CAPÍTULO 1. INTRODUCCIÓN

El problema de inferencia

$$X|\boldsymbol{\theta} \sim f(\cdot)$$

Dada una muestra de datos, $\mathbf{x} = (x_1, \ldots, x_n)^T$, se quiere hacer inferencia sobre $\boldsymbol{\theta}$, por ejemplo:

- estimación puntual,
- estimación por intervalos,
- contrastes.

Además, en muchos casos, se quiere predecir los valores de nuevas observaciones.

Métodos posibles

- inferencia clásica (frecuentista): (Neyman y Pearson 1933, Fisher 1925).

- métodos basados en la verosimilitud: (Barnard 1949, Birnbaum 1962).

- inferencia fiducial: (Fisher 1930), inferencia pivotal: (Barnard 1980), inferencia estructural (Fraser 1968).

- inferencia bayesiana

Inferencia clásica

- Sólo se puede cuantificar la información proporcionada por la muestra $\mathbf{x}$ ($\boldsymbol{\theta}$ es una cantidad fija).

- La inferencia se basa en la función de verosimilitud $l(\boldsymbol{\theta}|\mathbf{x})$.

- Los procedimientos se basan en el rendimiento asintótico (debido al concepto frecuentista de probabilidad).

  - Se define una función o estimador $\mathbf{t} = \mathbf{t}(\mathbf{x})$.

  - Se mide la similitud de un valor $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ con la función de verosimilitud $f(\mathbf{t}|\boldsymbol{\theta}_0)$.

  - Si $\mathbf{t}$ no está en la cola de $f(\mathbf{t}|\boldsymbol{\theta}_0)$ entonces $\boldsymbol{\theta}_0$ es un valor razonable de $\boldsymbol{\theta}$.

- La estimación puntual se basa en elegir estimadores con buenas propiedades: insesgadez, consistencia, suficiencia, mínima varianza etc.

  Observación 1 *Asintóticamente, el estimador máximo verosimil (EMV) es insesgado, eficiente y normalmente distribuido.*

- Para estimación de intervalos, se elige un intervalo $(l(\mathbf{x}), u(\mathbf{x}))$ donde

$$P(l(\mathbf{X}) < \boldsymbol{\theta} < u(\mathbf{X})|\boldsymbol{\theta}) = 1 - \alpha$$

  para algún valor fijo α.

- Los contrastes de hipótesis tratan de rechazar $\boldsymbol{\theta}_0$ si $P(t(\mathbf{X}) > t(\mathbf{x})|\boldsymbol{\theta}_0) < \alpha$.

Principios que Justifican el Enfoque Clásico

Se usan varios principios para justificar procedimientos clásicos.

- **Suficiencia**

  Definición 1 *Un estadístico* $\mathbf{t}$ *es suficiente para* $\boldsymbol{\theta}$ *si*

  $$l(\boldsymbol{\theta}|\mathbf{x}) = h(\mathbf{t}, \boldsymbol{\theta})g(\mathbf{x}).$$

  El **Principio de Suficiencia** (Fisher 1922) dice que si existe un estadístico suficiente, $\mathbf{t}$, dada dos muestras, $\mathbf{x}_1$, $\mathbf{x}_2$, que cumplen $\mathbf{t}(\mathbf{x}_1) = \mathbf{t}(\mathbf{x}_2)$, las conclusiones basadas en $\mathbf{x}_1$ y $\mathbf{x}_2$ deben ser iguales.

  Observación 2 *Todos los métodos estandar de inferencia estadística cumplen el principio de suficiencia.*

- **Muestreo Repetido**

  El principio dice que se debe evaluar un procedimiento en términos de su rendimiento en repeticiones idénticas del experimento.

Observación 3 *Este principio es más controversial. El principio implica que las medidas de incertidumbre son frecuencias hipotéticas asintóticas. Luego, suponiendo este principio, no es posible medir la precisión de un procedimiento a posteriori, dada la muestra* $\mathbf{x}$.

Críticas a la Inferencia Clásica

- La definición de probabilidad en términos de frecuencia es muy limitada.

- Problemas con estimación.

  En primer lugar, no siempre es posible elegir un estimador insesgado y además, en algunos problemas, los estimadores insesgados pueden ser "estúpidos".

  Ejemplo 1 *Sea X una observación de una distribución Poisson: $\mathcal{P}(\lambda)$. Se quiere estimar $\phi = e^{-2\lambda}$.*

  Sólo existe un estimador insesgado que es $(-1)^X = \pm 1$. Pero $0 < e^{-2\lambda} \leq 1$ para cualquier valor de λ.

  No existe ningún estimador insesgado para $\theta = 1/\lambda$.

En segundo lugar, los EMV's tienen una justificación asintótica pero ¿qué pasa en muestras pequeñas? Por ejemplo, el sesgo de un EMV puede ser muy grande.

