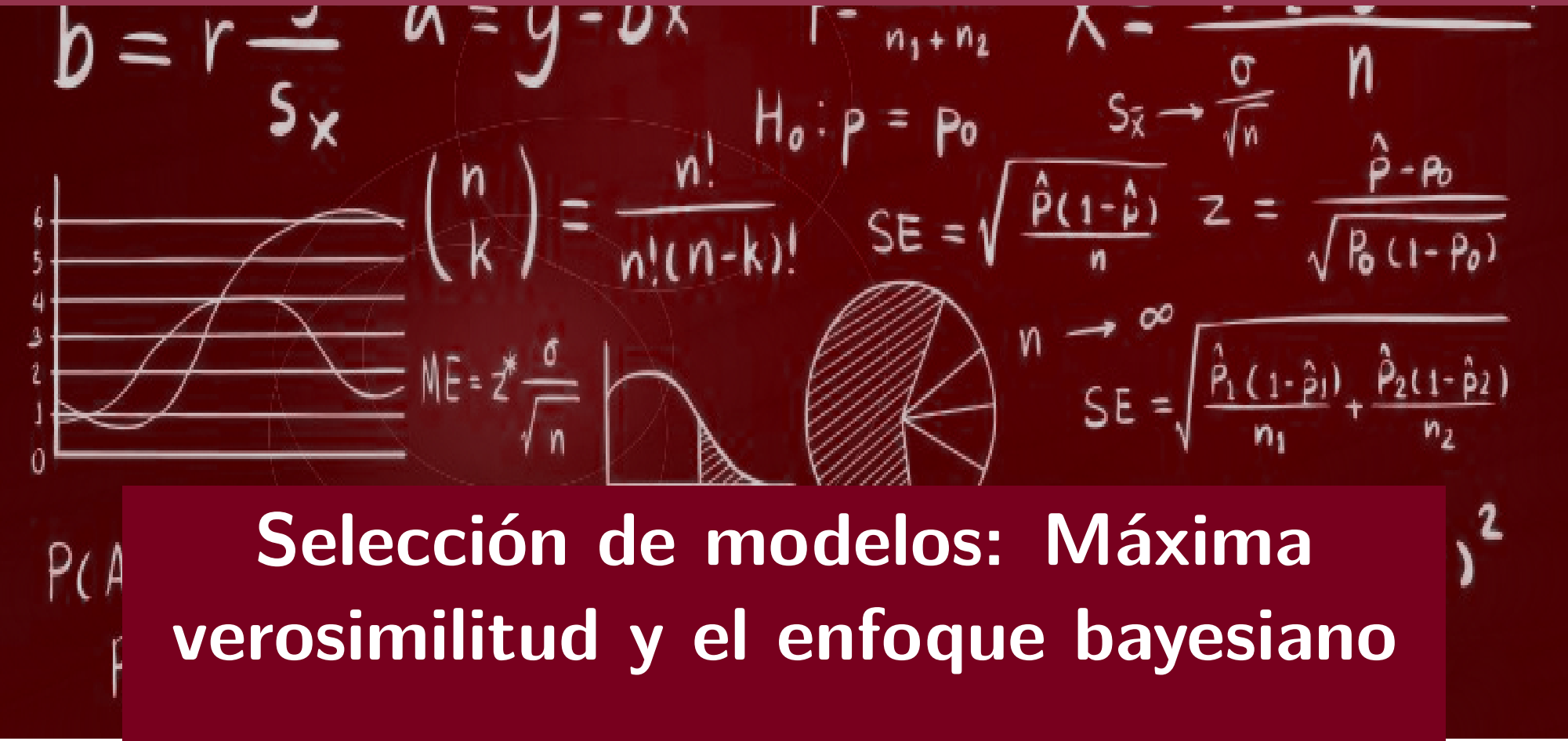


Selección de modelos: Máxima verosimilitud y el enfoque bayesiano

Grupo 1 - Andribet Sebastián, Mininni Tomás, Rodríguez Tomás

Laboratorio 4 - Facultad de Ciencas Exactas y Naturales - UBA

2 de Mayo, 2024

Introducción

- Buscaremos estudiar **cuál** es el modelo que mejor describe un conjunto de datos adquirido, en contraposición a estudiar si un modelo es adecuado.
- Hablaremos de la importancia de elegir un modelo correcto, y los parámetros que nos permiten definir cual es el más adecuado.
- Compararemos para un mismo set de datos dos enfoques diferentes. El enfoque de máxima verosimilitud (EMV) y el enfoque bayesiano.

Sobre-Ajuste

