inferencias:

1. Estimar θ a través de $\dfrac{1}{\bar{X}_n}$.

2. ¿Qué tan probable es que el promedio de las siguientes observaciones es al menos 2?

$$\frac{1}{10}\sum_{i=m+1}^{m+10} X_i > 2$$

3. ¿Qué tan cierto es que $\theta \leq 0{,}4$ después de observar la muestra?

- **Parámetro**: característica (s) que determinan la distribución conjunta de las variables aleatorias de interés.
- **Espacio paramétrico** Ω (espacio de parámetros, puede ser de probabilidad)

Ejemplos:

- $\theta > 0$ (ejemplo anterior); $\Omega = (0, +\infty)$.
- $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, (μ, σ^2) parámetros; $\Omega = \mathbb{R} \times [0, +\infty)$.

Ejemplo: Clientes de un banco

¿Qué tan probable es que un cliente no pague su crédito hoy?

- **Datos**: $X_i = \begin{cases} 1 & \text{el cliente } \#i \text{ no pagó} \\ 0 & \text{el cliente } \#i \text{ pagó} \end{cases}$.
- **Muestra**: $X_1, \ldots, X_{10000}$ (realización al día de hoy).
- **Modelos**: $X_1, \ldots, X_{10000} \overset{i.i.d}{\sim} \text{Ber}(p)$ con $p \in [0, 1]$.
- **Parámetro**: p, $\Omega = [0, 1]$.
- **Inferencias:**
    - Estimar p (probabilidad de impago).
    - Suponga que $L(X_i)$ es el saldo en la cuenta del cliente $\#i$.

$$\mathbb{P}\left(\sum_{i=1}^{10000} L(X_i) > u\right) = \text{Probabilidad de ruina}$$

2.3. Estadístico

Definición. Si $X_1, \ldots, X_n$ es una muestra observable. Sea r una función real de n variables:

$$T = r(X_1, \ldots, X_n)$$

es un estadístico.

Nota: T también es aleatorio.

Ejemplos:

- $\hat{p} = \dfrac{1}{10000} \sum_{i=1}^{10000} X_i = \dfrac{\#\text{ no pagan}}{\text{Total}} = r(X_1, \ldots, X_{10000})$
- $L_m = \max L(X_i)$ (saldo del cliente más riesgoso).
- $R_m = \max L(X_i) - \min L(X_i), 1 \leq i \leq 10000$

Capítulo 3

Distribución previa (distribución a priori)

Suponga que tenemos un modelo estadístico con parámetro θ. Su θ es aleatorio entonces su densidad (antes de observar cualquier muestra) se llama **densidad previa**: π.

Ejemplo: $X_1, \ldots, X_n \sim \operatorname{Exp}(\theta)$ y θ es aleatorio tal que $\theta \sim \Gamma(\overset{\alpha}{1}, \overset{\beta}{2})$ entonces

$$\pi(\theta) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} \theta^{\alpha-1} e^{\beta\theta} = 2e^{-2\theta}, \quad \theta > 0$$

Ejemplo: Sea θ la probabilidad de obtener cara al tirar una moneda.

- Moneda justa: $\theta = \dfrac{1}{2}$ con probabilidad previa 0,8 ($\pi(\frac{1}{2}) = 0{,}8$).
- Moneda de dos caras: $\theta = 1$ con probabilidad previa 0,2 ($\pi(1) = 0{,}2$).

Notas:

- π está definida en Ω (espacio paramétrico).
- π es definida antes de obtener la muestra.

Ejemplo (Componentes eléctricos)

Criterio experto: $\mathbb{E}[\theta] = 0{,}0002^K$, $\sqrt{\operatorname{Var}(\theta)} = 0{,}0001^K$.

