Examen resolt d'Introducció a la Probabilitat - Grau d'Estadística 26-gener-2010

1. (2 pts.) Un laboratori fabrica 3 fàrmacs A, B, C en les mateixes proporcions. En el procés d'etiquetatge, els fàrmacs B i C són correctament etiquetats en el 98% i el 99% dels casos respectivament. També és conegut que la probabilitat que a l'escollir un fàrmac a l'atzar, aquest es correspongui al fàrmac A i estigui correctament etiquetat és de 0.3. Es demana:

   (a) Calculeu la probabilitat que, en escollir un fàrmac a l'atzar, aquest estigui correctament etiquetat. (1 pt.)

   (b) Si escollim un fàrmac a l'atzar i resulta que està incorrectament etiquetat, quina és la probabilitat que pertanyi al tipus B? (1 pt)

   SOLUCIÓ

   En primer lloc escriurem la informació que ens dóna el problema:

   -Etiquetarem els successos com

   - CE com a **Correctament Etiquetat**
   - A,B,C com a fàrmacs **A**, **B** i **C** respectivament.

   -Escriurem en format probabilístic la informació proporcionada pel problema

   - $P(CE|B) = 0.98$,
   - $P(CE|C) = 0.99$,
   - $P(A \cap CE) = 0.3$
   - $P(A) = P(B) = P(C) = \frac{1}{3}$

   (a) Mitjançant el **Teorema de les Probabilitats Totals** obtindrem:

   $$P(CE) = P(CE|A)P(A) + P(CE|B)P(B) + P(CE|C)P(C)$$

   Disposem de tota la informació amb l'excepció de $P(CE|A)$ que es pot deduir en base a:

   $$P(CE|A) = \frac{P(A \cap CE)}{P(A)} = \frac{0.3}{\frac{1}{3}} = 0.9$$

   Per tant:

   $$P(CE) = P(CE|A)P(A) + P(CE|B)P(B) + P(CE|C)P(C) = 0.98 * 1/3 + 0.99 * 1/3 + 0.9 * 1/3 = 0.957$$

   (b) Mitjançant el **Teorema de Bayes** obtindrem:

   $$P(B|\overline{CE}) = \frac{P(\overline{CE}|B)P(B)}{1 - P(CE)} = \frac{0.02 * 1/3}{1 - 0.957} = 0.154$$

2. (3 pts.) Es pot suposar que el nombre de *typos* (errors tipogràfics que es produeixen en escriure un text a màquina) que surten als llibres d'una determinada editorial es distribueix com una llei de Poisson. Per terme mitjà, hi ha un *typo* per cada 5 pàgines.

   A tots els apartats, justifiqueu les respostes, explicant quines distribucions de probabilitat heu utilitzat.

   (a) Quines són les expressions matemàtiques de l'esperança i la variància d'una variable aleatòria discreta? Quin és el valor de l'esperança i la variància de la variable aleatòria "nombre de *typos* per pàgina"? (0.6 pts.)

   (b) Trobeu la probabilitat que, a una pàgina donada, no hi hagi cap *typo*. (0.4 pts.)

   (c) El tercer capítol té quinze pàgines, independents entre elles. Quina és la probabilitat que hi hagi més de quatre *typos* en aquest capítol? (0.5 pts.)

   (d) Per eliminar els maleïts *typos* es pot emprar un corrector d'estil, però aquesta eina només detecta un 80% dels errors tipogràfics. Si un cert llibre conté 20 *typos*, digueu quina és la probabilitat que se'n trobin, com a màxim, 19. (0.5 pts.)

   (e) La monumental obra, "Memòries completes d'un home amb molt bona memòria", amb 600 pàgines, és el major repte dels correctors de l'editorial. (1 pt.)

      (e.1) Quants *typos* es pot estimar que conté el text complet?

(e.2) Com es distribueix realment aquest nombre?

(e.3) Doneu una distribució alternativa (detalleu els paràmetres necessaris, justificant com els heu trobat).

(e.4) Apliqueu l'aproximació per calcular un interval de valors que inclogui amb una probabilitat del 80% el nombre de *typos* que pot contenir el text complet (deixeu a cada banda una cua del 10%).

SOLUCIÓ

(a) Si X és una v.a. discreta:

$$E(X) = \mu_X = \sum_{\forall k} kP(X = k)$$

$$Var(X) = \sigma_X^2 = \sum_{\forall k} (k - \mu_X)^2 P(X = k) = E(X^2) - (E(X))^2$$

Si N és el nombre de *typos* per pàgina, $N \sim P(\lambda = 1/5)$, i $\mu_N = \sigma_N^2 = 1/5$.

