Probabilidad Bayesiana: una lógica para razonar bajo incerteza

Sergio Davis[1,2] <sergio.davis@cchen.cl>

[1]Centro de Investigación en la Intersección de Física de Plasmas, Materia y Complejidad (P^{2}mc),
Comisión Chilena de Energía Nuclear

[2]Departamento de Física y Astronomía, Facultad de Ciencias Exactas,
Universidad Andrés Bello

Escuela de Computación Distribuida | ITM Medellín, Colombia | Agosto 2025

Motivación

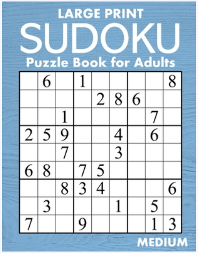
En Ciencia, deseamos obtener las mejores conclusiones usando información limitada.

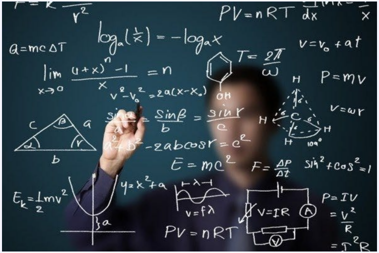
El pensamiento científico:

- Tiene el rol de una herramienta para **evitar sesgos**
- Debe aspirar a **ser racional**: creencias que contradicen observaciones del mundo real deben ser abandonadas
- Necesita poder **cuantificar el grado de incerteza** que tenemos respecto a observaciones o experimentos

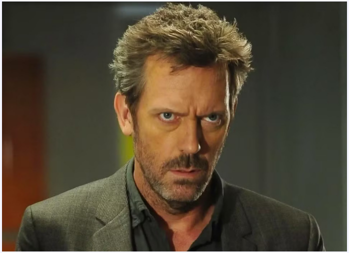
Dos modos de “razonamiento científico”

Deducción





Inferencia



Deducción

Premisas → **Lógica** → **Conclusiones**

En matemáticas:

Axiomas: Premisas que se asumen ciertas sin justificación, como la base de un sistema lógico

Teoremas: Conclusiones que se siguen de un conjunto de axiomas u otros teoremas establecidos previamente

> Si las premisas son ciertas y la lógica se aplica correctamente, las conclusiones **deben ser ciertas**.

Inferencia

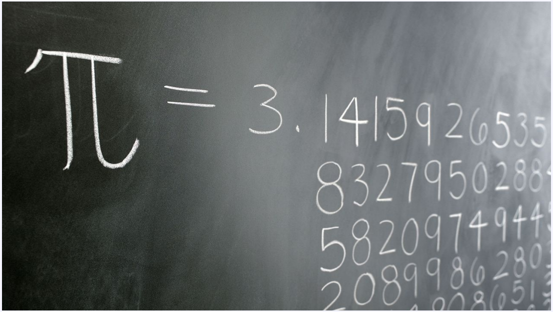
- Intenta entregar la mejor respuesta con información incompleta
- Construye y ajusta modelos en base a observaciones
- Evita entregar conclusiones fuera de lo justificado por los datos o experiencia

La inferencia **no puede asegurar** que sus respuestas sean correctas, sólo las afirma con un grado de credibilidad.

¿Vida en Marte?

¿Existe vida en Marte en este momento?
¿Cuán fuerte es su creencia respecto a esto?

