Estadística Bayesiana: Paradigma Bayesiano

Lizbeth Naranjo Albarrán
lizbethna@ciencias.unam.mx
Facultad de Ciencias, UNAM

Foro de Matemáticas y Actuaría
Universidad del Mar, Huatulco
24-25 Abril 2025

Contenido

Introducción

- En los últimos años ha habido un mayor interés en el desarrollo de métodos de inferencia Bayesiana.
- La razón principal es que la metodología Bayesiana proporciona un paradigma completo para la inferencia estadística bajo incertidumbre, que permite combinar información derivada de observaciones con información obtenida de expertos.
- Una gran ventaja del enfoque Bayesiano es que se puede usar información inicial, la cual se incorpora al modelo a través de distribuciones de probabilidad.
- Esto puede ser muy útil en situaciones específicas donde se puede obtener conocimiento de expertos o información histórica.

Limitaciones de la estadística frecuentista

Algunas limitaciones de la estadística frecuentista, y **que le dan una posible ventaja a la estadística Bayesiana**, son:

- No permite incorporar en el análisis estadístico información extra muestral disponible.
- Muchos de los métodos se apoyan en resultados asintóticos, la ley de los grandes números, el teorema central del límite, etc.
- Se necesitan muestras grandes para que los resultados sean confiables.
- Si no hay datos, la estadística frecuentista no es posible.
- Si hay pocos datos, la estadística frecuentista presenta problemas.
- La estadística Bayesiana usa información disponible a partir de los datos de muestras, y a partir de la información extra muestral.
- La estadística frecuentista solo usa dos hipótesis, mientras que en la Bayesiana es posible considerar k hipótesis a la vez.

Enfoque de la Probabilidad

- Estadística es un conjunto de técnicas para describir un fenómeno, a partir de un conjunto de datos que presentan variabilidad.
- Es un conjunto de métodos para alcanzar conclusiones acerca de una o varias características de interés de una población a partir de información parcial provista por una muestra de dicha población.
- La estadística se ocupa de los métodos científicos para recolectar, organizar, resumir, presentar y analizar datos usando modelos, así como de obtener conclusiones válidas y tomar desiciones con base en ese análisis.
- La estadística es la rama de la matemática que utiliza conjuntos de datos para obtener inferencias basadas en el cálculo de probabilidades.

Enfoque de la Probabilidad

Enfoque clásico

- Sea $(\Omega, \mathcal{F}, \mathbb{P})$ espacio de probabilidad, la probabilidad de que ocurra el evento A se define como:

  $$\mathbb{P}(A) = \frac{|A|}{n},$$

  donde $|A|$ denota la cardinalidad de A.
- Si un experimento o un fenómeno puede ocurrir de n maneras diferentes, mutuamente excluyentes e igualmente probables.
- Ej: lanzar un dado. Sea A el evento de obtener un 3, donde $n = 6$, entonces $\mathbb{P}(A) = \frac{1}{6}$.

Enfoque de la Probabilidad

Enfoque frecuentista

- La probabilidad de que ocurra el evento A se define como:

$$\mathbb{P}(A) = \lim_{n\to\infty} \frac{f_A(n)}{n},$$

  donde $f_A(n)$ es el número de veces que ocurre el evento A en n repeticiones idénticas e independientes. Se basa en datos observados.
- El límite $\lim_{n\to\infty} \frac{f_A(n)}{n}$ equivale a decir que $\forall \epsilon > 0$ existe k tal que si $n > k$ entonces $\left|\frac{f_A(n)}{n} - \mathbb{P}(A)\right| < \epsilon$.
- Ej: lanzar un dado. Sea A el evento de obtener un 3.
- Para calcular $\mathbb{P}(A)$ se lanzan $n = 600$ veces el dado, y se observa que A ocurre $f_A(600) = 106$ veces, entonces $\mathbb{P}(A) = \frac{106}{600}$.

Enfoque de la Probabilidad

Enfoque subjetivo

- La probabilidad de que ocurra el evento A es una medida del grado de creencia que tiene un individuo de la creencia de A, con base en la información K que dicho individuo posee, es decir,
$$\mathbb{P}(A) = \mathbb{P}(A|K).$$
- Ej: sea A el evento de que hoy llueva en la ciudad de México.
- Considerando que tenemos dos posibles escenarios de información:
  - K_1 es un individuo que vive en Wuhan (China) y No tiene acceso a internet (No tiene información);
  - K_2 es un individuo que vive en la Ciudad de México (tiene información, puede dar un vistazo).

