Tema 3
Estimación puntual

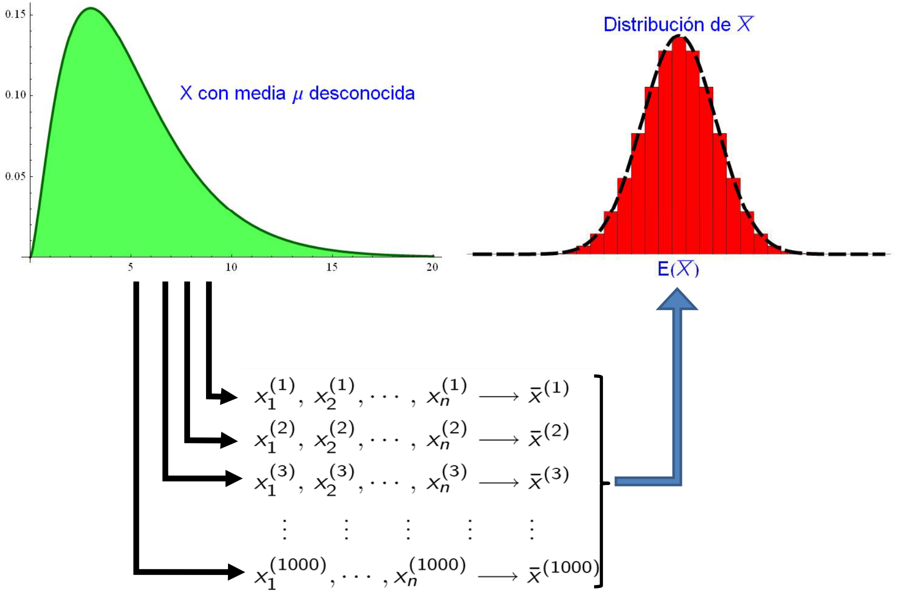
José R. Berrendero

Departamento de Matemáticas
Universidad Autónoma de Madrid

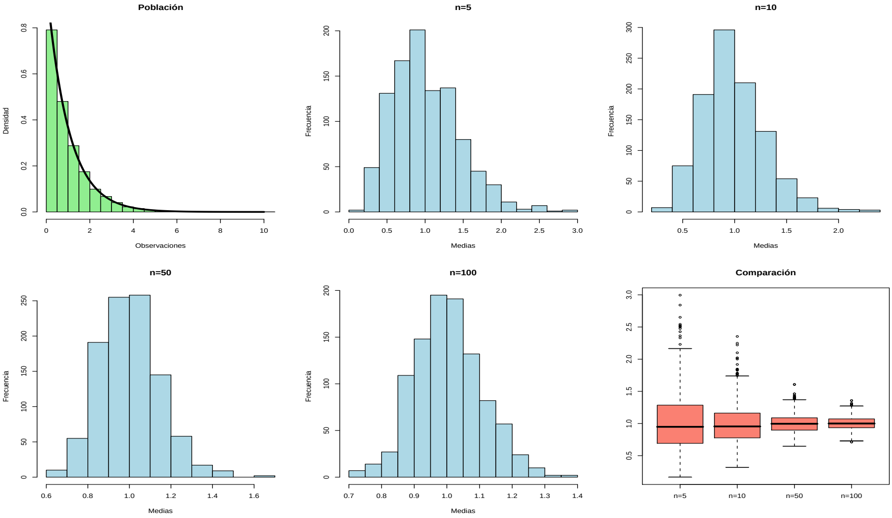
Estructura de este tema

- Estimación de la media poblacional.
- Estimación de la proporción poblacional.
- Sesgo, varianza y error cuadrático medio de un estimador.
- Métodos generales de obtención de estimadores:
  - Método de momentos.
  - Método de máxima verosimilitud.