Dígitos de Pi

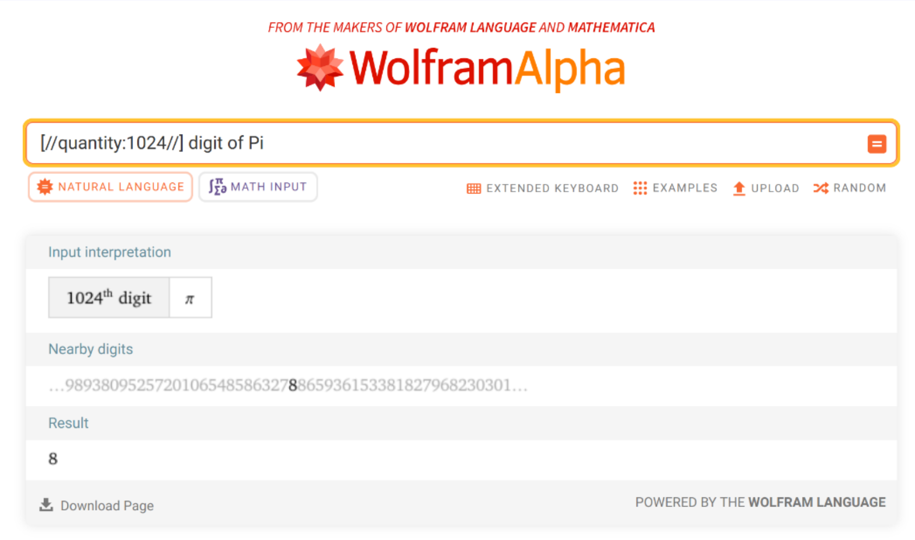


¿Cuán plausible es que el dígito número 1024 de π sea 7?

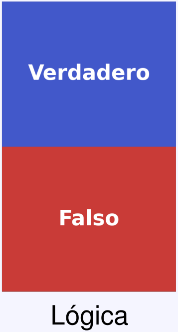
Dígito 1024 de π : Google AI Overviews

Dígito 1024 de π : Google AI Overviews

Dígito 1024 de π : Wolfram Alpha



Probabilidad Bayesiana: el lenguaje de la inferencia



$P(A|\mathcal{I})$: Probabilidad de A dado $\mathcal{I}$

Bayesianos versus frecuentistas

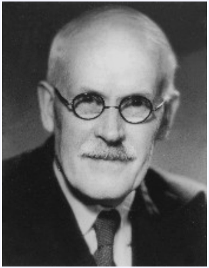
Frecuentista: La probabilidad es la frecuencia con que ocurre un evento aleatorio

Bayesiano: La probabilidad es un grado de creencia racional basado en evidencia

Fundadores de la probabilidad Bayesiana

Thomas Bayes (1701 - 1761)

Pierre-Simon Laplace (1749 - 1827)



Harold Jeffreys (1891 - 1989)

Edwin Jaynes (1922 - 1998)

Las reglas de la probabilidad

Regla de la suma

$$P(A|\mathcal{I}) + P(\text{not } A|\mathcal{I}) = 1$$

Regla del producto

$$P(A \text{ and } B|\mathcal{I}) = P(A|B \text{ and } \mathcal{I})P(B|\mathcal{I})$$

Regla extendida de la suma

$$P(A \text{ or } B|\mathcal{I}) = P(A|\mathcal{I}) + P(B|\mathcal{I}) - P(A \text{ and } B|\mathcal{I})$$

¿Cómo actualizamos nuestras creencias?

Definiciones

H : Una hipótesis que quisiéramos validar o refutar,

E : Una evidencia, posiblemente relevante para la validez de H,

$\mathcal{I}$: Un conocimiento base o previo, que no incorpora la evidencia E

Usando la regla del producto podemos escribir

$$P(H \text{ and } E|\mathcal{I}) = P(H|E \text{ and } \mathcal{I})P(E|\mathcal{I})$$

pero también es cierto que

$$P(E \text{ and } H|\mathcal{I}) = P(E|H \text{ and } \mathcal{I})P(H|\mathcal{I}).$$

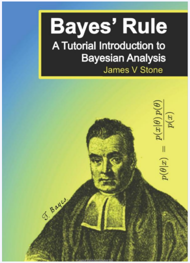
Como $(H \text{ and } E) = (E \text{ and } H)$, podemos igualar ambos lados derechos,

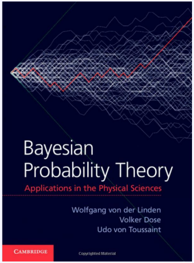
$$P(H|E \text{ and } \mathcal{I})P(E|\mathcal{I}) = P(E|H \text{ and } \mathcal{I})P(H|\mathcal{I})$$

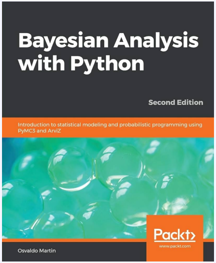
llegando finalmente a ...

