Problema 1. (Tots els apartats puntuen igual)

Es disposa de tres urnes, la 1ª conté 4 boles numerades de l'1 al 4, la 2ª conté les boles parell (2 i 4) i una 3ª amb els nombres senars (1 i 3).. Les probabilitats de cada extracció són:

	1ª Urna	**2ª Urna**	**3ª Urna**
Probabilitat surti 1	0.2	--	0.5
Probabilitat surti 2	0.1	0.4	--
Probabilitat surti 3	0.3	--	0.5
Probabilitat surti 4	0.4	0.6	--

S'extreu aleatòriament una bola de cada urna.

A1. Indiqueu el conjunt de resultats i les seves probabilitats

Ω	probabilitat	Event A	Suma	Event B
121	0,04	a	4	
123	0,04		6	
141	0,06	a	6	
143	0,06		8	
221	0,02	a	5	
223	0,02	a	7	b
241	0,03		7	b
243	0,03		9	
321	0,06		6	
323	0,06	a	8	
341	0,09		8	
343	0,09	a	10	
421	0,08		7	b
423	0,08		9	
441	0,12	a	9	
443	0,12	a	11	

A2.- Justifiqueu si els resultats de les tres urnes són independents o no.

Si que són independents perquè el resultat en una extracció no influeix en les probabilitats de les altres extraccions

A3. Definim els esdeveniments A= "Parell de xifres repetides" i B= "Suma de les xifres igual a 7". Quina es la probabilitat de l'event A? I de l'event B? Probabilitat de l'event B sabent que a la primera extracció ha sortit un nombre parell?

P(A)=0,53

P(B)=0,13

P(B|1a extracció ha sortit parell) =

=(0.02+0.03+0.08)/(0,02+0,02+0,03+0,03+0,08+0,08+0,12+0,12)=0.13/0.5=0.26

(càlcul de la probabilitat de B, pels subconjunt d'extraccions en que coincideix que la 1a extracció ha sortit parell)

B.

En una capsa hi ha 5000 xips; 1000 provenen de la fàbrica X que té una taxa del 10% de defectuosos, 4000 provenen de la fàbrica Y que té una taxa del 5% de defectuosos. S'escull un xip a l'atzar, i és defectuós.

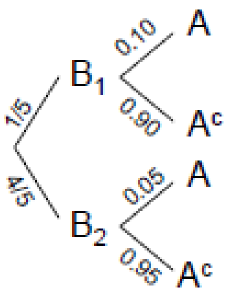
B1.- Representeu l'arbre de l'experiència aleatòria indicant-ne el conjunt de resultats i les seves probabilitats

Definim els successos:

A=xip defectuós

Bi=prové de la fàbrica i, on i=1 fàbrica X

i=2 fàbrica Y



B2. Calculeu la probabilitat que l'hagi fabricat X.

<u>A PRIORI</u>: $P(B_1)=1/5 \quad P(B_2)=4/5$

$P(A|B_1)=0.10 \quad P(A|B_2)=0.05$

$P(A^c|B_1)=0.90 \quad P(A^c|B_2)=0.95$

<u>A POSTERIORI</u>:

$$P(B_1|A) = \frac{P(A|B_1)\cdot P(B_1)}{P(A|B_1)\cdot P(B_1)+P(A|B_2)\cdot P(B_2)} =$$

$$= 0.02/(0.02+0.04) = \mathbf{1/3}$$

C.

S'han recollit dades de l'edat i sou (euros al mes) de 19 titulats en informàtica a la FIB. Aquest és el resultat de l'estadística descriptiva.

```
#Descriptiu univariant de la variable Sou.
> summary(dat$sou)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    400     750    1105    2378    2650   12040
#Desviació estandard
> sd(dat$sou)
[1] 2944.3
#Correlació entre Edat i Sou
> cor(dat$sou,dat$edat)
[1] 0.79
```

C1. Indica quin és el valor numèric dels dos indicadors de tendència central de la variable sou? Què volen dir aquests valors?

