Estadística y Optimización

Errores más habituales y cómo resolverlos

Índice

1. Errores globales, sobre teoría y consejos generales

- **Importante:** A algunos/as os convendria
  1) Repasar las propiedades de potencias, logaritmos y exponenciales.
  2) Repasar la integración tanto en una variable como en varias.

- "Definir" un objeto matemático (concepto, función, propiedad,...) significa: *"en términos generales, es delimitar, o sea, indicar, expresar el límite que separa este objeto de todos los demás"*. NO significa describir una idea de lo que (se cree que) este objeto significa, ó para qué se utiliza o algún ejemplo bien/mal avenido.
  - **P.ej.**, se pedía *"Definir el sesgo de un estimador para un parámetro poblacional"*:

Respuesta correcta

Siendo $T = T(X_1, \dots, X_n)$ un estimador para el parámetro θ, el sesgo de T se define mediante:

$$b(T) = E[T] - \theta.$$

  - **P.ej.**, se pedía *"Para X variable aleatoria discreta, definir función de verosimilitud de una muestra aleatoria simple de X"*:

Respuesta correcta

Siendo $(X_1, \dots, X_n)$ una m.a.s. de una var. aleatoria X discreta con parámetro desconocido θ y función de masa de probabilidad $p_\theta(x)$, la función de verosimilitud $L(X_1, \dots, X_n; \theta)$ de $(X_1, \dots, X_n)$ se define como la función de masa de probabilidad conjunta de la muestra. Además, como la muestra es aleatoria y simple, ésta se puede expresar como el producto

$$L(X_1, \dots, X_n; \theta) = p_\theta(x_1) \cdot p_\theta(x_2) \cdots p_\theta(x_n) = \prod_{i=1}^{n} p_\theta(x_i),$$

(Nota: Esta última expresión también valdria como definición, pero expresamente para muestras aleatorias simples)

☞ "Enunciar" una propiedad significa dar el enunciado matemático (correcto) de lo que se está pidiendo (hipótesis, condiciones y conclusión). NO significa describir una idea de lo que (se cree que) la propiedad significa y algún ejemplo bien/mal avenido.

P.ej., se pedía *"Enunciar la propiedad de falta de memoria para X una variable geométrica"*:

RESPUESTA CORRECTA

Sea $X \sim \mathcal{G}(p)$ de parámetro p. Para todo $m, k \in \mathbb{N}$, se tiene:

$$P\left(X = m + k \,|\, X > m\right) = P\left(X = k\right).$$

☞ "Demostrar" una propiedad significa dar el conjunto de pasos y silogísmos lógicos que nos llevan de las hipótesis y condiciones del enunciado a las consecuencias.

P.ej., se pedía *"Demostrar la propiedad de falta de memoria para X una variable geométrica"*:

RESPUESTA ("EXCESIVAMENTE") DETALLADA

Demostración. Para una v.a. geométrica $X \sim \mathcal{G}(p)$, sabemos que $P\left(X = k\right) = q^{k-1}p$ para todo $k \geq 1$, donde $q = 1 - p$. Del mismo modo, $P\left(X > k\right) = q^k$. Ahora, usando la definición de probabilidad condicional:

$$\begin{aligned} P\left(X = m + k \,|\, X > m\right) &= \frac{P\left(\{X = m + k\} \cap \{X > m\}\right)}{P\left(X > m\right)} = \frac{P\left(X = m + k\right)}{P\left(X > m\right)} \\ &= \frac{q^{m+k-1}p}{q^m} = q^{k-1}p = P\left(X = k\right). \end{aligned}$$

Donde la igualdad

$$P\left(\{X = m + k\} \cap \{X > m\}\right) = P\left(X = m + k\right)$$

proviene del hecho que el suceso $\{X = m + k\}$ implica el suceso $\{X > m\}$ (el que X sea $m+k$ ya implica automáticamente que X es mayor que m, o si se quiere: "primer éxito en el $(m+k)$-ésimo intento implica que fracasamos al menos m veces"), por lo que

$$\{X = m + k\} \cap \{X > m\} = \{X = m + k\}.$$

□

☞ **Importante:** Es conveniente:

a) Detallar los cálculos que no son "directos", para justificar los diversos resultados.

b) Decir que se está usando un cierto teorema o propiedad a la hora de aplicarlo para cálculos: p.e. el Teorema de la Probabilidad Total, el Teorema de Bayes,...

c) Definir los objetos que se usan, como por ejemplo las variables aleatorias que usamos en un problema:

- **P.ej.**, X ="núm. de mensajes que se recibe en una hora" ó Y ="núm de mensajes que se reciben en 1000 oficinas en una hora", etc...

