

PROBABILITAT, PROCESSOS ESTOCÀSTICS I ESTADÍSTICA

Solució de l'Examen final

14 de juny de 2013

1. En un centre de càlcul hi ha ordinadors connectats a una xarxa que té dues seccions. En la secció S_1 hi ha quatre ordinadors i en la secció S_2 n'hi ha cinc. Un virus infecta un ordinador de S_1 i dos ordinadors de S_2 . Quan un usuari utilitza un ordinador infectat hi ha una probabilitat $\frac{3}{5}$ que els seus arxius s'infectin.
- (a) Un procediment de verificació es basa en triar tres ordinadors a l'atzar dels nou disponibles. Calculeu la probabilitat que es triï com a màxim un ordinador infectat. Feu-ho pels dos esquemes següents: quan es trien els tres ordinadors diferents, per un costat, i quan es trien els tres ordinadors de forma independent (amb possible repetició, per tant), per l'altre. Indiqueu quina de les dues probabilitats és més alta i raoneu perquè.
- (b) Als usuaris que accedeixen a un ordinador se'ls assigna primer una secció, de manera que la probabilitat de triar S_1 val $\frac{2}{3}$, i després se'ls assigna a l'atzar un dels ordinadors de la secció elegida. Calculeu la probabilitat que s'infectin els arxius d'un usuari que accedeix al sistema. Si un usuari veu que els seus arxius s'han infectat, quina és la probabilitat que se li hagués assignat la secció S_1 ? Compareu amb la probabilitat a priori i comenteu si la variació és raonable.
- (c) Al sistema van accedint usuaris independents (seguint l'esquema de l'apartat (b)). Sigui N la variable que compta quants usuaris han accedit en el moment en que es produeix la primera infecció d'arxius. Trobeu el tipus de variable que és, la seva esperança i la probabilitat que $3 \leq N \leq 7$.

Solució:

- (a) Si X és la variable que dona el nombre d'ordinadors infectats dins dels tres triats, ens demanen $P(X \leq 1)$.
Quan en triem 3 diferents, en un conjunt de 9 ordinadors (3 infectats, 6 no infectats):

$$P(X \leq 1) = P(X=0) + P(X=1) = \frac{\binom{6}{3}}{\binom{9}{3}} + \frac{\binom{3}{1}\binom{6}{2}}{\binom{9}{3}} = \frac{65}{84} = 0,7738.$$

Quan els triem de manera independent, X és binomial amb $n = 3$ i $p = \frac{1}{3}$.

$$\begin{aligned} P(X \leq 1) &= P(X=0) + P(X=1) = \binom{3}{0} \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^3 + \binom{3}{1} \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^2 \\ &= \frac{8}{27} + \frac{12}{27} = \frac{20}{27} = 0,7407. \end{aligned}$$

(o, més directament, hi ha 9^3 maneres de triar els tres ordinadors, 6^3 de triar-ne zero infectats i $3 \cdot 3 \cdot 6^2$ de triar-ne un infectat. Llavors la probabilitat val $(216 + 324)/729 = 20/27$.)

Triar un nombre gran d'ordinadors infectats és més probable quan els triem independents ja que triant-los diferents l'elecció d'un ordinador infectat disminueix la probabilitat que els altres triats ho siguin. Així, els valors petits de X són menys probables en la tria independent.

- (b) Denotem O_I = “Ordinador infectat”, A_I = “Arxius infectats”. $P(S_1) = \frac{2}{3}$, $P(S_2) = \frac{1}{3}$. $P(O_I|S_1) = \frac{1}{4}$, $P(O_I|S_2) = \frac{2}{5}$. $P(A_I|O_I) = \frac{3}{5}$. Llavors:

$$\begin{aligned} P(A_I) &= P(A_I|O_I)P(O_I) = P(A_I|O_I)(P(O_I|S_1)P(S_1) + P(O_I|S_2)P(S_2)) \\ &= \frac{3}{5} \left(\frac{1}{4} \cdot \frac{2}{3} + \frac{2}{5} \cdot \frac{1}{3} \right) = \frac{9}{50} = 0,18. \end{aligned}$$

Per Bayes:

$$\begin{aligned} P(S_1|A_I) &= \frac{P(A_I|S_1)P(S_1)}{P(A_I)} = \frac{P(A_I|O_I)P(O_I|S_1)P(S_1)}{P(A_I)} \\ &= \frac{\frac{3}{5} \cdot \frac{1}{4} \cdot \frac{2}{3}}{\frac{9}{50}} = \frac{5}{9} = 0,55. \end{aligned}$$

A priori $P(S_1) = 0,66$. La probabilitat ha disminuït ja que la infecció és més probable en S_2 .