Hablando de polinomios, donde $M-1$ es el grado del polinomio, para $M = N$ uno podría lograr un "ajuste perfecto" donde el polinomio aproximado intersecta todos los puntos obtenidos en las mediciones. N es el número de puntos.

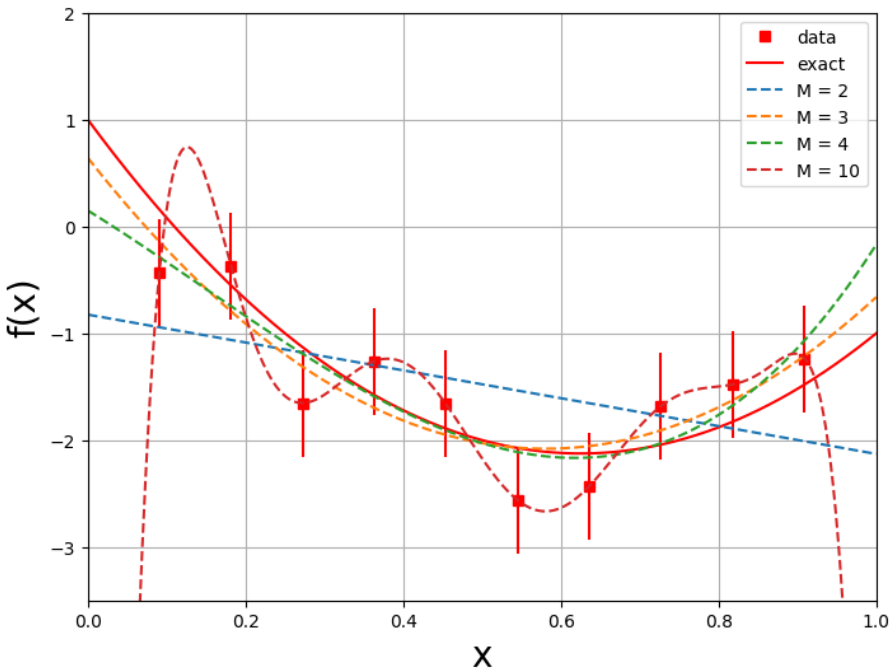


Fig.: parábola con ruido gaussiano ajustada por distintos polinomios

M=2	M=3	M=4	M=7	Exacto
-0.825	0.642	0.149	-2.139	1
-1.304	-9.376	-4.566	42.614	-10
-	8.072	-3.4	-352.976	8
-	-	7.648	1260.788	-
-	-	-	-2328.094	-
-	-	-	2140.064	-
-	-	-	-765.492	-

Tabla: Parámetros de ajuste para distintos polinomios

Máxima Verosimilitud (EMV)

El EMV se basa en una noción frecuentista donde:

$$\Omega = \lim_{n \to \infty} \frac{n_{casosfavorables}}{n_{casostotales}} \quad (1)$$

y por lo tanto el error obtenido al hacer un ajuste se deriva del scatter de los datos **asumiendo que el modelo usado es correcto**.

