Barcelona, 29 de gener de 2015

Examen d'Introducció a la Probabilitat

Atenció: és necessari definir prèviament els successos rellevants. Totes les variables aleatòries utilitzades s'han d'especificar amb claredat. Els càlculs hauran d'estar degudament justificats.

1. **(25%)** Sigui X una variable aleatòria absolutament contínua amb funció de densitat de probabilitat igual a

$$f(x) = \beta e^{-\alpha x} \quad \text{si} \quad x \in \mathbb{R}^+ \quad \text{i} \quad f(x) = 0 \quad \text{si} \quad x \notin \mathbb{R}^+$$

on $\alpha > 0$.

(a) Trobeu el valor de β per a que $f(x)$ sigui realment una densitat de probabilitat.(1.5 pts.)

Una funció de densitat d'una variable absolutament contínua és una funció no negativa tal que la integral sobre tota la recta real és igual a 1. Per tant, en el present cas, tindrem

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \beta e^{-\alpha x}\,dx = \beta\Big[-\frac{1}{\alpha}e^{-\alpha x}\Big]_0^{\infty} = \beta\Big(\lim_{x\to\infty} -\frac{1}{\alpha}e^{-\alpha x} + \frac{1}{\alpha}\Big) = \frac{\beta}{\alpha}$$

d'aquí es conclou que $\beta = \alpha$.

(b) Trobeu la funció de distribució de X i la funció de distribució de $Y = X^3$.(3.5 pts.)

La funció de distribució de X és

$$F_X(x) = P(X \le x) = \begin{cases} 0 & x \le 0 \\ \displaystyle\int_0^x \alpha e^{-\alpha t}\,dt = \big[-e^{\alpha t}\big]_0^x = 1 - e^{-\alpha x} & x > 0 \end{cases}$$

La funció de distribució de Y és

$$F_Y(y) = P(Y \le y) = P(X^3 \le x) = P(X \le \sqrt[3]{y}) = F_X(\sqrt[3]{y}) = \begin{cases} 0 & y \le 0 \\ 1 - e^{-\alpha\sqrt[3]{y}} & y > 0 \end{cases}$$

(c) Trobeu la funció de densitat de probabilitat de Y.(2.5 pts.)

La funció de densitat d'una variable absolutament contínua coincideix amb la derivada de la funció de distribució en tots els punts en que aquesta sigui derivable i podem definir-la arbitràriament igual a zero en tots els altres. Per tant

$$f_Y(y) = F_Y'(y) = \begin{cases} 0 & y < 0 \\ \dfrac{\alpha}{3} y^{-\frac{2}{3}} e^{-\alpha\sqrt[3]{y}} & y > 0 \end{cases}$$

y $f_Y(0) = 0$.

Una altra possibilitat consisteix en aplicar la fórmula del canvi de variable entre funcions de densitat, per a $Y = h(X)$, amb h que tingui inversa diferenciable, i resulta

$$f_Y(y) = f_X(h^{-1}(y)) \left|\frac{dh^{-1}(y)}{dy}\right|$$

en el nostre cas $h(x) = x^3$ y $h^{-1}(y) = \sqrt[3]{y}$ definida en tota la recta real, i diferenciable en tots els punts menys $y = 0$ (en aquest punt la funció de densitat podem definir-la arbitràriament igual a 0). Per tant tindrem

$$f_Y(y) = \alpha e^{-\alpha\sqrt[3]{y}} \frac{1}{3} y^{-\frac{2}{3}} = \frac{\alpha}{3} y^{-\frac{2}{3}} e^{-\alpha\sqrt[3]{y}} \quad \text{para} \quad y > 0$$

i

$$f_Y(y) = 0 \quad \text{para} \quad y \le 0$$

com hem vist anteriorment.

(d) Calculeu $\mathrm{E}(Y)$ i $\mathrm{var}(Y)$.(2.5 pts.)