Bajo la misma densidad π:

$$\mathbb{E}[\theta] = \frac{\alpha}{\beta}, \mathrm{Var}(\theta) = \frac{\alpha}{\beta^2}$$

$$\implies \begin{cases} \dfrac{\alpha}{\beta} = 2 \times 10^{-4} \\ \sqrt{\dfrac{\alpha}{\beta^2}} = 1 \times 10^{-4} \end{cases} \implies \beta = 20000, \alpha = 4$$

Notación:

- $X = (X_1, \dots, X_n)$: vector que contiene la muestra aleatoria.
- Densidad conjunta de X: $f_\theta(x)$.
- Densidad de X condicional en θ: $f_n(x|\theta)$.

Supuesto: x viene de una muestra aleatoria si y solo si X es condicionalmente independiente dado θ.

Consecuencia:

$$f_n(X|\theta) = f(X_1|\theta) \cdot f(X_2|\theta) \cdots f(X_n|\theta)$$

Ejemplo

Si $X = (X_1, \dots, X_n)$ es una muestra tal que $X_i \sim \mathrm{Exp}(\theta)$,

$$f_n(X|\theta) = \begin{cases} \prod_{i=1}^n \theta e^{-\theta X_i} & \text{si } X_i > 0 \\ 0 & \text{si no} \end{cases} = \begin{cases} \theta^n e^{-\theta \sum_{i=1}^n X_i} & X_i > 0 \forall i \\ 0 & \text{si no} \end{cases}$$

3.1. Densidad posterior

Definición. Considere un modelo estadístico con parámetro θ y muestra aleatoria $X_1, \dots, X_n$. La densidad condicional de θ dado $X_1, \dots, X_n$ se llama *densidad posterior*: $\pi(\theta|X)$

Teorema. Bajo las condiciones anteriores:

$$\pi(\theta|X) = \frac{f(X_1|\theta) \cdots f(X_n|\theta)\pi(\theta)}{g_n(X)}$$

para $\theta \in \Omega$, donde g_n es una constante de normalización.

Prueba:

$$\pi(\theta|X) = \frac{\pi(\theta, X)}{\text{marginal de X}} = \frac{\pi(\theta, X)}{\int \pi(\theta, X)\, d\theta} = \frac{P(X|\theta) \cdot \pi(\theta)}{\int \pi(\theta, X)\, d\theta}$$
$$\frac{f_n(X|\theta) \cdot \pi(\theta)}{g_n(X)} = \frac{f(X_1|\theta) \cdots f(X_n|\theta)\pi(\theta)}{g_n(X)}$$

Del ejemplo anterior,

$$f_n(X|\theta) = \theta^n e^{-\theta y}, y = \sum X_i \text{ (estadístico)}$$

Numerador:

$$f_n(X|\theta)\pi(\theta) = \underbrace{\theta^n e^{-\theta y}}_{f_n(X|\theta)} \cdot \underbrace{\frac{200000^4}{3!}\theta^3 e^{-20000\cdot\theta}}_{\pi(\theta)} = \frac{20000^4}{3!}\theta^{n+3} e^{(20000+y)\theta}$$

Denominador:

$$g_n(x) = \int_0^{+\infty} \theta^{n+3} e^{-(20000+y)\theta}\, d\theta = \frac{\Gamma(n+4)}{(20000+y)^{n+4}}$$

Entonces la posterior corresponde a

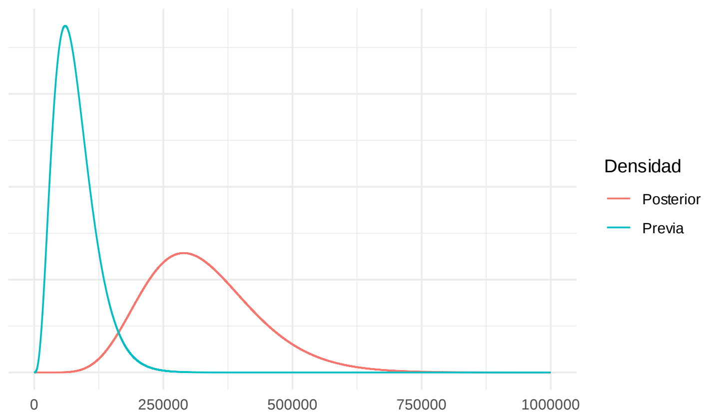
$$\pi(\theta|X) = \frac{\theta^{n+3} e^{-(20000+y)\theta}}{\Gamma(n+4)}(20000+y)^{n+4}$$

que es una $\Gamma(n+4, 20000+y)$.