Mitjana=2378
Mediana=1105
La mitjana i la mediana mostren una centralitat entorn 2378 i 1105 €/mes respectivament. La mitjana és més elevada, ja que és sensible als valors extrems (sous molt alts, màxim de 12040).

C2. Quins són els dos indicadors de dispersió? Feu la interpretació d'aquest valors.

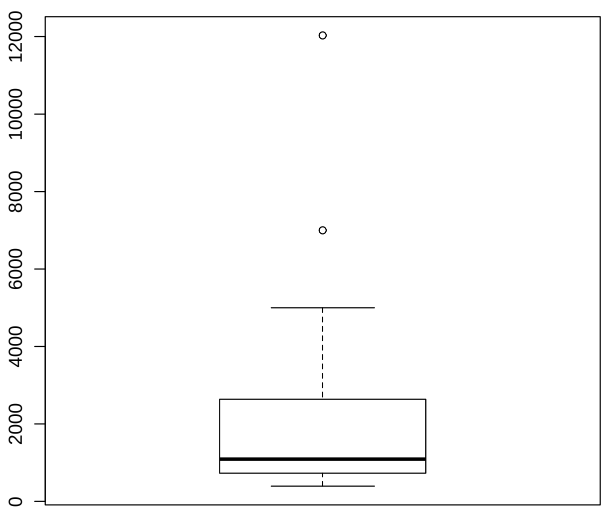
Desviació estàndard=2944.3
Rang interquartílic= 2650-750=1900
La desviació estàndard india una dispersió alta (2944 €) al voltant de la mitjana. El rang interquartílic també mostra que hi ha una dispersió elevada, però és menys sensibles a valors extrems que la SD.

C3. La correlació entre Sou i Edat és de 0.79. Com és la relació entre les 2 variables? Comenta quines són la magnitud i la direcció i què volen dir.

Relació positiva i forta.
Magnitud gran, propera a 1. La direcció és creixent (signe positiu), quan una variable creix l'altre també.
A més edat més sou.

C4. Creus que la distribució de la variable sou és simètrica? Dibuxa el boxplot que t'imagines per a aquesta variable.

No es simètrica, mitjana i mediana estan força allunyades.



Problema 2

Al llarg dels anys s'ha observat que el nombre d'exercicis setmanals d'estatus diferents que fan els estudiants d'una determinada assignatura fora de les classes segueix la següent distribució de probabilitat:

Taula 1: Distribució del 'Nombre d'exercicis d'estatus diferents realitzats a la setmana fora de clase'

$X = x$	0	1	2	3	4	5	≥ 6
$\mathrm{P}(X = x)$	0.1	0.25	0.25	0.15	0.15	0.1	0

a) (1 punt)
Quina és la funció de distribució de la variable X?

Solució:

$X = x$	0	1	2	3	4	5	≥ 6
$\mathrm{P}(X = x)$	0.1	0.25	0.25	0.15	0.15	0.1	0
$\mathrm{F}(x) = \mathrm{P}(X \leq x)$	0.1	0.35	0.6	0.75	0.9	1	1

Anomenarem X a la variable aleatòria 'Nombre setmanal d'exercicis diferents d'estatus realitzats fora de classe'.

b) (1 punt)
Quina és la probabilitat que un alumne faci entre 2 i 4 exercicis diferents a la setmana?

Solució:

$$\mathrm{P}(2 \leq X \leq 4) = \mathrm{P}(1 < X \leq 4) = \mathrm{F}(4) - \mathrm{F}(1) = 0.9 - 0.35 = 0.55.$$

c) (1,5 punt)
I quina és aquesta probabilitat si sabem que l'alumne ha fet al menys dos exercicis?

Solució:

$$\mathrm{P}(2 \leq X \leq 4 | X \geq 2) = \frac{\mathrm{P}(2 \leq X \leq 4)}{\mathrm{P}(X \geq 2)} = \frac{0.55}{1 - \mathrm{P}(X \leq 1)} = \frac{0.55}{0.65} = 0.85.$$

d) (2,5 punts)
Quin és el valor esperat i quina és la desviació estàndard d'X? Com s'interpreten aquests valors?