Distribución de la media muestral



Distribución de la media muestral



Distribución de la media muestral

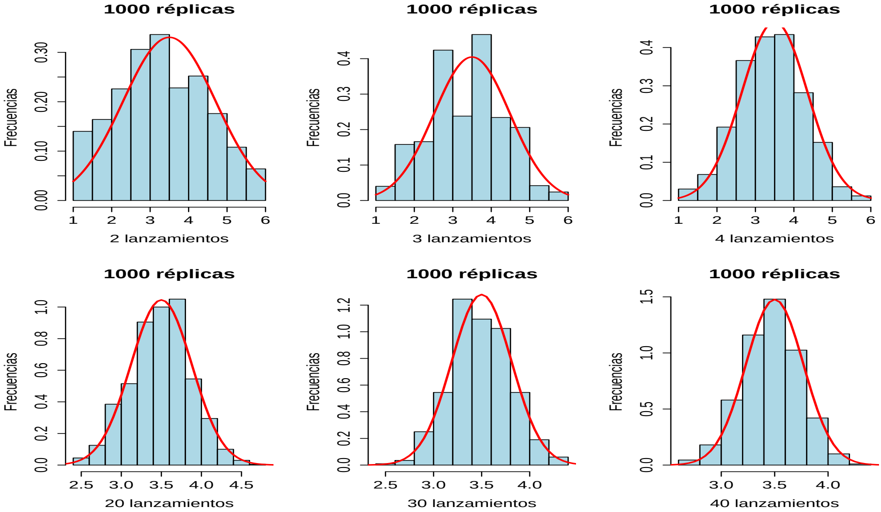
Teorema central del límite: Sea $\bar{x}$ la media de una muestra de tamaño n de una población con media μ y desviación típica σ. Entonces, si n es grande la distribución de los valores que toma $\bar{x}$ es aproximadamente normal de media μ y desviación típica $\sigma/\sqrt{n}$

En notación matemática, podemos escribir:

$$\bar{x} \cong \mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Si la población de partida es normal, el resultado anterior es cierto **de forma exacta para cualquier tamaño muestral n.**

Simulación del promedio al lanzar un dado



Ejemplos

- El tiempo de espera de los estudiantes de la UAM hasta que llega el tren a la estación de Cantoblanco es una variable aleatoria con distribución exponencial de media 10 minutos.

  (a) Calcula la probabilidad de que un estudiante que llega a la estación tenga que esperar entre 5 y 15 minutos.

  (b) Si se calcula el promedio de los tiempos de espera de 100 estudiantes (que llegan a la estación en días y horas diferentes, de manera que los tiempos se pueden considerar independientes), calcula la probabilidad aproximada de que este promedio sea superior a 11 minutos.

  (c) Calcula la probabilidad aproximada de que, entre los 100 estudiantes del apartado anterior, haya más de 45 cuyo tiempo de espera esté entre 5 y 15 minutos.

- El peso de los huevos producidos por una gallina tiene distribución normal de media $\mu = 65$ g y desviación típica $\sigma = 5$ g. ¿Cuál es la probabilidad de que una docena de huevos pese entre 750 y 825 g?

Error típico de la media muestral

El **error típico** de un estimador es un estimador de su desviación típica.

La desviación típica de la media es $\sigma/\sqrt{n}$, pero en la práctica σ es un parámetro poblacional desconocido.

Resulta natural estimar σ^2 con la cuasivarianza muestral:

$$S^2 = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}.$$

Se divide $n - 1$ ya que puede demostrarse que al dividir por n el estimador tiene una tendencia sistemática a infraestimar σ^2.

El error típico de la media muestral es

$$\frac{S}{\sqrt{n}}$$

Error típico de la media muestral

¿Sabes distinguir entre los conceptos siguientes? ¿Qué notación estamos usando para cada uno de ellos?

- La varianza de la población
- La desviación típica de la población.
- La varianza de la media muestral.
- La desviación típica de la media muestral.

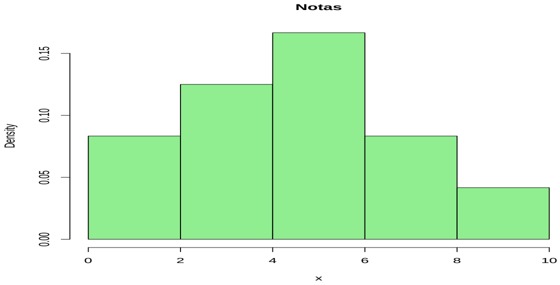
- La cuasivarianza muestral.
- La cuasidesviación típica muestral.
- El error típico de la media muestral.

¿En qué se diferencian los cuatro primeros de los tres últimos?