Teorema de Bayes

- Sea $(\Omega, \mathcal{F}, \mathbb{P})$ espacio de probabilidad. Sean $A, B \in \mathcal{F}$, tal que $\mathbb{P}(B) > 0$. La probabilidad condicional del evento A dado el evento B se define como

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

- Sea $(\Omega, \mathcal{F}, \mathbb{P})$ espacio de probabilidad. Sean $A, B \in \mathcal{F}$, tal que $\mathbb{P}(A) > 0$.
- El teorema de Bayes toma la forma:

$$\begin{aligned} \mathbb{P}(B|A) &= \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)} \end{aligned}$$

Teorema de Bayes

- De manera general, sea $\{B_1, \ldots, B_n\}$ un conjunto de sucesos mutuamente excluyentes y exhaustivos contenidos en $\mathcal{F}$, tales que $\mathbb{P}(B_j) > 0 \ \forall j = 1, \ldots, n.$. Sea $A \in \mathcal{F}$ un evento tal que $\mathbb{P}(A) > 0$. La Regla de Bayes se define como:

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{j=1}^{n} \mathbb{P}(A|B_j)\mathbb{P}(B_j)}$$

Ejemplo Teorema de Bayes: COVID-19

- Suponga que un grupo de ingenieros biomédicos han diseñado una prueba para el diagnóstico del COVID-19. Suponga que se elige a un individuo para la prueba, y consideremos los eventos de interés:

A: el evento de que la prueba diagnostique presencia de COVID-19
B: el evento de tener efectivamente COVID-19

- Los científicos presumen de que la prueba es muy buena. Probaron con un grupo de 100 portadores de COVID-19 y el 99 % dio positivo: $\mathbb{P}(A|B) = \frac{99}{100}$
- Probaron con un grupo de 100 sanos y el 98 % dio negativo: $\mathbb{P}(A^c|B^c) = \frac{98}{100}$
- Sea p la proporción de mexicanos con COVID-19: $p = \mathbb{P}(B)$.

Ejemplo Teorema de Bayes: COVID-19

- Utilizando la regla de Bayes, buscamos
  $\mathbb{P}(B|A) = \mathbb{P}($ *tener COVID-19* $|$ *diagnóstico positivo* $)$.

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)} = \frac{\frac{99}{100}p}{\frac{99}{100}p + \frac{2}{100}(1-p)}$$

Tabulando, para distintos valores de p:

p	$\mathbb{P}(B\|A)$
0.0001	0.0049
0.001	0.0472
0.01	0.3333
0.1	0.8462
0.5	0.9802

¿Qué sucedió? La prueba fue probada con personas que conocíamos su estado de salud. En la práctica esto NO sucede.

Ejemplo Teorema de Bayes: COVID-19

- Existen 2 tipos de pruebas de utilidad diagnóstica: las basadas en la detección del virus (RNA o antígeno viral) y las basadas en la detección de anticuerpos (IgM o IgG) frente al virus.[1]
- La tasa de detección positiva aumenta significativamente (98.6 %) cuando se combina la IgM con PCR para cada paciente en comparación con una sola prueba (Guo L, 2020).
- Repetir la prueba de manera independiente:

$$
\begin{aligned}
\mathbb{P}(B|AA) &= \frac{\mathbb{P}(AA|B)\mathbb{P}(B)}{\mathbb{P}(AA|B)\mathbb{P}(B) + \mathbb{P}(AA|B^c)\mathbb{P}(B^c)} \\
&= \frac{\mathbb{P}(A|B)\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(A|B^c)\mathbb{P}(B^c)} \\
&= \frac{\frac{99}{100}\frac{99}{100}p}{\frac{99}{100}\frac{99}{100}p + \frac{2}{100}\frac{2}{100}(1-p)}
\end{aligned}
$$

[1] https://www.fisterra.com/guias-clinicas/covid-19/

Ejemplo Teorema de Bayes: COVID-19

- Otra opción de calcularlo es usar la información de la primera prueba y solo actualizarlo. Sea

