Tema 6. ESTIMACIÓN PUNTUAL.

Objetivos

Conceptos:

- Comprender y distinguir los siguientes conceptos:
  - muestra aleatoria simple
  - estadístico
  - estimador de un parámetro
  - estimación de un parámetro
  - estimador centrado
  - estimador más eficiente que otro
- Comprender el significado del estimador de máxima verosimilitud y del modo de obtenerlo.

Saber hacer:

- Dada una variable aleatoria con modelo de distribución de probabilidad conocido, que depende de uno o más parámetros:
  - Hallar el estimador del parámetro por el método de los momentos
  - Hallar el estimador del parámetro por el método de máxima verosimilitud
- Dado un estimador de un parámetro:
  - Saber estudiar si es o no centrado
  - Saber estudiar si es más eficiente que otro estimador del mismo parámetro
  - Saber hallar su error cuadrático medio
- A partir de los resultados de una muestra aleatoria simple de una v.a.
  - Hallar una estimación de sus parámetros por el método de los momentos
  - Hallar una estimación de sus parámetros por el método de máxima verosimilitud

Problemas de exámenes (web):

SIN: Junio 2006: Problema 3 b)
Septiembre 2006: Problema 2
Febrero 2005: Problema 3 b)
Junio 2005: Problema 2 a)b)
Febrero 2004: Problema 2 c)d)
Junio 2004: Problema 2 c)
Junio 2003: Problema 2 e)
Septiembre 2003: Problema 3 a)b)c)

6.1.- Introducción

Hemos visto en prácticas cómo buscar un modelo probabilístico adecuado para un conjunto de datos. Recordemos lo visto al final del tema 4:

Ajuste de una variable estadística a un modelo teórico

Objetivo:

- elegir un modelo
- encontrar los parámetros del modelo

Medios:

- definición de la variable: ¿qué mide? ¿en qué condiciones?
- medidas descriptivas (media, varianza, simetrías, frecuencias, ...)
- representaciones gráficas

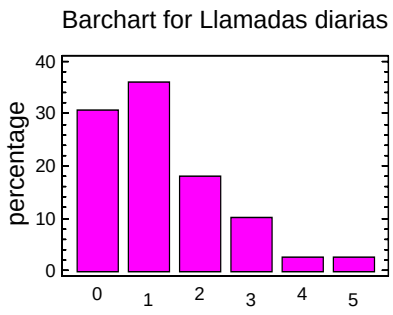
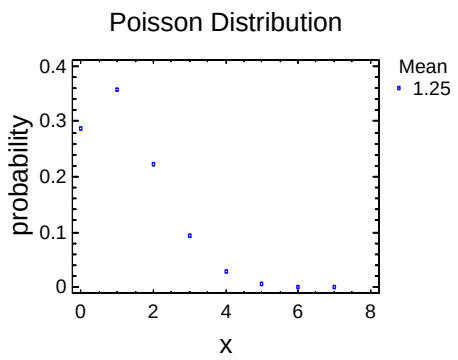
Verificación:

Contrastes no paramétricos con Statgraphics (Distribution Fitting):
p-valor > 0.3 para aceptar la hipótesis
(cuanto mayor sea, con más confianza se acepta el modelo propuesto)

y poníamos el ejemplo:

Ejemplo: X4: Número de llamadas diarias que se hacen por teléfono móvil

```
Average = 1.25641     ; Variance = 1.51147
```

Diagrama de barras de frecuencias relativas	**Función de masa P(1.25)**
	


```
Goodness-of-Fit Tests for Llamadas diarias          :

Fitted Poisson distribution:
     mean = 1.25641
Chi-Square = 0.555238 with 2 d.f.    P-Value = 0.757586
```

Luego aceptamos que la variable *número de llamadas diarias*, puede tener una distribución de Poisson de parámetro λ=1.25.

(Para otros ejemplos ver también las páginas 46 a 52 de la Guía Docente.)

Este tipo de trabajo se ha hecho a partir de los datos concretos de una variable estadística. Pero nos podemos plantear algunas cuestiones:

¿Se puede asegurar que esos resultados son válidos para todos los estudiantes de la EUI?
En caso afirmativo, ¿con qué confianza?; en caso contrario, ¿en qué condiciones podrían ser válidos?

En general, lo que se quiere estudiar es alguna característica de una población determinada y **encontrar un modelo probabilístico** para ella **a partir de unos datos concretos**. Este es el objetivo de la llamada ***Inferencia estadística***. Lo que veremos en este tema y los siguientes es cómo hacerlo y cómo medir la fiabilidad de los resultados obtenidos.

Damos una visión general de la ***Inferencia estadística***:

Para obtener los DATOS:

- **Censo** (toda la población)
- **Muestra** (subconjunto de elementos de la población)

  Es fundamental: **cómo** elegir los elementos de la muestra y el **tamaño** de la misma[1].

  - Población homogénea (**m.a.s.** :muestreo aleatorio simple[2])
  - Población heterogénea (muestreo estratificado[3])

Para obtener el MODELO PROBABILÍSTICO:

- Hipótesis sobre el modelo (por la definición de la variable, gráficas, datos, ...)
- **Inferencia paramétrica:** conocido el modelo, buscamos **información sobre los parámetros** (θ) del mismo, mediante:
  - **Estimación puntual:** $\theta = \theta_0$
  - **Estimación por intervalo:** $\theta \in (a,b)$ con un % de confianza
  - **Contraste de hipótesis:** aceptamos $\theta = \theta_0$ frente a $\theta \neq \theta_0$ (ó $\theta > \theta_0$ ó $\theta < \theta_0$), con nivel de significación α.
- **Inferencia no paramétrica:** para verificar las **hipótesis sobre el modelo** o sobre **independencia** de variables o muestras.
  - **Ajuste de distribuciones:** Aceptamos que **X~Modelo(θ)** con nivel de significación α.
  - **Independencia:** X e Y son independientes; los datos de la muestra son independientes.