Ejemplo 2 *Sea* $X \sim \mathcal{UD}[1,n]$ *donde* n *es desconocido. Dada una observación, el EMV de* n *es* X*.*

Luego el sesgo del EMV es

$$E[X] - n = \frac{n+1}{2} - n = \frac{1-n}{2}.$$

que puede ser enorme si n *es grande.*

En este caso, el EMV es siempre sesgado.

- ¿Qué es un intervalo de confianza?

  Si $(1, 3)$ es un intervalo de confianza al 95 %, significa que si se repite el procedimiento muchas veces y cada vez se calcula un intervalo de confianza, el 95 % de los intervalos incluirán θ, el verdadero valor del parámetro.

  No significa que la probabilidad de que θ esté en el intervalo $(1, 3)$ es 95 %.

  La definición es una consecuencia del punto de vista de la probabilidad como frecuencia.

- Existen problemas con parámetros de molestia (*nuisance parameters*).

Ejemplo 3 *Sean* $Y_{i,j} \sim \mathcal{N}\left(\phi_i, \sigma^2\right)$ *para* $i = 1, \ldots, n$ *y* $j = 1, 2$*.*

Supongamos que el parámetro de interés es la varianza y luego $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)$ *son parámetros de molestia.*

La verosimilitud es $l\left(\sigma^2, \boldsymbol{\phi} | \textit{data}\right) \propto$

$$\sigma^{-2n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_{i,1} - \phi_i)^2 + (y_{i,2} - \phi_i)^2\right).$$

Ahora, el método más natural para estimar σ^2 *es maximizar la verosimilitud perfíl (*profile likelihood*).*

En primer lugar, suponiendo σ^2 *fijo, se maximiza la verosimilitud con respeto a* $\boldsymbol{\phi}$*.*

$$
\begin{aligned}
l_P(\sigma^2|data) &= \sup_{\phi} l\left(\sigma^2, \phi|data\right) \\
&= \sigma^{-2n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{i,1}-\bar{y}_i)^2+(y_{i,2}-\bar{y}_i)^2\right) \\
&= \sigma^{-2n}\exp\left(-\frac{1}{4\sigma^2}\sum_{i=1}^{n}(y_{i,1}-y_{i,2})^2\right)
\end{aligned}
$$

Luego, maximizando la verosimilitud perfil con respeto a σ^2, se tiene el estimador

$$
\widehat{\sigma^2} = \frac{1}{2n}\sum_{i=1}^{n}(y_{i,1}-y_{i,2})^2.
$$

No obstante,

$$
E\left[\widehat{\sigma^2}\right] = \frac{\sigma^2}{2}
$$

para cualquier valor de n y luego el estimador es inconsistente.

- ¿Cómo se hacen predicciones?

  Sea $X \sim f(X|\theta)$. Luego, dada una muestra, la distribución utilizada para predicción es típicamente $f(X|\widehat{\theta})$. No obstante, esta distribución subestima la incertidumbre sobre la predicción.

- El método de muestreo es muy importante. Contrastes clásicos no cumplen el *principio de verosimilitud*.

Inferencia basada en la función de verosimilitud

Los principios básicos son:

1. La función de verosimilitud contiene toda la información pertinente proporcionada por los datos sobre la similitud de distintas valores de $\boldsymbol{\theta}$.

2. La razón de la verosimilitud para dos valores de $\boldsymbol{\theta}$, $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}_2$ es una medida de la evidencia a favor de $\boldsymbol{\theta}_1$ relativa a $\boldsymbol{\theta}_2$ proporcionada por los datos.

El primer punto equivale al **principio de verosimilitud** (Birnbaum 1962).

Existen muchos problemas en aplicar el punto 2 en presencia de parámetros de molestia. ¿Cómo marginalizar la verosimilitud?

El Principio de Verosimilitud

Para hacer inferencia sobre θ, después de haber visto $\mathbf{x}$, toda la información pertinente (proporcionada por $\mathbf{x}$) está contenida en la función de verosimilitud $l(\boldsymbol{\theta}|\mathbf{x})$. Además, dos funciones de verosimilitud $l(\boldsymbol{\theta}|\mathbf{x}_1)$ y $l(\boldsymbol{\theta}|\mathbf{x}_2)$ proporcionan la misma información (y entonces la misma inferencia) sobre $\boldsymbol{\theta}$ si

$$l(\boldsymbol{\theta}|\mathbf{x}_1) \propto l(\boldsymbol{\theta}|\mathbf{x}_2).$$

Observación 4 *Existen otras versiones menos fuertes de este principio. Ver Lindsey (2003).*

Observación 5 *Los EMV cumplen el principio de verosimilitud.*

Observación 6 *Contrastes clásicos de significación a nivel fijo (por ejemplo $\alpha = ,05$) e intervalos de confianza no cumplen este principio.*

Ejemplo 4 $\theta = P(\textit{cruz})$ *por una moneda. Se quiere contrastar* $H_0 : \theta = 1/2$ *frente a la alternativa* $H_1 : \theta > 1/2$ *al nivel de significación de 5 %.*

Supongamos que observamos 9 cruces y 3 caras. Luego no hay datos suficientes para determinar la función de verosimilitud. Dos posibilidades son las siguientes:

1. *Se ha fijado el número de tiradas de la moneda en doce. Luego* $X = \#\ \textit{cruces}|\theta \sim \mathcal{BI}(12, \theta)$ *y*

   $$l(\theta|x = 9) = \binom{12}{9} \theta^9 (1-\theta)^3$$

   Entonces, el p-valor es

   $$p_1 = \frac{1}{2}^{12} \left[\binom{12}{9} + \cdots + \binom{12}{12} \right] \approx ,075$$

   y no se rechaza la hipótesis nula.