Necessitem els moments de primer i segon ordre de Y, però com $Y = X^3$, aquests equivalen als moments de tercer i sisè ordre de X. El moment d'ordre k de X, fent el canvi $t = \alpha x\ dt = \alpha dx$, és

$$E(X^k) = \int_0^{\infty} x^k \alpha e^{-\alpha x}\,dx = \frac{1}{\alpha^k}\int_0^{\infty} (\alpha x)^k \alpha e^{-\alpha x}\,dx = \frac{1}{\alpha^k}\int_0^{\infty} t^k e^{-t}\,dt = \frac{\Gamma(k+1)}{\alpha^k}$$

Per tant

$$E(Y) = E(X^3) = \frac{\Gamma(4)}{\alpha^3} = \frac{3!}{\alpha^3} = \frac{6}{\alpha^3} \quad \text{y} \quad E(Y^2) = E(X^3) = \frac{\Gamma(7)}{\alpha^6} = \frac{6!}{\alpha^6} = \frac{720}{\alpha^6}$$

i

$$\mathrm{var}(Y) = E(Y^2) - E(Y)^2 = \frac{720}{\alpha^6} - \frac{36}{\alpha^6} = \frac{684}{\alpha^6}$$

També podríem haver obtingut els moments de Y a partir de

$$E(Y^k) = \int_0^\infty y^k \alpha \frac{1}{3} y^{-\frac{2}{3}} e^{-\alpha \sqrt[3]{y}} y^{-\frac{2}{3}}\, dy$$

que es resol amb el canvi $t = \alpha \sqrt[3]{y}$, $dt = \alpha \frac{1}{3} y^{-\frac{2}{3}} dy$, obtenint

$$E(Y^k) = \int_0^\infty \left(\frac{t}{\alpha}\right)^{3k} e^{-t} dt = \frac{1}{\alpha^{3k}} \int_0^\infty t^{3k} e^{-t} dt = \frac{\Gamma(3k+1)}{\alpha^{3k}}$$

i per tant

$$E(Y) = \frac{\Gamma(4)}{\alpha^3} = \frac{6}{\alpha^3} = \quad \text{y} \quad E(Y^2) = \frac{\Gamma(7)}{\alpha^6} = \frac{720}{\alpha^6}$$

com hem vist abans.

2. *OnOff Lights* és una empresa especialitzada en il·luminació amb leds,i fabrica un model concret de làmpada que incorpora 18 leds de 0.25 watts de potència. Es coneix molt bé com es comporta aquest tipus de led: per exemple, es sap que el 96% d'ells tenen una durada superior a 1500 hores.

   (a) Suposeu que el temps de vida útil d'un led es distribueix exponencialment. Amb la informació precedent, podeu trobar quina és la durada mitjana esperada d'un led?(2 pts.)

   Sigui L la durada d'un led, amb distribució Exp(λ). Es sap que: $P(L > 1500) = 0.96$. Expressant això amb la funció de distribució de L,

   $$1 - F_L(1500) = 0.96, \quad exp(-\lambda 1500) = 0.96, \quad \lambda = -log(0.96)/1500 = 2.721466 \cdot 10^{-5}$$

   Com el valor esperat de L és la inversa del paràmetre λ, durada mitjana = 36745 hores.

   (b) Quan més de 2 leds fallen, la làmpada pateix d'una sensible disminució de potència lumínica que no la fa pràctica. Quina és la probabilitat que passi això amb una làmpada que porta funcionant 1500 hores? Definiu amb precisió les variables aleatòries que es necessiten per respondre.(2 pts.)

   El nombre de leds que han fallat després de 1500 hores a una làmpada concreta es distribueix com una Binomial amb paràmetres $n = 18$ i $p = 0.04$, suposant que les fallides són independents d'uns leds i altres:

   $Y \sim B(18, 0.04)$. La probabilitat demanada és $P(Y > 2) = 1 - F_Y(2) = 1 - (P_Y(0) + P(Y(1) + P_Y(2))$.

   $$P(Y > 2) = 1 - (0.96^{18} + 18 \cdot 0.04 \cdot 0.96^{17} + 153 \cdot 0.04^2 \cdot 0.96^{16}) = 1 - (0.4796 + 0.3597 + 0.1274) = 0.0333$$

   (c) Assumint de nou que la durada d'un led és exponencial (amb mitjana 40000 hores, si no heu solucionat l'apartat 1), repetiu l'apartat anterior per una làmpada que porti funcionant 3000 hores. (2 pts.)