(b) $P(N = 0) = e^{-\lambda} = e^{-1/5} = 0.8187$

(c) C és el nombre de *typos* al tercer capítol, i es suposa que és la suma de quinze v.a. de Poisson independents: $C \sim P(\lambda = 15 \times 1/5)$, $C \sim P(\lambda = 3)$.

$$P(C > 4) = 1 - P(C \leq 4) = 1 - F_C(4) = 1 - \sum_{k=0}^{4} P(C = k) = 1 - \sum_{k=0}^{4} e^{-3}\frac{3^k}{k!}$$

k	$P(C = k)$
0	0.04979
1	0.14936
2	0.22404
3	0.22404
4	0.16803
	0.81526

Així doncs,
$P(C > 4) = 1 - 0.81526 = 0.18474$

(d) El nombre d'errors ja no és aleatori, el que és aleatori és la quantitat que el corrector d'estil pot identificar, dels que hi ha realment . Així, diguem que T és la v.a. que compta aquesta quantitat. Per tant, T es distribueix d'acord a la llei Binomial: $T \sim B(n = 20, p = 0.80)$. Llavors:

$$P(T \leq 19) = 1 - P(T > 19) = 1 - P(T = 20) = 1 - 0.80^{20} = 0.98847$$

(e) Ara, L denota la quantitat de *typos* en un llibre de 600 pàgines.

(e.1) $E(L) = 600 \times 1/5 = 120$

(e.2) Igual que per a un capítol, la distribució és de Poisson, amb paràmetre $\lambda = 120$.

(e.3) Com que el valor del paràmetre és molt alt, es pot aproximar per una distribució Normal, de paràmetres $\mu = \mu_L = 120$ i $\sigma = \sigma_L = \sqrt{120}$.

(e.4) Es busca el interval $[a, b]$ tal que $P(a \leq L \leq b) = 0.80$ i $P(L \leq a) = 0.10$. Com que la distribució Normal és simètrica, prenem com a centre el valor esperat de L:

$$[a, b] = 120 \pm Z_{0.10} \times \sqrt{120}$$

$Z_{0.10}$ és el valor d'una Normal estàndard que deixa per sota una probabilitat de 0.10. Com que el signe és indiferent, també podem utilitzar el valor oposat $Z_{0.90}$, que sí trobarem a les taules, de forma aproximada:

$$F_Z(1.28) = 0.8997 \qquad F_Z(1.29) = 0.9015$$

Encara que el valor exacte és 1.281552..., 1.28 ja serveix:

$$[a, b] = 120 \pm 1.28 \times \sqrt{120} = [105.98, 134.02]$$

3. (3 pts.) Donada una variable aleatòria X absolutament contínua amb la següent funció de densitat:

$$f(x) = \begin{cases} x^2 & \text{si} \quad 0 \leq x < 1 \\ \dfrac{-3x + 7}{4} & \text{si} \quad 1 \leq x \leq \dfrac{7}{3} \\ 0 & \text{altrament} \end{cases}$$

(a) Dibuixeu la gràfica de la funció de densitat, calculeu la funció de distribució i feu-ne també una gràfica aproximada. (1 pt.)

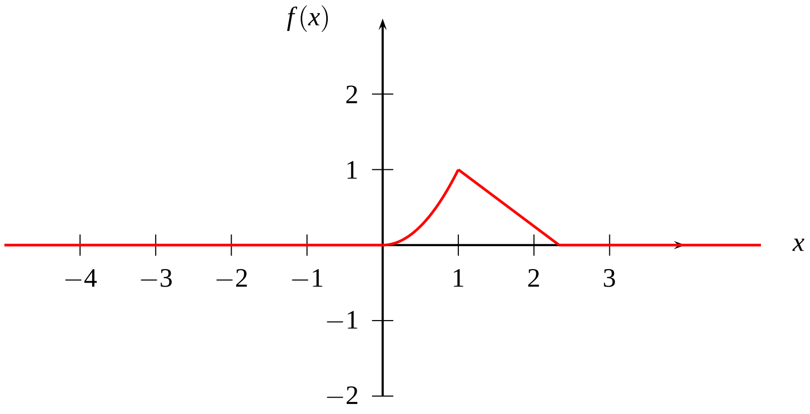
(b) Quin és el valor mitjà esperat de la v. a. X? I la seva desviació típica? (1 pt.)

(c) Calculeu $P(1/2 < X < 2)$. (0.5 pt)

(d) Si sabem que $X < 1$, quina és la probabilitat que $X > 1/3$? (0.5 pt)

SOLUCIÓ

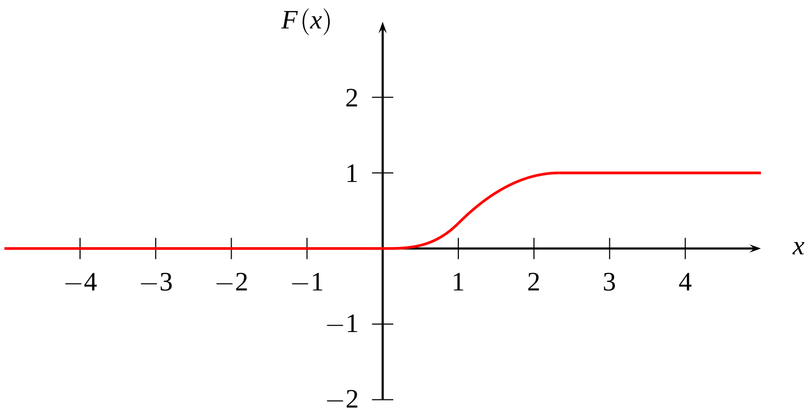
(a) Dibuix de la gràfica de la funció de densitat



