ETS d'Enginyeria de Telecomunicació de Barcelona

PROBABILITAT, PROCESSOS ESTOCÀSTICS I ESTADÍSTICA

Solució de l'examen final **21 de juny de 2011**

---

1 Les dades provinents de mesures astronòmiques arriben seqüencialment a un centre terrestre a través de tres satèl·lits que participen simultàniament en l'emissió de cada dada. Cada satèl·lit, de forma independents dels altres, pot fallar en la transmissió amb probabilitat $\dfrac{1}{2}$. A partir del senyal dels tres satèl·lits cada dada es reconstrueix amb una qualitat que pot ser alta o baixa. Si han fallat N satèl·lits, la probabilitat de tenir la dada amb qualitat alta val $\dfrac{1}{2N+2}$. El temps necessari (en mil·lisegons) per reconstruir la dada és una variable exponencial de paràmetre $\lambda = \dfrac{4-N}{10}$.

(a) En 80 dades independents, què valen el nombre mitjà i la desviació del nombre de dades reconstruïdes amb alta qualitat? Quina és la probabilitat de tenir més de 25 dades en alta qualitat?

(b) S'ha rebut una dada en alta qualitat. Quina és la probabilitat que hagin fallat exactament dos satèl·lits? Comenteu el canvi que hi ha hagut respecte la probabilitat a priori.

(c) Què val el temps mitjà per reconstruir una dada? Podem reconstruir-les a temps real si arriben 230 dades per segon?

Solució:

Denotem A l'esdeveniment "missatge recostruït amb qualitat alta". El nombre N de satèl·lits que fallen és una variable binomial amb $n = 3$ i $p = \frac{1}{2}$. Llavors, per $0 \le k \le 3$:

$P_N(k) = \dbinom{3}{k}\left(\dfrac{1}{2}\right)^k\left(\dfrac{1}{2}\right)^{3-k} = \dfrac{1}{8}\dbinom{3}{k}$. Per tant, $P_N(0) = \dfrac{1}{8}$, $P_N(1) = \dfrac{3}{8}$, $P_N(2) = \dfrac{3}{8}$, $P_N(3) = \dfrac{1}{8}$.

També tenim: $P(A|N{=}k) = \dfrac{1}{2k+2}$.

(a) El nombre X de dades en alta qualitat és una variable binomial amb $n = 80$ i

$$p = P(A) = \sum_{k=0}^{3} P(A|N{=}k)P_N(k) = \frac{1}{2}\cdot\frac{1}{8} + \frac{1}{4}\cdot\frac{3}{8} + \frac{1}{6}\cdot\frac{3}{8} + \frac{1}{8}\cdot\frac{1}{8} = \frac{15}{64} = 0{,}2343.$$

Així, $E[X] = np = 80\cdot\dfrac{15}{64} = \dfrac{75}{4} = 18{,}75$.

$$\sigma_X = \sqrt{npq} = \sqrt{80\cdot\frac{15}{64}\cdot\frac{49}{64}} = \frac{35\sqrt{3}}{16} = 3{,}7888.$$

Aproximant amb la gaussiana:

$$P(X > 25) = 1 - F_X(25{,}5) = 1 - \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{25{,}5 - 18{,}75}{\sqrt{2}\cdot 3{,}7888}\right)\right) = 0{,}5 - 0{,}5\operatorname{erf}(1{,}2597) = 0{,}0374.$$

(sense correcció del mig punt dóna 0,0495, el valor exacte és 0,0409.)

(b) Per Bayes:

$$P(N{=}2|A) = \frac{P(A|N{=}2)P(N{=}2)}{P(A)} = \frac{\frac{1}{6}\cdot\frac{3}{8}}{\frac{15}{64}} = \frac{4}{15} = 0{,}2666.$$

A priori, $P(N{=}2) = \frac{3}{8} = 0{,}375$. La probabilitat ha disminuït ja que un nombre alt de satèl·lits que fallen afavoreix la baixa qualitat.

(c) $E[T|N] = \dfrac{1}{\lambda} = \dfrac{10}{4-N}$.

$$E[T] = E\left[\frac{10}{4-N}\right] = 10\sum_{k=0}^{3}\frac{1}{4-k}P_N(k) = 10\left(\frac{1}{4}\cdot\frac{1}{8} + \frac{1}{3}\cdot\frac{3}{8} + \frac{1}{2}\cdot\frac{3}{8} + 1\cdot\frac{1}{8}\right) = \frac{75}{16} = 4{,}6875\text{ms}.$$

En un segon podem reconstruir $\dfrac{1000}{4{,}6875} \simeq 213$ dades. Per tant, no podem reconstruir les dades que arriben a temps real.