Con 5 observaciones (horas): 2911, 3403, 3237, 3509, 3118.

$$y = \sum_{i=1}^{5} X_i = 16478, \quad n = 5$$

por lo que $\theta|X \sim \Gamma(9, 36178)$



Es sensible al tamaño de la muestra (una muestra grande implica un efecto de la previa menor).

Hiperparámetros: parámetros de la previa o posterior.

3.2. Función de verosimilitud

Bajo el modelo estadístico anterior a $f_n(X|\theta)$ se le llama **verosimilitud** o **función de verosimilitud**.

Observación. En el caso de una función de verosimilitud, el argumento es θ.

Ejemplo.

Sea θ la proporción de aparatos defectuosos, con $\theta \in [0,1]$

$$X_i = \begin{cases} 1 & \text{falló} \\ 2 & \text{no falló} \end{cases}$$

$\{X_i\}_{i=1}^n$ es una muestra aleatoria y $X_i \sim Ber(\theta)$.

- **Verosimilitud**

$$f_n(X|\theta) = \prod_{i=1}^n f(X_i|\theta) = \begin{cases} \theta^{\sum X_i}(1-\theta)^{n-\sum X_i} & X_i = 0,1 \ \forall i \\ 0 & \text{si no} \end{cases}$$

- **Previa**:

$$\pi(\theta) = 1_{\{0\leq\theta\leq 1\}}$$

- **Posterior**:

Por el teorema de Bayes,

$$\pi(\theta|X) \propto \theta^y(1-\theta)^{n-y}\cdot 1 = \theta^{\overbrace{y+1}^{\alpha}-1}(1-\theta)^{\overbrace{n-y+1}^{\beta}-1} \implies \theta|X \sim \text{Beta}(y+1, n-y+1)$$

- **Predicción**.

Supuesto: los datos son secuenciales. Calculamos la distribución posterior secuencialmente:

$$\begin{aligned}
\pi(\theta|X_1) &\propto \pi(\theta)f(X_1|\theta) \\
\pi(\theta|X_1, X_2) &\propto \pi(\theta)f(X_1, X_2|\theta) = \pi(\theta)f(X_1|\theta)f(X_2|\theta) \text{ (por independencia condicional)} \\
&= \pi(\theta|X_1)f(X_2|\theta) \\
&\vdots \\
\pi(\theta|X_1, \ldots, X_n) &\propto f(X_n|\theta)\pi(\theta|X_1, \ldots, X_{n-1})
\end{aligned}$$

Bajo independencia condicional no hay diferencia en la posterior si los datos son secuenciales.

Luego,

$$\begin{aligned}
g_n(X) &= \int_\Omega f(X_n|\theta)\pi(\theta|X_1, \ldots, X_{n-1})\, d\theta \\
&= P(X_n|X_1, \ldots, X_{n-1}) \text{ (Predicción para } X_n)
\end{aligned}$$

Continuando con el ejemplo de los artefactos, $P(X_6 > 3000|X_1, X_2, X_3, X_4, X_5)$. Se necesita calcular $f(X_6|X)$. Dado que

$$\pi(\theta|X) = 2{,}6 \times 10^{36}\theta^8 e^{-361178\theta}$$

se tiene

$$f(X_6|X) = 2{,}6 \times 10^{36} \int_0^1 \underbrace{\theta e^{-\theta X_6}}_{\text{Densidad de } X_6} \theta^8 e^{-361780} \, d\theta = \frac{9{,}55 \times 10^{41}}{(X_6 + 36178)^{10}}$$

Entonces,

$$P(X_6 > 3000) = \int_{3000}^{\infty} \frac{9{,}55 \times 10^{41}}{(X_6 + 36178)^{10}} \, dX_6 = 0{,}4882$$

La vida media se calcula como $\frac{1}{2} = P(X_6 > u|X)$.