2. Errores en la primera parte del curso (Examen 17/04/2021)

- Estudiar la independencia de sucesos A y B requiere verificar que $P(A \cap B) = P(A)\,P(B)$ o, equivalentemente, que $P(B \mid A) = P(B)$.

- Tres sucesos A, B, C son independientes si <u>lo son dos a dos</u> y además

$$P(A \cap B \cap C) = P(A)\,P(B)\,P(C)\,.$$

Es decir, que lo sean dos a dos es una condición necesaria, pero no suficiente.

P.ej., en el Ejercicio 1-(b) del examen, se comprobaba que los sucesos S_1, S_2, S_3 no eran independientes dos a dos, por lo que *“no son independientes conjuntamente, ya que no lo son dos a dos”.*

- **Cuidado:** No confundir los sucesos y sus operaciones (*conjuntos*) con las probabilidades (*números.*)

“¿Qué va a ocurrir?”		*“¿Con qué probabilidad?”*
$\boxed{\text{Sucesos}}$	$\longrightarrow$	$\boxed{\text{Probabilidad}}$
	A	$P(A)$
“no A”	$\overline{A}$	$P(\overline{A}) = 1 - P(A)$
“A o B”	$A \cup B$	$P(A \cup B) = P(A) + P(B) + P(A \cap B)$
“A y B”	$A \cap B$	$P(A \cap B) = P(A \mid B)\,P(B) = P(A \mid B)\,P(B)$
“A o B”	$A \cup B$	$P(A \cup B) = P(A) + P(B) + P(A \cap B)$
“algo ocurre”	Ω	$P(\Omega) = 1$
“A y B incompatibles”	$A \cap B = \emptyset$	$P(\emptyset) = 0$
“A y B” independientes		$P(A \cap B) = P(A)\,P(B)$

- En los vectores aleatorios (X, Y):
  - $E[X]$, $E[X^2]$ se calculan mediante una integral simple de x o x^2 contra la función de densidad marginal $f_X(x)$, <u>no</u> sobre la conjunta $f(x, y)$:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\mathrm{d}x \qquad E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\mathrm{d}x.$$

  - El que $\mathrm{Cov}[X, Y] = 0$ es una <u>condición necesaria</u> para que las v.a. X e Y sean independientes, pero <u>no suficiente</u>. Es decir, si $\mathrm{Cov}[X, Y] \neq 0$ entonces X e Y no son independientes, pero si $\mathrm{Cov}[X, Y] = 0$ no podemos concluir nada en principio.

- Si vais a usar una v.a. X en un contexto concreto, definirla correctamente: p.ej. en el Ejercicio 3-(a) del examen, $X =$ “núm. de bolas negras en 9 extracciones”.

- Una binomial $X \sim \mathcal{B}(n, p)$ se puede <u>aproximar bien</u> por una Poisson $Y \sim \mathcal{P}(\lambda)$ de parámentro $\lambda = np$ cuando n es grande y p es pequeño. Pero $p = \frac{1}{3}$ <u>no</u> es pequeño.

☞ Al aproximar una va.a discreta X por una normal $Y \sim \mathcal{N}(\mu, \sigma)$, se ha de aplicar la corrección de Yates para obtener una mejor aproximación <u>sobre los extremos de los intervalos cerrados en la X</u>, no después cuando se ha tipificado (consultar el Resumen sobre el TCL en MOODLE). En el Ejercicio de examen:

$$P(X > 30) = P(X \geq 31) \approx P(Y \geq 30.5) = \text{tipificación} = \cdots$$

Bestiario de burradas avistadas:

- ✖ $\dfrac{1}{3} + \dfrac{1}{4} = \dfrac{1}{12}$.
- ✖ $P(A) \cup P(B)$.
- ✖ A y B son independientes si $P(A \cap B) = 0$.
- ✖ $\displaystyle\int_0^1 x + y \,\mathrm{d}y = x\left[\frac{y^2}{2}\right]_0^1$.
- ✖ $\left(x - \dfrac{5}{9}\right)^2 = \left(x^2 - \dfrac{25}{81}\right)$.
- ✖ $f_{X,Y}(x,y) = x+y \Longrightarrow \begin{cases} f_X(x) = x, & x \in [0,1] \\ f_Y(y) = y, & y \in [0,1] \end{cases}$.

