UNIVERSITE DE LIEGE
INSTITUT D'ELECTRICITE MONTEFIORE

INFORMATIQUE
(systèmes et modélisation)
Professeur BANH Tri An

SART TILMAN, B28
B-4000 Liège - (Belgique)

☎ 041-66.27.09
Telefax : 66.29.08
PARKING 32

Liège, le 26 mai 1995.

EXAMEN ECRIT - Cours de "Calcul des Probabilités".

Partie Théorie (à livres fermés).

L'étudiant mentionnera, à la partie supérieure de la feuille, de gauche à droite : son numéro d'ordre dans la liste de 2CI, 2CI(AR), 2C. Inf., le numéro de la question (à encadrer), ses nom et prénom.

Les réponses aux questions de l'examen de théorie doivent être brèves, claires et bien structurées.

---

1) Définir l'écart-type.
2) Énoncer le théorème des probabilités composées
3) Énoncer le théorème de BAYES
4) Définir la distribution de χ^2 (Helmert-Pearson)
5) Définir la puissance d'un test
6) Définir la fonction de puissance d'un test
7) Donner l'équation de la droite de régression de y en x par la méthode des moindres carrés.
8) Donner l'équation de la droite des moindres rectangles
9) Définir le coefficient de corrélation
10) Définir la notion d'entropie et d'entropie conditionnelle
11) Définir le concept d'information
12) Énoncer et démontrer la propriété de "l'information mutuelle des deux expériences α et β"

26 mai 1995

Exercice 1 -

Dans une entreprise, la durée des communications téléphoniques des employés est en moyenne de 5 minutes et suit une loi exponentielle.

1) Sachant qu'un employé occupe sa ligne depuis au moins 5 minutes, quelle est la probabilité que sa communication dure au moins trois minutes de plus.

$$P(X>8/X>5)=\frac{P(X>8)}{P(X>5)}=\frac{1-P(X\le 8)}{1-P(X\le 5)}$$

Comme $P(X\le x)=\int_0^x \lambda e^{-\lambda t}dt = 1-e^{-\lambda x}$,

on a $P(X>8/X>5)=\frac{e^{-8\lambda}}{e^{-5\lambda}}=e^{-3\lambda}$ où $\lambda=\frac{1}{m}=0,2$

Donc $P(X>8/X>5)=e^{-0,6}=0,5488\cong 55\%$

2) Suite à une recommandation quant à la durée des communications, on voudrait tester si la durée moyenne des communications a significativement diminué, avec un risque d'erreur de 10%. Voici les observations relevées sur un échantillon aléatoire de 200 communications, passées après cette recommandation.

Durée en minutes	Nombre de communications observées
de 0 à 2 au plus	49
de 2 à 4 au plus	62
de 4 à 6 au plus	46
de 6 à 8 au plus	19
de 8 à 10 au plus	16
de 10 à 20 au plus	8

Soit H_0 = la durée moyenne n'a pas changé, c'est-à-dire $m=5$

et H_1 = la durée moyenne a diminué c'est-à-dire $m<5$. (test unilatéral)

Comme n=200 (>30), la distribution d'échantillonnage de la moyenne $\overline{X}$ est normale, de moyenne m et d'écart-type $\sigma/\sqrt{n}$, sous l'hypothèse H_0. On estime σ par l'écart-type corrigé de l'échantillon car n>30.
Sur l'échantillon, les calculs conduisent, en prenant les centres de classe égaux à 1, 3, 5, 7, 9 et 15 aux résultats suivants :
$\overline{X}$ = 4,31 min et s_{NC}= 3,216 min.

La valeur estimée de $\sigma=s=\sqrt{\frac{n}{n-1}}s_{NC}=3,22$ min.

Par suite $P(\overline{X}<X_{crit})=0,10$ conduit à la valeur critique, dans la loi normale centrée réduite, égale à $z_c=-1,28$ donc $X_{crit}=5-1,28\frac{3,22}{\sqrt{200}}=4,63$ min.

Comme $\overline{X}=4,31<4,63$, on rejette H_0 et on accepte l'hypothèse que la moyenne a significativement diminué, avec un risque d'erreur de 10%.

