Tema 7. ESTIMACIÓN PUNTUAL.

7.1.- 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

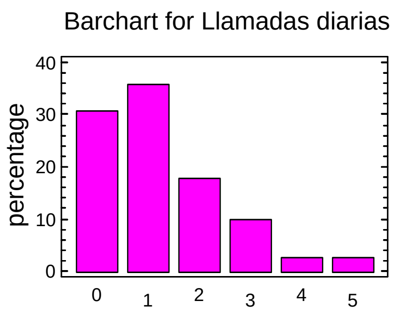
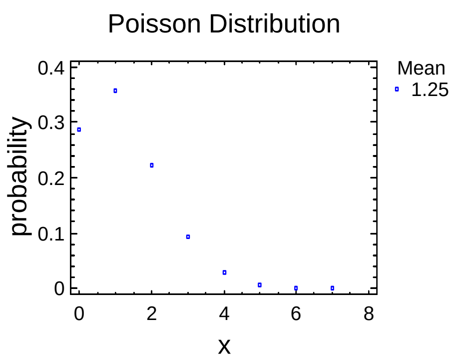
Verificación:

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
```

<table>
<tr><th>Diagrama de barras
de frecuencias relativas</th><th>Función de masa
P(1.25)</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

```
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 es válido cuando se quiere analizar 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, ¿qué hemos de hacer para que sí fueran 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** ($\boldsymbol{\theta}$) del mismo, mediante:

  - **Estimación puntual:** $\boldsymbol{\theta} = \theta_0$
  - **Estimación por intervalo:** $\boldsymbol{\theta} \in (a,b)$ con un % de confianza
  - **Contraste de hipótesis:** aceptamos $\boldsymbol{\theta} = \theta_0$ frente a $\boldsymbol{\theta} \neq \theta_0$ (ó $\boldsymbol{\theta} > \theta_0$ ó $\boldsymbol{\theta} < \theta_0$), 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($\boldsymbol{\theta}$)** con nivel de significación α.

---

[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.

- **Independencia:** X e Y son independientes; los datos de la muestra son independientes.

Ejemplo: Se quiere estudiar el número de llamadas diarias por el móvil que hace un joven madrileño (18-26 años).

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

  - **Estimación puntual:** $\theta = \theta_0$
    λ = 1.25641 (como λ es la media de una Poisson, una posible estimación es la media muestral.

  - **Estimación por intervalo:** $\theta \in$ (a,b) con un % de confianza

```
Confidence Intervals for Llamadas diarias
-----------------------------------------
```

    95.0% confidence interval for mean: 1.25641 +/- 0.39853 **[0.857878;1.65494]**

    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.

    Con los datos de la muestra:

    Rechazamos la hipótesis de que λ = 1 frente a λ > 1, con un 85% confianza (nivel α=0.15)

    Aunque la media de las mujeres es λ_M = 1.5 y la de los hombres es λ_V = 1.15, **no podemos rechazar** la hipótesis de que sean iguales ($\lambda_M = \lambda_V$). (Se necesitaría un tamaño de la muestra mayor para poder asegurar que son distintas a partir de esos resultados.)

- **Inferencia no paramétrica:** para verificar las **hipótesis sobre el modelo** o sobre **independencia** de variables o muestras.
    - Ajuste de distribuciones:
      Aceptamos que X~P(1.25461) con p-valor 0.75.
    - **Independencia:** X e Y son independientes; los datos de la muestra son independientes.

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.

7.2.- 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 $M(\theta)$

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** $M(\theta)$.	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: Es un estadístico (por tanto, una **v.a.**) $\theta^* = \theta^*(X_1, \dots, X_n)$, que **se utiliza para dar un valor estimado de un parámetro** θ.	Estimación de θ: $\theta \sim \theta^*(x_1, \dots, x_n) \in \mathbb{R}$

Ejemplo :

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

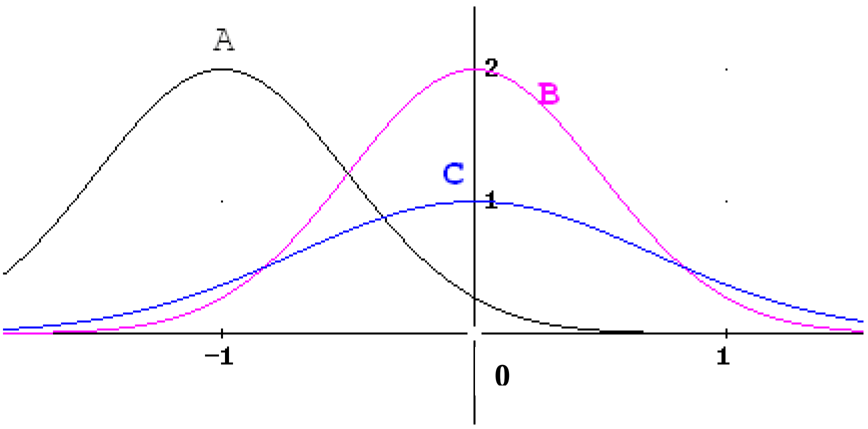
Antes de obtener los datos (n)	Después de obtener los datos (n=39)
$X_1, \dots, X_n$ son v.a. independientes con distribución $P(\lambda)$.	Resultado de la muestra: $\{1,0,3,2,\dots,2\} \in \mathbb{R}$
Estadísticos: Media muestral: $\bar{X} = \dfrac{X_1+\dots+X_n}{n}$ Varianza muestral: $V = \dfrac{(X_1-\bar{X})^2+\dots+(X_n-\bar{X})^2}{n}$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