Ejemplo con una población pequeña

- Población: Los 12 alumnos de una clase.
- Variable: Nota que un alumno obtiene en un examen

Estudiante	1	2	3	4	5	6	7	8	9	10	11	12
Nota	1	0	3	10	8	7	5	5	5	6	4	3



Parámetros poblacionales

- Media poblacional:

$$\mu = \frac{1+0+3+10+8+7+5+5+5+6+4+3}{12} = 4.75$$

- Varianza poblacional:

$$\sigma^2 = \frac{(1-4.75)^2 + (0-4.75)^2 + \cdots + (3-4.75)^2}{12} = 7.3542$$

- Desviación típica poblacional:

$$\sigma = \sqrt{7.3542} = 2.7119$$

Una muestra de tamaño $n = 4$

- Una posible muestra de tamaño 4 es:

Estudiante	1	2	3	4	5	6	7	8	9	10	11	12
Nota	1	0	3	10	8	7	5	5	5	6	4	3

$$x_1 = 4, \quad x_2 = 3, \quad x_3 = 5, \quad x_4 = 6$$

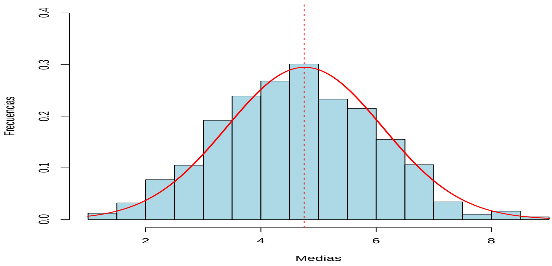
- A partir de estos datos, un estimador de μ (que sería útil si no conociéramos μ) es:

$$\hat{\mu} = \bar{x} = \frac{4+3+5+6}{4} = 4.5$$

- ¿Cómo se evalúa la precisión de $\bar{x}$, sin conocer μ?

2000 muestras de tamaño 4

- Extraemos 2000 muestras de tamaño 4.
- Todos los valores son equiprobables y se extraen con reemplazamiento (muestreo aleatorio simple).
- Un histograma de las correspondientes 2000 medias muestrales:



Características de la distribución de $\bar{x}$

- Las propiedades de $\bar{x}$ como estimador de μ se corresponden con las propiedades del histograma anterior.
- La forma del histograma es la de una distribución normal.
- Los valores de $\bar{x}$ se centran alrededor del verdadero valor de μ. El estimador es **centrado o insesgado**.
- La desviación típica de $\bar{x}$ es menor que σ. Se puede demostrar que la desviación típica de $\bar{x}$ es:

$$\frac{\sigma}{\sqrt{n}} = \frac{2.7119}{2} \approx 1.356.$$

Conclusiones de las observaciones anteriores

- Como $\bar{x}$ es insesgado, no hay tendencia sistemática a infraestimar o sobreestimar el valor de μ.

- Como $\bar{x} \cong \mathrm{N}(\mu, \sigma/\sqrt{n})$, con probabilidad aproximada 0.95 el error cometido al estimar μ mediante $\bar{x}$ es menor o igual que $2 \times \sigma/\sqrt{n} \approx 2.7119$

- Es decir, que podemos tener bastante confianza en que el valor de μ se encuentra en el intervalo:

$$[4.5 \mp 2.7119]$$

- Como en la práctica σ^2 es desconocida se usa S^2 en su lugar:

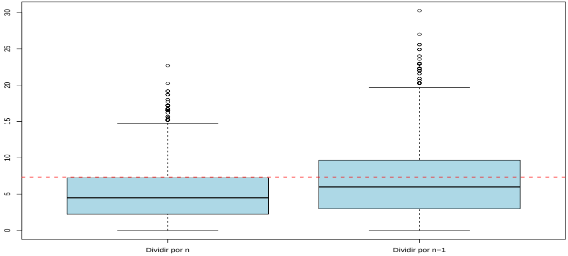
$$S^2 = \frac{(4-4.5)^2 + (3-4.5)^2 + (5-4.5)^2 + (6-4.5)^2}{3} = 1.666.$$

¿Por qué se divide por $n - 1$ en lugar de n?

- Puede comprobarse que la varianza muestral (dividiendo por n) presenta una tendencia sistemática a infraestimar σ^2.

- Para corregir este sesgo se incrementa ligeramente el valor del estimador dividiendo por $n - 1$ en lugar de n.

- Diagramas de cajas de las 2000 varianzas y cuasivarianzas muestrales. La línea roja corresponde a $\sigma^2 = 7.3542$.