2 A un punt d'una xarxa arriben missatges formats per dos blocs de mides X i Y (Mb). La densitat de la variable bidimensional (X, Y) és

$$f(x,y) = \begin{cases} K(1-xy), & \text{si } 0 \le x \le 1 \text{ i } 0 \le y \le 1 \\ \\ 0, & \text{altrament.} \end{cases}$$

on K és una constant.

(a) Calculeu el valor de K. Trobeu les densitats marginals de X i de Y. Calculeu el coeficient de correlació ρ entre X i Y. Són independents X i Y?

(b) Calculeu l'esperança de X condicionada a $Y = y$. Quina és la millor estimació possible del valor de X si sabem que $Y = 1$?

(c) La mida del missatge és $Z = X+Y$. Calculeu la seva esperança i variància. Trobeu la probabilitat que $Z > 1$.

(Indicació: $\displaystyle\int_0^1 t^n dt = \frac{1}{n+1}$.)

Solució:

(a) La variable (X, Y) viu en el quadrat $[0,1] \times [0,1]$.

$1 = \displaystyle\int_0^1 dx \int_0^1 dy\; K(1-xy) = K \int_0^1 dx \left(1 - \frac{x}{2}\right) = \frac{3K}{4}$. Així, $K = \frac{4}{3}$.

Densitat marginal de X: $f(x) = \displaystyle\int_0^1 dy\; \frac{4}{3}(1-xy) = \frac{4}{3}\left(1 - \frac{x}{2}\right)$, per $0 < x < 1$.

Per simetria $x \leftrightarrow y$, la densitat marginal de Y: $f(y) = \dfrac{4}{3}\left(1 - \dfrac{y}{2}\right)$, per $0 < y < 1$.

Com $f(x,y) \neq f(x)f(y)$, les variables *no* són independents.

$E[X] = E[Y] = \displaystyle\int_0^1 \frac{4}{3}x\left(1 - \frac{x}{2}\right) dx = \frac{4}{9} = 0{,}4444.$

$E[X^2] = E[Y^2] = \displaystyle\int_0^1 \frac{4}{3}x^2\left(1 - \frac{x}{2}\right) dx = \frac{5}{18}$. $V[X] = V[Y] = \dfrac{5}{18} - \left(\dfrac{4}{9}\right)^2 = \dfrac{13}{162} = 0{,}0802.$

$E[XY] = \displaystyle\int_0^1 dx \int_0^1 dy\; xy\frac{4}{3}(1-xy) = \frac{4}{3}\int_0^1 dx \left(\frac{x}{2} - \frac{x^2}{3}\right) = \frac{5}{27}.$

$C[X,Y] = \dfrac{5}{27} - \left(\dfrac{4}{9}\right)^2 = -\dfrac{1}{81} = -0{,}01234$. $\rho = \dfrac{C[X,Y]}{\sigma_X \sigma_Y} = -\dfrac{2}{13} = -0{,}1538.$

(b) $f(x|y) = \dfrac{f(x,y)}{f(y)} = \dfrac{1-xy}{1-\frac{y}{2}}$ per $0 < x < 1$.

$E[X|y] = \displaystyle\int_0^1 dx\; x\frac{1-xy}{1-\frac{y}{2}} = \frac{3-2y}{6-3y}.$

La millor estimació possible de X donada $Y = 1$ la dóna l'esperança condicionada:

$E[X|Y\!=\!1] = \dfrac{1}{3}.$

(c) $E[Z] = E[X+Y] = E[X] + E[Y] = \dfrac{4}{9} + \dfrac{4}{9} = \dfrac{8}{9} = 0{,}8888.$

$V[Z] = V[X+Y] = V[X] + V[Y] + 2C[X,Y] = \dfrac{11}{81} = 0{,}1358.$

$P(Z > 1) = 1 - P(Z < 1) = 1 - \displaystyle\int_0^1 dx \int_0^{1-x} dy\; \frac{4}{3}(1-xy) = 1 - \int_0^1 dx\; \frac{4}{3}(1 - x - \frac{x}{2}(1-x)^2) =$