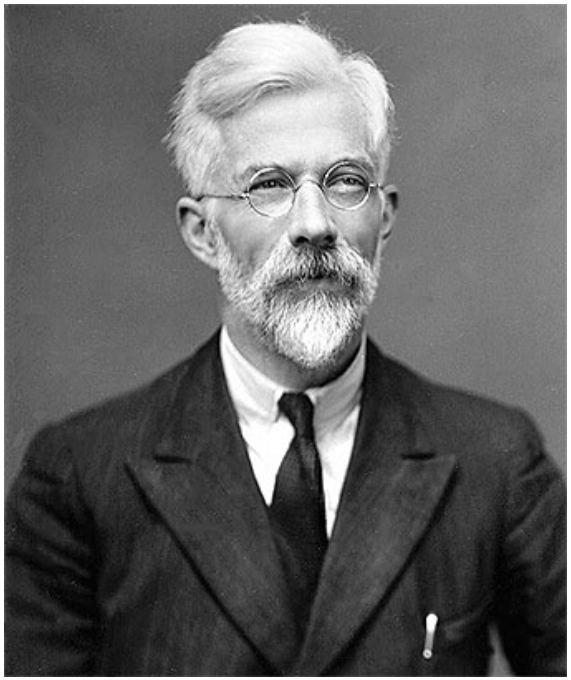


Fig.: Ronald Fisher

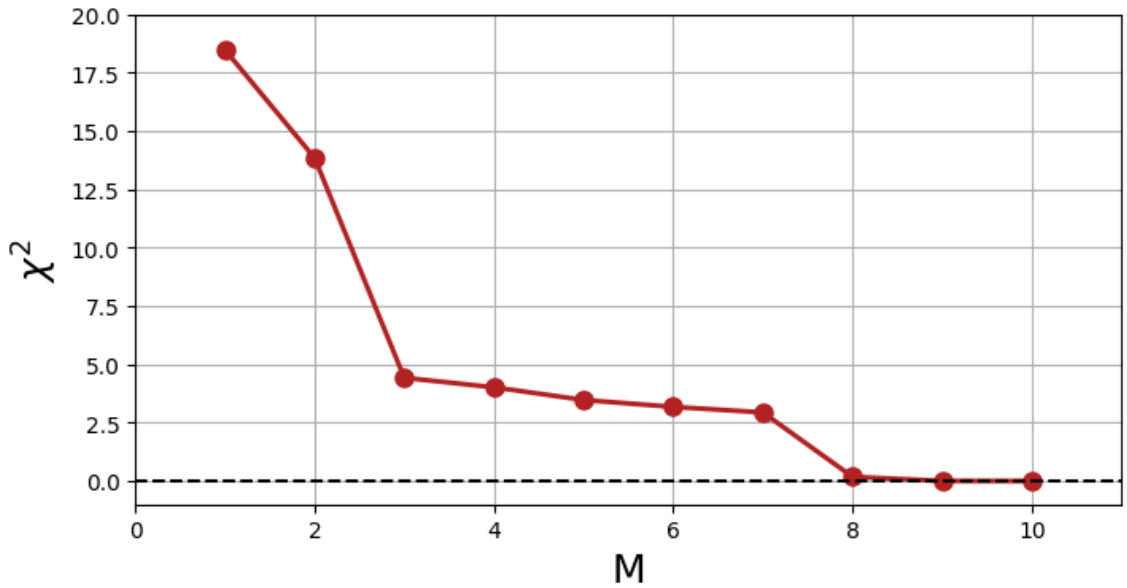
Estudio de χ^2

Inferimos del gráfico de $\chi^2(M)$ que el modelo más adecuado para realizar un ajuste es aquel que corresponda a el primer cambio considerable en el valor de χ^2.