2. *Se ha decidido recordar el número de cruces X hasta que salga la tercera cara. Luego $X|\theta \sim \mathcal{BN}(3, \theta)$ y*

$$l(\theta|x) = \binom{11}{9} \theta^9 (1-\theta)^3 \quad \textit{y el p valor es}$$

$$\begin{aligned} p_2 &= \binom{11}{9} \theta^9 (1-\theta)^3 + \binom{12}{10} \theta^{10} (1-\theta)^3 + \cdots \\ &= ,0325 \quad \textit{y se rechaza la hipótesis nula.} \end{aligned}$$

La razón es que para hacer un contraste de significación, se tiene que especificar el espacio muestral. Es diferente en los casos diferentes que se han visto:

1. $\Omega = \{(u, d) : u + d = 12\}$

2. $\Omega = \{(u, d) : d = 3\}$

El Principio de Condicionalidad

Suponiendo que se puede hacer uno de dos experimentos E_1 y E_2 sobre $\boldsymbol{\theta}$ y que se elige uno con probabilidad 0,5, entonces la inferencia sobre $\boldsymbol{\theta}$ debe depender sólo del resultado del experimento seleccionado.

Teorema 1 *(Birnbaum 1962). El principio de verosimilitud = el principio de suficiencia + el principio de condicionalidad.*

Demostración Parcial

Demostramos que suficiencia más condicionalidad $\Rightarrow$ *verosimilitud. Ver Birnbaum (1962) para una demostración completa.*

Sea $EV(E, \mathbf{x})$ *la información sobre* $\boldsymbol{\theta}$ *proveniente del experimento* E*.*

Definimos el experimento $E^\star = E_1$ *con probabilidad* 0,5 *o* E_2 *con probabilidad* 0,5*. Los resultados del experimento* $E^\star$ *son el número del experimento y la observación* $(i, \mathbf{x}_i)$*.*

Condicionalidad implica que

$$EV[E^\star, (i, \mathbf{x})] = EV[E_i, \mathbf{x}_i].$$

Elegimos dos valores $\mathbf{x}_1^0$, $\mathbf{x}_2^0$ *donde*

$$l(\theta|\mathbf{x}_1^0) = c l(\theta|\mathbf{x}_2^0) \ \forall \ \theta$$

Definimos T *como*

$$T(i, \mathbf{x}_i) = \begin{cases} (1, \mathbf{x}_1^0) & \textit{si } i = 2, \mathbf{x}_2 = \mathbf{x}_2^0 \\ (i, \mathbf{x}_i) & \textit{sino} \end{cases}$$

Se quiere demostrar que T *es suficiente para* $\boldsymbol{\theta}$.

Si $t \neq (1, \mathbf{x}_1^0)$ *entonces*

$$P(\mathbf{X}^\star = (i, \mathbf{x}_i) | T = t, \boldsymbol{\theta}) = I_{t=(i,\mathbf{x}_i)}.$$

Si $t = (1, \mathbf{x}_1^0)$ *entonces*

$$\begin{aligned} P(\mathbf{X}^\star = (1, \mathbf{x}_1^0) | T = (1, \mathbf{x}_1^0), \boldsymbol{\theta}) &= \frac{0{,}5cl(\boldsymbol{\theta}|\mathbf{x}_2^0)}{0{,}5l(\boldsymbol{\theta}|\mathbf{x}_1^0) + 0{,}5cl(\boldsymbol{\theta}|\mathbf{x}_2^0)} \\ &= \frac{c}{1+c} \ \forall\, \boldsymbol{\theta}. \end{aligned}$$

En ambos casos, la probabilidad no depende de $\boldsymbol{\theta}$ *y entonces* T *es suficiente para* $\boldsymbol{\theta}$.

Ahora, suficiencia implica que

$$EV(E^{\star},(1,\mathbf{x}_1)) = EV(E^{\star},(2,\mathbf{x}_2))$$

es decir el principio de verosimilitud.

$\diamond$

Observación 7 *Igualmente, se puede demostrar que verosimilitud + suficiencia ⇒ condicionalidad o que verosimilitud + condicionalidad ⇒ suficiencia.*

Inferencia Fiducial etc.

Están motivadas en definir una medida probabilística a posteriori para $\boldsymbol{\theta}$ sin la necesidad de definir una distribución a priori.