[1] Si el **tamaño de la muestra** es demasiado pequeño la información obtenida puede no ser representativa y si es excesivamente grande podemos estar derrochando recursos sin obtener más información relevante.

[2] **m.a.s.:** todos los elementos de la población tienen la **misma probabilidad** de estar en la muestra y son **independientes** entre ellos. Para ello, la elección de los elementos se debería hacer con reemplazamiento; sin embargo, cuando el tamaño de la muestra es menor del 5% del tamaño de la población se considera válido hacerlo sin reemplazamiento. Se suele hacer mediante enumeración de los elementos y elección aleatoria de los mismos.

[3] Se divide la población en clases o **estratos** homogéneos (repecto a las variables que afectan a la característica estudiada) y dentro de cada uno de ellos se hace una m.a.s. El tamaño de la muestra de cada estrato ha de ser proporcional al tamaño del estrato dentro de la población general.

Ejemplo: Se quiere estudiar el número de llamadas diarias por el móvil que hace un estudiante de la EUI.

Para obtener los DATOS:

- **Censo** (toda la población) Podría hacerse pero con un coste excesivo y no es imprescindible para obtener la información que se precisa. (Sí es necesario hacerlo en unas elecciones, la evaluación de una asignatura, ...)
- **Muestra** (subconjunto de elementos de la población)
  - Población homogénea (**m.a.s.** :muestreo aleatorio simple)
    Entre los estudiantes de la EUI se podrían elegir 100 (menos del 5%) de forma aleatoria.
  - Población heterogénea (muestreo estratificado)
    Se podrían hacer estratos por grupos de edad, nivel educativo, zona de residencia, sexo, ... y en cada uno de ellos se hace una m.a.s.

Para obtener el MODELO PROBABILÍSTICO:

(Para el ejemplo consideramos como resultado de la muestra los datos obtenidos con n=39 en el grupo GM23)

- Hipótesis sobre el modelo (por la definición de la variable, gráficas, datos, ...)
  Por la definición de la misma podría ajustarse a un modelo de Poisson; el diagrama de barras, y la similitud de media y varianza lo confirman.

- **Inferencia no paramétrica:** para verificar las **hipótesis sobre el modelo** o sobre **independencia** de variables o muestras.

  - Ajuste de distribuciones:
    A partir de un test (Chi-cuadrado) **aceptamos que X~P(1.25461)** con un p-valor de 0.75 (que implica bastante confianza en dicha decisión).

- **Inferencia paramétrica:** como suponemos que es un modelo de Poisson **buscamos información sobre su parámetro** λ:

  - **Estimación puntual:** $\theta = \theta_0$
    Se asigna al parámetro **un valor concreto**: $\lambda = 1.25641$ (como λ es la media de una Poisson, una posible estimación es la media muestral $\overline{X} = 1.25641$)

  - **Estimación por intervalo:** $\theta \in (a,b)$ con un % de confianza
    Se encuentra un **intervalo en el que está incluido el parámetro** con una determinada confianza:

    ```
    Confidence Intervals for Llamadas diarias
    ----------------------------------------
    ```
    95.0% confidence interval for mean: 1.25641 +/- 0.39853 **[0.857878;1.65494]**

    Según explica Statgraphics: *In practical terms we state with 95.0% confidence that the true mean Llamadas diarias is somewhere between 0.857878 and 1.65494.*

  - **Contraste de hipótesis:** aceptamos $\theta = \theta_0$ frente a $\theta \neq \theta_0$ (ó $\theta > \theta_0$ ó $\theta < \theta_0$), con nivel de significación α.

    Se plantean dos hipótesis y se estudia si los datos de la muestra aportan la suficiente evidencia para rechazar una de ellas frente a la otra.

Con los datos de la muestra, y tras un análisis adecuado con Statgraphics:

Rechazamos la hipótesis de que $\lambda = 1$ frente a $\lambda > 1$, con un 85% confianza (nivel α=0.15). **La muestra aporta datos evidentes** de que el número medio de llamadas (de toda la población) es mayor que uno.

Aunque la media de las mujeres es $\lambda_M = 1.5$ y la de los hombres es $\lambda_V = 1.15$, **no podemos rechazar** la hipótesis de que sean iguales ($\lambda_M = \lambda_V$). **La muestra no aporta datos evidentes** para poder asegurar que el número medio de llamadas en las mujeres es mayor que el de los varones. El tamaño muestral es muy pequeño y la diferencia entre las medias muestrales puede deberse a factores aleatorios y no puede garantizarse que sean distintas.

En este tema estudiaremos cómo realizar la **estimación puntual**, en concreto veremos diversas formas de obtener dichas estimaciones y qué propiedades son deseables que verifiquen.

6.2.- Estimación puntual

Dada una v.a. X, de la que suponemos conocido el modelo de distribución que sigue, buscamos información sobre los parámetros (θ) del mismo; en el caso de la **estimación puntual**, **buscamos un valor concreto para** θ.

Antes, hemos de definir los conceptos matemáticos y la notación que utilizaremos.

