ETS d'Enginyeria de Telecomunicació de Barcelona

PROBABILITAT, PROCESSOS ESTOCÀSTICS I ESTADÍSTICA

Examen Parcial **2 de maig de 2013**

1. A un compte de correu arriben un 40% de missatges d'spam. El sistema disposa d'un filtre que desvia els missatges d'spam a la carpeta d'spam (C_S) deixant la resta en la carpeta d'entrada (C_E). El filtre, però, pot fallar i hi ha una probabilitat 0,07 que un missatge d'spam (S) vagi a C_E, i una probabilitat 0,05 que un missatge bo (B) vagi a C_S.

   (a) Amb 6 missatges entrants, (abans de l'acció del filtre) quina és la probabilitat que hi hagi 3 o més missatges d'spam? Si hi ha 3 missatges d'spam, quina és la probabilitat que en els tres primers entrants n'hi hagi un d'spam?

   (b) Quin tant per cent de missatges d'spam conté la carpeta d'entrada? Quin tant per cent de missatges bons conté la carpeta d'spam?

   (c) Siguin N i M el nombre de missatges que cal obrir fina a trobar-ne un d'spam en C_E i un de bo en C_S, respectivament. Digues quin tipus de variable són i què valen les seves esperances i desviacions. Si la llista de missatges es visualitza a 20 per pàgina. Quina és la probabilitat en cada cas que la primera pàgina no mostri cap missatge anòmal?

   (d) L'usuari rep uns 40 missatges diaris. La carpeta d'spam els manté durant un més. És un valor acceptable una quota de 500 missatges en C_S? Quants mesos haurien de passar fins que notéssim problemes amb la quota?

   Solució:

   Segons l'enunciat, $P(S) = 0{,}4$, $P(B) = 0{,}6$, $P(C_E|S) = 0{,}07$, $P(C_S|B) = 0{,}05$.

   (a) Sigui X el nombre d'spams en els 6 missatges (binomial $n = 6$ $p = 0{,}4$) i X_1, X_2 els nombres d'spams en els tres primers missatges i en els tres últims missatges, respectivament (binomials $n = 3$ $p = 0{,}4$, independents).

$$P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$$

$$= 1 - 0{,}6^6 - 6 \cdot 0{,}4 \cdot 0{,}6^5 - \binom{6}{2} 0{,}4^2 \cdot 0{,}6^4 = 0{,}4557.$$

$$P(X_1 = 1|X = 3) = \frac{P(X_1 = 1)P(X_2 = 2)}{P(X = 3)} = \frac{\binom{3}{1}0{,}4 \cdot 0{,}6^2 \binom{3}{2}0{,}4^2 \cdot 0{,}6}{\binom{6}{3}0{,}4^3 \cdot 0{,}6^3}$$

$$= \frac{\binom{3}{1}\binom{3}{2}}{\binom{6}{3}} = \frac{9}{20} = 0{,}45.$$

   (b)

$$P(S|C_E) = \frac{P(C_E|S)P(S)}{P(C_E|S)P(S) + P(C_E|B)P(B)} = \frac{0,07 \cdot 0{,}4}{0,07 \cdot 0{,}4 + (1 - 0,05) \cdot 0{,}6} = 0{,}0468.$$

$$P(B|C_S) = \frac{P(C_S|B)P(B)}{P(C_S|B)P(B) + P(C_S|S)P(S)} = \frac{0,05 \cdot 0{,}6}{0,05 \cdot 0{,}6 + (1 - 0,07) \cdot 0{,}4} = 0{,}0746.$$

   Així, un $4,7\%$ d'spam en la carpeta d'entrada i un $7,4\%$ de missatges bons en la carpeta d'spam.

(c) Són variables geomètriques. N amb $p = P(S|C_E) = 0{,}0468$ i, per tant, valor mitjà $\frac{1}{p} = 21{,}3$ i desviació $\frac{\sqrt{q}}{p} = 20{,}8$; M amb $p = P(B|C_S) = 0{,}0746$ i, per tant, valor mitjà 13,4 i desviació 12,9.

Les probabilitats de no trobar missatges anòmals valen: en C_E, $P(N > 20) = (1 - 0{,}0468)^{20} = 0{,}38$; en C_S, $P(M > 20) = (1 - 0{,}0746)^{20} = 0{,}21$.