3. Errores en la segunda parte del curso (Examen 04/06/2021)

- Una función de densidad $f_\theta(x)$ de una población X de parámetro desconocido θ debe verificar tanto que ES NO NEGATIVA como que ESTÁ NORMALIZADA:

  a) $f_\theta(x) \geq 0$, para todo $x \in \mathbb{R}$.

  b) $\displaystyle\int_{-\infty}^{+\infty} f_\theta(x)\, \mathrm{d}x = 1.$

  Si $f_\theta(x)$ depende de un parámetro θ, estas condiciones nos dan eventuales restricciones sobre los posibles valores que puede tomar θ.

- Si las anteriores restricciones nos dan que "$\theta =$un cierto valor particular", no hay nada que estimar ya que el valor de θ sólo puede ser ese. (Por lo que si obtenemos una condición de este estilo en un ejercicio de estimación de un examen, ¡algo habremos hecho mal!)

- El número $1^{\theta+1}$ SIEMPRE ES 1, ya que elevar 1 a cualquier número es siempre 1. Aunque esto es intuitivo, se puede justificar formalmente mediante la definición de potencia por números reales para bases positivas $a > 0$:

  $$a^b := e^{b \ln(a)}$$

  Por lo que:

  $$1^x = e^{x \ln(1)} = e^0 = 1, \quad \text{para todo } x \in \mathbb{R}.$$

- Las distribuciones χ^2_n y $F_{n,m}$ no toman valores negativos, (es decir $\chi^2_n \geq 0$ y $F_{n,m} \geq 0$), por lo que la correspondiente región de aceptación o región crítica (dependiendo de si el contraste es de una o dos colas), comienza en el 0:

  $$\text{“}R_0 = [0, \dots\text{”} \qquad \text{ó} \qquad \text{“}R_1 = [0, \dots\text{”}$$

- Para dos poblaciones normales e independientes $X \sim \mathcal{N}(\mu_X, \sigma_X)$ e $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$, podemos usar el estadístico para la diferencia de medias $\mu_X - \mu_Y$ dado por:

  $$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} \times \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \sim t_{n_X + n_Y - 2}$$

  solo SI SABEMOS QUE LAS VARIANZAS DE LAS POBLACIONES SON IGUALES: $\sigma_X^2 = \sigma_Y^2$. Es importante justificar esta condición para poder usar este estimador T.

  - **P.ej.:** en el caso del 2º apartado del EJ. 2, esta igualdad entre varianzas se podia asumir gracias al resultado del contraste de hipótesis del primer apartado.

- Cuando se quiere demostrar que "*la esperanza de la población X es mayor que la de Y con $\alpha = 0.1$*", queremos contrastar la hipótesis

  $$H: \ \mu_X > \mu_Y$$

  frente a los datos. Como esta hipótesis no incluye una desigualdad (o igualdad) estricta, debemos tomarla como la hipótesis alternativa H_1 frente a su contraria $\mu_X \leq \mu_Y$ la cual sí viene expresada de esta forma. Así,

  $$\begin{array}{l} H_0 : \mu_X \leq \mu_Y \\ H_1 : \mu_X > \mu_Y \end{array} \qquad \text{o también} \qquad \begin{array}{l} H_0 : \mu_X - \mu_Y \leq 0 \\ H_1 : \mu_X - \mu_Y > 0 \end{array}$$

Por lo que nos queda un contraste de hipótesis sobre el parámetro $\theta = \mu_X - \mu_Y$ con una cola a derecha (ya que los "valores raros" vendrán dados por estimaciones de θ positivos y alejados del 0, como indica la H_1).