Asumiendo ruido gaussiano en los datos y varianza σ_i, la probabilidad queda:

$$P(y) = \frac{1}{(2\pi)^{N/2}\left(\prod_{i=1}^{N}\sigma_i\right)} e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - \sum_{\alpha=1}^{M} a_\alpha X_\alpha(x_i)}{\sigma_i}\right)^2} \tag{2}$$

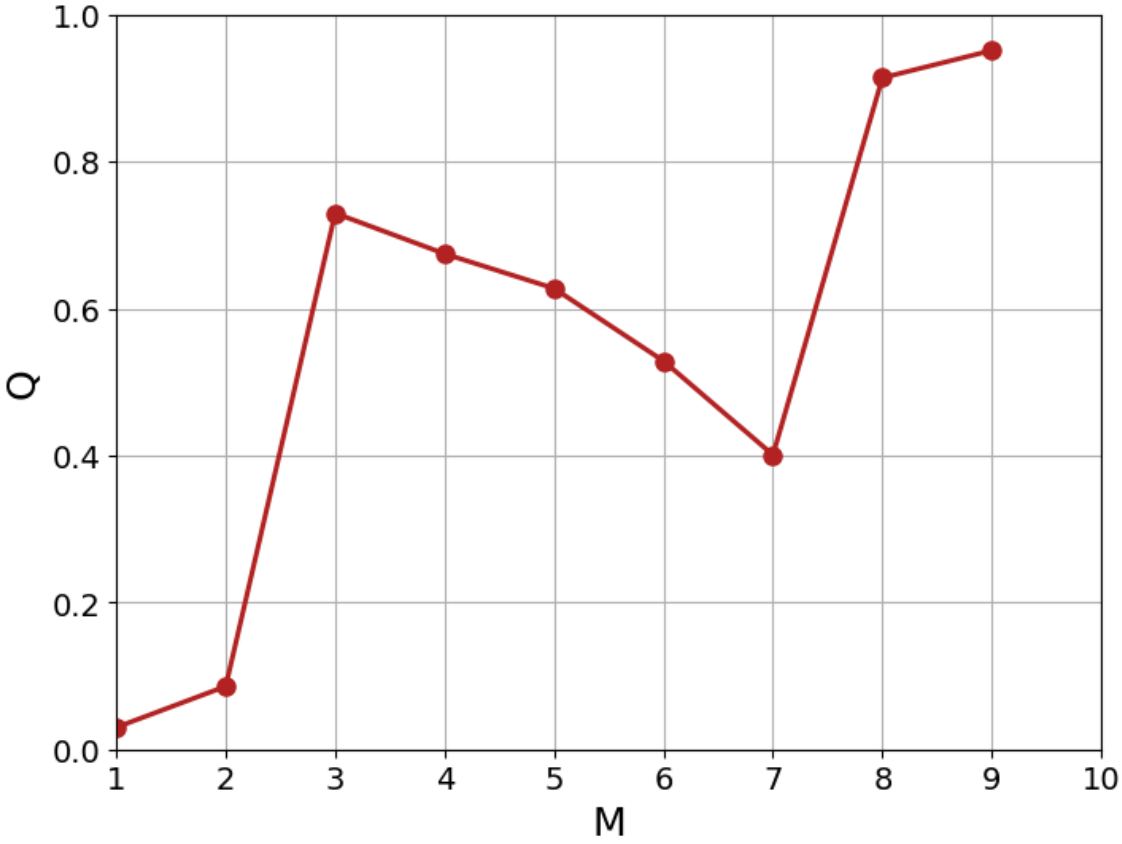
Vamos a buscar maximizar esta probabilidad, proporcional a $-\frac{1}{2}\chi^2$, equivalente a minimizar χ^2.