   La clau és calcular la proporció de fallides per a un led funcionant 3000 hores, perquè després d'aquest temps el nombre de leds que ja no funcionen es distribueix diferent (encara que segueix sent una binomial, la p és distinta). $p = P(L < 3000) = 1 - exp(-\lambda 3000)$:

   - Si $\mu = 40000$, $p = 0.07226$
   - Si $\mu = 36745$, $p = 0.07840$

   Per tant, es calcula $P(W > 2)$, on $W \sim B(n = 18, p)$:

   - Si $\mu = 40000$, 0.1367
   - Si $\mu = 36745$, 0.1631

   Després de 1500 hores, una de cada 30 làmpades ja no il·lumina bé; després de 3000 hores és una de cada 6!

   (d) *OnOff Lights* disposa d'un laboratori on pot col·locar 100 làmpades i observar com van fallant. Doneu un model de probabilitat per la variable "Nombre de làmpades que després de 1500 hores de funcionament ja no són pràctiques[1]". Quin seria el valor esperat i la desviació típica? Algun model alternatiu?(2 pts.)

   Sigui T la variable anterior. Podem veure que és una altra vegada una variable Binomial, suposant que les làmpades del laboratori són independents, amb $n = 100$ i p la probabilitat que a una làmpada amb 1500 hores de funcionament hagin fallant almenys 3 leds (0.0333).

   El valor èsperat és $E(T) = n \cdot p = 3.33$, i la desviació tipus $\sigma_T = \sqrt{n \cdot p \cdot (1-p)} = 1.7942$.

   Com que n és bastant gran i p petit, un model alternatiu seria una distribució de Poisson amb $\lambda = 3.33$.

[1]... perquè la seva potència lumínica ha baixat sensiblement.

(e) Com podeu fàcilment comprovar, a l'inici de l'experiment, quan tots els leds funcionen, es consumeix una potència elèctrica igual a 450 watts. Però quan ja han transcorregut 1500 hores, quina és la potència consumida? Expresseu el resultat amb una distribució apropiada de probabilitat (pot ser aproximada), i trobeu la probabilitat que el consum sigui inferior a 410 watts.(2 pts.)

El consum elèctric al principi és 450 W perquè tots els leds (1800) funcionen i cadascú consumeix 0.25W. Després de 1500 hores el nombre de leds que funcionen depén de l'atzar, però es pot expressar com. $N = 1800 - \sum Y_i, i = 1...100$, és a dir, 1800 leds restant la suma de tots els que s'han espatllat a cada làmpada (Y_i).

La suma es pot aproximar pel TCL a una Normal, recordant que $E(Y_i) = 18 \cdot 0.04 = 0.72$ i $V(Y_i) = 18 \cdot 0.04 \cdot 0.96 = 0.6912$. Llavors: $\sum Y_i \approx N(\mu = 72, \sigma^2 = 69.12)$.

La potencia consumida P és simplement $0.25N$: per tant $P \approx N(0.25(1800-72), \sigma = \sqrt{0.25^2 69.12}) = N(432, 2.078)$

El nombre total de leds que fallaran estarà al voltant de 72. El consum serà de aproximadament 432 W, amb una desviació de 2 W. Es molt poc probable que sigui inferior a 410 W:

$$P(P < 410) = P(Z < (410-432)/2.078) = P(Z < -10.6) \approx 0$$

3. **(40%)** El nivell en sèrum de 1,25 dihidroxivitamina D (1,25DD) en noies adolescents es distribueix segons una distribució Normal de mitjana $\mu = 65$ i desviació típica $\sigma = 12.5$. El nivell de 1,25DD pot tenir una influència en el desenvolupament de certes anomalies com ara la osteoporosi.

(a) Quina és la probabilitat que una noia escollida a l'atzar de la població tingui un nivell de 1,25DD inferior a 40? (1 pt.)