Ejemplo 5 *Sea* $X|\mu,\sigma^2 \sim \mathcal{N}(\mu,\sigma^2)$*. Se quiere hacer inferencia sobre* μ *dada la muestra* $\mathbf{x}$*.*

Se sabe que $T = \frac{(\bar{X}-\mu)\sqrt{n}}{S} \sim t_{n-1}$*. Entonces para cualquier* t*,* $P(T > t) = p(t)$ *donde* $p(t)$ *es conocido.*

La idea de Fisher es

$$\begin{aligned}
p(t) &= P(T > t) \\
&= P\left(\frac{(\bar{X}-\mu)\sqrt{n}}{S} > t\right) \\
&= P\left(\mu < \bar{X} - \frac{St}{\sqrt{n}}\right)
\end{aligned}$$

y, dados los datos, se define

$$p(t) = P\left(\mu < \bar{x} - \frac{st}{\sqrt{n}}\right)$$

es la probabilidad fiducial de que μ esté menor que $\bar{x} - \frac{st}{\sqrt{n}}$.

Problemas:

- Se transfiere la medida de probabilidad del espacio muestral al espacio paramétrico. ¿Cuál es la justificación?

- ¿Qué pasa si no existe ninguna pivota?

- Muy difícil de aplicar en problemas multidimensionales.

Observación 8 *En muchos casos, intervalos fiduciales son iguales a intervalos bayesianos dadas distribuciones a priori* no informativos. *En particular, inferencia estructural (Fraser 1968) corresponde a inferencia bayesiana con distribuciones a priori de Haar.*

No obstante, la justificación bayesiana de estos intervalos es más coherente.

Inferencia Bayesiana

- Se define una medida a priori $f(\boldsymbol{\theta})$.

  Todos tenemos nuestras propias probabilidades para cualquier suceso:

  P(cruz), P(lloverá mañana), P(nací en 1962).

  Nuestras probabilidades pueden ser diferentes, son nuestras propias medidas de verosimilitud para sucesos. La unica restricción es que sean coherentes.

- Dada la muestra x, se modifican sus creencias sobre $\boldsymbol{\theta}$ mediante **el teorema de Bayes**:

$$
\begin{aligned}
f(\boldsymbol{\theta}|\mathrm{x}) &= \frac{f(\mathrm{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathrm{x})} \\
&\propto f(\mathrm{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}) = l(\boldsymbol{\theta}|\mathrm{x})f(\boldsymbol{\theta})
\end{aligned}
$$

- siendo $l(\boldsymbol{\theta}|\mathbf{x})$ la función de verosimilitud,
- siendo $f(\boldsymbol{\theta})$ su **distribución a priori** (inicial) y
- $f(\boldsymbol{\theta}|\mathbf{x})$ su **distribución a posteriori** (final).

Observación 9 *Existe una medida directa a priori y una medida directa a posteriori del incertidumbre sobre θ.*

- El método de muestreo no importa. La inferencia bayesiana cumple el principio de verosimilitud.

Demostración

Sea $l(\boldsymbol{\theta}|\mathbf{x}_1) \propto l(\boldsymbol{\theta}|\mathbf{x}_2)$ *con una distribución a priori* $f(\boldsymbol{\theta})$*. Entonces*

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{x}_1) &\propto f(\boldsymbol{\theta})l(\boldsymbol{\theta}|\mathbf{x}_1) \\ &\propto f(\boldsymbol{\theta})l(\boldsymbol{\theta}|\mathbf{x}_2) \\ &\propto f(\boldsymbol{\theta}|\mathbf{x}_2) \\ &= f(\boldsymbol{\theta}|\mathbf{x}_2) \end{aligned}$$

$\diamond$

Observación 10 *Se supone aquí que la distribución a priori es personal y subjetiva, es decir que no se la selecciona a través de unas reglas automáticas, por ejemplo* siempre uniforme*.*

- Un intervalo de credibilidad de 95 % para θ es un intervalo en que se tiene una probabilidad del 95 % de que esté θ.

- La estimación es un problema de decisión. En situaciones diferentes se elegirían estimadores diferentes. Se utiliza **la teoría de utilidad** para elegir. Ver el tema 5.

- No hay problemas con parámetros de molestia.

  Si $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ donde $\boldsymbol{\theta}_2$ son parámetros de molestía, se puede expresar

  $$f(\boldsymbol{\theta}) = f(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) f(\boldsymbol{\theta}_2)$$

  y luego

  $$f(\boldsymbol{\theta}_1|\mathbf{x}) = \int f(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x}) f(\boldsymbol{\theta}_2|\mathbf{x})\, d\boldsymbol{\theta}_2.$$

  Observación 11 *Si $f(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \mathbf{x})$ varia mucha con valores diferentes de $\boldsymbol{\theta}_2$, entonces existe mucha sensibilidad y se debe tomarlo en cuenta. Ver Box y Tiao (1973), Sección 1.6.*

- Predicción es fácil.