El teorema de Bayes

$$P(H|E \text{ and } \mathcal{I}) = P(H|\mathcal{I}) \times \frac{P(E|H \text{ and } \mathcal{I})}{P(E|\mathcal{I})}$$







Es la regla universal para actualizar creencias **previas** a creencias **posteriores**,

$$P(H|\mathcal{I}) \quad \rightarrow \quad P(H|E \text{ and } \mathcal{I})$$

Dos casos interesantes

Caso 1: La hipótesis H es **falsable**, y de hecho predice que E es imposible.


¿Qué ocurre si E es observada? Se tiene que $P(E|H \text{ and } \mathcal{I}) = 0$

y por lo tanto

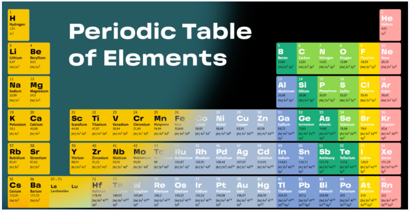
$$P(H|E \text{ and } \mathcal{I}) = P(H|\mathcal{I}) \times \frac{\cancelto{0}{P(E|H \text{ and } \mathcal{I})}}{P(E|\mathcal{I})} = 0 \implies \boxed{H \text{ es refutada por } E}$$

En este caso el teorema de Bayes se reduce a la aplicación de la lógica:

$$\boxed{(H \Rightarrow \text{not } E) \ \text{ es equivalente a } \ (E \Rightarrow \text{not } H)}$$

Dos casos interesantes

Caso 2: La hipótesis H predice que E será observada.



¿Qué ocurre cuando E es de hecho observada? En este caso se tiene $P(E|H \text{ and } \mathcal{I}) = 1$ y $P(E|\mathcal{I}) < 1$, luego

$$P(H|E \text{ and } \mathcal{I}) = P(H|\mathcal{I}) \times \frac{\cancelto{1}{P(E|H \text{ and } \mathcal{I})}}{P(E|\mathcal{I})} > P(H|\mathcal{I})$$

$$\Longrightarrow \boxed{H \text{ resulta favorecida por } E}$$

E no demuestra que H es cierta! Sólo la hace más probable.

Es imposible demostrar H definitivamente acumulando evidencia.

Distribución previa, posterior y verosimilitud

θ **:** Parámetros de un modelo

D **:** Datos que el modelo intenta describir

$\mathcal{I}$ **:** Información previa a ver los datos

Distribución previa

Función verosimilitud

$$P(\boldsymbol{\theta}|D,\mathcal{I}) = \frac{P(\boldsymbol{\theta}|\mathcal{I})P(D|\boldsymbol{\theta},\mathcal{I})}{P(D|\mathcal{I})}$$

Distribución posterior

Ejemplo 1: Lanzamiento de la moneda

Lanzamiento	Resultado
1	Cara
2	Sello
3	Cara
4	Cara
5	Sello
6	Cara
7	Sello
8	Cara
9	Cara
10	Cara
11	???

En $n = 10$ lanzamientos independientes se obtuvo 7 caras y 3 sellos.

- ¿Cuál es la probabilidad de obtener cara en el siguiente lanzamiento?
- ¿Cómo cambia la respuesta para 70 caras y 30 sellos? ($n = 100$)

Lanzamiento de la moneda: solución frecuentista

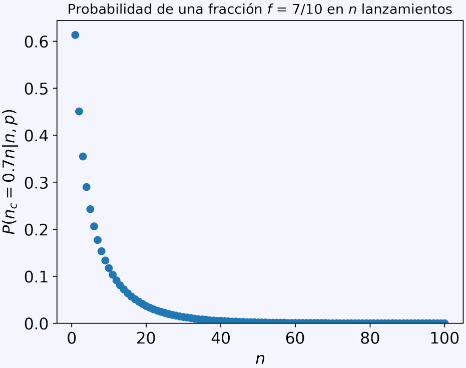
Se obtendrá cara con probabilidad p **en cada lanzamiento, independiente de los resultados previos**, y sello con probabilidad $1 - p$.