3) On veut tester s'il est raisonnable d'ajuster u à 12 au pluscette série d'observations avec un risque d'erreur de 5%

La nouvelle loi exponentielle serait de ddp $\lambda e^{-\lambda x}$ avec $\lambda = \frac{1}{\overline{X}} = \frac{1}{4,31} = 0,232.$

La fréquence relative de la première classe est donnée par

$f_1 = P(0 < X \leq 2) = \int_0^2 \lambda e^{-\lambda x} dx = 1 - e^{-2\lambda} = 1 - 0,6287 = 0,3713$

La fréquence relative de la deuxième classe est donnée par

$f_2 = P(2 < X \leq 4) = \int_2^4 \lambda e^{-\lambda x} dx = e^{-2\lambda} - e^{-4\lambda} = e^{-2\lambda}\left(1 - e^{-2\lambda}\right) = e^{-2\lambda} f_1 = 0,2334$

Celle de la 3ème classe :

$f_3 = P(4 < X \leq 6) = e^{-4\lambda}\left(1 - e^{-2\lambda}\right) = e^{-2\lambda} f_2 = 0,1467$

Celle de la 4ème classe :

$f_4 = P(6 < X \leq 8) = e^{-6\lambda}\left(1 - e^{-2\lambda}\right) = e^{-2\lambda} f_3 = 0,0923$

Celle de la 5ème classe :

$f_5 = P(8 < X \leq 10) = e^{-8\lambda}\left(1 - e^{-2\lambda}\right) = e^{-2\lambda} f_4 = 0,0580$

Enfin, pour la dernière classe :

$f_6 = P(X > 10) = e^{-10\lambda} = e^{-2,32} = 0,0983$

Le tableau des effectifs théoriques est alors :

de 0 à 2 au plus	74,26
de 2 à 4 au plus	46,68
de 4 à 6 au plus	29,34
de 6à 8 au plus	18,46
de 8 à 10 au plus	11,60
plus de 10	19,66
	200

$$\chi^2_{obs} = \frac{(74,26-49)^2}{74,26} + \frac{(46,68-62)^2}{46,68} + \frac{(29,34-46)^2}{29,34} + \frac{(18,46-19)^2}{18,46} + \frac{(11,60-16)^2}{11,60} + \frac{(19,66-8)^2}{19,66}$$

$= 31,68$

Or, le nombre de degrés de liberté étant $\nu = 6 - 1 - 1$ (pour le paramètre λ évalué à partir de $\overline{X}$), on obtient $\chi^2_{crit} = 9,49$ pour $1 - \alpha = 0,95$. Donc $\chi^2_{obs} > \chi^2_{crit}$, on rejette l'ajustement.

Exercice 2 -

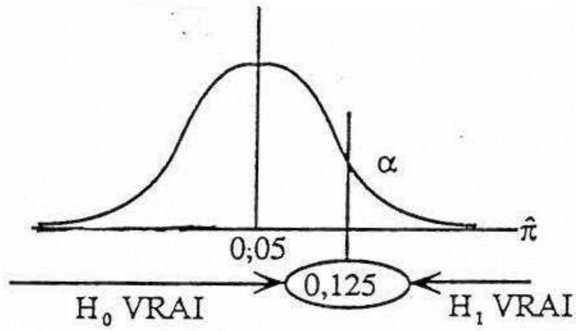
Dans la fabrication d'un composant pour ordinateurs, on veut contrôler la qualité en observant le nombre de défectueux sur 30 composants différents issus de la fabrication d'une journée.

On désire tester si la proportion des mauvais composants est supérieure à 5% au seuil de signification α=3%.

a) Établir la règle de décision.