  Sea Y una nueva observación. Entonces se puede calcular la distribución predictiva:

$$f(y|\mathbf{x}) = \int f(y|(\mathbf{x}), \boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta}$$

Críticas a la Teoría Bayesiana

- $\boldsymbol{\theta}$ no tiene que ser variable.

  $\boldsymbol{\theta}$ puede ser fijo pero la distribución $f(\boldsymbol{\theta})$ muestra los conocimientos de $\boldsymbol{\theta}$. Los conocimientos y luego las creencias cambian con datos.

- Falta de objetividad. ¿Cómo se puede elegir la distribución a priori?

  - El modelo no es objetivo.
  - A menudo, un análisis clásico corresponde a un análisis Bayesiano con una distribución a priori "no informativa".
  - Los aspectos subjetivos son explícitos en un análisis Bayesiano. Los bayesianos tienen que justificar sus elecciones.

- Falta de robustez. Si pequeños cambios en la distribución a priori provocan grandes cambios en la distribución a posteriori, ¿qué hacemos?

  Es imprescindible hacer **análisis de sensibilidad**. Si los resultados cambian cuando se cambia la distribución inicial, implica que la verosimilitud no proporciona mucha información y la elección de la distribución a priori es fundamental.

  Con el uso de la metodología robusta y distribuciones a priori robustas, se pueden reducir los problemas. Ver el capítulo 12.

Análisis Bayesiano del Ejemplo 4.

Se recuerde la definición de la distribución beta.

Definición 2 *θ tiene una distribución beta con parámetros α, β ($\theta \sim \mathcal{B}(\alpha, \beta)$) si*

$$f(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

por $0 < \theta < 1$ donde $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

La media y varianza son $E[\theta] = \frac{\alpha}{\alpha+\beta}$ y $V[\theta] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Ejemplo 6 *Para ilustrarlo, se supone una distribución a priori uniforme $\theta \sim \mathcal{U}(0, 1)$.*

Observación 12 *La distribución uniforme es una distribución beta con $\alpha = \beta = 1$.*

Observación 13 *Es poco real, significa que no se sabe mucho de la moneda. Sería mejor una distribución beta simétrica, por ejemplo* $\mathcal{B}(5,5)$.

La distribución a posteriori es

$$\begin{aligned} f(\theta|x) &\propto \binom{12}{9}\theta^9(1-\theta)^3 \\ &\propto \theta^9(1-\theta)^3 \\ &\propto \theta^{10-1}(1-\theta)^{4-1} \end{aligned}$$

que significa que $\theta|\mathbf{x} \sim \mathcal{B}(10,4)$.

Se puede demostrar que $P(\theta \leq 1/2|\mathbf{x}) \approx ,046$.

Observación 14 *No es un contraste formal. Se estudiaran métodos formales más adelante.*

Observación 15 *Tanto la distribución a priori como la distribución es una distribución beta. Se trata situaciones parecidas en el Tema 2.*

La media a posteriori está entre la media a priori y la EMV

Ejemplo 7 *Retomando el Ejemplo 6 también se puede calcular la media a posteriori de θ.*

Mediante las propiedades de la distribución beta,

$$E[\theta|x] = \frac{10}{10+4} = \frac{5}{7} \quad \text{y además}$$

$$\frac{5}{7} = \frac{1}{7} \times \frac{1}{2} + \frac{6}{7} \times \frac{9}{12}$$

que implica que

$$E[\theta|x] = \frac{1}{7}E[\theta] + \frac{6}{7}\widehat{\theta}$$

donde $E[\theta] = 1/(1+1) = 1/2$ es la media a priori y $\widehat{\theta} = 9/12$ es el EMV de θ.

Una interpretación es que los datos tienen seis veces la importancia de la distribución a priori en la determinación de la distribución a posteriori o igualmente, que la información en la distribución a priori equivale a la información contenida en un experimento de dos tiradas de la moneda donde se observa una cara y una cruz.

Predicción

Dados los datos $\mathbf{x} = (x_1, \ldots, x_n)^T$, supongamos que se quiere predecir el valor de X_{n+1}. Luego, se calcula **la distribución predictiva**

$$f(x_{n+1}|\mathbf{x}) = \int f(x_{n+1}|\boldsymbol{\theta}, \mathbf{x}) f(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta}.$$

En nuestra situación (los $X_i|\boldsymbol{\theta}$ son **intercambiables** $\approx$ independientes) esta fórmula simplifica a

$$f(x_{n+1}|\mathbf{x}) = \int f(x_{n+1}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta}.$$

Ejemplo 8 *Volviendo al Ejemplo 6, se quiere predecir el número de cruces en diez tiradas más de la misma moneda.*