X será la **variable aleatoria** de la que queremos estudiar su modelo de distribución M(θ)

Antes de obtener los datos (n)	**Después de obtener los datos (n)**
Muestra aleatoria simple (m.a.s.): $X_1, \dots, X_n$ son **v.a. independientes con igual distribución** M(θ).	Resultado de la muestra: $x_1, \dots, x_n \in \mathbb{R}$
Estadístico: $T=T(X_1, \dots, X_n)$ es una **v.a.** que depende de la m.a.s.	Valor del estadístico: $t=T(x_1, \dots, x_n) \in \mathbb{R}$
Estimador: $\theta^* = \theta^*(X_1, \dots, X_n)$ Es un estadístico (por tanto, una **v.a.**) que **se elige para encontrar un valor para el parámetro** θ.	Estimación de θ: $\theta^*(x_1, \dots, x_n) \in \mathbb{R}$

Ejemplo :

X: número de llamadas diarias por el móvil que hace un estudiante de la EUI. Suponemos $X \sim P(\lambda)$.

Antes de obtener los datos (n)	**Después de obtener los datos (n=39)**
$X_1, \dots, X_n$ son **v.a.** independientes con distribución $P(\lambda)$.	Resultado de la muestra: $\{1,0,3,2,\dots,2\} \in \mathbb{R}$
Estadísticos: Media muestral: $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ Varianza muestral: $V = \frac{(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}$ Máximo: $Máx(X_1, \dots, X_n)$	Valor del estadístico: $\bar{X} = 1.25641 \in \mathbb{R}$ $V = 1.47368 \in \mathbb{R}$ $Máx(X_i) = 5 \in \mathbb{R}$

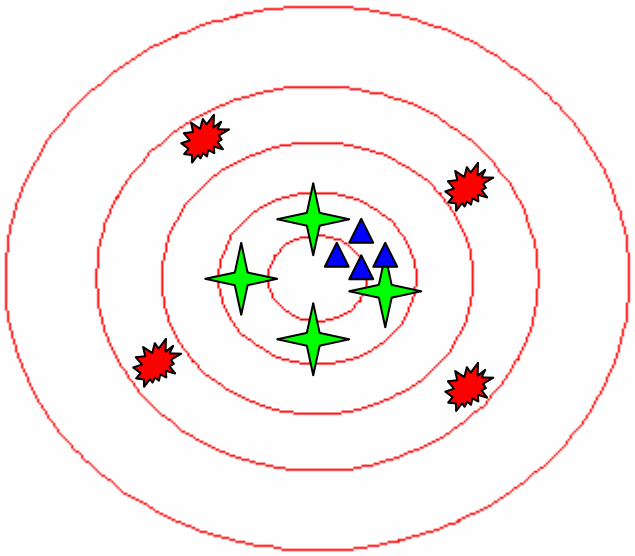
Estimador:	Estimación de λ:
Como $X \sim P(\lambda)$, sabemos que $E(X)=\lambda$ y que $V(X)=\lambda$. En este caso, los estadísticos $\bar{X}$ y V, podrían ser adecuados para obtener un valor para λ: $\lambda^* = \bar{X}$, $\lambda' = V$	En este caso, podemos obtener dos estimaciones para λ: $\lambda \sim \lambda^* = 1.25641 \in \mathbb{R}$ $\lambda \sim \lambda' = 1.47368 \in \mathbb{R}$

6.4.- Propiedades de los estimadores

En el ejemplo anterior, podemos elegir dos estimadores distintos para el parámetro λ de una distribución de Poisson, y ambos con criterios "razonables": la media de la Poisson es λ y la varianza de la Poisson también es λ. ¿Con cuál nos quedamos? ¿Qué criterios vamos a utilizar para determinar qué estimadores son los más adecuados?

Ejemplo:

El objetivo es dar en el centro de la diana, y los impactos de distintas escopetas están señalados con diferentes señales.



¿Cuál parece más adecuada? ¿Qué ventajas tienen unas frentes a otras?

Los "triángulos-azules" están muy agrupados entre sí (buena puntería), pero entorno a un punto que no es el objetivo (punto de mira algo desviado).

Las "estrellas-verdes" tienen bien el punto de mira y una puntería moderadamente buena.

Los "impactos-rojos" parecen tener bien el punto de mira (están repartidos homogéneamente alrededor del objetivo), pero una puntería pésima.

Lo deseable sería tener bien el punto de mira y una buena puntería.

Si lo "traducimos" a los criterios que vamos a utilizar para seleccionar un estimador, buscaremos estimadores en los que el valor esperado coincida con el deseado (tener bien el punto de mira) y con la menor varianza posible (buena puntería).

Definición.

Decimos que θ^* es un **estimador centrado (o insesgado) de** θ, si $\mathbf{E(\theta^*)=\theta}$, para cualquier valor de θ.

En caso contrario decimos que es **sesgado** y se define **sesgo de** θ^* como $b(\theta^*)=E(\theta^*)-\theta$.

Ejemplos:

- La **media muestral** $\bar{X}$ es un estimador centrado de E(*X*)(demo[4]). Por tanto,
  - Si $X \sim P(\lambda)$, $\lambda^* = \bar{X}$ es un estimador centrado para λ .
  - Si $X \sim N(\mu, \sigma)$, $\mu^* = \bar{X}$ es un estimador centrado para μ..
- La **varianza muestral** $V = \frac{(X_1 - \bar{X})^2 + ... + (X_n - \bar{X})^2}{n}$ **no** es un estimador centrado para V(*X*)=σ^2. De hecho, E(***V***)=$\frac{n-1}{n}\sigma^2$ (demo[5]). Por tanto, si $X \sim P(\lambda)$, $\lambda' = V$ no es un estimador centrado para λ .
- La **cuasivarianza muestral** $S^2 = \frac{(X_1 - \bar{X})^2 + ... + (X_n - \bar{X})^2}{n-1}$ **sí** es un estimador centrado para V(*X*)=σ^2 (demo[6]). Por tanto, si $X \sim P(\lambda)$, S^2 es un estimador centrado para λ .

Observación:

- No siempre es fácil estudiar si un estimador es centrado.
- No siempre existe un estimador centrado para cualquier parámetro.