$H_0 : \pi = 0,05 \quad H_1 : \pi > 0,05$

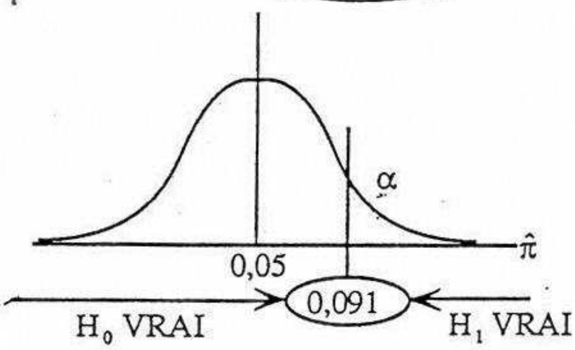
$$\hat{\pi} = f \text{ est } N\left(\pi^{*} ; \sqrt{\frac{\pi^{*}(1-\pi^{*})}{n}}\right) = N(0,05 ; 0,0398)$$



Si on a plus de 12,5% de défectueux sur 30, c'est-à-dire au moins 4 défectueux sur 30, on déclarera H_1 vrai.

b) Faire de même si on prend un échantillon de 100 composants.

Pour n=100, f est $N(0,05;0,0218)$



Si on a plus de 9,1% de défectueux sur 100, c'est-à-dire au moins 10 défectueux sur 100, on déclarera H_1 vrai.

c) Montrer que dans ce dernier cas, le test est plus puissant.

On va calculer β pour $\pi = 0{,}06\,;\,0{,}08\,;\,0{,}15$ dans les deux cas.

$n = 30$

$\pi = 0{,}06$: f est $N(0{,}06\,;\,0{,}0434)$

$\beta = Pr_N[f < 0{,}125]$

$= Pr_g[Z < 1{,}5] = 0{,}9332$

$\pi = 0{,}08$: f est $N(0{,}08\,;\,0{,}0495)$

$\beta = Pr_g[Z < 0{,}91] = 0{,}8186$

$\pi = 0{,}15$: f est $N(0{,}15\,;\,0{,}0652)$

$\beta = Pr_g[Z < -0{,}38] = 0{,}3520$

$n = 100$

$\pi = 0{,}06$: f est $N(0{,}06\,;\,0{,}0237)$

$\beta = Pr_N[f < 0{,}091]$

$= Pr_g[Z < 1{,}31] = 0{,}9049$

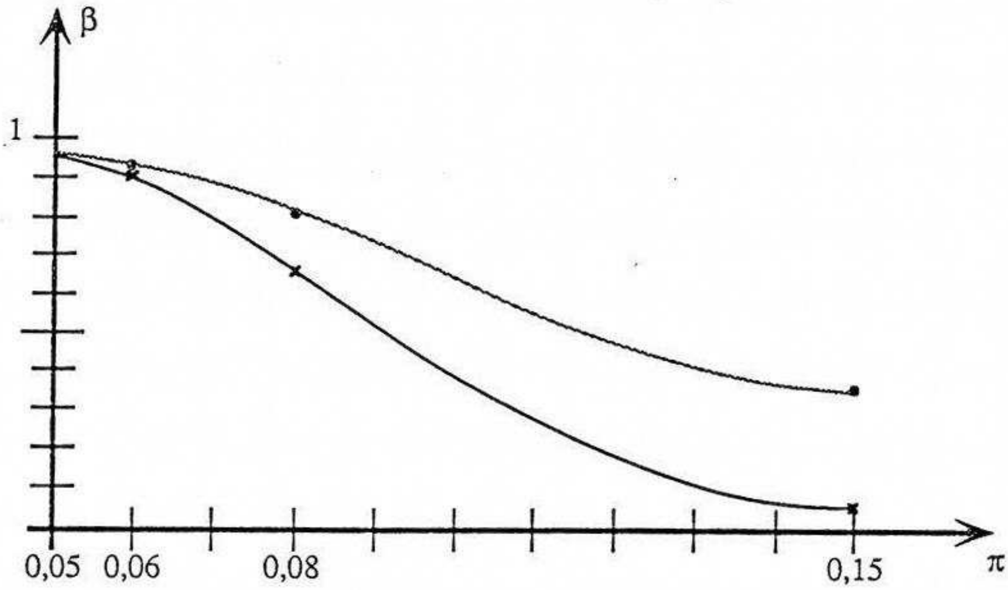
$\pi = 0{,}08$: f est $N(0{,}08\,;\,0{,}0271)$

$\beta = Pr_g[Z < 0{,}41] = 0{,}6591$

$\pi = 0{,}15$: f est $N(0{,}15\,;\,0{,}0357)$

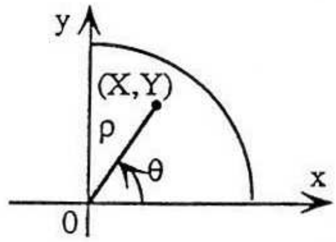
$\beta = Pr_g[Z < -1{,}65] = 0{,}0495$

Ici β toujours <



Exercice 3 -

Deux variables continues X, Y sont distribuées uniformément dans un quart de cercle de rayon R centré à l'origine.