Si la moneda no está cargada, se tiene que $p = 0.5$.

La probabilidad de n_c caras en n intentos es binomial:

$$\boxed{P(n_c|n,p) = \binom{n}{n_c} p^{n_c}(1-p)^{n-n_c}}$$

y se cumple que $n_c \approx np$ para $n \gg 1$.



Lanzamiento de la moneda: solución Bayesiana

Existe una probabilidad p desconocida de obtener cara, a la cual asignamos una **creencia inicial** $\varnothing$ **plana**,

$$P(p|\varnothing) = \begin{cases} 1 & \text{si } 0 \leq p \leq 1, \\ 0 & \text{en caso contrario.} \end{cases}$$

Esta creencia cambiará al incluir los lanzamientos observados por medio del teorema de Bayes,

$$\boxed{P(p|n_c, n) = P(p|\varnothing) \times \frac{P(n_c|n, p)}{P(n_c|n, \varnothing)}}$$

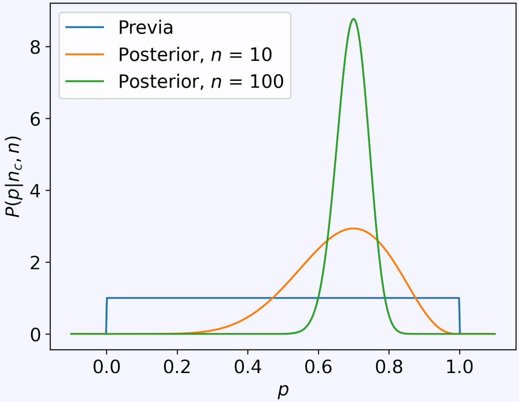
Usamos la probabilidad binomial de n_c caras en n intentos,

$$P(n_c|n, p) = \binom{n}{n_c} p^{n_c}(1-p)^{n-n_c}.$$

Lanzamiento de la moneda: distribución posterior

La distribución posterior de p es entonces

$$P(p|n_c, n) = \frac{p^{n_c}(1-p)^{n-n_c}}{B(n_c+1, n-n_c+1)}$$

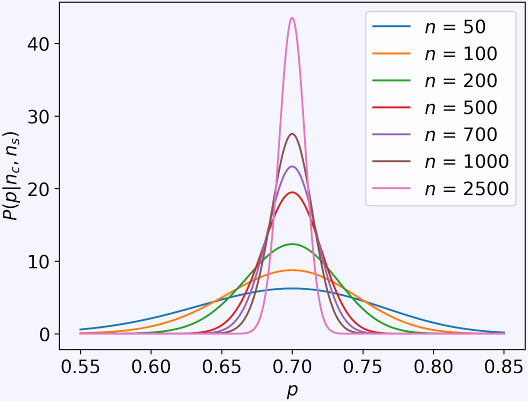


Valor más probable de p:

$$p^* = f := \frac{n_c}{n}$$

Lanzamiento de la moneda: probabilidad y frecuencia

$\big\langle p \big\rangle_{n_c,n} \approx f$	$\dfrac{\big\langle (\delta p)^2 \big\rangle_{n_c,n}}{\big\langle p \big\rangle^2_{n_c,n}} \approx \left(\dfrac{1-f}{f}\right)\dfrac{1}{n} \to 0$

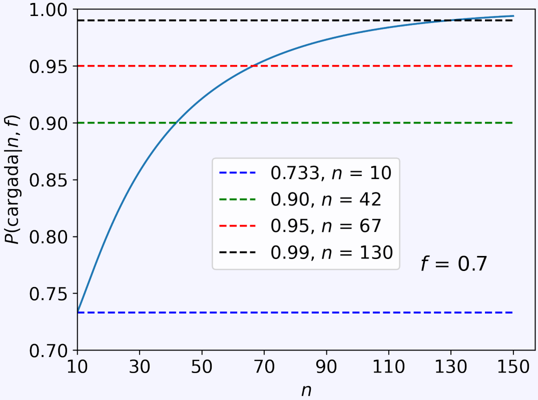


$$\lim_{n\to\infty} P(p|n_c, n) = \delta\big(p - f\big)$$

¿Está cargada la moneda?