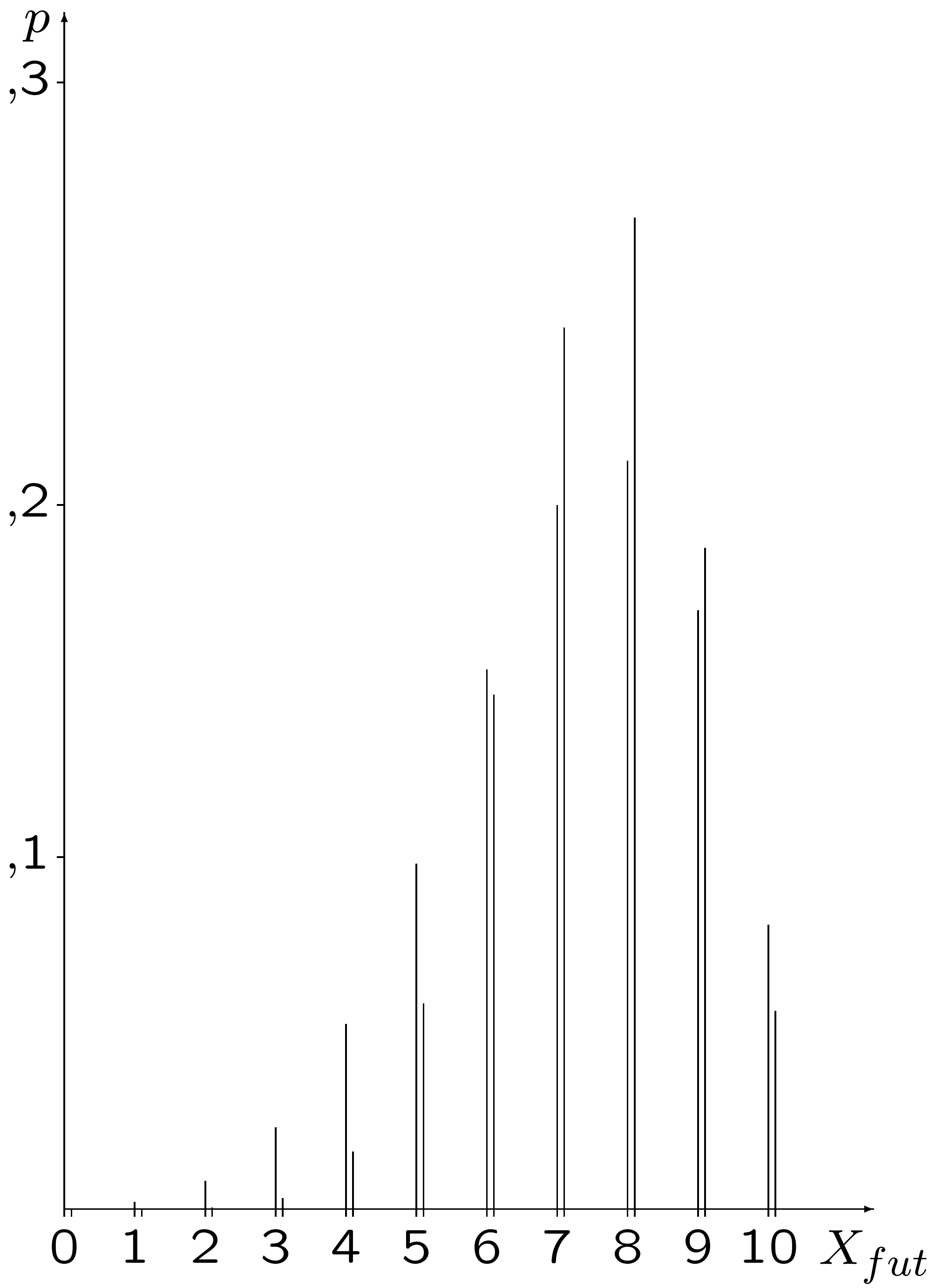
$$X_{fut}|\theta \sim \mathcal{BI}(10, \theta)$$

y entonces (usando $B(a, b)$ para representar la función beta)

$$
\begin{aligned}
f(x_{fut}|\mathbf{x}) &= \int_0^1 \binom{10}{x_{fut}} \theta^{x_{fut}}(1-\theta)^{10-x_{fut}} \times \\
&\quad \times \frac{1}{B(10,4)} \theta^{10-1}(1-\theta)^{4-1}\, d\theta \\
&= \binom{10}{x_{fut}} \frac{1}{B(10,4)} \times \\
&\quad \times \int_0^1 \theta^{10+x_{fut}-1}(1-\theta)^{14-x_{fut}-1}\, d\theta \\
&= \binom{10}{x_{fut}} \frac{B(10+x_{fut}, 14-x_{fut})}{B(10,4)}
\end{aligned}
$$

la llamada **distribución beta-binomial.**

El diagrama ilustra las probabilidades predictivas de X_{fut} y las probabilidades calculadas usando la EMV ($\mathcal{BI}(10,,75)$).



p
,3
,2
,1
0
1
2
3
4
5
6
7
8
9
10
X_{fut}

La Media y Varianza Predictiva

Se puede evaluar la media de $X_{fut}|\mathbf{x}$ sin tener que evaluar la distribución predictiva utilizando

$$E[Z] = E[E[Z|Y]]$$

para variables Z e Y.