Estimación de una proporción poblacional

Queremos estimar la proporción p de personas en una población que han seguido una dieta en los últimos 5 años. Para ello, preguntamos a 10 personas y definimos

$$x_i = \begin{cases} 0, & \text{si la persona } i \text{ no ha seguido una dieta;} \\ 1, & \text{si la persona } i \text{ ha seguido una dieta.} \end{cases}$$

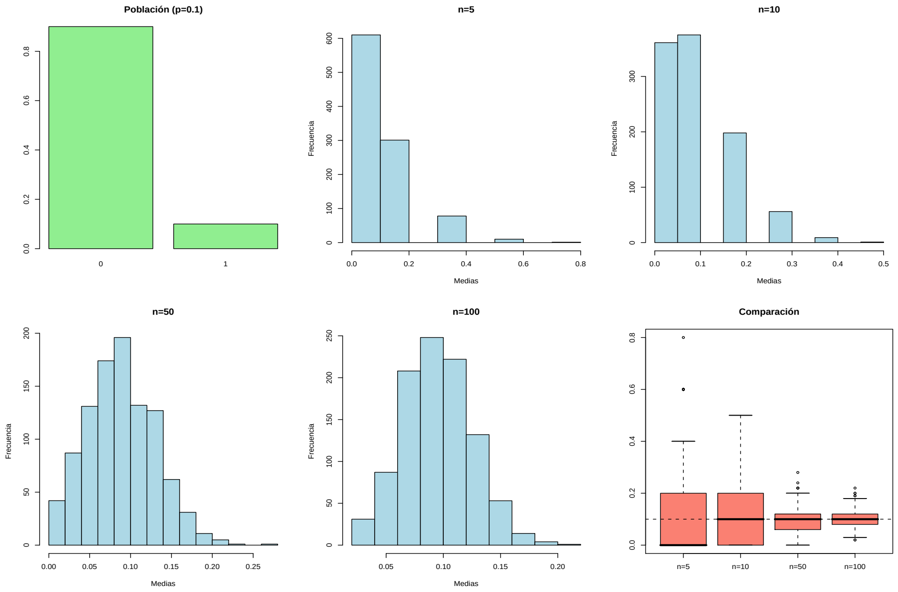
Obtenemos los siguientes datos:

$$1,\ 0,\ 0,\ 1,\ 1,\ 0,\ 0,\ 0,\ 1,\ 0$$

Estos datos son 10 observaciones de una v.a. de Bernoulli con parámetro p.

¿Cuál es el estimador más natural de p?

Distribución de la proporción muestral



Distribución de la proporción muestral

Según el TCL, ¿cómo se distribuye aproximadamente la proporción muestral $\hat{p}$?

¿Cuál es la desviación típica de $\hat{p}$?

¿Cuál es el máximo (mínimo) valor posible de esta desviación típica?

¿En qué situación se va a dar ese valor?

En general, ¿cuál es el error típico de $\hat{p}$?

Calcula el error típico de $\hat{p}$ para los datos de la encuesta sobre la dieta.

Estimación puntual: planteamiento general

Disponemos de una muestra aleatoria simple $X_1, \ldots, X_n$ de una v.a. X:

- Las observaciones $X_1, \ldots, X_n$ son **independientes**.
- Todas ellas tienen **la misma distribución** que X

Se supone que la distribución de X es conocida salvo por el valor de un conjunto de **parámetros** que denotamos θ.

Objetivo: Aproximar el valor de θ a partir de la muestra. Para ello necesitamos calcular un **estimador** $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$.

¿Que propiedades debe tener un buen estimador?

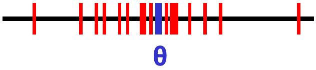
¿Existen métodos generales para obtener estimadores?

Sesgo

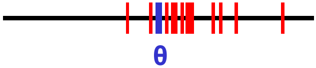
$$\text{Sesgo}(\hat{\theta}) = \mathsf{E}(\hat{\theta}) - \theta.$$

Un buen estimador debe ser insesgado o tener un sesgo pequeño.

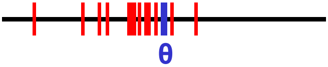
Estimador insesgado:



Sesgo positivo:

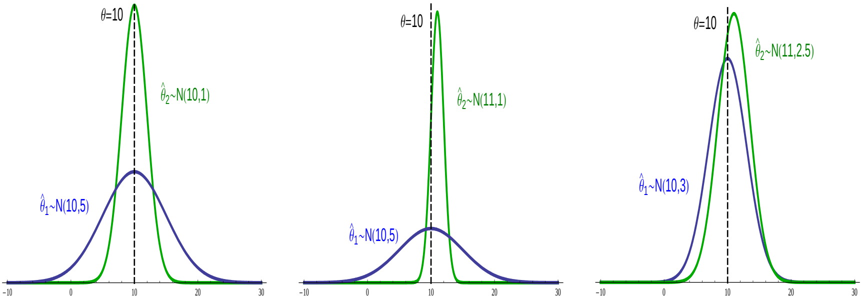


Sesgo negativo:



Varianza

La varianza de un estimador debe ser lo menor posible.