Estudio del factor Q

Intuitivamente, buscamos penalizar el aumento de M, por lo que definimos la función Q. Esta brinda una noción del valor de χ^2 por grado de libertad. El Q aumenta en si χ^2 (M) presenta un salto importante entre puntos consecutivos. El valor Q está restringido entre 0 y 1.

$$Q = \frac{1}{\Gamma(N_{GL}/2)} \int_{\chi^2/2}^{\infty} y^{(N_{GL}/2)-1} e^{-y} dy \quad (3)$$



"Se considera como modelo correcto el de grado asociado al primer máximo de P(M)"

Bayesiano

Noción probabilística donde se consideran "probabilidades condicionales" y priors sobre la distribución en base a conocimientos o suposiciones previas a la medición.
Sean x e y dos variables, la probabilidad de que $x = X$ es

$$P(X) = \sum_{y} P(X, Y) \qquad (4)$$

y buscamos relacionarlo con los condicionales $P(X \mid Y)$ y $P(Y \mid X)$.

Fig.: Thomas Bayes

$$P(X) = P(Y \mid X)P(X) = P(X \mid Y)P(Y)$$

$$\Rightarrow P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \tag{5}$$

- P(Y) es un dato conocido previo al análisis estadístico y se lo denomina un "**prior**".
- $P(X \mid Y)$ se determina en función a la data recolectada.
- $P(Y \mid X)$ lo provee el teorema de Bayes. Se la denomina distribución "**posterior**".

El bayesiano incluye información adicional en la forma de "priors" que hacen que sea útil para el problema de selección de modelos. Buscamos estudiar $P(M \mid D)$ trabajando con el prior $P(F \mid \gamma)$. Asumimos una distribución Gaussiana.

$$P(M \mid D) = \sum_{\gamma} P(\gamma \mid D) = \sum_{\gamma} \frac{P(D \mid \gamma) P(\gamma)}{P(\gamma)} \tag{6}$$

$$P(D \mid \gamma) = \sum_{F} P(D \mid F) P(F \mid \gamma) \tag{7}$$

$$P(D \mid F) = \frac{1}{(2\pi)^{N/2} \prod_{i=1}^{N} \sigma_i} e^{-\frac{1}{2} \sum_{i=1}^{N} \left(\frac{y_i - \sum_{\alpha} a_\alpha X_\alpha}{\sigma_i} \right)^2} \tag{8}$$

$$P(F \mid \gamma) = \left(\frac{\gamma}{2\pi} \right)^{M/2} e^{-\frac{\gamma}{2} \sum_{\alpha=1}^{M} a_\alpha^2} \tag{9}$$

$$P(D \mid \gamma) = (\frac{\gamma}{2\pi})^2 \frac{1}{(2\pi)^{N/2} \prod_{i=1}^{N} \sigma_i} \prod_{\alpha=1}^{M} (\int_{-\infty}^{\infty} da_\alpha) e^{-\frac{1}{2} E_0([a_\alpha])} \quad (10)$$

donde:

$$E_0([a_\alpha]) = \sum_{i=1}^{N} \left(\frac{y_i - \sum_{\alpha=1}^{M} a_\alpha X_\alpha(x_i)}{\sigma_i} \right)^2 + \gamma \sum_{\alpha} a_\alpha^2 \quad (11)$$

Factor que incluye a χ^2 y a un factor de corrección que penaliza valores de M grandes. Vamos a buscar minimizar la expresión.

Se plantea un problema similar al que se obtiene para cuadrados mínimos, se obtienen los parámetros optimizados de a_α, y se los reemplaza en E_0. Se obtiene la función costo.

$$\bar{E}_0(\{\hat{a}_\alpha\}, \gamma) =$$

$$\sum_{i=1}^{N} \left(\frac{y_i - \sum_{\alpha=1}^{M} \hat{a}_\alpha X_\alpha(x_i)}{\sigma_i} \right)^2 + \sum_{\alpha=1}^{M} \hat{a}_\alpha^2 + \sum_{I=1}^{M} \left(\frac{\lambda_I^{(0)} + \gamma}{\gamma} \right) + \sum_{i=1}^{N} ln(2\pi\sigma_i^2)$$