Ejemplo 9 *Retomando el Ejemplo 8, tenemos* $E[X_{fut}|\theta] = 10\theta$.

$E[\theta|\mathbf{x}] = \frac{5}{7}$, *luego*

$$E[X_{fut}|\mathbf{x}] = 10 \times \frac{5}{7} \approx 7{,}141$$

Para evaluar la varianza predictiva, se usa la fórmula

$$V[Z] = E[V[Z|Y]] + V[E[Z|Y]]$$

La verosimilitud escalada

De vez en cuando (si θ es unidimensional), se quiere ver la influencia de la distribución inicial y de la verosimilitud en la distribución final. Para hacer eso, es útil calcular la verosimilitud escalada

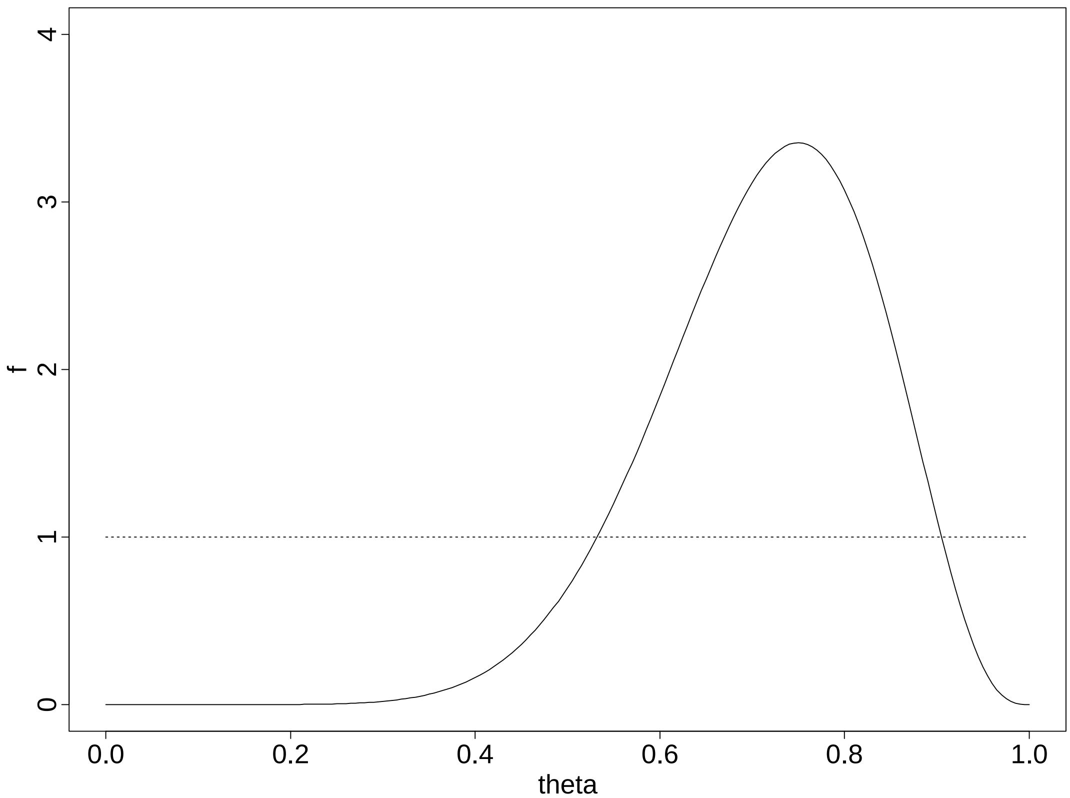
$$\frac{l(\theta|\mathbf{x})}{\int l(\theta|\mathbf{x})\, d\theta}$$