3.3. Familias conjugadas

Definición. Sea $X_1, \dots, X_n$ i.i.d. condicional dado θ con densidad $f(X|\theta)$. Sea ψ la familia de posibles densidades previas sobre Ω. Si, sin importar los datos, la posterior pertenece a ψ, entonces decimos que ψ es una familia conjugada de previas.

Ejemplos:

- La familia Beta es familia conjugada para muestras según una Bernoulli.
- La familia Gama es familia conjugada para muestras exponenciales.
- Para el caso Poisson, si $X_1, \dots, X_n \sim Poi(\lambda)$,entonces la familia Gamma es familia conjugada.

La función de densidad de una Poisson es $P(X_i = k) = e^{-\lambda} \frac{\lambda^k}{k!}$. La verosimilitud corresponde a

$$f_n(X|\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda_i^X}{X_i!} = \frac{e^{-n\lambda \lambda^y}}{\prod_{i=1}^n X_i}.$$

La previa de λ está definida por $\pi(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda}$. Por lo tanto, la posterior es

$$\pi(\lambda|X) \propto \lambda^{y+\alpha-1} e^{-(\beta+n)\lambda} \implies \lambda|X \sim \Gamma(y+\alpha, \beta+n)$$

- En el caso normal, si $X_1, \dots, X_n \sim N(\theta, \sigma^2)$,entonces la familia normal es conjugada si σ^2 es conocido.

Si $\theta \sim N(\mu_0, V_0^2) \implies \theta|X \sim N(\mu_1, V_1^2)$ donde,

$$\mu_1 = \frac{\sigma^2\mu_0 + nV_0^2\bar{X}_n}{\sigma^2 + nV_0^2} = \frac{\sigma^2}{\sigma^2 + nV_0^2}\mu_0 + \frac{nV_0^2}{\sigma^2 + nV_0^2}\bar{X}_n$$

Combina de manera ponderada la previa y la de los datos.

Ejemplo 7.3.6

Considere una verosimilitud Poisson(λ) y una previa

$$\pi(\lambda) = \begin{cases} 2e^{-2\lambda} & \lambda > 0 \\ 0 & \lambda \geq 0 \end{cases} \quad \lambda \sim \Gamma(1,2)$$

Supongamos que es una muestra aleatoria de tamaño n. ¿Cuál es el número de observciones para reducir la varianza, a lo sumo, a 0.01?

Por teorema de Bayes, la posterior $\lambda|x \sim \Gamma(y+1, n+2)$. Luego, la varianza de la Gamma es

$$\frac{\alpha}{\beta^2} = \frac{\sum x_i + 1}{(n+2)^2} \leq 0{,}01 \implies \frac{1}{(n+2)^2} \leq \frac{\sum x_i + 1}{(n+2)^2} \leq 0{,}01 \implies 100 \leq (n+2)^2 \implies n \geq 8$$

Teorema. Si $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ con σ^2 conocido y la previa es $\theta \sim N(\mu_0, V_0^2)$, entonces $\theta|X \sim N(\mu_1, V_1^2)$ donde

$$\mu_1 = \frac{\sigma^2\mu_0 + nV_0^2\bar{X}_n}{\sigma^2 + nV_0^2}, \quad V_1^2 = \frac{\sigma^2 V_0^2}{\sigma^2 + nV_0^2}$$

Prueba:

- **Verosimilitud**:

$$f_n(X|\theta) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i\theta)^2\right]$$

Luego,

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \theta)^2 &= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \theta)^2 \\ &= n(\bar{X} + \theta)^2 + \sum_{i=1}^{n}(X_i - \bar{X})^2 + \underbrace{2\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \theta)}_{=0 \text{ pues } \sum Xi = n\bar{X})}\end{aligned}$$

Entonces
$$f_n(X|\theta) \propto \exp\left[-\frac{n}{2\sigma^2}(\bar{X}-\theta)^2\right].$$