D: el evento de tener efectivamente COVID-19, dado que la prueba diagnosticó presencia de COVID-19 $\equiv B \mid A$

$$\mathbb{P}(D) = \mathbb{P}(B \mid A)$$

$$\mathbb{P}(D|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(D)}{\mathbb{P}(A|B)\mathbb{P}(D) + \mathbb{P}(A|B^c)\mathbb{P}(D^c)}$$

`Código R:` Bayes1_1ReglaBayes.R

Ejemplo Teorema de Bayes: COVID-19

1 prueba

p	$\mathbb{P}(B\|A)$
0.0001	0.0049
0.001	0.0472
0.01	0.3333
0.1	0.8462
0.5	0.9802

2 pruebas independientes

p	$\mathbb{P}(B\|AA)$
0.0001	0.1986
0.001	0.7104
0.01	0.9612
0.1	0.9963
0.5	0.9996

3 pruebas independient

p	$\mathbb{P}(B\|AAA)$
0.0001	0.9238
0.001	0.9918
0.01	0.9992
0.1	0.9999
0.5	0.9999

Distribución *a priori* (distribución inicial)

- Sea $X_1, \ldots, X_n$ una muestra de variables aleatorias (v.a.) de una función de densidad de probabilidad $\mathbb{P}(\boldsymbol{X}|\boldsymbol{\theta})$ con $\boldsymbol{\theta} \in \Theta$, $\Theta \subset \mathbb{R}^d$, donde $\boldsymbol{\theta}$ es el vector de parámetros y Θ es el espacio paramétrico, cuya función de *densidad conjunta* o *funtión de verosimilitud* (*likelihood*) es $\mathbb{P}(\boldsymbol{X}|\boldsymbol{\theta})$.
- En Estadística es común considerar una familia paramétrica de densidades,

$$\mathcal{P} = \{\mathbb{P}(\boldsymbol{X}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\},$$

  y proceder como si la distribución F correspondiera a alguno de los modelos en $\mathcal{P}$.
- El problema en estadística se reduce a hacer inferencias sobre el supuesto valor de $\boldsymbol{\theta}$.

Distribución *a priori* (distribución inicial)

- Particularmente, en estadística Bayesiana se modela la incertidumbre de $\boldsymbol{\theta}$ usando métodos probabilísticos, y se considera a $\boldsymbol{\theta}$ como aleatoria, en específico, se usa el Teorema de Bayes:

$$\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{\mathbb{P}(\boldsymbol{\theta})\mathbb{P}(\boldsymbol{X}|\boldsymbol{\theta})}{\mathbb{P}(\boldsymbol{X})},$$

donde $\mathbb{P}(\boldsymbol{X}) = \int_{\Theta} \mathbb{P}(\boldsymbol{\theta})\mathbb{P}(\boldsymbol{X}|\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ no depende de $\boldsymbol{\theta}$, por lo que es común escribir:

$$\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X}) \propto \mathbb{P}(\boldsymbol{\theta})\mathbb{P}(\boldsymbol{X}|\boldsymbol{\theta}).$$

Distribución *a priori* (distribución inicial)

- La distribución de probabilidad inicial (*a priori*) $\mathbb{P}(\boldsymbol{\theta})$ describe la información inicial que se tiene de $\boldsymbol{\theta}$.
- Esta distribución se basa en experiencia previa (información histórica, experiencia de expertos en los datos u otra información adicional).
- Entonces, la distribución de probabilidad final (*a posteriori*) $\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X})$ permite incorporar información contenida en la muestra a través de $\mathbb{P}(\boldsymbol{X}|\boldsymbol{\theta})$ e información inicial a través de $\mathbb{P}(\boldsymbol{\theta})$.
- Por tanto, en estadística Bayesiana las inferencias sobre el parámetro $\boldsymbol{\theta}$ se basan en calcular la distribución final $\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X})$.

Proporción de enfermos

- Se examina una población, y θ es la proporción de enfermos. Un día, se examinan 10 sujetos, denotados por $X_1, \ldots, X_{10}$, donde $X_i = 1$ indica que el sujeto i está enfermo y $X_i = 0$ está sano.
- Esto puede verse como una muestra aleatoria con distribución Bernoulli de parámetro θ, con fdp $f_X(x; \theta) = \theta^x (1-\theta)^{1-x} I_{\{0,1\}}(x)$ para $0 \leq \theta \leq 1$.
- ¿Cuál es el valor de θ?
  Personas enfermas: $X_i \sim \mathrm{Bernoulli}(\theta)$, con $\theta \in [0, 1]$.