On désigne par $\rho(X,Y)$ et $\theta(X,Y)$ les coordonnées polaires du point (X,Y).

a) Déterminer la fonction de répartition et la densité de probabilité de ρ.
   Calculer $E(\rho)$ et $ET(\rho)$.

ρ est une variable aléatoire continue prenant ses valeurs dans $]0,R[$.
Calculons $F_\rho(t)$ lorsque $t \in]0,R[$. Il vient

$$F_\rho(t) = P(\rho \le t) = P\left(\sqrt{X^2+Y^2} \le t\right) = P\left(X^2+Y^2 \le t^2\right) = P((X,Y) \in E)$$

où E désigne le cercle de rayon t centré à l'origine.

Comme les variables X,Y sont distribuées uniformément dans un quart de cercle de rayon R centré à l'origine, que nous désignons par Ω, on obtient

$$F_\rho(t) = \frac{\text{aire de } E \cap \Omega}{\text{aire de } \Omega} = \frac{\pi t^2/4}{\pi R^2/4} = \frac{t^2}{R^2}$$

De là,

$$F_\rho(t) = \begin{cases} 0 & \text{si } t \le 0 \\ \dfrac{t^2}{R^2} & \text{si } 0 < t < R \\ 1 & \text{si } t \ge R \end{cases}$$

Il s'ensuit que la densité de probabilité de ρ est donnée par

$$f_\rho(t) = \begin{cases} \dfrac{2t}{R^2} & \text{si } 0 < t < R, \\ 0 & \text{sinon.} \end{cases}$$

L'espérance de ρ vaut $E(\rho) = \int_{-\infty}^{+\infty} t\, f_\rho(t)dt = \int_0^R \frac{2t^2}{R^2} dt = \frac{2}{R^2}\left[\frac{t^3}{3}\right]_0^R = \frac{2}{3}R$

tandis que sa variance est donnée par

$$\operatorname{Var}(\rho) = \int_{-\infty}^{+\infty} t^2 f_\rho(t)dt - [E(\rho)]^2$$

$$= \int_0^R \frac{2t^3}{R^2}dt - \frac{4}{9}R^2$$

$$= \frac{2}{R^2}\left[\frac{t^4}{4}\right]_0^R - \frac{4}{9}R^2$$

$$= \frac{R^2}{2} - \frac{4}{9}R^2$$

$$= \frac{R^2}{18}$$

Par conséquent,

$$ET(\rho) = \frac{R}{3\sqrt{2}}$$

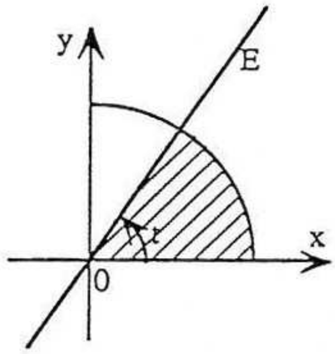
b) Montrer que θ est distribué uniformément dans $\left]0, \frac{\pi}{2}\right[$.

θ est distribuée dans $\left]0, \frac{\pi}{2}\right[$.

Calculons $F_\theta(t)$ lorsque $t \in \left]0, \frac{\pi}{2}\right[$.

Il vient $F_\theta(t) = P(\theta \le t) = P\left(\operatorname{arctg}\frac{Y}{X} \le t\right) = P\left(\frac{Y}{X} \le \operatorname{tg} t\right) = P((X,Y) \in E)$

où $E = \{(x,y) : y - x \operatorname{tg} t \le 0\}$ est le demi-plan situé en-dessous de la droite $y - x\operatorname{tg} t = 0$ passant par l'origine et faisant une angle t avec le demi-axe des x positifs.