Solució:

$$\begin{aligned}
\mu_X &= 0 \cdot 0.1 + \cdots + 5 \cdot 0.1 = 2.3, \\
\sigma_X &= \sqrt{(0 - 2.3)^2 \cdot 0.1 + \cdots + (5 - 2.3)^2 \cdot 0.1} \\
&= \sqrt{0^2 \cdot 0.1 + \cdots + 5^2 \cdot 0.1} - 2.3^2 = \sqrt{2.21} = 1.49.
\end{aligned}$$

Després de "moltes" setmanes, la mitjana d'exercicis fets per alumne seria un valor a prop de 2.3 amb desviació estàndard a prop de 1.49.

e) (1 punt)
Per quines raons es podria suposar que X segueix una distribució de Poisson?

Solució:
Per una banda, X recull el nombre d'events durant un interval de temps i per altra banda tenim: $\mu_X \approx \sigma_X^2$.

En total hi ha 90 alumnes que fan l'assignatura i suposem que cadascun utilitza l'estatus independentment dels altres. Definim llavors una nova variable: X_{90} 'Nombre setmanal d'exercicis diferents d'estatus realitzats fora de classe per 90 alumnes'.

f) (2 punts)
Quins són el valor esperat i la desviació estàndard d'X_{90}?

Solució:

$$X_{90} = \underbrace{X + X + \cdots + X}_{\text{90 alumnes}},$$

on X segueix la distribució de la Taula 1. Per tant,

$$\mathrm{E}(X_{90}) = \sum_{i=1}^{90} \mathrm{E}(X) = 90 \cdot 2.3 = 207$$

$$\sigma_{X_{90}} \overset{\text{Indep.}}{=} \sqrt{\sum_{i=1}^{90} \mathrm{Var}(X)} = \sqrt{90 \cdot 2.21} = 14.1.$$

g) (1 punt)
Sigui Y la variable 'Setmanes que falten per a l'examen'. Si suposem que X canvia en funció d'Y, quin signe seria d'esperar que tingués la correlació entre X i Y. Per què?

Solució:
És molt probable que es facin més exercicis quan menys falta per l'examen. Per tant, seria d'esperar que el signe de la correlació fos negatiu.

Problema 3.

El servei d'informàtica d'un departament disposa d'una funcionalitat per recuperar continguts des d'un backup actualitzat cada dia. Eventualment, els usuaris del departament sol·liciten un backup, i s'ha observat que el procés pel qual es precisa de la funcionalitat *no té memòria*: és independent de l'usuari i del moment en que es demana. Se sap també que una quarta part dels usuaris sol·licita un backup en un període inferior a tres mesos des de la darrera petició.

Anomenarem T a la variable aleatòria "temps entre backups successius per al mateix usuari, en mesos".

1. Justifiqueu que T segueix una llei exponencial de paràmetre λ i calculeu quant val λ. (1 pt)
   La sol·licitud de backups es descriu com un procés aleatori, eventual (no és gaire freqüent), independent en el temps i entre usuaris, i amb una taxa d'ocurrència fixada, per tant, correspon a l'anomenat procés de Poisson. I el temps entre dos events correspon en conseqüència a una distribució Exponencial. El paràmetre λ es pot calcular per la información proporcionada:
   $P(T < 3) = 0.25 = 1 - \exp(-\lambda \cdot 3)$
   $\lambda = -1/3 \log(1-0.25) = 0.0959$. És una taxa d'events mensual.

Si no heu pogut trobar λ, pels següents apartats assumiu el valor 0.1.