- **Previa**:

$$\pi(\theta) \propto \exp\left[-\frac{1}{2V_0^2}(\theta-\mu_0)^2\right].$$

- **Posterior**:

$$\pi(\theta|X) \propto \exp\left[-\frac{n}{2\sigma^2}(\bar{X}-\theta)^2-\frac{1}{2V_0^2}(\theta-\mu_0)^2\right].$$

Con μ_1 y V_1^2 definidos anteriormente, se puede comprobar la siguiente identidad:

$$-\frac{n}{\sigma^2}(\bar{X}-\theta)^2-\frac{1}{V_0^2}(\theta-\mu_0)^2=\frac{1}{V_1^2}(\theta-\mu_1)^2+\underbrace{\frac{n}{\sigma^2+nV_0^2}(\bar{X}_n-\mu_0)^2}_{\text{Constante con respecto a }\theta}$$

Por lo tanto,
$$\pi(\theta|X) \propto \exp\left[-\frac{n}{2V_1^2}(\theta-\mu_1)^2\right]$$

Media posterior:

$$\mu_1=\underbrace{\frac{\sigma^2}{\sigma^2+nV_0^2}}_{W_1}\mu_0+\underbrace{\frac{nV_0^2}{\sigma^2+nV_0^2}}_{W_2}\bar{X}_n$$

Afirmaciones:

1) Si V_0^2 y σ^2 son fijos, entonces $W_1 \xrightarrow[n\to\infty]{} 0$ (la importancia de la media empírica crece conforme aumenta n).
2) Si V_0^2 y n son fijos, entonces $W_2 \xrightarrow[\sigma^2\to\infty]{} 0$ (la importancia de la media empírica decrece conforme la muestra es menos precisa).

3) Si σ^2 y n son fijos, entonces $W_2 \xrightarrow[V_0^2 \to \infty]{} 1$ (la importancia de la media empírica crece conforma la previa es menos precisa).

Ejemplo (determinación de n)

Sean $X_1, \ldots, X_n \sim N(\theta, 1)$ y $\theta \sim N(\mu_0, 4)$. Sabemos que

$$V_1^2 = \frac{\sigma^2 V_0^2}{\sigma^2 + n V_0^2}.$$

Buscamos que $V_1 \leq 0{,}01$, entonces

$$\frac{4}{4n+1} \leq 0{,}01 \implies n \geq 99{,}75 \text{ (al menos 100 observaciones)}$$

3.4. Densidades previas impropias

Definición. Sea π una función positiva cuyo dominio está en Ω. Suponga que $\int \pi(\theta)\, d\theta = \infty$. Entonces decimos que π es una **densidad impropia**.

Ejemplo: $\theta \sim \text{Unif}(\mathbb{R})$, $\lambda \sim \text{Unif}(0, \infty)$.

Una técnica para seleccionar distribuciones impropia es sustituir los hiperparámetros previos por 0.

Ejemplo:

Se presenta el número de soldados prusianos muertos por una patada de caballo (280 conteros, unidades de combate en 20 años).

Unidades	Ocurrencias
144	0
91	1
32	2
11	3
2	4

- Muestra de Poisson: $X_1 = 0, X_2 = 1, X_3 = 1, \ldots, X_{280} = 0 \sim \text{Poi}(\lambda)$.
- Previa: $\lambda \sim \Gamma(\alpha, \beta)$.

- Posterior: $\lambda|X \sim \Gamma(y+\alpha, n+\beta) = \Gamma(196+\alpha, 280+\beta)$.

Sustituyendo,

$$\alpha = \beta = 0 \implies \pi(\lambda) \propto \lambda_{\alpha-1} e^{-\lambda\beta} = \frac{1}{\lambda}$$

donde $\int_0^\infty \frac{1}{\lambda} d\lambda = \infty$.

Por teorema de Bayes,

$$\theta|X \sim \Gamma(196, 280)$$

3.5. Funciones de pérdida

Definición. Sean $X_1, \ldots, X_n$ datos observables cuyo modelo está indexado por $\theta \in \Omega$. Un estimador de θ es cualquier estadístico $\delta(X_1, \ldots, X_n)$.