- (c) N és geomètrica de paràmetre $p = P(A_I) = 0,18$. La seva esperança val $E[N] = \frac{1}{p} = 5,5$. La funció de distribució de N és $F_N(n) = 1 - q^n$ on $q = 1 - p = 0,82$.

$$P(3 \leq N \leq 7) = P(2 < N \leq 7) = F_N(7) - F_N(2) = 0,82^2 - 0,82^7 = 0,4231.$$

2. Donada la variable aleatòria bidimensional (X, Y) amb funció de densitat

$$f_{XY}(x, y) = \begin{cases} e^{-(x+y)}, & x, y > 0 \\ 0, & \text{altrament,} \end{cases}$$

es consideren les variables aleatòries

$$S = X + Y, \quad Q = \frac{Y}{X}.$$

Es demana:

- Les funcions de distribució marginals de S i Q .
- La funció de densitat conjunta de (S, Q) .
- Les funcions de densitat marginals de S i Q . Obtingueu aquestes densitats tant a partir de (a) com de (b). Són S i Q independents?

Solució:

Les variables aleatòries X, Y són exponencials de paràmetre 1. En efecte,

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \begin{cases} 0, & x < 0 \\ \int_0^{\infty} e^{-x} e^{-y} dy = e^{-x}, & x > 0 \end{cases}.$$

Anàlogament, $f_Y(y) = e^{-y}$ si $y > 0$ i $f_Y(y) = 0$ altrament. Com que $f_{XY}(x, y) = f_X(x)f_Y(y)$ tenim que X, Y són independents

(a) Sabem que $F_S(s) = P(S \leq s) = P(X + Y \leq s)$. Si $s < 0$ aquesta probabilitat val 0. Sigui $s > 0$ i considerem la regió $D = \{(x, y) \in \mathbb{R}^2 : x + y \leq s\}$. Tenim

$$\begin{aligned} P(X + Y \leq s) &= \iint_D f_{XY}(x, y) dx dy = \int_0^s \left(\int_0^{s-x} e^{-y} dy \right) e^{-x} dx \\ &= \int_0^s \left(1 - e^{-(s-x)} \right) e^{-x} dx = \int_0^s (e^{-x} - e^{-s}) dx = 1 - e^{-s} - se^{-s}. \end{aligned}$$

D'altra banda, $F_Q(q) = P(Q \leq q) = P(Y/X \leq q)$. Si $q < 0$ aquesta probabilitat val 0 ja que X i Y prenen valors positius. Sigui $q > 0$ i considerem la regió $T = \{(x, y) \in \mathbb{R}^2 : y/x \leq q\}$. Tenim

$$\begin{aligned} P(Y/X \leq q) &= \iint_T f_{XY}(x, y) dx dy = \int_0^{\infty} \left(\int_0^{qx} e^{-y} dy \right) e^{-x} dx \\ &= \int_0^{\infty} (1 - e^{-qx}) e^{-x} dx = \int_0^{\infty} (e^{-x} - e^{-(1+q)x}) dx = 1 - \frac{1}{1+q} = \frac{q}{1+q}. \end{aligned}$$

(b) Aplicarem el teorema del canvi de variable. Podem comprovar que la transformació definida pel sistema $s = x + y$, $q = y/x$, aplica bijectivament l'interior del primer quadrant del pla (x, y) en l'interior del primer quadrant del pla (s, q) . Donats $s > 0$, $q > 0$, la única solució del sistema és

$$x = \frac{s}{1+q}, \quad y = \frac{sq}{1+q}.$$

El jacobià de la transformació val

$$J(x, y) = \begin{vmatrix} 1 & 1 \\ -y/x^2 & 1/x \end{vmatrix} = \frac{x+y}{x^2}.$$

Per tant,

$$f_{SQ}(s, q) = \frac{f_{XY}(x, y)}{|J(x, y)|} \Big|_{x=s/(1+q), y=sq/(1+q)} = \frac{se^{-s}}{(1+q)^2}, \quad s, q > 0.$$

(c) Les densitats marginals de S i Q valen

$$f_S(s) = \int_{-\infty}^{\infty} f_{SQ}(s, q) dq = \int_0^{\infty} \frac{se^{-s}}{(1+q)^2} dq = se^{-s}, \quad s > 0;$$

$$f_Q(q) = \int_{-\infty}^{\infty} f_{SQ}(s, q) ds = \int_0^{\infty} \frac{se^{-s}}{(1+q)^2} ds = \frac{1}{(1+q)^2}, \quad q > 0.$$