Observación 16 *No es cierto que exista.*

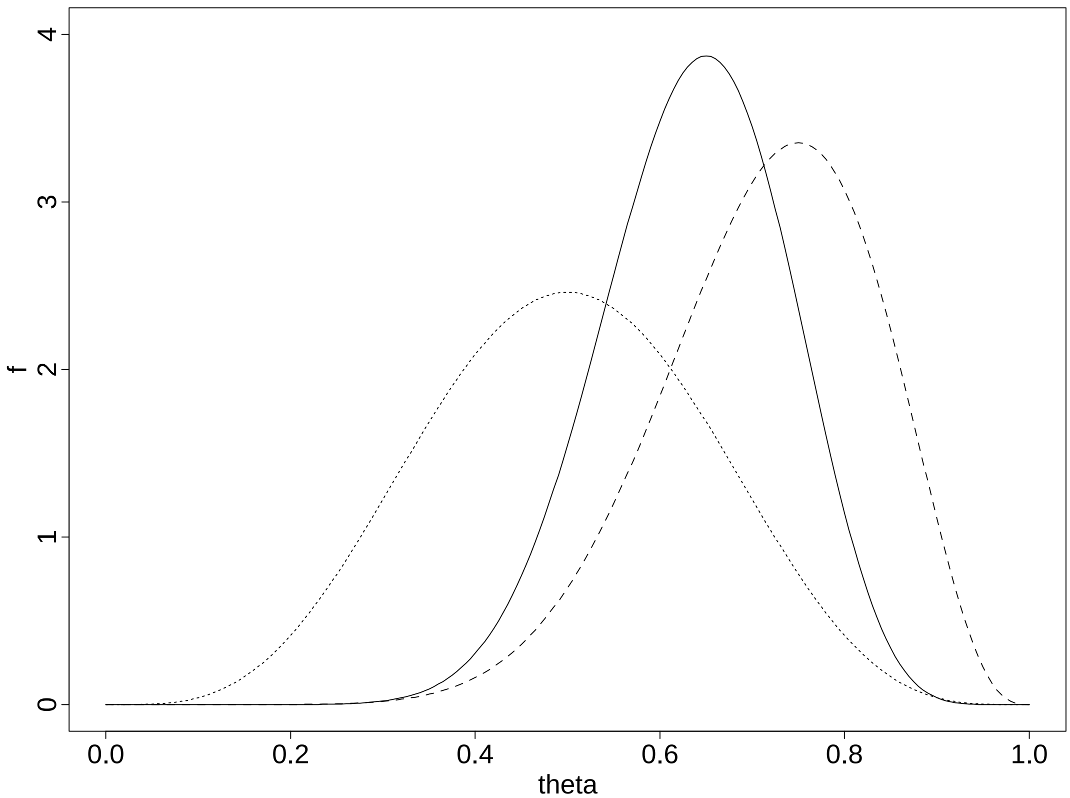
Dada la verosimilitud escalada, se puede hacer un diagrama mostrando la distribución a priori, la distribución a posteriori y la verosimilitud escalada para ver la relación entre las tres.

Ejemplo 10 *Veamos unos diagramas de las distribuciones iniciales y finales y la verosimilitud escalada usando distribuciones iniciales diferentes.*

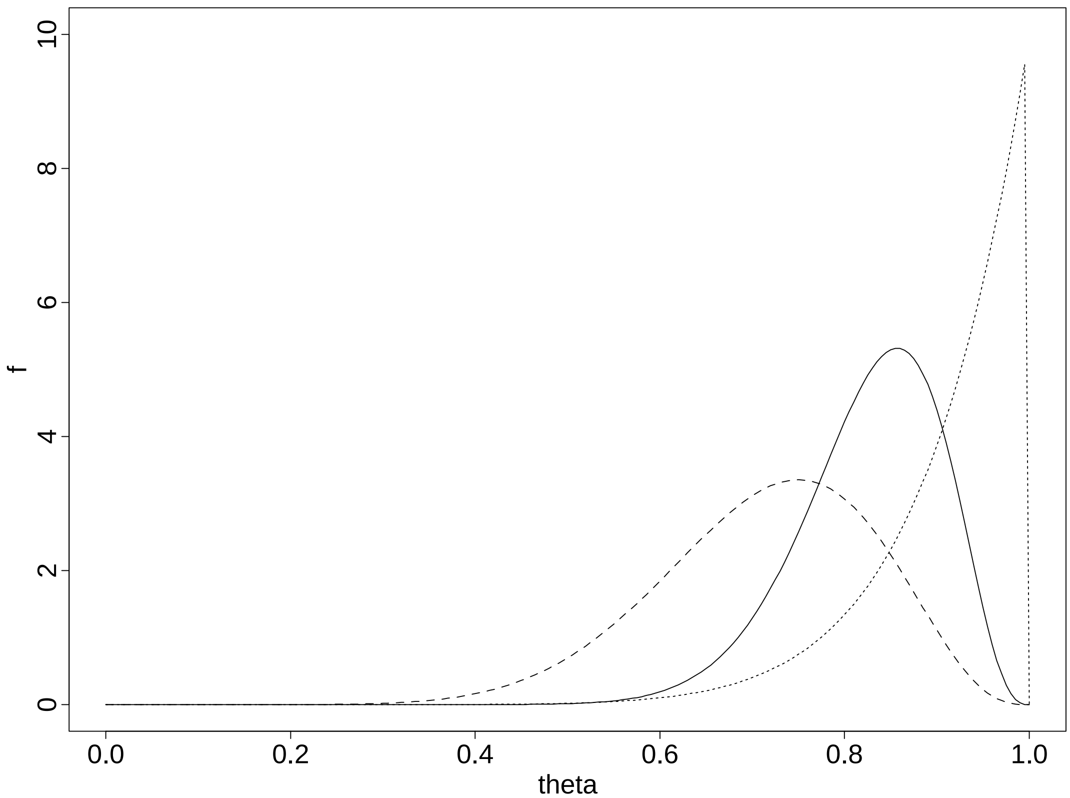
1) con distribución a priori uniforme.



2) con distribución a priori $\mathcal{B}(5,5)$.



3) con distribución a priori $\mathcal{B}(10, 1)$.



4) con distribución a priori $\mathcal{B}(1,10)$.