Proporción de enfermos

- Suponga que el experto tiene información inicial acerca de θ, y que ha observado que a lo largo de los días la proporción de enfermos cambia como una v.a. con fdp $\pi(\theta) = 6\theta(1-\theta)I_{[0,1]}(\theta)$, es decir, $\theta \sim Beta(2,2)$ (distribución a priori).
- Entonces la distribución a posteriori de θ es:

$$
\begin{aligned}
P(\theta|x_1,\ldots,x_n) &= \frac{P(\boldsymbol{\theta})P(x_1,\ldots,x_n|\boldsymbol{\theta})}{P(x_1,\ldots,x_n)} \\
&= \frac{\prod_{i=1}^{10}\theta^{x_i}(1-\theta)^{1-x_i}\times 6\theta(1-\theta)}{\int_0^1\prod_{i=1}^{10}\theta^{x_i}(1-\theta)^{1-x_i}\times 6\theta(1-\theta)d\theta} \\
= \cdots &= \frac{\Gamma(12)\ \times\ \theta^{\sum_{i=1}^{10}x_i+1}(1-\theta)^{11-\sum_{i=1}^{10}x_i}}{\Gamma(\sum_{i=1}^{10}x_i+1)\Gamma(11-\sum_{i=1}^{10}x_i)}
\end{aligned}
$$

que tiene una densidad $Beta\left(\sum_{i=1}^{10}x_i+2,\ 12-\sum_{i=1}^{10}x_i\right)$.

Proporción de enfermos

- Porque la distribución $\theta \sim Beta(a, b)$ tiene una función de densidad

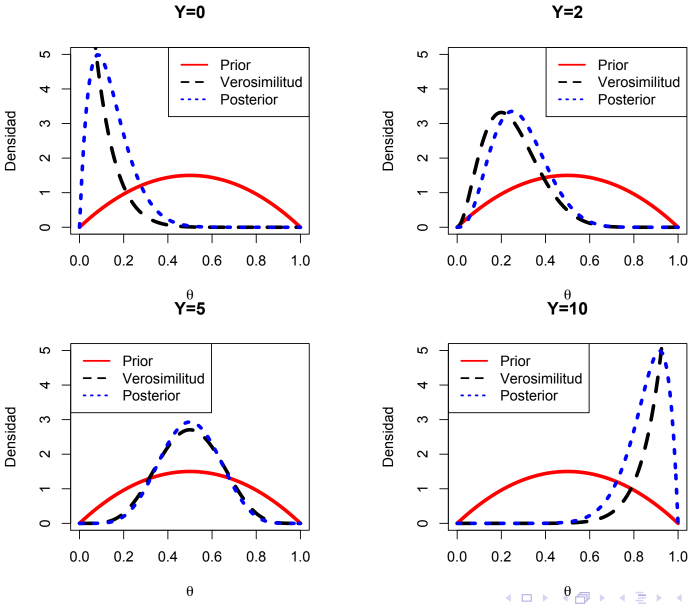
$$
\begin{aligned}
P(\theta) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \\
\int_0^1 P(\theta)d\theta &= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}d\theta \;=\; 1
\end{aligned}
$$

- De manera análoga:

$$
\begin{aligned}
P(\theta|x_1,\ldots,x_n) &\propto P(\boldsymbol{\theta})P(x_1,\ldots,x_n|\boldsymbol{\theta}) \\
&\propto \prod_{i=1}^{10} \theta^{x_i}(1-\theta)^{1-x_i} \times 6\theta(1-\theta) \\
&\propto \theta^{\sum_{i=1}^{10} x_i+1}(1-\theta)^{10-\sum_{i=1}^{10} x_i+1} \\
&= \frac{\Gamma(12)}{\Gamma(\sum_{i=1}^{10} x_i+1)\Gamma(11-\sum_{i=1}^{10} x_i)}\theta^{\sum_{i=1}^{10} x_i+1}(1-\theta)^{11-\sum_{i=1}^{10} x_i}
\end{aligned}
$$

`Código R:` Bayes2_1DistribucionBernoulli.R

Proporción de enfermos



Análisis Conjugado

- Tanto $p(\theta)$ como $p(\theta|x_1, \ldots, x_n)$ son distribuciones de probabilidad sobre el parámetro θ.
- La primera distribución sólo describe la información inicial y la segunda distribución actualiza dicha información usando también la información muestral que se pueda obtener.
- Resulta conveniente tanto para el análisis como desde el punto de vista computacional, que $p(\theta)$ y $p(\theta|x_1, \ldots, x_n)$ pertenezcan a la misma familia paramétrica.