També podem obtenir aquests resultats derivant les funcions de distribució obtingudes a l'apartat (a):

$$f_S(s) = F'_S(s) = (1 - e^{-s} - se^{-s})' = se^{-s}, \quad s > 0;$$

$$f_Q(q) = F'_Q(q) = \left(\frac{q}{1+q} \right)' = \frac{1}{(1+q)^2}, \quad q > 0.$$

Les variables S i Q són independents ja que es compleix $f_{SQ}(s, q) = f_S(s)f_Q(q)$.

3. El nombre de viatgers que arriba a una terminal d'un aeroport segueix un procés de Poisson $X(t)$ amb paràmetre $\lambda = 15$ viatgers per minut.

- (a) Trobeu la probabilitat que en 1 minut, en els primers 20 segons arribin 3 viatgers i en els últims 16 segons n'arribin 2.
- (b) Trobeu l'esperança i la variància del temps fins que arriba el desè viatger.
- (c) Trobeu la millor estimació lineal no homogènia de $X(t)$ en funció de $X(t - 10)$ (temps expressat en segons). Calculeu l'error quadràtic mitjà que es comet amb aquesta estimació.

Solució:

El promig d'arribades en segons és $\lambda = 15/60 = 1/4$.

- (a) Com que les arribades de viatgers segueixen un procés de Poisson,

$$\begin{aligned} P[X(20) = 3, X(60) - X(44) = 2] &= P[X(10) = 3]P[X(60 - 44)] = 2 \\ &= P[X(20) = 3]P[X(16)] = 2 \\ &= \frac{5^3 e^{-5}}{3!} \frac{4^2 e^{-4}}{2!} = \frac{500}{3e^9}. \end{aligned}$$

- (b) Sigui X_i la variable aleatoria que compta el temps en segons entre la arribada del $i - 1$ i del i -èsim viatger. Aleshores, les variables X_i són exponencials independents i per tant el temps d'arribada del desè viatger és, $T_{10} = X_1 + X_2 + \dots + X_{10}$. Per tant,

$$\begin{aligned} E[T_{10}] &= 10E[X_i] = 10/\lambda = 40, \\ Var[T_{10}] &= 10Var[X_i] = 10/\lambda^2 = 160. \end{aligned}$$

- (c) Estimem $X(t)$ per $\widehat{N}(t) = aX(t - 10) + b$. Calculem a, b segons el Principi d'ortogonalitat,

$$\begin{aligned} E[(\widehat{N}(t) - X(t))X(t - 10)] &= 0 \\ E[(\widehat{N}(t) - X(t)) \cdot 1] &= 0 \end{aligned}$$

D'on,

$$\begin{aligned} aE[X(t - 10)X(t - 10)] + bE[X(t - 10)] &= E[X(t)X(t - 10)] \\ aE[X(t - 10)] + b &= E[X(t)] \end{aligned}$$

Fent servir del valor mitjà i l'autocorrelació d'un procés de Poisson,

$$E[X(t)] = \lambda t, \quad R_X(t_1, t_2) = \lambda^2 t_1 t_2 + \lambda \min\{t_1, t_2\},$$

tenim,

$$\begin{aligned} a(\lambda^2(t - 10)^2 + \lambda(t - 10)) + b\lambda(t - 10) &= \lambda^2 t(t - 10) + \lambda(t - 10) \\ a\lambda(t - 10) + b &= \lambda t \end{aligned}$$

De forma equivalent,

$$\begin{aligned} a\lambda(t - 10) + a + b &= \lambda t + 1 \\ a\lambda(t - 10) + b &= \lambda t \end{aligned}$$

Per tant, $a = 1$ i $b = 5/2$.

Calculem l'error quadràtic mitjà d'aquesta estimació lineal,

$$\begin{aligned}\bar{\epsilon} &= E[(N(t) - \widehat{N}(t))^2] = E[(N(t) - \widehat{N}(t))N(t)] \\ &= E[(N^2(t) - N(t-10)N(t) - 5N(t))/2] \\ &= R_N(t, t) - R_N(t-10, t) - 5E[N(t)]/2 \\ &= \lambda^2 t^2 + \lambda t - \lambda^2 t(t-10) - \lambda(t-10) - 5\lambda t/2 \\ &= 10\lambda(\lambda t + 1 - t/4) \\ &= 5/2.\end{aligned}$$