Biometría

Estimación por Máxima verosimilitud

Métodos de estimación de parámetros

Mínimos cuadrados ordinarios (MCO u OLS por sus siglas en inglés *ordinary least squares*)

- Consiste en estimar los parámetros de manera tal de minimizar

$$\sum (y_i - \hat{y}_i)^2 = \sum e_i^{\ 2}$$

- Método “clásico” para estimar parámetros de modelos
- Cálculos sencillos
- Para la estimación no requiere supuestos sobre la distribución de probabilidades de la variable, pero sí para la inferencia
- Sensible al desbalanceo

Supuestos del modelo

$$\varepsilon_{ijk} \sim NID(0, \sigma^2)$$

- independientes

  correlación temporal, espacial, anidamiento

- distribución normal

  otras distribuciones

- varianza constante

  heterocedasticidad

Y además tenemos el karma del desbalanceo

- Las estimaciones por cuadrados mínimos en el análisis de varianza de 2 o más factores se complican en diseños desbalanceados
- Pérdida de ortogonalidad en las SC
- Parches:
  - Forzar el balanceo
  - Reemplazar dato faltante
  - Ajustar la SC: SC tipo II, III, IV…

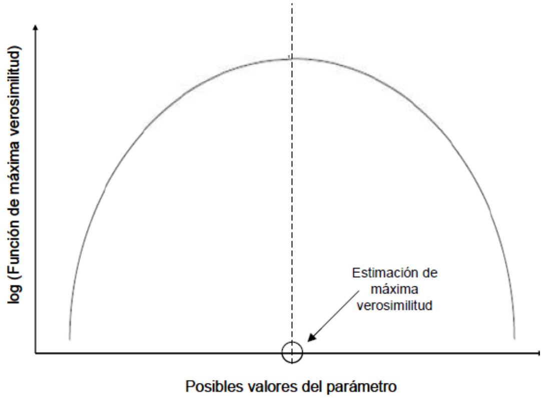
Y ni hablar de celdas vacías!

Métodos de estimación de parámetros

Máxima verosimilitud (MV o ML por sus siglas en inglés *maximum likelihood*)

- Consiste hallar los valores de los parámetros (asumiendo una distribución de probabilidades) que maximicen la función de verosimilitud

$$L(\theta / datos, modelo)$$



- Cálculos complejos
- Para la estimación es fundamental asumir una distribución de probabilidades de la variable
- Apropiado cuando fallan los supuestos o hay desbalanceo
- Si se cumplen los supuestos de independencia, normalidad y homocedasticidad y el diseño es balanceado, las estimaciones coinciden

Entendiendo la máxima verosimilitud…

- Se arrojan 5 monedas equilibradas y se cuenta la cantidad de caras

Nro de caras	Probabilidad
0	0,03125
1	0,15625
2	0,31250
3	0,31250
4	0,15625
5	0,03125
Total	**1**

Distribución binomial

$$P(x) = {}_nC_x \pi^x (1-\pi)^{n-x}$$

Siendo

x = número de caras

π = probabilidad de cara =0,5

n = cantidad de repeticiones = 5

Parámetros del modelo

Función de distribución conjunta

- Se arrojan 5 monedas equilibradas 10 veces y se cuenta la cantidad de caras: 5 3 2 2 4 1 1 3 3 4
- Para cada tirada se puede calcular, utilizando la distribución binomial, la probabilidad de obtener ese resultado:

0.03125, 0.31250 , 0.31250 , 0.31250, 0.15625, 0.15625 , 0.15625 , 0.31250, 0.31250, 0.15625

Nro de caras	Probabilidad
0	0,03125
1	0,15625
2	0,31250
3	0,31250
4	0,15625
5	0,03125
Total	1

- La **función de distribución conjunta** de los valores observados, asumiendo independencia, se calcula como el producto de las probabilidades individuales:

$$f(x/n, \pi, \text{modelo binomial}) = P_{(5)} \times P_{(3)} \times \dots \times P_{(4)} =$$

$$= 0.03125 \times 0.31250 \times \dots = 5.55 \times 10^{-8}$$

Máxima verosimilitud

- La **función de verosimilitud** L de los valores observados se calcula de la misma manera que la distribución conjunta:

  L = 0.03125 x 0.31250 x …. = 5.55×10^{-8}

- Pero mientras que en la función de distribución conjunta, x es variable y los parámetros fijos, en la función de verosimilitud x es fijo y el/los parámetro/s son variables:

  L *(π/x, modelo binomial)*

  Parámetro/s del modelo

- En general:

$$L(\theta / datos, modelo)$$

Otro ejemplo

- Una planta leguminosa de la selva amazónica produce frutos que contienen siempre 5 semillas
- Interesa estimar la probabilidad de que una semilla germine
- Una muestra aleatoria de 10 frutos arrojó la siguiente cantidad de semillas germinadas por fruto:

  3 5 2 4 4 5 4 3 3 4

- ¿Cuál es la probabilidad de los valores observados? Podríamos calcularlas asumiendo una distribución de probabilidades y dado cierto valor para la probabilidad de que una semilla germine. Pero… ¿cuánto vale?
- No lo sabemos. Estimemos por ensayo y error

- Asumiendo que la probabilidad de que una semilla germine es independiente de que lo haga otra, podemos suponer que la distribución de probabilidades es binomial
- Supongamos que la probabilidad de que una semilla germine es $\pi = 0,5$
- La probabilidad estimada de cada uno de los resultados observados (3 5 2 4 4 5 4 3 3 4) es:

0.31250, 0.03125, 0.31250, 0.15625, 0.15625, 0.03125, 0.15625, 0 .31250, 0.31250, 0.15625

Nro de germinadas	Probabilidad
0	0,03125
1	0,15625
2	0,31250
3	0,31250
4	0,15625
5	0,03125
Total	1

- Para $\pi = 0,5$ la función de verosimilitud de los valores observados que se estima como el producto de las probabilidades individuales es:

$$L = \mathbf{5.55 \times 10^{-9}}$$

- Probemos otro valor para la probabilidad de que una semilla germine. Por ejemplo $\pi = 0,7$
- Aplicando la distribución binomial, la probabilidad estimada de los mismos resultados observados es (3 5 2 4 4 5 4 3 3 4) es distinta:

  0.30870, 0.16807, 0.13230, 0.36015, 0.36015,

  0.16807, 0.36015, 0 .30870, 0.30870, 0.36015
- Dado $\pi = 0,7$ la función de verosimilitud de los valores observados que se estima como el producto de las probabilidades individuales es:

  L = **1.85×10^{-6}**

¿Máxima verosimilitud?

- Para $\pi = 0.8$ la verosimilitud de los valores observados es:

  $L = 1.33 \times 10^{-6}$

- Para $\pi = 0.75$ la verosimilitud de los valores observados es:

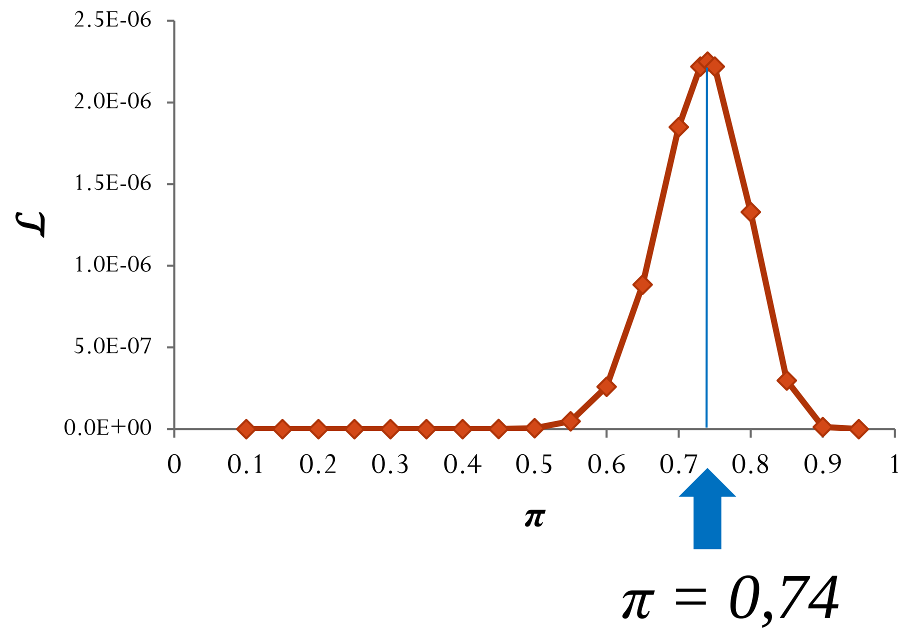
  $L = 2.22 \times 10^{-6}$

- Para $\pi = 0.73$ la verosimilitud de los valores observados es:

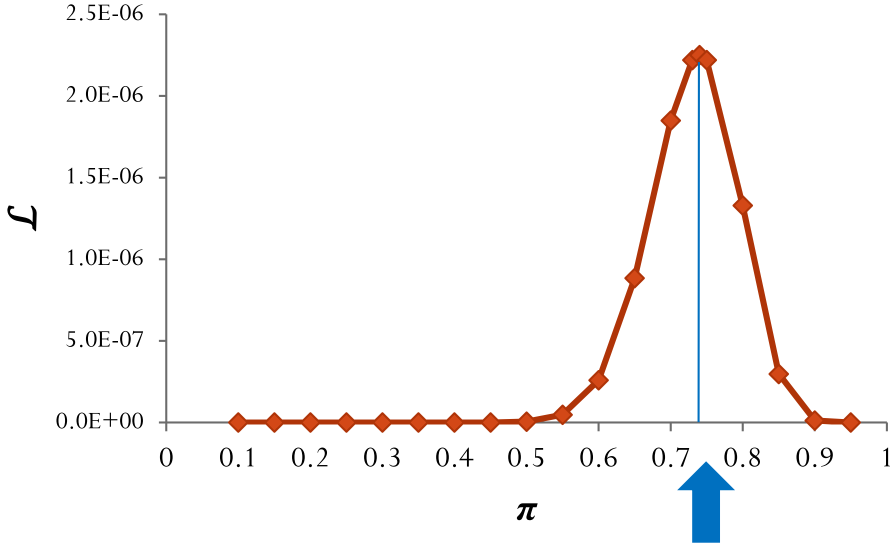
  $L = 2.22 \times 10^{-6}$

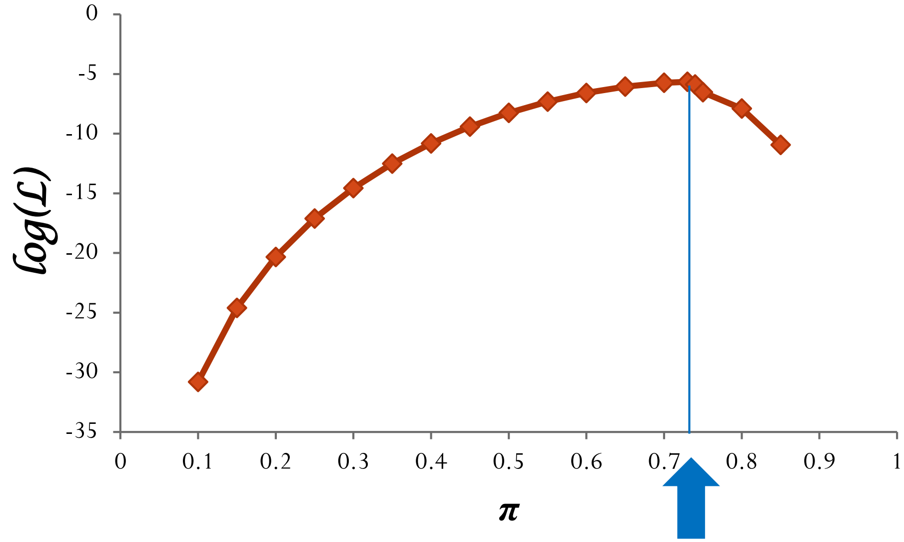
- Para $\pi = 0.74$ la verosimilitud de los valores observados es:

  $L = 2.25 \times 10^{-6}$



- El valor donde la función de verosimilitud L(θ/x) alcanza un máximo determina el valor del estimador por MV
- Al ser un producto de probabilidades, L(θ/x) toma valores entre 0 y 1





$\pi = 0{,}74$

- Para facilitar los cálculos se aplica logaritmo natural (el *ln* del producto se convierte en suma de *ln;* mejor trabajar con sumas que con productos):

$$log L(\theta / x) = \Sigma \, ln \, (f(x_i;\theta))$$

- Se halla el máximo derivando con respecto a θ e igualando a 0

$$\frac{\partial \log L(\theta / x)}{\partial(\theta_j)} = 0$$

- *log*L(θ/*x*) toma siempre valores negativos dado que 0<L<1

Estimación de parámetros por MV

- Se busca el valor del parámetro θ que haga más verosímil (más probable) el resultado que hemos obtenido. Es decir, que maximice la probabilidad de obtener la muestra observada
- Para ello usaremos la función de verosimilitud L(θ) definida como la función de distribución conjunta, función del parámetro desconocido dada una muestra y una distribución de probabilidades:

$$L(\theta / datos, modelo)$$

- Elegiremos el valor de θ para el cual L(θ) (o su log) nos de el máximo
- Si el tamaño de la muestra es grande, las estimaciones por MV proveen estimadores insesgados (la esperanza del estimador coincide con el parámetro) y consistentes (su varianza tiende a cero cuando n tiende a infinito)
- Sin embargo, las estimaciones de las varianzas están sesgadas, por lo que en la práctica se aplica una corrección a los GL: MV restringida o REML

Modelos lineales generales, generalizados y mixtos

- Estimaciones de los parámetros por máxima verosimilitud
- Exigen cierta distribución de probabilidades
- No afectados por desbalanceo
- Permiten modelar la estructura de la matriz de varianzas-covarianzas $\Rightarrow$
  - heterocedasticidad
  - anidamiento, correlación espacial y temporal