Anomenarem:

$Z \sim N(0,1)$

X_D = nivell de sèrum dones $\sim N$ (65,12.5)

$$P(X_D < 40) = P(\frac{X_D - 65}{12.5} < \frac{40-65}{12.5}) = P(Z < -2) = 1 - P(Z \le 2) = 0.02275$$

(b) Quin és l'interval central en el que es troba el 95 % de la concentració de 1,25DD en la població de noies?(1.5 pts.)

$$P(65 - k < X_D < 65 + k) = 0.95 \Rightarrow P(\frac{X_D - 65}{12.5} < \frac{65 + k - 65}{12.5}) = 0.975$$

$$P(Z < \frac{k}{12.5}) = 0.975 \Rightarrow \frac{k}{12.5} = 1.96 \Rightarrow k = 24.5$$

$$P(40.5 < X_D < 89.5) = 0.95$$

(c) Si la població total està formada per un 60 % de noies i la resta són nois, i en els nois la distribució de la 1,25DD és una Normal de mitjana $\mu = 70$ i desviació típica $\sigma = 12.5$, quina és la probabilitat que triat un individu a l'atzar de la població, el seu nivell de 1,25DD sigui superior a 40?(1.5 pts.)

Definim:

X= nivell sèrum individu

X_H = nivell de sèrum homes $\sim N$ (70,12.5)

$$P(X_H > 40) = P(\frac{X_H - 70}{12.5} > \frac{40-70}{12.5}) = P(Z > -2.4) = P(Z \le 2.4) = 0.9918$$

Sabem

$$P(X_D > 40) = 1 - P(X_D \le 40) = 1 - P(X_D < 40) = 1 - 0.02275 = 0.97725$$

$$P(X > 40) = P(X_H > 40) * P(H) + P(X_D > 40) * P(D) = 0.9918 * 0.4 + 0.97725 * 0.6 = 0.98307$$

(d) Quina és la probabilitat que en una mostra de 20 noies, almenys 2 presentin un valor de 1,25DD inferior a 40?(1.5 pts.) Sigui

Y= Nombre de noies de les 20 que tenen un valor de 1,25DD inferior a 40

$$Y \sim B(n = 20, p = 0.02275)$$

$$P(Y \ge 2) = 1 - P(Y < 2) = 1 - P(Y = 0) - P(Y = 1) = 0.075$$

(e) Considerem una mostra de 50 noies escollides a l'atzar, quina és la probabilitat que com a màxim 5 d'elles tinguin una concentració de 1,25DD inferior a 40?(1.5 pts.)

Sigui

W= Nombre de noies de les 50 que tenen un valor de 1,25DD inferior a 40

$W \sim B(n = 50, p = 0.02275) \approx Pois(\lambda = 1.1375)$

$$P(W \leq 5) = \sum_{i=0}^{5} e^{-\lambda} \frac{\lambda^i}{i!} = 0.998853$$

(f) Es sabut que en les noies que presenten un valor de 1,25DD inferior a 40 la probabilitat de desenvolupar osteoporosi és del 15 %, mentre que en les que presenten un valor superior tan sols és del 5 %. Quina és la probabilitat que una dona que ha desenvolupat osteoporosi tingués un nivell de 1,25DD inferior a 40?(1.5 pts.)

$P(\text{Osteoporosi}) = P(\text{Osteoporosi}|X_D < 40) * P(X_D < 40) + P(\text{Osteoporosi}|X_D \geq 40) * P(X_D \geq 40) =$

$0.15 * 0.02275 + 0.05 * 0.97725 = 0.052275$

$$P(X_D < 40|\text{Osteoporosi}) = \frac{P(X_D < 40 \bigcap \text{Osteoporosi})}{P(\text{Osteoporosi})} = \frac{0.15 * 0.02275}{0.052275} = 0.06528$$

(g) Escollim ara a l'atzar un noi i una noia, quina és la probabilitat que la concentració de 1,25DD de la noia sigui superior a la del noi?(1.5 pts)

Hem de calcular: $P(X_D > X_H) = P(X_D - X_H > 0)$

Com X_D i X_H són v.a. independents amb distribucions Normals

$E(X_D - X_H) = E(X_D) - E(X_H) = 65 - 70 = -5$

$Var(X_D - X_H) = Var(X_D) + Var(X_H) = (12.5)^2 + (12.5)^2 = 2 * (12.5)^2$

Aleshores sabem que:

$X_D - X_H \sim N(-5, (12.5)\sqrt{2})$

$$P(X_D > X_H) = P(X_D - X_H > 0) = P(\frac{X_D - X_H + 5}{(12.5)\sqrt{2}} > \frac{5}{(12.5)\sqrt{2}}) =$$

$P(Z > 0.2828) = 1 - P(Z \leq 0.2828) = 0.3897$