La funció de distribució és:

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0.8 \\ \displaystyle\int_0^x t^2\,dt = \frac{x^3}{3} & \text{si} \quad 0 \leq x < 1 \\ \displaystyle\int_0^1 x^2\,dx + \int_1^x \left(-\frac{3}{4}t + \frac{7}{4}\right)dt = -\frac{3}{8}x^2 + \frac{7}{4}x - \frac{25}{24} & \text{si} \quad 1 \leq x < \frac{7}{3} \\ 1 & \text{si} \quad x \geq \frac{7}{3} \end{cases}$$

Dibuix de la gràfica de la funció de distribució:



(b)

$$E(X) = \int_{-\infty}^{-\infty} x\,f(x)\,dx = \int_0^1 x x^2\,dx + \int_1^{7/3} x\left(-\frac{3}{4}x + \frac{7}{4}\right)dx = 1.213$$

$$E(X^2) = \int_{-\infty}^{-\infty} x^2\,f(x)\,dx = \int_0^1 x^2 x^2\,dx + \int_1^{7/3} x^2\left(-\frac{3}{4}x + \frac{7}{4}\right)dx = 1.6565$$

$$Var(X) = E(X^2) - (E(X))^2 = 1.6565 - (1.213)^2 = 0.1851 \qquad \sqrt{Var(X)} = 0.43$$

(c)
$$P(1/2 < X < 2) = \int_{1/2}^{1} x^2\,dx + \int_{1}^{2} (-\frac{3}{4}x + \frac{7}{4})\,dx = \frac{11}{12} = 0.9167$$

(d)
$$P(X > 1/3 \,|\, X < 1) = \frac{P(1/2 < X < 2)}{P(X < 1)} = \frac{\int_{1/3}^{1} x^2\,dx}{\int_0^1 x^2\,dx} = \frac{26}{27} = 0.963$$

4. (2 pts.) Contesteu els següents apartats justificant, en tots els casos, les vostres respostes. Suposarem que X i Y són dues v.a. definides sobre el mateix espai de probabilitat.

(a) Donades dues variables aleatòries amb funció de distribució conjunta $F(x,y)$, és certa la següent igualtat? (0.6 pts.)
$$F(x,y) = F_X(x)F_Y(y)$$
amb $F_X(x)$ i $F_Y(y)$ funcions de distribució marginals de X i Y respectivament.

(b) Donades dues variables aleatòries X i Y, es compleix la següent igualtat? (0.6 pts.)
$$Var(X-Y) = Var(X) - Var(Y)$$

(c) Donada una variable aleatòria amb distribució Binomial de paràmetres n i p, en quines condicions es pot aproximar a una variable Poisson? I a una Normal? Doneu els paràmetres d'aquestes distribucions aproximades. (0.8 pts.)

SOLUCIÓ

(a) Aquesta igualtat només és certa si i només si X i Y són v.a. independents donat que:
$$\begin{aligned} F(x,y) &= P(X \leq x, Y \leq y) = P(X \in (-\infty, x] \cap Y \in (-\infty, y]) \\ &= P(X \in (-\infty, x])P(Y \in (-\infty, y]) = P(X \leq x)P(Y \leq y) \\ &= F_X(x)F_Y(y) \end{aligned}$$

(b) En el cas que X i Y siguin independents sabem que:
$$Var(X-Y) = Var(X+(-Y))) = Var(X) + Var(-Y) = Var(X) + (-1)^2 Var(Y) = Var(X) + Var(Y)$$
Però en el cas general, i donat que l'esperança és lineal:
$$\begin{aligned} Var(X-Y) &= E((X-Y)^2) - (E(X-Y))^2 = E(X^2 - 2XY + Y^2) - (E(X) - E(Y))^2 \\ &= E(X^2) - 2E(XY) + E(Y^2) - (E(X))^2 + 2E(X)E(Y) - (E(Y))^2 \\ &= Var(X) + Var(Y) - 2(E(XY) - E(X)E(Y)) \\ &= Var(X) + Var(Y) - 2Cov(X,Y) \end{aligned}$$

(c) Suposem $X \sim B(n,p)$, aleshores:

- $X \approx N(\mu = np, \sigma = \sqrt{np(1-p)})$ quan $n \geq 30$ i $0.1 < p < 0.9$. Si $p < 0.1$ o bé $n < 30$, l'aproximació és acceptable per si $np > 5$. Si $p \simeq 0.5$, l'aproximació segueix sent vàlida si $np > 3$, fins i tot per a valors molt moderats de n.
- $X \approx Poiss(\lambda = np)$ quan $n \geq 30$, $p < 0.1$ i $np < 5$.