χ^2

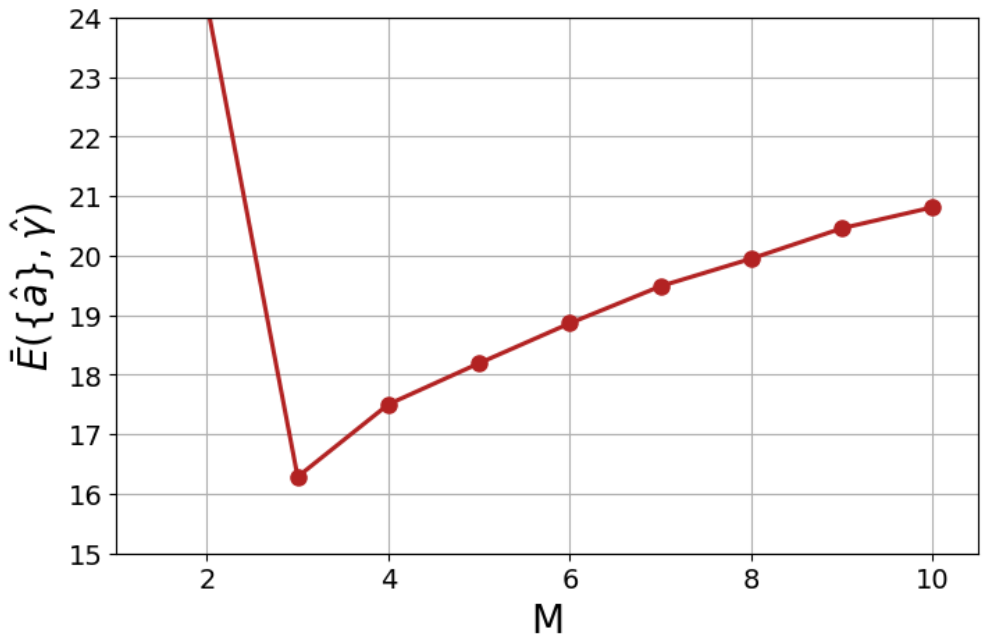
Penalizan el aumento en M

Evita que el mínimo sea γ=0 (CM)

- Se decide un valor de M y de γ.
- Se define el valor óptimo de a_α mediante (12).
- Se reemplazan los valores en la ecuación (13) y se obtiene el valor óptimo de γ.
- Se repite para diferentes valores de M comparando los valores de $E(M)$, buscando el mínimo.
- De ser necesario, se deben normalizar los datos x e y entre -1 y 1.

$$\hat{a}_\alpha = \sum_{\beta=1}^{M} \left(U^{-1}\right)_{\alpha\beta} v_\beta \tag{12}$$

$$\gamma \sum_{\alpha=1}^{M} \hat{a}_\alpha^2 = \sum_{I=1}^{M} \ln\left(\frac{\lambda_I^{(0)} + \gamma}{\gamma}\right) \tag{13}$$



"Se considera como modelo correcto el asociado al primer mínimo asociado a $\bar{E}(M)$"

Diferencia entre probabilidades

EMV:

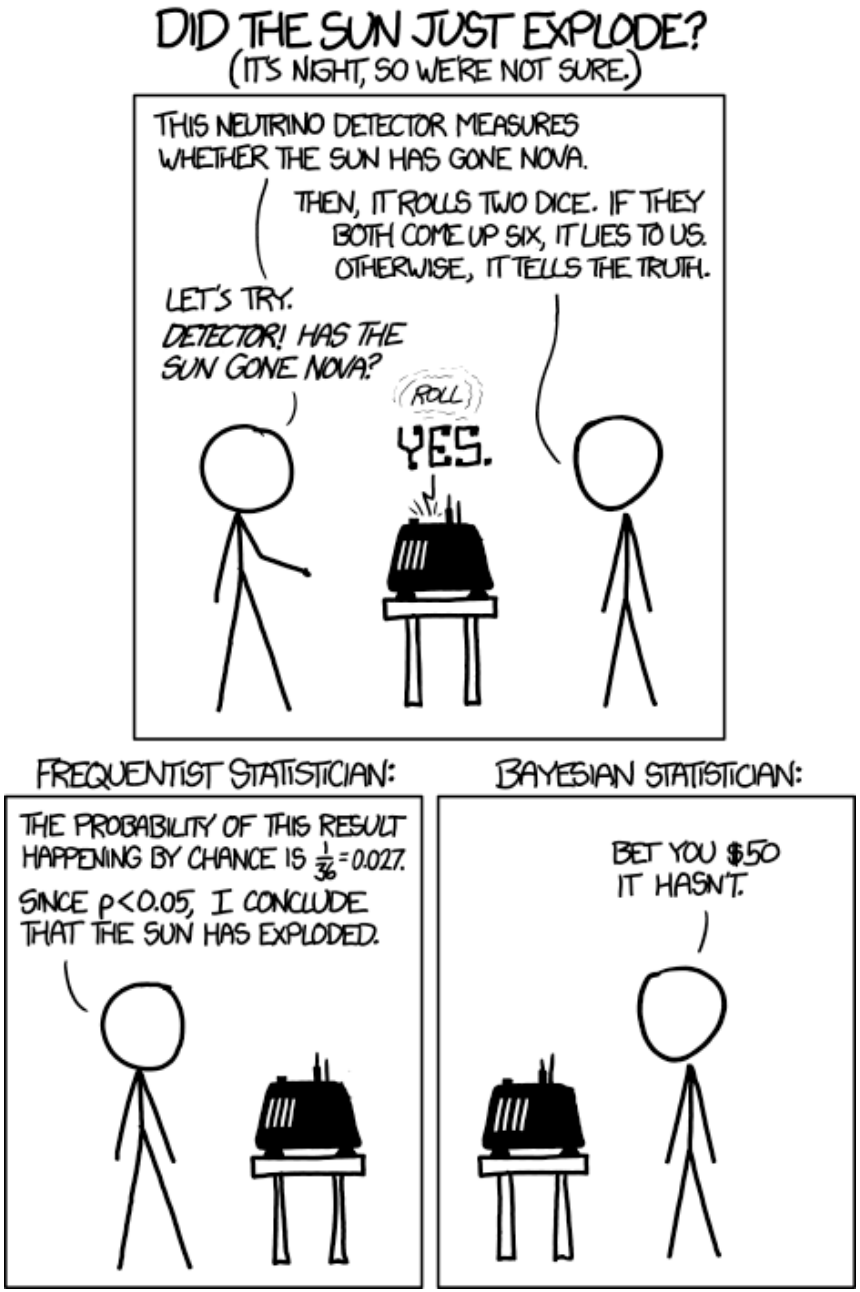
- No asume conocimientos previos sobre los parámetros.
- Asume que las variables son constantes desconocidas.
- " Asigna probabilidades a valores " .
- Calcula los valores que maximizan la función de probabilidad (probabilidad de observar resultados compatibles dados los parámetros).

Bayesiano:

- Asume conocimientos/creencias previas sobre los parámetros (priors).
- Trata a los parámetros como variables con distribuciones de probabilidad.
- " Asigna probabilidades a hipótesis " .
- Calcula la distribución de probabilidad posterior de los parámetros.

Bibliografía

- Peter Young (2014), 'Everything you wanted to know about Data Analysis and Fitting but were afraid to ask', arXiv:1210.3781 [physics.data-an], pp 38-52.
- Isabella Fornacon-Wood, Corinne Johnson-Hart, Corinne Faivre-Finn, James P.B. O ' Connor, and Gareth J. Price (2021), 'Understanding the Differences Between Bayesian and Frequentist Statistics', Statistics for the people, volume 112 issue 5, pp 1076-1082.
- Código de referencia:



Muchas Gracias