Error cuadrático medio

Es una medida de la calidad de un estimador que tiene en cuenta tanto el sesgo como la varianza.

El **error cuadrático medio** de un estimador $\hat{\theta}$ es:

$$\text{ECM}(\hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

Puede comprobarse que

$$\text{ECM}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{Sesgo}^2(\hat{\theta})$$

De acuerdo con el ECM, ¿qué estimador es mejor en cada gráfico de la transparencia anterior?

Método de momentos

Paso 1. Calculamos la media poblacional μ para determinar la relación entre μ y el parámetro θ que queremos estimar:

$$\theta = f(\mu)$$

Paso 2. Calculamos la media muestral $\bar{x}$, y estimamos μ mediante $\bar{x}$.

Paso 3. El estimador de θ es

$$\hat{\theta} = f(\bar{x}).$$

La idea es suponer que la relación entre $\bar{x}$ y $\hat{\theta}$ es la misma que la relación entre μ y θ.

Método de momentos: algunas distribuciones conocidas

- Si $X \equiv P(\lambda)$, entonces $\hat{\lambda} = \bar{X}$.
- Si $X \equiv \mathrm{B}(1; p)$, entonces $\hat{p}$ es la proporción muestral.
- Si $X \equiv \mathrm{Exp}(\lambda)$, entonces $\hat{\lambda} = 1/\bar{X}$.
- Si $X \equiv \mathrm{N}(\mu; \sigma)$, entonces $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = V_X$.

Observación: si hay dos parámetros desconocidos, se plantea un sistema de dos ecuaciones usando la media de los datos y la media de los datos al cuadrado.

Método de momentos: ejemplo

La distribución de una variable aleatoria X es

Valores	0	1	2
Probabilidades	p	p	$1-2p$

- ¿Qué valores puede tomar el parámetro desconocido p?
- Calcula la esperanza de X.
- Se observa la muestra 1,0,0,1,2,2,2,2. Calcula el estimador de momentos de p a partir de estos datos.
- Calcula el estimador de momentos de una muestra cualquiera $X_1, \ldots, X_n$.
- Calcula el estimador de momentos correspondiente a la muestra 0,0,0,0,0,1,1,2. ¿Crees que el resultado obtenido es bueno?

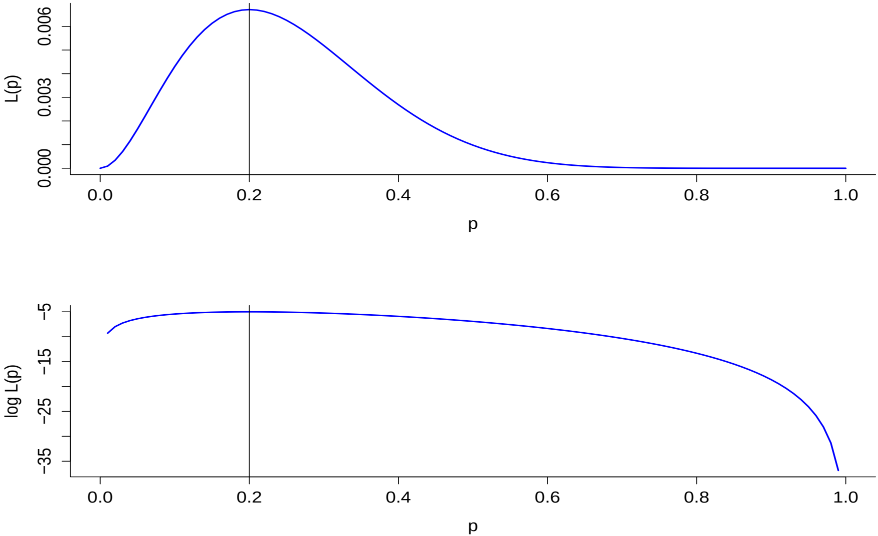
Método de máxima verosimilitud

En un estudio sobre sus efectos secundarios se administró un medicamento a 10 personas, con los resultados siguientes: 0,0,0,0,1,0,0,0,0,1 (0 significa que no hubo efectos y 1 que sí).

