Considerar una moneda de la que no se tiene **ninguna información**. Se quiere determinar la probabilidad μ de que la moneda caiga de uno de los lados (digamos cara).

a. ¿Cuál sería un prior razonable para el parámetro μ dada esta información?
b. Se lanza la moneda *N* veces, y cae *c* veces cara (C) y *e* = *N* - *c* veces ceca (escudo, E).

- Escribir la función de verosimilitud para estos datos.
- ¿Qué función de distribución es la apropiada para modelar este proceso?
- ¿Cuál es el estimador de máxima verosimilitud, μ_{ml} ?

Prior, likelihood y EMV

- ¿Sabemos algo de μ ?

  Solo que se mueve entre 0 y 1 ¿Qué distribución conocida nos puede servir?

Prior, likelihood y EMV

- ¿Sabemos algo de μ ?

  Solo que se mueve entre 0 y 1 ¿Qué distribución conocida nos puede servir?

$$\mu \sim U(0, 1)$$

Prior, likelihood y EMV

- Cada vez que se toma un dato es una Bernoulli $Bern(x \,|\, \mu) = \mu^x (1-\mu)^{1-x}$
- Nuestro conjunto de datos $D = \{x_1, x_2, \ldots, x_n\}$ son *n* tiradas (independientes).
- Construimos la likelihood como:

$$P(x_1, x_2, \ldots, x_n \,|\, \mu) \xrightarrow[\text{Independencia}]{}$$

Prior, likelihood y EMV

- Cada vez que se toma un dato es una Bernoulli $Bern(x \mid \mu) = \mu^x (1-\mu)^{1-x}$
- Nuestro conjunto de datos $D = \{x_1, x_2, \ldots, x_n\}$ son *n* tiradas (independientes).
- Construimos la likelihood como:

$$P(x_1, x_2, \ldots, x_n \mid \mu) \xrightarrow[\text{Independencia}]{} \prod_{n=1}^{N} P(x_n \mid \mu)) \longrightarrow$$

Prior, likelihood y EMV

- Cada vez que se toma un dato es una Bernoulli $Bern(x \mid \mu) = \mu^x (1-\mu)^{1-x}$
- Nuestro conjunto de datos $D = \{x_1, x_2, \ldots, x_n\}$ son *n* tiradas (independientes).
- Construimos la likelihood como:

$$P(x_1, x_2, \ldots, x_n \mid \mu) \xrightarrow[\text{Independencia}]{} \prod_{n=1}^{N} P(x_n \mid \mu)) \xrightarrow[\text{Identicamente distribuidas}]{} \prod_{n=1}^{N} \mu^{x_n} (1-\mu)^{1-x_n}$$

Prior, likelihood y EMV

- Estimando el valor de μ maximizando la likelihood (EMV):

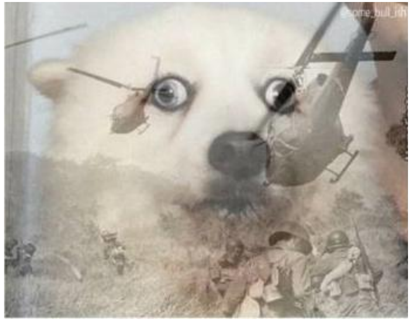
$$\prod_{n=1}^{N} P(x_n \mid \mu) \longrightarrow$$

Prior, likelihood y EMV

- Estimando el valor de μ maximizando la likelihood (EMV):

WAR FLASHBACK DE ANÁLISIS 1

$$\prod_{n=1}^{N} P(x_n \mid \mu) \longrightarrow$$

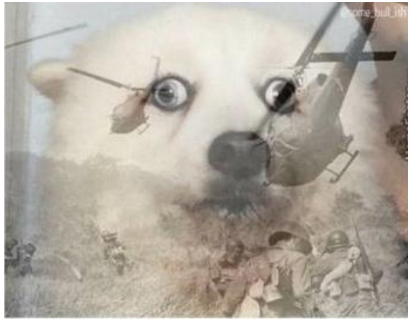


Prior, likelihood y EMV

- Estimando el valor de μ maximizando la likelihood (EMV):