☞ **Cuidado con el léxico**: "algo" poblacional vs "algo" muestral,

Nociones asociadas a una población X	vs	Nociones asociadas a una muestra aleatoria simple $(X_1, \ldots, X_n)$
Esperanza poblacional $\mu_X = E[X]$ (parámetro, es decir, un número fijo determinado o a estimar propio de la población).	vs	**Media muestral** $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (estadístico, es decir, una variable aleatoria obtenida a partir de la muestra que estima un parámetro). A partir de la observación 👁 de la muestra, toma un determinado valor, p.ej. $\overline{x} = 15$
Varianza poblacional $\sigma_X^2 = V[X]$ (parámetro, es decir, un número fijo determinado o a estimar propio de la población).	vs	**Varianza muestral** $S_X^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \overline{X}^2$ (estadístico, es decir, una variable aleatoria obtenida a partir de la muestra que estima un parámetro). A partir de la observación 👁 de la muestra, toma un determinado valor, p.ej. $s_X^2 = 4$

Bestiario de burradas avistadas:

✘ Razonamientos del tipo:

$$1^{\theta+1} = 1 \Longrightarrow \theta + 1 = 1 \Longrightarrow \theta = 0$$

$$1^{\theta+1} = 1 \Longrightarrow \theta + 1 = 0 \Longrightarrow \theta = -1$$

✔ Eventual razonamiento correcto:

$$1^{\theta+1} = 1 \Longrightarrow (\theta+1)\ln(1) = \ln(1) \Longrightarrow 0 = 0,$$

obteniendo una relación trivial, por lo que no nos da una restricción sobre los valores de θ.

✘ $x \cdot (\theta+1)x^\theta = (\theta+1)x^{2\theta}$.

✘ $\int_0^1 x^\theta \, \mathrm{d}x = \left[\frac{x^{2\theta}}{2}\right]_0^1$.

✘ $\left[\frac{\theta x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta^{\theta+1}}{\theta+1}$.

✘ $(\theta+1) = \overline{X} \cdot (\theta+2) \Longrightarrow 1 = \frac{\overline{X} \cdot (\theta+2)}{\theta}$.

✖ $(\theta+1)x_1^\theta \cdots (\theta+1)x_n^\theta = \begin{cases} (\theta+1)^n x^{n\theta}, \\ (\theta+1)^n (x_1 \cdots x_n)^{n\theta}, \\ (\theta+1)^n \left(\sum_i x_i\right)^\theta, \\ (\theta+1)^n \sum_i x_i^{n\theta}. \end{cases}$

✖ $\ln\left((\theta+1)^n \cdot \prod_i x_i^\theta\right) = n\ln(\theta+1) \cdot \theta \ln\left(\prod_i x_i\right).$

✖ La varianza muestral es:

- ✖ σ_X^2.
- ✖ $\widehat{S}_X^2$.
- ✖ S_X.
- ✖ μ_X.

✖ $P(A) \cup P(B)$ ó $P(A) \cap P(B)$.

4. Errores en la convocatoria extraordinaria (Examen 07/07/2021)

☞ **Importante:** Revisar las listas de errores aparecidas anterioremente, porque algunos/as SEGUÍS EMPECINADOS/AS EN TROPEZAR n VECES CON LA MISMA PIEDRA.
En particular:

a) No plantear el problema, justificar los resultados/teoremas que os permiten usar una fórmula...

b) No describir las variables aleatorias que definís.

c) El uso (correcto) de de la corrección de Yates a la hora de aproximar por la Normal.

d) En un contraste, plantearlo con los datos del problema, plantear las hipótesis, determinar las colas, describir el estadístico, detallar cómo se obtienen las regiones de aceptación y rechazo...

☞ Confundís nociones de variables estadísticas (datos) con las de variables aleatorias. En el caso de la covarianza, se pedía la definición para dos variables aleatorias X, Y.

- Si X, Y son VARIABLES ALEATORIAS, se define: $$\boxed{\mathrm{Cov}[X,Y] = \begin{cases} E\Big[\big(X - E[X]\big)\cdot\big(Y - E[Y]\big)\Big] \\ \text{o alternativamente,} \\ E[XY] - E[X]\cdot E[Y] \end{cases}}$$

- Si X e Y VARIABLES ESTADÍSTICAS sobre una misma población, con datos

$$\begin{array}{c|cccc} X & x_1 & x_2 & \cdots & x_n \\ \hline Y & y_1 & y_2 & \cdots & y_n \end{array} \quad ,$$

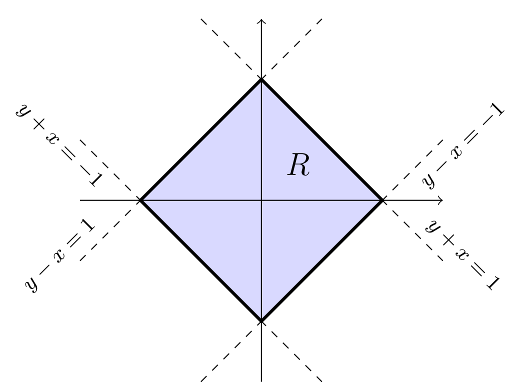
se define: $$\boxed{\mathrm{Cov}[X,Y] = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \\ \text{o alternativamente,} \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i y_i - \overline{x}\cdot\overline{y} \end{cases}}$$

☞ El que la covarianza sea nula (es decir $\mathrm{Cov}[X,Y] = 0$) NO IMPLICA la independencia de X e Y.