2. Trobeu la desviació estàndard de T. (0.5 pt)
   $V(T) = 1 / \lambda^2$, i l'arrel cuadrada = 10.428 [mesos]

3. Calculeu la probabilitat que un usuari trigui entre 6 i 12 mesos a sol·licitar un backup. (1 pt)
   $P(6 < T < 12) = F_T(12) - F_T(6)$
   $= (1- \exp(-\lambda \cdot 12)) - (1- \exp(-\lambda \cdot 6)) = \exp(-\lambda \cdot 6) - \exp(-\lambda \cdot 12)$
   $= 0.24609$

4. Quina seria la probabilitat que un usuari demani en un any almenys dos backups? Amb quin model de probabilitat heu hagut de treballar en aquesta qüestió? (1.5 pt)
   Per aquesta qüestió hem de treballar amb una variable aleatòria que compti el nombre de backups demanats en un any, amb distribució de Poisson. S'ha de trobar la taxa que correspon a un periode de 1 any:
   $N \sim P(\lambda_A)$; $\lambda_A = 12 \cdot \lambda = 12 \cdot 0.0959 = 1.1507$ [mitjana d'events per any i persona]
   $P(N \geq 2) = 1 - P(N < 2) = 1 - F_N(1)$
   Amb les taules, podem prendre un valor intermedi per les taxes 1.1 i 1.2:
   $F_N(1) \approx (0.699 + 0.663)/2 = 0.681$ $P(N \geq 2) \approx 0.319$
   El càlcul exacte també és possible perquè solament s'ha de calcular $1-(P_N(0)+P_N(1)) = 0.319496$

5. Si considerem 8 usuaris (independents entre sí), quina és la probabilitat que com a molt un d'ells necessiti del backup després d'un període superior a 12 mesos? Com abans, expliciteu els models utilitzats. (1.5 pt)
   La variable M ens dirà quants dels 8 usuaris trigaran més d'un any a demanar un backup. Segueix una distribució Binomial B(8, p), perquè són independents entre ells.
   p val $P(T>12) = \exp(-\lambda \cdot 12) = 0.3164$
   $P(M \leq 1) = P_M(0)+P_M(1) = (1-p)^8 + 8(1-p)^7 p = 0.224257$
   Amb taules es pot respondre aproximadament, prenent la p més propera 0.3:
   $P(M \leq 1) \approx$ Binomial(n=8, p=0.3, k=1) = 0.2553, que té un error no despreciable.

Si tenim 50 usuaris independents del departament, doneu:

6. la distribució, valor esperat i desviació estàndard de la variable "nombre de backups demanats en 12 mesos". (1 pt)

És un model de Poisson, i la taxa λ_{50} és $50\lambda_A$ = 57.54
El valor esperat és de 57.54 backups.
La desviació estàndard és l'arrel de 57.54: 7.59 backups.

7. la distribució, valor esperat i desviació estàndard de la variable "nombre d'usuaris que precisen alguna vegada un backup durant un any". (1 pt)
   diguem L aquesta variable. Segueix un model binomial; la n és 50, i la p la probabilitat $P(N>0) = P(T<12) = 1 - 0.3164 = 0.6836$.
   El valor esperat és de 50 0.6836 = 34.18 usuaris.
   La desviació estàndard és l'arrel de 50 0.6836 0.3164 = $\sqrt{10.815}$ = 3.2886 usuaris.

8. la probabilitat que, d'aquests 50, 40 o més precisin algun backup durant un any. Justifiqueu totes les passes de la resposta. (1.5 pt)
   Tindriem que trobar $P(L \geq 40)$.
   Com no podem usar les taules, i no és fàcil fer al càlcul a ma, ens plantegem utilitzar una aproximació.
   $L \sim B(50, 0.6836)$. Com que n és prou gran i la p no és molt petita (ni molt propera a 1), proposem la Normal com a distribució aproximada.
   $L \approx N(\mu=34.18, \sigma=3.2886)$
   $P(L \geq 40) \approx P(Z \geq (40 - 34.18)/ 3.2886) = P(Z > 1.77) = 1 - F_Z(1.77) = 1 - 0.9616$
   $= 0.0384$

9. Representeu gràficament la distribució del punt 7, amb la probabilitat del punt 8. (1 pt)