Análisis Conjugado

- Tanto $p(\theta)$ como $p(\theta|x_1, \dots, x_n)$ son distribuciones de probabilidad sobre el parámetro θ.
- Sea $\mathcal{P} = \{p(x_1, \dots, x_n|\theta) : \theta \in \Theta\}$ una familia paramétrica. Una clase (o colección) de distribuciones de probabilidad $\mathcal{F}$ es una familia conjugada para $\mathcal{P}$ si para todo $p(x_1, \dots, x_n|\theta) \in \mathcal{P}$ y $p(\theta) \in \mathcal{F}$ se cumple que $p(\theta|x_1, \dots, x_n) \in \mathcal{F}$.
- Para garantizar que $p(\theta)$ y $p(\theta|x_1, \dots, x_n)$ pertenezcan a la misma familia general de funciones de distribución, se elige a $p(\theta)$ de tal manera que tenga la misma "estructura" de $p(x_1, \dots, x_n|\theta)$ vista como una función de θ.

Familia conjugada: Bernoulli

- Muchas aplicaciones hacen referencia a un número fijo n de observaciones binarias.
- Sean $y_1, \ldots, y_n$ los resultados de n ensayos independientes e idénticos tal que $p(Y_i = 1) = \pi$ y $p(Y_i = 0) = 1 - \pi$.
- Generalmente se etiquetan como "éxitos" y "fracasos" los resultados 1 y 0 respectivamente.
- Las variables aleatorias $\{Y_i\}$ independientes e idénticamente distribuidas se conocen como variables aleatorias Bernoulli.
- El número total de éxitos, $Y = \sum_{i=1}^{n} Y_i$, tiene una distribución binomial con índice n y parámetro π, y se denota por $Bin(y|n, \pi)$.

Familia conjugada: Bernoulli

- La función de masa de probabilidad de Y es

$$p(y|\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \qquad y = 0, 1, \ldots, n, \; 0 < \pi < 1,$$

donde el coeficiente binomial es

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}.$$

- La verosimilitud es

$$\mathcal{L}(\pi|y) \propto \pi^y (1-\pi)^{n-y}.$$

Familia conjugada: Bernoulli

- Una distribución inicial conjugada para una distribución binomial es la distribución beta con parámetros α y β (ambos positivos), $\pi \sim Beta(\pi|\alpha,\beta)$, cuya función de densidad de probabilidad de π está dada por

$$p(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}, \qquad 0<\pi<1,\ \alpha>0,\ \beta>0,$$

donde $\Gamma(\cdot)$ es la función gamma dada por

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt.$$

- Note que una distribución inicial simétrica sobre π se obtiene haciendo $\alpha=\beta$ y cuando $\alpha=\beta=1$ se reduce a una distribución inicial uniforme.

Familia conjugada: Bernoulli

- La distribución final de π también es una distribución beta, con parámetros $\alpha + y$ y $\beta + n - y$, cuya función de densidad de probabilidad es

$$p(\pi|y,n) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+y)\Gamma(\beta+n-y)}\pi^{\alpha+y-1}(1-\pi)^{\beta+n-y-1}, \qquad 0 < \pi$$

Familia conjugada: Poisson

- Los datos de conteo resultan de un número de ensayos y su rango son los números enteros no negativos.
- En estos casos, y bajo ciertas condiciones, un modelo que puede utilizarse es la distribución Poisson.
- Sea X una variable aleatoria con distribución Poisson con parámetro $\theta > 0$; entonces su función de densidad de probabilidad es:

$$\mathbb{P}(X = x) \quad = \quad e^{-\theta}\frac{\theta^x}{x!}, \qquad x = 0, 1, 2, \ldots, \text{ para } \theta > 0.$$

- Suponga que $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ son n variables aleatorias con distribución Poisson(θ).