P. ej. vimos en clase que si (X, Y) tienen función de densidad conjunta:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{2}, & (x,y) \in R, \\ 0, & (x,y) \notin R, \end{cases}$$

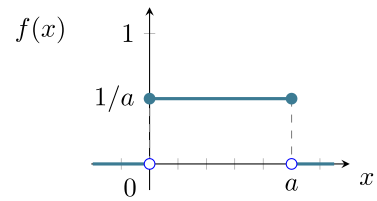
siendo R la región acotada del plano delimitada por las rectas $y + x = -1$, $y - x = 1$, $x + y = 1$ y $x - y = 1$; es fácil comprobar que $\mathrm{Cov}[X,Y] = 0$, pero que X e Y no son independientes (basta ver que R no viene dado por un rectángulo del tipo $[a,b]\times[c,d]$).



☞ Lo que SÍ ES CIERTO es que la independencia de X e Y IMPLICA que $\mathrm{Cov}[X,Y] = 0$. (Por lo tanto, si $\mathrm{Cov}[X,Y] \neq 0$ entonces X e Y no son independientes)

La función de densidad de una var. aleatoria uniforme $X \sim \mathcal{U}([0, a])$ es:

$$f(x) = \begin{cases} \dfrac{1}{a} & \textsf{(constante)}, \quad x \in [0, a] \\ 0, & \text{resto.} \end{cases}$$



ya que la densidad debe de ser constante sobre $[0, a]$ y estar normalizada:

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_0^a \frac{1}{a}\,\mathrm{d}x = \frac{a}{a} = 1.$$

Burrada conceptual MUY recurrente: Para estudiar si un estadístico T para un parámetro θ desconocido de una población X es: *centrado, consistente, etc...*

> Se han de estudiar las propiedades (esperanza, varianza,...) del estadístico T, <u>NO</u> las de la población X.

- *Sesgo de* T: $\quad b(T) = E[T] - \theta$,
- *Riesgo (error cuadrático medio) de* T: $\quad r(T) = E\left[(T-\theta)^2\right] = V[T] + b(T)^2$.

En el caso del Ej. 2 del examen, el estadístico es $T = \widehat{a}$ dado por $\boxed{\widehat{a} = 2\overline{X}}$ para el parámetro desconocido $\theta = a$ de la población $X \sim \mathcal{U}([0, a])$. Por lo que se pedía estudiar:

- *Sesgo de* $\widehat{a}$: $\quad b(\widehat{a}) = E[\widehat{a}] - a$,
- *Riesgo (error cuadrático medio) de* $\widehat{a}$: $\quad r(\widehat{a}) = E\left[(\widehat{a} - a)^2\right] = V[\widehat{a}] + b(\widehat{a})^2$,

y todo esto se puede calcular para $\widehat{a} = 2\overline{X}$ usando propiedades de $\overline{X}$, la esperanza y la varianza, como está detallado en la solución.

En el constraste de la proporción de la variedad G del Ej. 3, habeis tomado el estadístico

$$T = \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1),$$

el cual es correcto, <u>SIN EMBARGO</u> la media muestral $\overline{X}$ en este caso corresponde a la frecuencia muestral de aparición de la variedad G, es decir:

$$\overline{x} = \widehat{p} = \frac{20}{100} = 0.2.$$

¿Por qué? Pues porque este problema se plantea considerando:

- <u>Población</u>: $X =$ “Individuo de la especie es de la variedad G”$\sim \mathcal{B}er(p)$, dónde p es la proporción teórica de aparición de la variedad G.
- <u>Muestra</u>: $(X_1, \ldots, X_{100})$ con $X_i \sim \mathcal{B}er(p)$ independientes (es decir, si el individuo observado es de la variedad G o no). La media muestral por tanto corresponde a la frecuencia muestral observada:

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{\text{“núm. observado de la variedad G de 100 individuos”}}{100} = \widehat{p}.$$