(d) En un més arriben uns $40 \cdot 30 = 1200$ missatges. El nombre de missatges d'spam té mitjana $1200 \cdot 0{,}4 = 480$ i desviació $\sqrt{1200 \cdot 0{,}4 \cdot 0{,}6} = 17$. Com el marge que ens dóna la quota és de poc més d'una desviació, la capacitat resulta insuficient. La probabilitat que un més és superi la quota és, aproximant per la normal, $P(X > 500) = 1 - F(500) = 1 - \frac{1}{2}(1 + \operatorname{erf}(\frac{500-480}{\sqrt{2}\cdot 17})) = 0{,}5 - 0{,}5 \operatorname{erf}(0{,}83) = 0{,}12$. Podem esperar tenir problemes en uns $\frac{1}{0{,}12} = 8$ mesos.

2. L'ocupació d'ample de banda en una línia de comunicació es descriu amb una variable aleatòria X amb valors en l'interval $\Omega_X = [0,1]$ i funció de densitat

$$f_X(x) = Kx(x+a)(1-x), \qquad 0 \le x \le 1,$$

on K i a són constants positives.

(a) Calcula el valor de K, en funció de a. Quins valors pot tenir l'esperança, m, de X?

(b) Per $a = 1$: Calcula la desviació estàndard, σ, de X. Calcula la probabilitat que $X \in (m-\sigma, m+\sigma)$ i compara-la amb el valors que tindria si X fos uniforme o gaussiana.

(c) Per $a = 1$: Calcula la funció de densitat i l'esperança de la variable $Y = X^2 + 1$.

Solució:

(a) $1 = \int_0^1 K(ax+(1-a)x^2-x^3)dx = K\left(\frac{a}{2}+\frac{1-a}{3}-\frac{1}{4}\right) = K\frac{2a+1}{12} \Rightarrow K = \frac{12}{2a+1}.$

$E[X] = \frac{12}{2a+1}\int_0^1 x(ax+(1-a)x^2-x^3)dx = \frac{12}{2a+1}\left(\frac{a}{3}+\frac{1-a}{4}-\frac{1}{5}\right) = \frac{5a+3}{10a+5}.$

És una funció monòtona que varia entre $\frac{3}{5} = 0{,}6$ $(a=0)$ i $\frac{1}{2} = 0{,}5$ $(a=\infty)$. Per tant, $0{,}5 < E[X] \le 0{,}6$.

(b) En aquest cas, $f_X(x) = 4(x-x^3)$ per $0 < x < 1$ i $m = E[X] = \frac{8}{15} = 0{,}53$.

$E[X^2] = \int_0^1 x^2 4(x-x^3)dx = \frac{1}{3}$. $\sigma = \sqrt{E[X^2]-E[X]^2} = \sqrt{\frac{1}{3}-\left(\frac{8}{15}\right)^2} = \frac{\sqrt{11}}{15} = 0{,}2211$.

$P(X \in (m-\sigma, m+\sigma)) = \int_{m-\sigma}^{m+\sigma} 4(x-x^3)dx = \left[2x^2-x^4\right]_{0{,}3122}^{0{,}7544} = 0{,}629.$

En el cas d'una variable uniforme (en $[0,1]$) és $m = \frac{1}{2}$ i $\sigma = \frac{1}{\sqrt{12}}$ amb el que l'anterior probabilitat val $\frac{2}{\sqrt{12}} = 0{,}577$. En el cas de la gaussiana la probabilitat val $F(m+\sigma) - F(m-\sigma) = \mathrm{erf}\left(\frac{1}{\sqrt{2}}\right) = 0{,}683$.

Les probabilitats són similars però la de X és superior a la de la uniforme i inferior a la de la gaussiana ja que la uniforme té la densitat més dispersa i la gaussiana la té més concentrada.

(c) Com X varia entre 0 i 1, la variable Y pren valors en $[1,2]$ i la seva densitat és:

$$f_Y(y) = f_X(x)\frac{1}{\left|\frac{dy}{dx}\right|} = 4(x-x^3)\frac{1}{2x} = 2(1-x^2) = 2(2-y), \qquad 1 < y < 2.$$

(per a cada $y \in (1,2)$ hi ha un únic valor $x = \sqrt{y-1}$ que hi va a parar.)

$$E[Y] = \int_1^2 y \cdot 2(2-y)dy = \frac{4}{3}.$$

Alternativament, $E[Y] = E[X^2] + 1 = \frac{1}{3} + 1 = \frac{4}{3}$.