Familia conjugada: Poisson

- La distribución inicial de θ se puede tomar como una distribución gamma con parámetros α_0 y β_0, cuya función de densidad de probabilidad está dada por:

$$\mathbb{P}(\theta) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\theta^{\alpha_0-1}e^{-\beta_0\theta}, \qquad \theta > 0,\ \alpha_0 > 0,\ \beta_0 > 0.$$

- La verosimilitud de la distribución Poisson es:

$$\mathbb{P}(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} e^{-\theta}\frac{\theta^{x_i}}{x_i!} \propto e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}.$$

Familia conjugada: Poisson

- La distribución final de θ es:

$$\begin{aligned}\mathbb{P}(\theta|\boldsymbol{x}) \quad \propto \quad \mathbb{P}(\boldsymbol{x}|\theta)\mathbb{P}(\theta) & \propto e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}\theta^{\alpha_0-1}\exp(-\beta_0\theta) \\ &= \theta^{\alpha_0+\sum_{i=1}^{n} x_i-1}\exp[-\theta(\beta_0+n)],\end{aligned}$$

  de tal manera que la distribución final para θ es también una distribución gamma, específicamente,

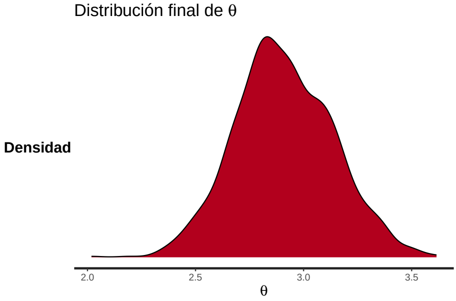
$$\begin{aligned} p(\theta|\boldsymbol{x}) &= \mathrm{Gamma}\left(\theta|\alpha_0+\sum_{i=1}^{n} x_i, \beta_0+n\right) \\ &= \mathrm{Gamma}(\theta|\alpha_1,\beta_1),\end{aligned}$$

  donde $\alpha_1 = \alpha_0 + \sum_{i=1}^{n} x_i$ y $\beta_1 = \beta_0 + n$.

- Debido a que la distribución inicial y la final pertenecen a la misma familia de distribuciones, entonces la familia de distribuciones gamma es conjugada para la familia de distribuciones Poisson.

Familia conjugada: Poisson

Poisson: distribución final de θ



Familia conjugada: Poisson

- Si $p(\theta)$ es conjugada a una familia paramétrica $f(x|\theta)$, entonces las distribuciones a priori y a posteriori, y las distribuciones predictivas a priori y a posteriori pertenecen a la familia de distribuciones.
- ¿Bajo qué condiciones es posible construir una familia conjugada?
- Usualmente dos distribuciones pretenecientes a la Famililа Exponencial:

$$f(x|\theta) = a(\theta)b(x)\exp\left\{c(\theta)d(x)\right\}$$

donde $a(\theta)$ y $c(\theta)$ son funciones de θ, y $b(x)$ y $d(x)$ son funciones solo de x, tales que,

$$\int f(x|\theta)\mathrm{d}x = \int a(\theta)b(x)\exp\left\{c(\theta)d(x)\right\}\mathrm{d}x = 1$$

Familia Paramétrica & Familia Conjugada

Familia Paramétrica	Familia Conjugada
Bernoulli(θ)	Beta($\theta \vert a, b$)
Binomial(k,θ)	Beta($\theta \vert a, b$)
Poisson(θ)	Gamma($\theta \vert a, b$)
Binomial-Negativa(α,θ)	Beta($\theta \vert a, b$)
Geométrica(θ)	Beta(θ)
Normal(μ)	Normal($\mu \vert m_0, s_0^2$)
Normal(μ, λ)	Normal-Gamma($\mu, \lambda \vert m_0, n_0, a, b$)
Exponencial(θ)	Gamma($\theta \vert a, b$)
Uniforme($0, \theta$)	Pareto($\theta \vert a, b$)

Predicción

- En ocasiones el propósito de un análisis estadístico es hacer predicciones de una observación futura $X^\star$.
- La distribución de probabilidad que describe el comportamiento de una observación futura $X^\star$ es $\mathbb{P}(X^\star|\boldsymbol{\theta})$, sin embargo $\boldsymbol{\theta}$ es desconocido, así que es necesario estimarlo.
- Desde la perspectiva Bayesiana, la distribución marginal de $X^\star$:

$$\mathbb{P}(X^\star) = \int_\Theta \mathbb{P}(X^\star|\boldsymbol{\theta})\mathbb{P}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta},$$

conocida como distribución predictiva inicial (**predictiva a priori**), describe la información acerca de $X^\star$ dada la información inicial disponible por $\mathbb{P}(\boldsymbol{\theta})$.