Definición.

Decimos que un estimador $\hat{\theta}_1$ es **más eficiente que** $\hat{\theta}_2$ cuando $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.
Si un estimador tiene la mínima varianza posible se dice que es **eficiente**.

Observación:

- No siempre es fácil estudiar la varianza de un estimador, y como consecuencia, no será fácil comparar o estudiar la eficiencia de los estimadores.
- No siempre existe un estimador eficiente para cualquier parámetro.

Ejemplos:

- La varianza muestral ***V*** es más eficiente que la cuasivarianza muestral S^2:
  Como $S^2 = \left(\frac{n}{n-1}\right)V \Rightarrow V(S^2) = \left(\frac{n}{n-1}\right)^2 V(V)$, por tanto $V(S^2) > V(V)$.
- Si $X \sim P(\lambda)$, $\bar{X}$ y S^2 son estimadores centrados para λ , ¿cuál es más eficiente? Hallamos sus varianzas:

$$V(\bar{X}) = \frac{\lambda}{n} \text{ (demo[7])} \quad V(S^2) = \frac{2}{n-1}\lambda^2 + \frac{\lambda}{n} \text{ (demo[8])}$$

---

[4] $E(\bar{X}) = E\left(\frac{X_1 + ... + X_n}{n}\right) = \frac{1}{n}\left(E(X_1) + ... + E(X_n)\right) = \frac{1}{n}\left(\mu + ... + \mu\right) = \frac{1}{n}(n\mu) = \mu$

[5] Como es más extensa, se desarrolla al final del tema, ANEXO 1.

[6] $S^2 = \left(\frac{n}{n-1}\right)V \Rightarrow E(S^2) = E\left(\left(\frac{n}{n-1}\right)V\right) = \left(\frac{n}{n-1}\right)E(V) = \left(\frac{n}{n-1}\right)\left(\frac{n-1}{n}\right)\sigma^2 = \sigma^2$

[7] $V\left(\frac{X_1 + ... + X_n}{n}\right) = \frac{1}{n^2}\left(V(X_1) + ... + V(X_n)\right) = \frac{1}{n^2}\left(\lambda + ... + \lambda\right) = \frac{1}{n^2}(n\lambda) = \frac{\lambda}{n}$

[8] Se basa en la fórmula general de la varianza de la varianza muestral. Los detalles en el ANEXO 2.

Como $\frac{2}{n-1}\lambda^2 > 0$, se tiene que $V(\bar{X}) < V(S^2)$, y por tanto $\bar{X}$ es más eficiente que S^2.

Una medida que recoge estas dos propiedades de los estimadores (centrado y eficiente) es la siguiente:

Definición.

Si **θ*** es un estimador de θ, se define su **error cuadrático medio** como:

$$ECM(\theta^*) = E\left((\theta^* - \theta)^2\right)$$

(Mide la "distancia esperada" entre el estimador y el parámetro que queremos estimar.)

Propiedades:

- Se verifica que $ECM(\theta^*) = V\left(\theta^*\right) + \left(E(\theta^*) - \theta\right)^2$
- Si θ^* es un estimador centrado, entonces $ECM(\theta^*) = V\left(\theta^*\right)$.
- Si un estimador es centrado y eficiente (mínima varianza), tendrá el menor ECM posible.

Ejemplos:

- En una muestra aleatoria simple de tamaño 3 de una variable aleatoria normal de media μ y desviación típica 1, se consideran los estimadores para μ:

$$U_3 = \frac{1}{8}X_1 + \frac{3}{8}X_2 + \frac{1}{2}X_3 \; ; \; U_4 = \frac{1}{9}X_1 + \frac{2}{9}X_2 + \frac{4}{9}X_3 .$$

Estudiamos cuál de ellos tiene menor error cuadrático medio:

$$E(U_3) = E\left(\frac{1}{8}X_1 + \frac{3}{8}X_2 + \frac{1}{2}X_3\right) = \frac{1}{8}\mu + \frac{3}{8}\mu + \frac{1}{2}\mu = \mu$$

$$V(U_3) = V\left(\frac{1}{8}X_1 + \frac{3}{8}X_2 + \frac{1}{2}X_3\right) = \left(\frac{1}{8}\right)^2 \cdot 1 + \left(\frac{3}{8}\right)^2 \cdot 1 + \left(\frac{1}{2}\right)^2 \cdot 1 = \frac{13}{32}$$

Por tanto $ECM(U_3) = V\left(U_3\right) + \left(E(U_3) - \mu\right)^2 = \frac{13}{32} + \left(\mu - \mu\right)^2 = \frac{13}{32}$.

$$E(U_4) = E\left(\frac{1}{9}X_1 + \frac{2}{9}X_2 + \frac{4}{9}X_3\right) = \frac{1}{9}\mu + \frac{2}{9}\mu + \frac{4}{9}\mu = \frac{7}{9}\mu$$

$$V(U_4) = V\left(\frac{1}{9}X_1 + \frac{2}{9}X_2 + \frac{4}{9}X_3\right) = \left(\frac{1}{9}\right)^2 \cdot 1 + \left(\frac{2}{9}\right)^2 \cdot 1 + \left(\frac{4}{9}\right)^2 \cdot 1 = \frac{7}{27}$$

Por tanto $ECM(U_4) = V\left(U_4\right) + \left(E(U_4) - \mu\right)^2 = \frac{7}{27} + \left(\frac{7}{9}\mu - \mu\right)^2 = \frac{7}{27} + \frac{4}{81}\mu^2$.

Para compararlos, resolvemos la desigualdad $\frac{13}{32} < \frac{7}{27} + \frac{4}{81}\mu^2$, que es cierta si $|\mu| > 1.725$.