Si decimos que la moneda está cargada si $p < 0.4$ o si $p > 0.6$ ...

$$P\left(0.4 \leq p \leq 0.6 \middle| n, f\right) = \int_{0.4}^{0.6} dp\, \frac{p^{nf}(1-p)^{n(1-f)}}{B(nf+1,(1-f)n+1)}$$



Ejemplo 2: Apagones

$$P(n|\lambda) = \frac{\exp(-\lambda)\lambda^n}{n!}$$

- En una cuadra ocurrieron $n = 5$ apagones durante el mes de febrero.
- Suponga que cada mes el número de apagones sigue la misma distribución de Poisson.

¿Cuál es la probabilidad de tener uno o más apagones durante marzo?

Apagones

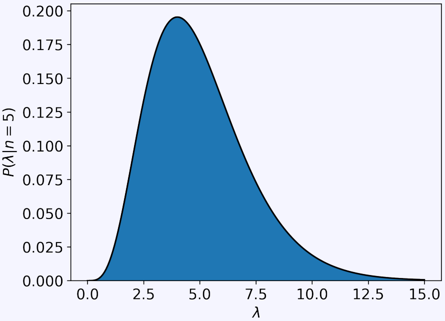
La probabilidad de n apagones para una distribución Poisson es

$$P(n|\lambda) = \frac{\exp(-\lambda)\lambda^n}{n!}$$

donde λ es el parámetro que controla el valor medio de n.

Usamos el teorema de Bayes para obtener la **distribución posterior** de λ

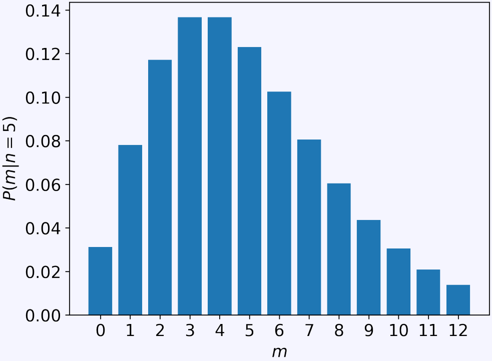
$$\begin{aligned} P(\lambda|n) &= P(\lambda|\varnothing) \times \frac{P(n|\lambda)}{P(n|\varnothing)} \\ &= \frac{\exp(-\lambda)\lambda^{n-1}}{(n-1)!} \end{aligned}$$



Apagones

Finalmente, la **distribución predictiva** de m apagones en marzo dado n en febrero considera todos los valores de λ posibles,

$$P(m|n) = \int_0^\infty d\lambda\, P(m|\lambda)P(\lambda|n) = \int_0^\infty d\lambda \frac{\exp(-2\lambda)\lambda^{m+n-1}}{(m!)(n-1)!}$$



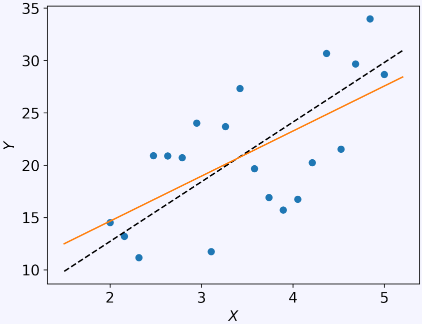
Distribución binomial negativa

$$\boxed{P(m|n) = \frac{(m+n-1)!}{2^{m+n}(m!)(n-1)!}}$$

$$\langle m \rangle_n = n$$

$$P(m=0|n) = \frac{1}{2^n} \quad \Longrightarrow \quad P(m \geq 1|n) = 1 - \frac{1}{2^n} \quad (96.875\%)$$

Costo total $=$ (Costo por kWh) $\times$ Consumo $+$ (Costo base) $+$ Variaciones



$$Y = \underbrace{AX + B}_{\text{dependencia lineal}} + \underbrace{\mathcal{E}}_{\text{incerteza}}$$

$$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

¿Cuál es la mejor recta que representa a los datos?