7.4.- Propiedades de los estimadores

En el ejemplo anterior vimos que 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:

3 camareros tienen distintos métodos de hacer una estimación de la demanda semanal de café. A lo largo de un año se anota el valor de: *demanda real – demanda estimada* (es decir el error entre la estimación y la realidad). La distribución de frecuencias se puede representar mediante las siguientes "líneas de densidad":



¿Qué camarero ha hecho mejores estimaciones a lo largo del año?
¿Qué aspectos tenemos en cuenta?

El valor esperado coincida con el deseado y que tenga la mínima varianza posible.

Definición.

Decimos que θ^* es un **estimador centrado (o insesgado) de** θ, si $E(\theta^*) = \theta$, 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** $\overline{X}$ es un estimador centrado de $E(X)$ (demo). Por tanto,
  - Si $X \sim P(\lambda)$, $\lambda^* = \overline{X}$ es un estimador centrado para λ.
  - Si $X \sim N(\mu, \sigma)$, $\mu^* = \overline{X}$ es un estimador centrado para μ..
- La **varianza muestral** V no es un estimador centrado para $V(X)=\sigma^2$. De hecho, $E(V) = \frac{n-1}{n}\sigma^2$ (demo). Por tanto,
  - Si $X \sim P(\lambda)$, $\lambda' = V$ no es un estimador centrado para λ.

- La **cuasivarianza muestral** $S^2 = \dfrac{(X_1-\bar{X})^2+...+(X_n-\bar{X})^2}{n-1}$ sí es un estimador centrado para $V(X)=\sigma^2$. 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 facil 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 ***S²***:

  Como $V = \left(\dfrac{n-1}{n}\right)S^2 \Rightarrow V(V) = \left(\dfrac{n-1}{n}\right)^2 V(S^2)$, por tanto $V(V) < V(S^2)$.
- Ver estimadores del Problema 4.

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:

- Se puede demostrar que ECM(***V***) > ECM (***S²***) para $n>2$.
- Ver estimadores del Problema 4.

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 (mínima varianza).

¿Cómo los buscamos? Ahora veremos algunos métodos.

7.3.- Obtención de estimadores

Método de los momentos.

Es el método más intuitivo, que consiste en utilizar la relación que tienen los parámetros con los "momentos" de la población (esperanza, varianza, ...). El ejemplo, más claro es 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 *k*** como $\alpha_k = E(X^k)$.

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

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

En concreto:

i) Si se quiere estimar un parámetro, se resuelve el sistema $\alpha_1 = a_1$ (es decir: $E(X) = \bar{X}$).

ii) Si se quiere estimar dos parámetros, se resolvería el sistema $\begin{cases} \alpha_1 = a_1 \\ \alpha_2 = a_2 \end{cases}$, aunque 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 p por el método de los momentos.
- Si $X \sim B(n,p)$, buscar estimadores para n y p por el método de los momentos.

Método de la máxima verosimilitud.

Problema 3 b)

Definición

Sea X una v.a. que sigue un modelo de distribución $M(\theta)$ 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.

Es decir, dado un resultado concreto de una muestra, *(x_1, ... , x_n)*, el estimador de máxima verosimilitud de θ, nos da el valor de θ que hace máxima la probabilidad de obtener dicho resultado.

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)).* (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 $x,y \in$ 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 $\ln(L(\theta))$
ii) 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 λ, siendo *X~P(λ).*
- Estimador de máxima verosimilitud de (μ, σ) siendo *X~N(μ,σ)*
- Estimador de máxima verosimilitud de *b,* siendo *X~ U(0,b)*

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:

- Si $g(x)$ es biyectiva, $g(\hat{\theta}_n)$ es estimador de máxima verosimilitud de $g(\theta)$. (Problema 6)
- Si $n \to \infty$, $E(\hat{\theta}_n) \to \theta$ (es asintóticamente centrado)
- Si $n \to \infty$, $V(\hat{\theta}) \to v_m$, siendo v_m la mínima varianza posible (es asintóticamente eficiente)
- Si $n \to \infty$, $\frac{\hat{\theta}_n - \theta}{\sqrt{V(\hat{\theta}_n)}} \approx N(0,1)$ (es asintóticamente normal)

Es decir, asintóticamente tienen todas las propiedades "deseables" de los estimadores.