Por tanto, si sabemos que $|\mu| > 1.725$, el estimador U_3 tiene menor ECM; si $|\mu| < 1.725$, el estimador U_4 tiene menor ECM.

- Cuando se trata de poblaciones normales, se verifica que ECM($\boldsymbol{V}$) < ECM ($\boldsymbol{S^2}$) (demo[9]). En general se utiliza $\boldsymbol{S^2}$ por ser centrado.

Conclusiones:

En general, y teniendo en cuenta que no siempre será posible encontrarlos, se buscarán estimadores que sean centrados y, entre ellos, se elegirá el más eficiente (menor varianza).

¿Cómo los buscamos? Ahora veremos algunos métodos.

6.3.- Obtención de estimadores

Método de los momentos.

Es el método más intuitivo. Consiste en utilizar la relación que tienen los parámetros con los "momentos" de la población (esperanza, varianza,...).

Por ejemplo, cuando $X \sim P(\lambda)$, como $E(X) = \lambda$, un estimador "natural" de λ será la media muestral.

Esta idea se generaliza de la siguiente forma:

Definición.

Dada una v.a. X se define su **momento de orden** $\boldsymbol{k}$ como $\boldsymbol{\alpha_k = E(X^k)}$.

Dada una variable estadística X se define su **momento de orden** $\boldsymbol{k}$ como $\boldsymbol{a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k}$.

El método de los momentos consiste en resolver el sistema de ecuaciones $\begin{cases} \alpha_1 = a_1 \\ \quad \ldots \\ \alpha_k = a_k \end{cases}$.

En concreto:

i) Si se quiere estimar **un** parámetro, se resuelve la ecuación $\alpha_1 = a_1$ (es decir: $\boldsymbol{E(X) = \bar{X}}$), despejando de dicha ecuación el parámetro que queremos estimar.

ii) Si se quiere estimar **dos** parámetros, se resolvería el sistema de **dos** ecuaciones $\begin{cases} \alpha_1 = a_1 \\ \alpha_2 = a_2 \end{cases}$. (En la práctica se resuelve el sistema equivalente $\begin{cases} E(X) = \bar{X} \\ V(X) = V \end{cases}$.)

Ejemplos:

- Si $X \sim G(p)$, buscar un estimador de $\boldsymbol{p}$ por el método de los momentos:

  Resolvemos el sistema $\boldsymbol{E(X) = \bar{X} \Leftrightarrow \frac{1-p}{p} = \bar{X}}$, y despejamos $\boldsymbol{p}$:

  $\boldsymbol{p = \frac{1}{\bar{X}+1}}$, sería el **estimador de** $\boldsymbol{p}$ por el método de los momentos.

[9] [7] PEÑA, D.: *"Fundamentos de Estadística"*. Alianza Editorial, 2001.

Por ejemplo, tenemos una muestra de una v.a. geométrica: 3,0,2,0,1,5,2,5,15,1, y $\bar{X} = \frac{3+0+\ldots+15+1}{10} = 3.4$.

Una **estimación de *p*** por el método de los momentos, sería:

$$p = \frac{1}{\bar{X}+1} = \frac{1}{3.4+1} = 0.23,$$

o bien la solución del sistema: $\frac{1-p}{p} = 3.4 \Rightarrow p = 0.23$.

- Si $X \sim U(a,b)$, buscar estimadores para ***a*** y ***b*** por el método de los momentos.

Como son dos parámetros, resolvemos el sistema: $\begin{cases} E(X) = \bar{X} \\ V(X) = V \end{cases} = \begin{cases} \frac{a+b}{2} = \bar{X} \\ \frac{(b-a)^2}{12} = V \end{cases}$, y despejando ***a*** y ***b***, obtenemos: $\begin{cases} a = \bar{X} - \sqrt{3V} \\ b = \bar{X} + \sqrt{3V} \end{cases}$, **estimadores de *a* y *b*** por el método de los momentos

Por ejemplo, tenemos una muestra de una v.a. uniforme:

1.28,1.165, 0.465, 1.09, 1.05, 0.0045, 0.52, 0.88, 0.76, 1.37

y se tiene: $\bar{X} = \frac{1.28+\ldots+1.37}{10} = 0.86$ y $V = \frac{(1.28-0.86)^2+\ldots+(1.37-0.86)^2}{10} = 0.18$

Unas **estimaciones de *a* y *b*** por el método de los momentos, serían:

$$\begin{cases} a = \bar{X} - \sqrt{3V} \\ b = \bar{X} + \sqrt{3V} \end{cases} = \begin{cases} a = 0.86 - \sqrt{3 \cdot 0.18} = 0.125 \\ b = 0.86 + \sqrt{3 \cdot 0.18} = 1.594 \end{cases}$$

o bien las soluciones del sistema: $\begin{cases} \frac{a+b}{2} = 0.86 \\ \frac{(b-a)^2}{12} = 0.18 \end{cases} = \begin{cases} a = 0.125 \\ b = 1.594 \end{cases}$.

¿Es **coherente** ese resultado con los datos de la muestra? ¿Es posible que una v.a. con distribución U(0.125, 1.594) tome el valor 0.0045?[10]

El método de los momentos, es sencillo, en muchos casos permite obtener buenos estimadores, pero en otros tiene las limitaciones vistas anteriormente u otras. Por ello, existen otros métodos de obtención de estimadores. Uno de los métodos que permite obtener estimadores con buenas propiedades es el método de máxima verosimilitud.

Método de la máxima verosimilitud.

En el ejemplo anterior, que ***a***=0.125 hace imposible obtener el valor 0.0045 en la muestra.

Si suponemos que ***a***=−5 y ***b***=5, es posible obtener los valores obtenidos en la muestra, pero es extraño que no aparezca ningún valor negativo, por lo que es "poco probable" que los valores de ***a*** y ***b*** sean esos.