Notación:

- Estimador $\to \delta(X_1, \ldots, X_n)$.
- Estimación o estimado: $\delta(X_1, \ldots, X_n)(\omega) = \delta(\overbrace{x_1, \ldots, x_n}^{datos})$

Definición. Una **función de pérdida** es una función de dos variables:

$$L(\theta, a), \quad \theta \in \Omega$$

con a un número real.

Interpretación: es lo que pierde un analista cuando el parámetro es θ y el estimador es a.

Asuma que θ tiene una previa. La pérdida esperada es

$$\mathbb{E}[L(\theta, a)] = \int_\Omega L(\theta, a)\pi(\theta)\, d\theta$$

la cual es una función de a, que a su vez es función de $X_1, \ldots, X_n$. Asuma que a se selecciona el minimizar esta esperanza. A ese estimador $a = \delta^*(X_1, \ldots, X_n)$ se le llama **estimador bayesiano**, si ponderamos los parámetros con respecto a la posterior.

$$\mathbb{E}[L(\theta, \delta^*)|X] = \int_\Omega L(\theta, a)\pi(\theta)\, d\theta = \min_a \mathbb{E}[L(\theta|a)X].$$

3.5.1. Función de pérdida cuadrática

$$L(\theta, a) = (\theta - a)^2$$

En el caso en que θ es real y $\mathbb{E}[\theta|X]$ es finita, entonces

$$\delta^*(X_1, \ldots, X_n) = \mathbb{E}[\theta|X] \text{ cuando } L(\theta, a) = (\theta - a)^2.$$

Ejemplo: $X_1, \ldots, X_n \sim \text{Ber}(\theta)$, $\theta \sim \text{Beta}(\alpha, \beta) \implies \theta|X \sim \text{Beta}(\alpha + y, \beta + n - y)$.

El estimador de θ es

$$\delta^*(X_1, \ldots, X_n) = \frac{\alpha + y}{\alpha + \beta + n} = \overbrace{\frac{\alpha}{\alpha + \beta}}^{\text{Esperanza previa}} \cdot \frac{\alpha + \beta}{\alpha + \beta + n} + \overbrace{\frac{y}{n}}^{\bar{X}} \cdot \frac{n}{\alpha + \beta + n}.$$

3.5.2. Función de pérdida absoluta

$$L(\theta, a) = |\theta - a|$$

La pérdida esperada es

$$f(a) = \mathbb{E}[L(\theta, a)|X] = \int_{-\infty}^{+\infty} |\theta - a| \pi(\theta|X)\, d\theta = \int_{a}^{+\infty} (\theta - a) \pi(\theta|X)\, d\theta + \int_{-\infty}^{a} (a - \theta) \pi(\theta|X)\, d\theta$$

Usando el teorema fundamental del cálculo,

$$F_\pi(a|X) = \int_{-\infty}^{\hat{a}} \pi(\theta|X)\, d\theta = \frac{1}{2} \Leftrightarrow \hat{a} = \underset{a}{\operatorname{argmin}}\, f(a)$$

La **mediana** es el punto de $X_{0,5}$ tal que $F(X_{0,5}) = \dfrac{1}{2}$.

Corolario. Bajo la función de pérdida absoluta, el estimador bayesiano es la mediana posterior.

Ejemplo: Bernoulli.

$$\frac{1}{\text{Beta}(\alpha + y, \beta + n - y)} \int_{-\infty}^{X_{0,5}} \theta^{\alpha + y - 1} (1 - \theta)^{\beta + n - y - 1}\, d\theta = \frac{1}{2}$$

Resuelva para $X_{0,5}$.