WAR FLASHBACK DE ANÁLISIS 1

$$\prod_{n=1}^{N} P(x_n \mid \mu) \longrightarrow$$



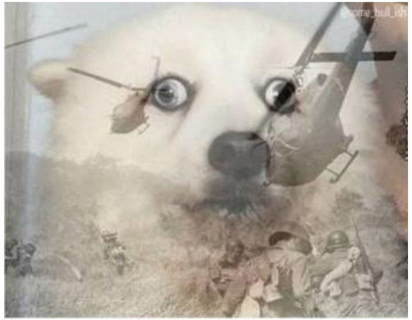
- Derivar en función de μ e igualar a cero despejándola.
- Horrible con la productoria -> El truco del logaritmo

Prior, likelihood y EMV

- Estimando el valor de μ maximizando la likelihood (EMV):

WAR FLASHBACK DE ANÁLISIS 1

$$\prod_{n=1}^{N} P(x_n \mid \mu) \longrightarrow$$



$$\longrightarrow \sum_{n=1}^{N} \log P(x_n \mid \mu)$$

- Derivar en función de μ e igualar a cero despejándola.
- Horrible con la productoria -> El truco del logaritmo

Prior, likelihood y EMV

$$\sum_{n=1}^{N} \log P(x_n \mid \mu) \longrightarrow \sum_{n=1}^{N} \log(\mu^{x_n}(1-\mu)^{1-x_n})$$

Prior, likelihood y EMV

$$\sum_{n=1}^{N} \log P(x_n \mid \mu) \longrightarrow \sum_{n=1}^{N} \log(\mu^{x_n}(1-\mu)^{1-x_n})$$

$$\sum_{n=1}^{N} x_n \, log(\mu) + (1 - x_n) \, log(1 - \mu)$$

Prior, likelihood y EMV

$$\sum_{n=1}^{N} \log P(x_n \mid \mu) \longrightarrow \sum_{n=1}^{N} \log(\mu^{x_n}(1-\mu)^{1-x_n})$$

$$\sum_{n=1}^{N} x_n \, log(\mu) + (1-x_n)\, log(1-\mu)$$

- Ahora si, derivando, igualando a 0 y despejando...

Prior, likelihood y EMV

$$\mu_{ml} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

Comentario post clase: Esta es una forma frecuentista de aproximar a μ, no tiene que ver con lo bayesiano que viene después.

Prior, likelihood y EMV

- Si denotamos *x=1* cada éxito en el conjunto de datos y tenemos *m* éxitos. La probabilidad de "cara" está dada por:

$$\mu_{ml} = \frac{m}{N}$$

(La media aritmética)

- ¿Qué pasa si, x ej, se tira la moneda 3 veces y todas dan cara? (Overfitting).
- Se puede trabajar en el número de *m* observaciones de *x=1,* dado N tiradas

Prior, likelihood y EMV

- Si denotamos *x=1* cada éxito en el conjunto de datos y tenemos *m* éxitos. La probabilidad de “cara” está dada por:

$$\mu_{ml} = \frac{m}{N}$$ (La media aritmética)

- ¿Qué pasa si, x ej, se tira la moneda 3 veces y todas dan cara? (Overfitting).
- Se puede trabajar en el número de *m* observaciones de *x=1,* dado N tiradas

$$Bin(m \,|\, N, \mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m}$$

Prior, likelihood y EMV

Binomial

Uniforme

$$P(\mu \,|\, (N, X)) = \frac{P((N, X) \,|\, \mu) \times P(\mu|N)}{P(N, X)}$$

Dados los experimentos, estimar la distribución de probabilidad \mu

Evidencia
(Factor de normalización)

Prior, likelihood y EMV

Binomial

Uniforme

$$P(\mu \,|\, (N,X)) = \frac{P((N,X) \,|\, \mu) \times P(\mu|N)}{P(N,X)}$$

Dados los experimentos,
estimar la distribución de
probabilidad \mu

Evidencia
(Factor de normalización)

$$P(\mu \,|\, (N,X) \propto P((N,X) \,|\, \mu) = Binomial(X,N,\mu)$$

Beta

- Recibe α y β, definida en [0,1]

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

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt$$

- Beta(1,1) = ?