[10] Recordad que si $X \sim U(a,b)$, entonces X toma valores sólo en el intervalo (a,b).

Sería "más probable" obtener esa muestra si a=0 y b=2, o quizás "más probable aún" si a=0 y b=1.5, pues los valores de la muestra están todos en el intervalo (0, 1.5), en concreto entre 0.0045 y 1.37.

El **estimador de máxima verosimilitud de** θ, nos dará el valor de θ que hace máxima la probabilidad de obtener un resultado concreto de una muestra, $(x_1, \dots, x_n)$.
Para encontrarlo, necesitamos poder medir la probabilidad de obtener un resultado concreto de una muestra, $(x_1, \dots, x_n)$. Para ello definimos la denominada **función de verosimilitud**.

Definición

Sea X una v.a. que sigue un modelo de distribución M(θ) y sea $(x_1, \dots, x_n)$ el resultado de una m.a.s. $(X_1, \dots, X_n)$ de dicha v.a.

i) Si X es discreta se define su **función de verosimilitud**

$$L(\theta) = P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

ii) Si X es continua, con función de densidad $f(x)$, se define su **función de verosimilitud**

$$L(\theta) = \prod_{i=1}^{n} f(x_i)$$

El método de máxima verosimilitud consiste en buscar el valor de θ que hace que la función de verosimilitud tome su valor máximo.

Para hallar el máximo utilizaremos algunas técnicas de análisis matemático. En concreto utilizaremos que:

- La función $ln(x)$ es una función creciente, por lo que el máximo de $f(x)$ se alcanzará en el mismo punto que el máximo de $ln(f(x))$. Esto permite derivar más fácilmente, pues el $ln(x)$ transforma los productos en sumas. (Suponemos $f(x)>0$, para cualquier x.)
- El máximo de una función $f(x)$ en un intervalo $[a,b]$ se alcanza en algún punto crítico (puntos que anulan la derivada de $f(x)$) o en los extremos del intervalo.
- El máximo de una función $f(x,y)$, con $x,y \in \mathbb{R}$, se alcanza en algún punto crítico (puntos que anulan las derivadas parciales de $f(x,y)$: $\frac{\partial}{\partial x} f(x,y) = 0, \frac{\partial}{\partial y} f(x,y) = 0$).

Por ello, para **hallar el estimador de máxima verosimilitud de** θ, seguiremos los siguientes pasos:

i) Definir la función $L(\theta)$
ii) Definir la función $\ln(L(\theta))$
iii) Hallar el máximo de $\ln(L(\theta))$:

- Resolver la ecuación $\frac{d}{d\theta}\ln(L(\theta)) = 0$.
- Si θ sólo puede tomar valores en un intervalo, estudiar si el máximo se alcanza en los extremos de dicho intervalo.

Ejemplos:

- **Estimador de máxima verosimilitud de *p*, siendo *X~G(p)*.**

  i) Definimos $L(p)$:

$$L(p)=\prod_{i=1}^{n} P(X_i=x_i)=\prod_{i=1}^{n}(1-p)^{x_i}p=(1-p)^{x_1}(1-p)^{x_2}\ldots(1-p)^{x_n}\cdot p\cdot p\cdot \overset{n\text{veces}}{\ldots}\cdot p=(1-p)^{\sum_1^n x_i}p^n$$

ii) Definimos $\ln(L(p))$:

$$\ln(L(p))=\ln\left((1-p)^{\sum_1^n x_i}p^n\right)=\ln\left((1-p)^{\sum_1^n x_i}\right)+\ln\left(p^n\right)=\left(\sum_1^n x_i\right)\ln\left((1-p)\right)+n\ln\left(p\right)$$

iii) Hallamos el máximo de $\ln(L(p))$, resolviendo la ecuación $\frac{d}{dp}\ln(L(p))=0$:

$$\frac{d}{dp}\ln(L(p))=\frac{d}{dp}\left(\left(\sum_1^n x_i\right)\ln\left((1-p)\right)+n\ln\left(p\right)\right)=\left(\sum_1^n x_i\right)\left(\frac{-1}{1-p}\right)+n\left(\frac{1}{p}\right)=0$$

Despejando p en la anterior ecuación se obtiene:

$$p=\frac{n}{\left(\sum_1^n x_i\right)+n}\overset{\text{dividiendo arriba y abajo por } n}{=}\frac{1}{\frac{\left(\sum_1^n x_i\right)}{n}+1}=\frac{1}{\bar{X}+1}.$$

Decimos que, si $X\sim G(p)$, el estimador de máxima verosimilitud de p es $\hat{p}=\frac{1}{\bar{X}+1}$.

(Observad que es el mismo estimador que se obtuvo por el método de los momentos.)

- **Estimador de máxima verosimilitud de *b*, siendo *X*~ *U(0,b)***

i) Definimos $L(b)$:

$$L(b)=\prod_{i=1}^{n} f(x_i)=\prod_{i=1}^{n}\frac{1}{b}=\frac{1}{b}\cdot\frac{1}{b}\cdot\overset{n\text{veces}}{\ldots}\cdot\frac{1}{b}=\left(\frac{1}{b^n}\right),\quad con\ x_i\in(0,b)$$

ii) Definimos $\ln(L(b))$:

$$\ln(L(b))=\ln\left(\frac{1}{b^n}\right)=-n\ln\left(b\right)$$

iii) Hallamos el máximo de $\ln(L(b))$, resolviendo la ecuación $\frac{d}{db}\ln(L(b))=0$:

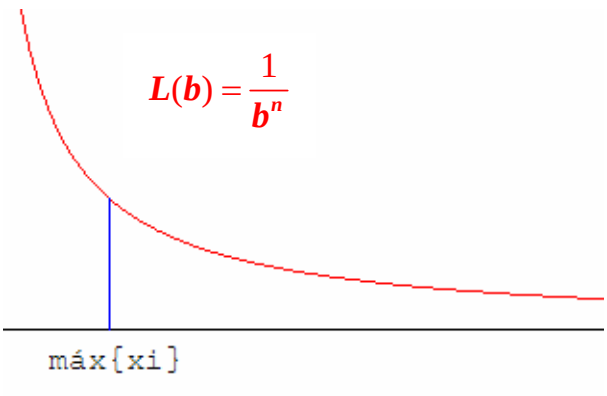
$$\frac{d}{db}\ln(L(b))=\frac{d}{db}\left(-n\ln\left(b\right)\right)=-n\left(\frac{1}{b}\right)=0$$

Pero en este caso, esa ecuación no tiene solución para ningún valor de b.

Estamos en el caso en que el parámetro sólo puede tomar valores en un intervalo, por lo que **hay que estudiar si el máximo se alcanza en los extremos de dicho intervalo**.

En este caso, como los valores de la muestra $x_i\in[0,b]$, se tiene que

$$x_i \in [0,\beta],\ \forall x_i \Rightarrow\ x_i \le \beta,\ \forall x_i \Rightarrow \max\{x_i\} \le \beta\ \Rightarrow \beta \in [\max\{x_i\}, +\infty)$$



Como $L(\beta) = \frac{1}{\beta^n}$ es decreciente, tomará su mayor valor en el extremo inferior del intervalo, en este caso $\max\{x_i\}$.

Decimos que, si $X \sim U(0,\beta)$, el estimador de máxima verosimilitud de β es $\hat{\beta} = \max\{x_i\}$.

Observad que **no es el mismo** estimador que se obtuvo por el método de los momentos.

Considerando ahora el ejemplo que vimos antes en el caso Uniforme, si suponemos $X \sim U(0,\beta)$ y tenemos los resultados de una muestra:

1.28, 1.165, 0.465, 1.09, 1.05, 0.0045, 0.52, 0.88, 0.76, 1.37

Ahora, si hacemos la estimación por el método de máxima verosimilitud, tenemos

$$\beta = \max\{x_i\} = \max\{0.0045, 0.52, ..., 1.28, 1.37\} = 1.37$$

Por lo que suponemos $X \sim U(0, \mathbf{1.37})$, que es coherente con los datos obtenidos en la muestra.

Propiedades de los estimadores de máxima verosimilitud

Si $\hat{\theta}_n$ es el estimador de máxima verosimilitud de θ para una m.a.s. de tamaño *n*, entonces:

1. Si $g(x)$ es biyectiva, $g(\hat{\theta}_n)$ es estimador de máxima verosimilitud de $g(\theta)$.
2. Si $n \to \infty$, $E(\hat{\theta}_n) \to \theta$ (es *asintóticamente* centrado)
3. Si $n \to \infty$, $V(\hat{\theta}) \to v_m$, siendo v_m la mínima varianza posible (es *asintóticamente* eficiente)
4. Si $n \to \infty$, $\frac{\hat{\theta}_n - \theta}{\sqrt{V(\hat{\theta}_n)}} \approx N(0,1)$ (es *asintóticamente* normal)
5. Si existe un estimador centrado y eficiente para θ, entonces coincide con el de máxima verosimilitud.

Es decir, para valores suficientemente grandes de *n* (*asintóticamente)* tienen todas las propiedades "deseables" de los estimadores, por lo que suelen ser los más utilizados.

Ejemplos:

1. Como $\bar{X}$ es el estimador de máxima verosimilitud para la media de una v.a. *X* con distribución exponencial, y tenemos que $E(X)=\frac{1}{\beta}$. Entonces, el estimador de máxima verosimilitud para β es $\hat{\beta}=\frac{1}{\bar{X}}$.

   Si *X*~ *N*(μ,σ), el estimador de máxima verosimilitud de σ^2 es V. Como $\sigma=\sqrt{\sigma^2}$, entonces, el estimador de máxima verosimilitud de σ es $\hat{\sigma}=\sqrt{V}$.

2. Si *X*~*G*(p), el estimador de máxima verosimilitud de p es $\hat{p}=\frac{1}{\bar{X}+1}$, y calcular $E(\hat{p})$ no es sencillo. La propiedad 2 nos asegura que, en caso de no ser centrado, sí lo es *asintóticamente.*

   En otros casos, sí se puede calcular la esperanza y es fácil verificar esta propiedad:

   Si *X*~ *U*(0,b), el estimador de máxima verosimilitud de b es $\hat{b}=\max\{x_i\}$, pero no es un estimador centrado, pues $E(\hat{b})=\frac{n}{n+1}b$ (demo[11]). Lo que sí se verifica es que, si $n\to\infty$, entonces $E(\hat{b})\to b$.

3-4-5. El ejemplo más sencillo de las últimas propiedades es $\hat{\lambda}=\bar{X}$, estimador de máxima verosimilitud de λ para una v.a. con distribución de Poisson.

3. Se ha visto que $V(\bar{X})=\frac{V(X)}{n}$, por lo que siempre se verifica que $V(\hat{b})=\frac{V(X)}{n}\to 0$, cuando $n\to\infty$.

4. Además, por el TCL, sabemos que si *n* es suficientemente grande, entonces $\bar{X}$ sigue una distribución normal.

5. El estimador $\bar{X}$ es centrado y eficiente para λ, y precisamente el estimador de máxima verosimilitud que se obtiene para λ es $\hat{\lambda}=\bar{X}$.

Comprobar estas propiedades para otros casos suele ser bastante complicado y se escapa de los objetivos de este curso.