Predicción

- Una vez obtenida la muestra $\boldsymbol{X}$, la distribución final del parámetro $\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X})$ junto con el modelo $\mathbb{P}(X^\star|\boldsymbol{\theta})$, inducen un distribución conjunta de $(X^\star, \boldsymbol{\theta})$ condicional a los valores observados:

$$\mathbb{P}(X^\star, \boldsymbol{\theta}|\boldsymbol{X}) = \mathbb{P}(X^\star|\boldsymbol{\theta}, \boldsymbol{X})\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X}) = \mathbb{P}(X^\star|\boldsymbol{\theta})\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X}),$$

  donde la igualdad $\mathbb{P}(X^\star|\boldsymbol{\theta}, \boldsymbol{X}) = \mathbb{P}(X^\star|\boldsymbol{\theta})$ se cumple siempre que haya independencia condicional entre la observación futura $X^\star$ y la muestra observada $\boldsymbol{X}$.
- Por tanto, la distribución de probabilidad:

$$\mathbb{P}(X^\star|\boldsymbol{X}) = \int_\Theta \mathbb{P}(X^\star|\boldsymbol{\theta})\mathbb{P}(\boldsymbol{\theta}|\boldsymbol{X})\mathrm{d}\boldsymbol{\theta},$$

  conocida como distribución predictiva final (**predictiva a posteriori**), describe el comportamiento de $X^\star$ dada toda la información disponible de $\boldsymbol{X}$ y de $\boldsymbol{\theta}$.

Predicción: Bernoulli

- La distribución predictiva final está dada por

$$
\begin{aligned}
p(x^\star|x_1,\ldots,x_n) &= \int_0^1 p(x^\star|\theta)p(\theta|x_1,\ldots,x_n)\mathrm{d}\theta \\
&= \int_0^1 \theta^{x^\star}(1-\theta)^{1-x^\star}\theta^{\alpha_1-1}(1-\theta)^{\beta_1-1}\mathrm{d}\theta \\
&= \tfrac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)}\int_0^1 \theta^{x^\star+\alpha_1-1}(1-\theta)^{1-x^\star+\beta_1-1}\mathrm{d}\theta \\
&= \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)}\frac{\Gamma(\alpha_1+x^\star)\Gamma(\beta_1+1-x^\star)}{\Gamma(\alpha_1+\beta_1+1)},
\end{aligned}
$$

- Esta distribución se conoce como *Beta-Binomial*, $x = 0, 1$.

Predicción: Poisson

- La distribución predictiva final está dada por:

$$
\begin{aligned}
p(x^\star|x_1,\ldots,x_n) &= \int_0^\infty p(x^\star|\theta)p(\theta|x_1,\ldots,x_n)\mathrm{d}\theta \\
&= \int_0^\infty e^{-\theta}\frac{\theta^{x^\star}}{x^\star!}\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\theta^{\alpha_1-1}e^{-\beta_1\theta}\mathrm{d}\theta \\
&= \frac{1}{x^\star!}\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\int_0^\infty e^{-\theta-\beta_1\theta}\theta^{x^\star+\alpha_1-1}e^{-\beta_1\theta}\mathrm{d}\theta \\
&= \frac{1}{x^\star!}\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\frac{\Gamma(\alpha_1+x^\star)}{(\beta_1+1)^{(\alpha_1+x^\star)}}.
\end{aligned}
$$

- Esta distribución se conoce como *Poisson-Gamma*$(\alpha_1,\beta_1,1)$.
- Considerando un conjunto de $n=50$ datos simulados a partir de una distribución Poisson$(\theta=3)$, se obtienen las estimaciones con métodos Bayesianos usando Stan.

Predicción: Poisson

Poisson: resultados de la distribución final de θ, y distribución predictiva final de $x^\star$ usando `Stan`.

	Media	SE media	SD	2.5 %	25 %	50 %	75 %	97.5 %	neff	F
θ	2.91	0.01	0.23	2.48	2.76	2.9	3.07	3.36	719	
$x^\star$	2.90	0.05	1.71	0.00	2.00	3.0	4.00	6.00	1051	

Poisson: distribución predictiva final de $x^{\star}$.