En raisonnant comme précédemment, on obtient

$$F_\theta(t) = \frac{\text{aire de } E \cap \Omega}{\text{aire de } \Omega} = \frac{tR^2/2}{\pi R^2/4} = \frac{2t}{\pi}$$

puisque $E \cap \Omega$ est le secteur circulaire hachuré d'angle au centre t.

De là, $F_\theta = \begin{cases} 0 & \text{si } t \le 0 \\ \dfrac{2t}{\pi} & \text{si } 0 < t < \dfrac{\pi}{2} \\ 1 & \text{si } t \ge \dfrac{\pi}{2} \end{cases}$

et la densité de probabilité de θ vaut $f_\theta(t) = \begin{cases} \dfrac{2}{\pi} \text{ si } 0 < t < \dfrac{\pi}{2}, \\ 0 \text{ sinon} \end{cases}$

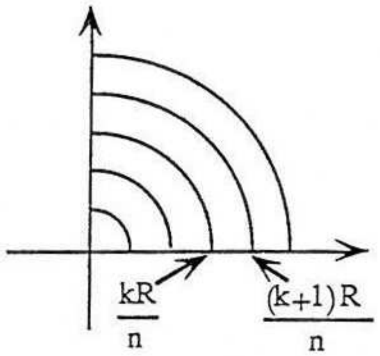
ce qui montre que θ est distribué uniformément dans $\left]0, \dfrac{\pi}{2}\right[$.

Le quart de cercle ci-dessus sert de cible pour un jeu de fléchettes. Un joueur non-chevronné lance des fléchettes au hasard. On suppose qu'il atteint la cible à chaque fois.

On lui attribue le score de $1 - \dfrac{k}{n}$ si la distance du point d'impact au point 0 appartient à l'intervalle

$\left[\dfrac{kR}{n}, \dfrac{(k+1)R}{n}\right[\qquad (k = 0,\ldots,n-1).$

c) Calculer le score moyen du joueur. Que devient-il si le joueur a la probabilité p de toucher la cible avec chaque fléchette.



Désignons par S_n le score du joueur. Il s'agit d'une variable aléatoire discrète prenant les valeurs $1 - \dfrac{k}{n}$ $(k = 0,\ldots,n-1)$ avec la probabilité

$$\begin{aligned} P\left(S_n = 1 - \frac{k}{n}\right) &= P\left(\frac{kR}{n} \le \rho < \frac{(k+1)R}{n}\right) \\ &= F_\rho\left(\frac{(k+1)R}{n}\right) - F_\rho\left(\frac{kR}{n}\right) \\ &= \frac{(k+1)^2R^2}{n^2R^2} - \frac{k^2R^2}{n^2R^2} \\ &= \frac{1}{n^2}(2k+1) \end{aligned}$$

Le score moyen vaut donc

$$
\begin{aligned}
E(S_n) &= \sum_{k=0}^{n-1}\left(1-\frac{k}{n}\right)\frac{1}{n^2}(2k+1) \\
&= \frac{1}{n^2}\sum_{k=0}^{n-1}\left(1+\left(2-\frac{1}{n}\right)k-\frac{2}{n}k^2\right) \\
&= \frac{1}{n^2}\left[n+\left(2-\frac{1}{n}\right)\sum_{k=0}^{n-1}k-\frac{2}{n}\sum_{k=0}^{n-1}k^2\right] \\
&= \frac{1}{n^2}\left[n+\left(2-\frac{1}{n}\right)\frac{n(n-1)}{2}-\frac{2}{n}\frac{n(n-1)(2n-1)}{6}\right] \\
&= \frac{1}{n^2}\left[n+\frac{1}{6}(n-1)(2n-1)\right] \\
&= \frac{(n+1)(2n+1)}{6n^2}
\end{aligned}
$$

Si le joueur a la probabilité p de toucher la cible à chaque fois, le score moyen est multiplié par p, puisque S_n prend les valeurs précédentes avec des probabilités multipliées par p ainsi que la valeur 0 avec des probabilité $1-p$.

$$\left(\text{N.B.}\quad \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}\right)$$