[11] Ver [1] CORONADO, J.L.; CORRAL, A.; GÓMEZ, J.I.; LÓPEZ, P.; RUIZ, B.; VILLÉN, J.: "Estadística". Servicio de Publicaciones de la E.U. de Informática, 2004. Ejemplo 8.12

ANEXO 1. Esperanza del estadístico *V*: varianza muestral.

$$E(V)=E\left(\frac{(X_1-\bar{X})^2+\ldots+(X_n-\bar{X})^2}{n}\right)=\frac{1}{n}E\left((X_1-\bar{X})^2+\ldots+(X_n-\bar{X})^2\right)=$$

$$=\frac{1}{n}E\left((X_1^2+\bar{X}^2-2X_1\bar{X})+\ldots+(X_n^2+\bar{X}^2-2X_n\bar{X})\right)=$$

$$=\frac{1}{n}E\left((X_1^2+\ldots+X_n^2)+(\bar{X}^2+\ldots+\bar{X}^2)-2X_1\bar{X}-\ldots-2X_n\bar{X}\right)$$

$$\frac{1}{n}\left(\underset{(1)}{E(X_1^2+\ldots+X_n^2)}+\underset{(2)}{E(\bar{X}^2+\ldots+\bar{X}^2)}+\underset{(3)}{E(-2\bar{X}(X_1+\ldots+X_n))}\right)$$

(1) $V(X_i)=E(X_i^2)-E(X_i)^2 \Rightarrow E(X_i^2)=V(X_i)+E(X_i)^2=\sigma^2+\mu^2$

Por tanto, $E(X_1^2+\ldots+X_n^2)=nE(X_i^2)=n(\sigma^2+\mu^2)=n\sigma^2+n\mu^2$

(2) $V(\bar{X})=E(\bar{X}^2)-E(\bar{X})^2 \Rightarrow E(\bar{X}^2)=V(\bar{X})+E(\bar{X})^2=\frac{\sigma^2}{n}+\mu^2$

Por tanto, $E(\bar{X}^2+\ldots+\bar{X}^2)=nE(\bar{X}^2)=n\left(\frac{\sigma^2}{n}+\mu^2\right)=\sigma^2+n\mu^2$

(3) $\bar{X}=\frac{(X_1+\ldots+X_n)}{n}\Rightarrow(X_1+\ldots+X_n)=n\bar{X}$

Por tanto, $E(-2\bar{X}(X_1+\ldots+X_n))=E(-2\bar{X}\left(n\bar{X}\right))=-2nE(\bar{X}^2)=-2n\left(\frac{\sigma^2}{n}+\mu^2\right)=-2\sigma^2-2n\mu^2$

Volviendo al desarrollo original, tenemos:

$$E(V)=\frac{1}{n}\left(\underset{(1)}{(n\sigma^2+n\mu^2)}+\underset{(2)}{(\sigma^2+n\mu^2)}-\underset{(3)}{2\sigma^2-2n\mu^2}\right)=\frac{1}{n}\left((n-1)\sigma^2\right)=\frac{n-1}{n}\sigma^2.$$

ANEXO 2. Varianza del estadístico S^2 *para una distribución de Poisson.*

En general, la varianza de la varianza muestral es $\boldsymbol{V(V)} = \frac{(\boldsymbol{n}-1)^2}{\boldsymbol{n}^2}\frac{1}{\boldsymbol{n}}\left(\mu_4 - \frac{\boldsymbol{n}-3}{\boldsymbol{n}-1}\sigma^4\right)$ siendo $\mu_4 = \boldsymbol{E}\left((\boldsymbol{X}-\mu)^4\right)$. (ver Teorema 8.3.3 en [1] CORONADO, J.L.; CORRAL, A.; GÓMEZ, J.I.; LÓPEZ, P.; RUIZ, B.; VILLÉN, J.: "Estadística". Servicio de Publicaciones de la E.U. de Informática, 2004)

Por tanto,

$$\boldsymbol{V(S^2)} = \boldsymbol{V}\left(\left(\frac{\boldsymbol{n}}{\boldsymbol{n}-1}\right)\boldsymbol{V}\right) = \left(\frac{\boldsymbol{n}}{\boldsymbol{n}-1}\right)^2 \boldsymbol{V(V)} = \left(\frac{\boldsymbol{n}}{\boldsymbol{n}-1}\right)^2 \frac{(\boldsymbol{n}-1)^2}{\boldsymbol{n}^2}\frac{1}{\boldsymbol{n}}\left(\mu_4 - \frac{\boldsymbol{n}-3}{\boldsymbol{n}-1}\sigma^4\right) = \frac{1}{\boldsymbol{n}}\left(\mu_4 - \frac{\boldsymbol{n}-3}{\boldsymbol{n}-1}\sigma^4\right).$$

Si X~P(λ), $\sigma^2 = \lambda$ y μ_4 = $\sum_{r=0}^{\infty} (r-\lambda)^4 \cdot \frac{\hat{e}^{-\lambda}\cdot\lambda^r}{r!} = 3\cdot\lambda^2 + \lambda$

Por tanto, en el caso de un modelo de Poisson, como $\sigma^2 = \lambda$, tenemos:

$$\boldsymbol{V(S^2)} = \frac{1}{\boldsymbol{n}}\left(3\lambda^2 + \lambda - \frac{\boldsymbol{n}-3}{\boldsymbol{n}-1}\lambda^2\right) = \frac{1}{\boldsymbol{n}}\left(\frac{2\boldsymbol{n}}{\boldsymbol{n}-1}\lambda^2 + \lambda\right) = \frac{2}{\boldsymbol{n}-1}\lambda^2 + \frac{\lambda}{\boldsymbol{n}}.$$