$1 - \dfrac{4}{3}\displaystyle\int_0^1 dx\; (1 - \frac{3}{2}x + x^2 - \frac{x^3}{2}) = \frac{7}{18} = 0{,}3888.$

3 El soroll en un canal de comunicació ve donat pel procés

$$X(t) = A\cos t + B\sin t + C,$$

on A, B, C són variables aleatòries gaussianes incorrelades, d'esperances $m_A = m_B = 0$, $m_C = 1$ i desviacions $\sigma_A = \sigma_B = \sigma_C = 1$

(a) Calculeu les funcions de valor mitjà, d'autocorrelació, d'autocovariància i la potència del procés $X(t)$. És estacionari en sentit ampli?

(b) Trobeu la millor estimació de $X(\pi)$ per una constant. Calculeu l'error mitjà associat a aquesta estimació.

(c) Trobeu la millor estimació lineal homogènia de $X(\pi)$ donades $X(0)$ i $X(\frac{\pi}{2})$. Compareu l'error mitjà d'aquesta estimació amb el de l'anterior apartat. Quina estimació és millor? Com es podria millorar l'estimació lineal anterior si només disposem d'aquestes dues dades?

Solució:

(a) Notem primer que $E[A^2] = \sigma_A^2 + m_A^2 = 1$, $E[B^2] = \sigma_B^2 + m_B^2 = 1$, $E[C^2] = \sigma_C^2 + m_C^2 = 2$. Al ser incorrelades $E[AB] = m_A m_B = 0$, $E[AC] = m_A m_C = 0$, $E[BC] = m_B m_C = 0$.
$m(t) = E[X(t)] = E[A]\cos t + E[B]\sin t + E[C] = 1$.
$R(t_1, t_2) = E[X(t_1)X(t_2)] = E[(A\cos t_1 + B\sin t_1 + C)(A\cos t_2 + B\sin t_2 + C)] = E[A^2]\cos t_1 \cos t_2 + E[B^2]\sin t_1 \sin t_2 + E[C^2] +$ termes creuats d'esperança zero $= \cos t_1 \cos t_2 + \sin t_1 \sin t_2 + 2 = 2 + \cos(t_2 - t_1)$.
$C(t_1, t_2) = R(t_1, t_2) - m(t_1)m(t_2) = 1 + \cos(t_2 - t_1)$.
Potència $= R(t, t) = 2 + \cos 0 = 3$.
El procés és estacionari en sentit ampli ja que $m(t) = 1$ és constant i $R(t, t + \tau) = 2 + \cos\tau$ no depèn de t.

(b) La millor estimació per una constant la dóna l'esperança. Així, és $C = E[X(\pi)] = m(\pi) = 1$. L'error mitjà en aquest cas és la variància de la variable: $\overline{\epsilon} = V[X(\pi)] = C(\pi, \pi) = 2$.

(c) L'estimador és $\widehat{X(\pi)} = \alpha X(0) + \beta X(\frac{\pi}{2})$. Les equacions del principi d'ortogonalitat són:

$$\left\{\begin{array}{rcl} E[(\alpha X(0) + \beta X(\frac{\pi}{2}))X(0)] & = & E[X(\pi)X(0)] \\ E[(\alpha X(0) + \beta X(\frac{\pi}{2}))X(\frac{\pi}{2})] & = & E[X(\pi)X(\frac{\pi}{2})]. \end{array}\right.$$

Utilitzant $E[X(t_1)X(t_2)] = R(t_1, t_2) = 2 + \cos(t_2 - t_1)$:

$$\left\{\begin{array}{rcl} 3\alpha + 2\beta & = & 1 \\ 2\alpha + 3\beta & = & 2. \end{array}\right.$$

La solució és $\alpha = -\dfrac{1}{5}$, $\beta = \dfrac{4}{5}$.
L'error mitjà: $\overline{\epsilon} = E\left[\left(X(\pi) - \left(-\frac{1}{5}X(0) + \frac{4}{5}X(\frac{\pi}{2})\right)\right)X(\pi)\right] = R(\pi, \pi) + \frac{1}{5}R(0, \pi) - \frac{4}{5}R(\frac{\pi}{2}, \pi) = \dfrac{8}{5} = 1{,}6$.
Aquesta estimació és millor que la constant ja que l'error és menor. Una manera de millorar-la és fer-la no homogènia. Com les variables són gaussianes aquesta seria ja la millor de les estimacions possibles.