Se quiere estimar p, la probabilidad de que el medicamento tenga efectos secundarios.

- Calcula la probabilidad de observar la muestra anterior como función de p (es decir, la función de verosimilitud $L(p)$ correspondiente a la muestra).
- Calcula el logaritmo de la función anterior, $\ln L(p)$.
- ¿En qué punto se maximizan $L(p)$ y $\log L(p)$?

Método de máxima verosimilitud



Método de máxima verosimilitud

Sea $X_1, \ldots, X_n$ una muestra de una v.a. X cuya distribución tiene función de densidad o de probabilidad f_θ. Para calcular el EMV,

Paso 1. Calculamos la función de verosimilitud

$$L(\theta) = f_\theta(X_1) \cdots f_\theta(X_n)$$

Paso 2. Calculamos el logaritmo de la función de verosimilitud:

$$\ln L(\theta) = \ln f_\theta(X_1) + \cdots + \ln f_\theta(X_n)$$

Paso 3. Derivamos $\ln L(\theta)$ respecto a θ e igualamos a cero:

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

Bajo ciertas condiciones de regularidad, la solución es el EMV de θ.

Método de máxima verosimilitud: distribución continua

El tiempo de supervivencia en años de una enfermedad sigue una distribución exponencial de parámetro λ. Se han registrado los tiempos de supervivencia de 4 individuos resultando 5.5, 6, 10, 1. Calcula el correspondiente estimador de máxima verosimilitud de λ.

Función de verosimilitud:

$$L(\lambda) = \lambda e^{-5.5\lambda} \cdot \lambda e^{-6\lambda} \cdot \lambda e^{-10\lambda} \cdot \lambda e^{-1\lambda} = \lambda^4 e^{-22.5\lambda}$$

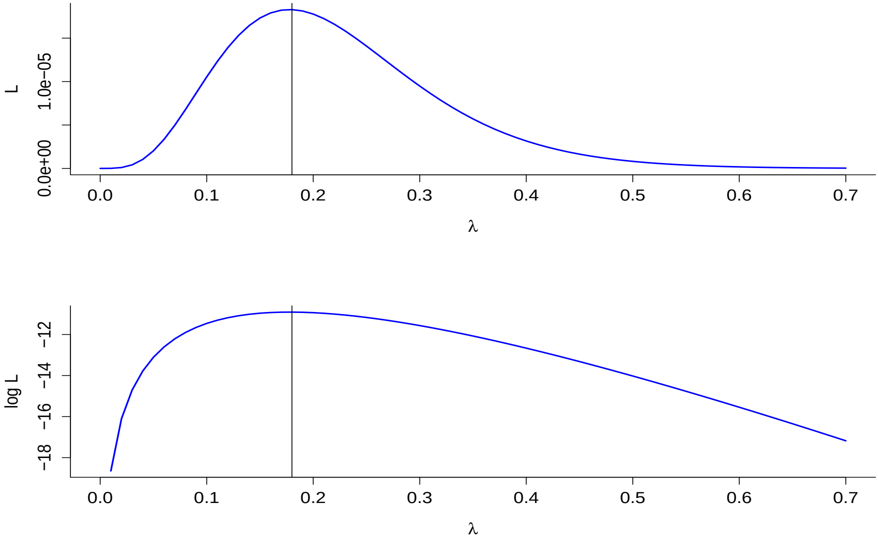
Tomamos logaritmos:

$$\ln L(\lambda) = 4 \ln \lambda - 22.5\lambda$$

Derivamos e igualamos a cero:

$$\frac{4}{\hat{\lambda}} - 22.5 = 0 \;\Rightarrow\; \hat{\lambda} = \frac{4}{22.5}$$

Método de máxima verosimilitud



Método de máxima verosimilitud y de momentos

La distribución de una variable aleatoria X es

Valores	0	1	2
Probabilidades	p	p	$1-2p$

- Se observa la muestra 1,0,0,1,2,2,2,2. Calcula el estimador de máxima verosimilitud de p a partir de estos datos.
- Calcula el estimador de máxima verosimilitud de p correspondiente a la muestra 0,0,0,0,0,1,1,2.
- Compara los resultados con los obtenidos con el estimador de momentos.
- Si en una muestra de tamaño n no se ha observado ningún 2, ¿cuál es el estimador de máxima verosimilitud?
- Si en una muestra de tamaño n se han observado k valores iguales a 2, ¿cuál es el estimador de máxima verosimilitud?