Regresión lineal Bayesiana

Suponiendo que la incerteza $\mathcal{E}$ sobre Y sigue una distribución normal,

$$P(\mathcal{E}|\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\mathcal{E}^2}{2\sigma^2}\right)$$

luego las cantidades desconocidas son A, B y σ, y el teorema de Bayes dice

$$\boxed{P(A, B, \sigma|D) = P(A, B, \sigma|\varnothing) \times \frac{P(D|A, B, \sigma)}{P(D|\varnothing)}}$$

Como las observaciones son independientes, la función verosimilitud es

$$P(D|A, B, \sigma) = \prod_{i=1}^{n} P(y_i|x_i, A, B, \sigma) \times P(x_i|\varnothing).$$

Regresión lineal Bayesiana

$$P(D|A, B, \sigma) = \prod_{i=1}^{n} P(y_i|x_i, A, B, \sigma) \times P(x_i|\varnothing).$$

La probabilidad de y_i dado x_i es la probabilidad de observar una desviación

$$\mathcal{E}_i = y_i - \big(Ax_i + B\big),$$

que es normal, luego se tiene

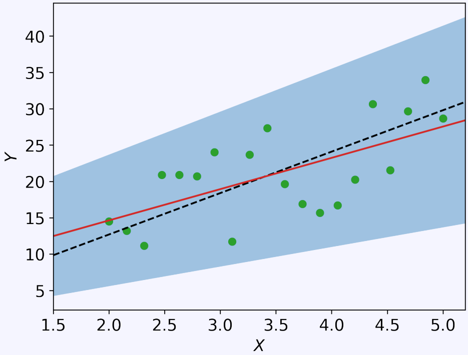
$$P(y_i|x_i, A, B, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\big(y_i - Ax_i - B\big)^2\right).$$

y la función verosimilitud es

$$P(D|A, B, \sigma) = \frac{1}{(\sqrt{2\pi})^n \sigma^n} \left[\prod_{i=1}^{n} P(x_i|\varnothing)\right] \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\big(y_i - Ax_i - B\big)^2\right).$$

Regresión lineal Bayesiana

$$P(A, B, \sigma | D) \propto \frac{1}{\sigma^{n+1}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \Big(y_i - \big[Ax_i + B\big]\Big)^2\right)$$



$$\mathbb{E}^2(A, B) := \sum_{i=1}^{n} \Big(\underbrace{y_i}_{\text{dato}} - \underbrace{\big[Ax_i + B\big]}_{\text{predicción}} \Big)^2$$

¡La recta más probable es la que minimiza el error cuadrático!

Regresión lineal

Definiendo $s_x := \sqrt{\overline{x^2} - (\overline{x})^2}$, la distribución de A y B dado σ es normal,

$$P(A|\sigma, D) = \frac{1}{\sqrt{2\pi}\sigma_A} \exp\left(-\frac{1}{2\sigma_A^2}\big(A - A^*\big)^2\right)$$
$$P(B|\sigma, D) = \frac{1}{\sqrt{2\pi}\sigma_B} \exp\left(-\frac{1}{2\sigma_B^2}\big(B - B^*\big)^2\right)$$

donde los valores más probables de A y B corresponden a la solución tradicional de mínimos cuadrados,

$$\boxed{\begin{aligned} A^* &= \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} \\ B^* &= \overline{y} - A^*\overline{x} \end{aligned}}$$

La incerteza sobre A y B decrece como $1/\sqrt{n}$ al aumentar el número de puntos,

$$\sigma_A := \frac{\sigma}{\sqrt{n}s_x}, \qquad \sigma_B := \frac{\sigma}{\sqrt{n}s_x} \times \sqrt{\overline{x^2}}.$$

Conclusiones

- La probabilidad Bayesiana es una extensión de la lógica tradicional que permite incorporar incerteza
- Es posible construir y analizar modelos usando inferencia Bayesiana en todas las áreas de la Ciencia
- Este enfoque es particularmente útil cuando se trata con un número pequeño de observaciones, o con eventos que no se repiten

Agencia Nacional de Investigación y Desarrollo

Agradecimientos al proyecto
ANID FONDECYT 1220651 (Chile)