3.5.3. Otras funciones de pérdida

- $L(\theta, a) = |\theta - a|^k$, $k \neq 1, 2$, $0 < k < 1$.
- $L(\theta, a) = \lambda(\theta)|\theta - a|^2$ ($\lambda(\theta)$ penaliza la magnitud del parámetro).
- $L(\theta, a) = \begin{cases} 3(\theta - a)^2 & \theta \leq a \text{ (sobreestima)} \\ (\theta - a)^2 & \theta \geq a \text{ (subestima)} \end{cases}$

3.6. Efecto de muestras grandes

Ejemplo: ítemes malos (proporción: θ), $\theta \in [0, 1]$. Función de pérdida cuadrática. El tamaño de muestra son $n = 100$ ítemes, de los cuales $y = 10$ están malos.

$$X_1, \ldots, X_n \sim \text{Ber}(\theta)$$

- Primer previa. $\alpha = \beta = 1$ (Beta). El estimador bayesiano corresponde a

$$\mathbb{E}[\theta|X] = \frac{\alpha + y}{\alpha + \beta + n} = \frac{1 + 10}{2 + 100} = 0{,}108$$

- Segunda previa. $\alpha = 1, \beta = 2 \implies \pi(\theta) = 2e^{-2\theta}, \theta > 0$.

$$\mathbb{E}[\theta|X] = \frac{1 + 10}{1 + 2 + 100} = \frac{11}{103} = 0{,}107$$

La media es $\bar{X}_n = \dfrac{10}{100} = 0{,}1$.

3.7. Consistencia

Definición. Un estimador de θ $\delta(X_1, \ldots, X_n)$ es consistente si

$$\delta(X_1, \ldots, X_n) \xrightarrow[n\to\infty]{\mathbb{P}} \theta.$$

Bajo pérdida cuadrática, $\mathbb{E}[\theta|X] = W_1\mathbb{E}[\theta] + X_2\bar{X}_n = \delta^*$. Sabemos, por ley de grandes números, que $\bar{X}_n \xrightarrow[n\to\infty]{\mathbb{P}} \theta$. Además, $W_1 \xrightarrow[n\to\infty]{} 0$ y $W_2 \xrightarrow[n\to\infty]{} 1$.

En los ejemplos que hemos analizado

$$\delta^* \xrightarrow[n\to\infty]{\mathbb{P}} \theta$$

Teorema. Bajo condiciones generales, los estimadores bayesianos son consistentes.

Estimador. Si $X_1, \ldots, X_n$ es una muestra en un modelo indexado por θ, $\theta \in \Omega$ (k-dimensiones), sea

$$h : \Omega \to H \subset \mathbb{R}^d.$$

Sea $\psi = h(\theta)$. Un **estimador** de ψ es un estadístico $\delta^*(X_1, \ldots, X_n) \in H$. A $\delta^*(X_1, \ldots, X_n)$ estimador de ψ se puede evaluar y construir estimadores nuevos.

Ejemplo. $X_1, \ldots, X_n \sim \text{Exp}(\theta)$, $\theta|X \sim \Gamma(\alpha, \beta) = \Gamma(4, 8{,}6)$. La característica de interés es $\psi = \dfrac{1}{\theta}$, el valor esperado del tiempo de fallo.

Es estimador se calcula de la siguiente manera:

$$\begin{aligned}
\delta^*(x) = \mathbb{E}[\psi|x] &= \int_0^\infty \frac{1}{\theta} \pi(\theta|x)\, d\theta \\
&= \int_0^\infty \frac{1}{\theta} \frac{8{,}6^4}{\Gamma(4)} \theta^3 e^{-8{,}6\theta}\, d\theta \\
&= \frac{8{,}6^4}{6} \underbrace{\int_0^\infty \theta^2 e^{-8{,}6\theta}\, d\theta}_{\frac{\Gamma(3)}{8{,}6^3}} \\
&= \frac{8{,}6^4}{6} \frac{2}{8{,}6^3} = 2{,}867 \text{ unidades de tiempo.}
\end{aligned}$$

Por otro lado, vea que $\mathbb{E}(\theta|X) = \dfrac{4}{8{,}6}$. El estimador *plug-in* correspondería a

$$\frac{1}{\mathbb{E}(\theta|X)} = \frac{8{,}6}{4} = 2{,}15.$$