Beta

- Recibe α y β, definida en [0,1]

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

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt$$

- Beta(1,1) = Uniforme (notebook)

Usando Beta como prior

- Habíamos supuesto que que, a priori, la distribución sobre μ era uniforme.
- Pero ahora tenemos algo que va más allá de la uniforme

$$P(\mu \mid N) = f(\mu; \alpha_0, \beta_0) = \frac{1}{B(\alpha_0, \beta_0)} x^{\alpha_0 - 1}(1 - x)^{\beta_0 - 1}$$

$$B(\alpha, \beta) = \int_0^1 t^{\alpha - 1}(1 - t)^{\beta - 1}\, dt$$

Usando Beta como prior

Binomial

Beta

$$P(\mu \,|\, (N, X)) = \frac{P((N, X) \,|\, \mu) \times P(\mu | N)}{P(N, X)}$$

Dados los experimentos, estimar la distribución de probabilidad \mu

Evidencia
(Factor de normalización)

Usando Beta como prior

- Binomial x Beta

$$\binom{N}{x}\mu^{x}(1-\mu)^{N-x}\times\mu^{\alpha_0-1}(1-\mu)^{\beta_0-1}\frac{1}{B(\alpha_0,\beta_0)}$$

$$\binom{N}{x}\mu^{x+(\alpha_0-1)}(1-\mu)^{N-x+(\beta_0-1)}\frac{1}{B(\alpha_0,\beta_0)}$$

$$\alpha' = x+\alpha_0$$
$$\beta' = N-x+\beta_0$$

$$\binom{N}{x}\mu^{\alpha'-1}(1-\mu)^{\beta'-1}\frac{1}{B(\alpha_0,\beta_0)}$$

Usando Beta como prior

- Normalizando con: $$\int_0^1 \binom{N}{x} t^{\alpha'-1}(1-t)^{\beta'-1}\frac{1}{B(\alpha_0,\beta_0)}\,dt$$

Es el término de arriba, pero integramos sobre TODOS los μ

$$\frac{\binom{N}{x}\mu^{\alpha'-1}(1-\mu)^{\beta'-1}\frac{1}{B(\alpha_0,\beta_0)}}{\int_0^1 \binom{N}{x} t^{\alpha'-1}(1-t)^{\beta'-1}\frac{1}{B(\alpha_0,\beta_0)}\,dt}$$

Usando Beta como prior

- Normalizando con $$\int_0^1 \binom{N}{x} t^{\alpha'-1}(1-t)^{\beta'-1}\frac{1}{B(\alpha_0,\beta_0)}\,dt$$

$$\frac{\cancel{\binom{N}{x}}\mu^{\alpha'-1}(1-\mu)^{\beta'-1}\cancel{\frac{1}{B(\alpha_0,\beta_0)}}}{\int_0^1 \cancel{\binom{N}{x}}t^{\alpha'-1}(1-t)^{\beta'-1}\cancel{\frac{1}{B(\alpha_0,\beta_0)}}\,dt}$$

Usando Beta como prior

$$\frac{\mu^{\alpha'-1}(1-\mu)^{\beta'-1}}{\int_0^1 t^{\alpha'-1}(1-t)^{\beta'-1}\,dt}$$

$$Beta(\alpha', \beta')$$

En 03p_playground_Ej_3.ipynb pueden jugar un rato...