Université de Liège
Faculté des Sciences Appliquées
Informatique
Systèmes et modélisation
Professeur Banh Tri An

Liège, le 03/05/96

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) Enoncer l'inégalité de Bienaymé Tchebychev et interpréter les différents termes.
Enoncer le théorème de la limite centrale et la loi des grands nombres.
Signification physique.

2) Définir la notion de puissance d'un test bilatéral.
Illustrer sur une figure
Est-il possible de réduire simultanément les risques de première et deuxième espèce? Expliquer.

3) Expliquer le test de Kolmogorov-Smirnov.
En pratique, dans quel contexte l'utilise-t-on (comparer avec χ^2) ?

4) Définir les notions d'entropie, d'entropie conditionnelle et le concept d'information, en veillant à indiquer les relations entre ces différentes notions et en donnant à chaque fois l'interprétation physique correspondante.

Université de Liège
Faculté des Sciences Appliquées
Informatique
Systèmes et modélisation
Professeur Banh Tri An

Liège, le 24/05/96

EXAMEN ECRIT - Cours de "Calcul des Probabilités" - Solutions.

Exercice 1 -

Soient deux échantillons de dix éprouvettes chacun. On mesure la résistance de ces éprouvettes et on obtient les résultats suivants :

échantillon I	31,90	31,88	32,14	32,85	31,82
	32,11	31,63	31,79	31,05	31,86
échantillon II	31,51	31,30	31,50	31,81	31,66
	31,65	31,57	31,42	31,76	31,71

Peut-on admettre que ces deux échantillons proviennent de la même production, dans laquelle la résistance suit une loi normale, en acceptant un risque d'erreur de 2% ?

Solution

Il s'agit d'un test de comparaison de moyennes.
Soient l'hypothèse $H_0 : m_1 - m_2 = 0$
et l'hypothèse $H_1 : m_1 - m_2 \neq 0$ (test bilatéral).

L'écart-type σ est inconnu et la taille des échantillons est petite $(n_1 = n_2 = 10)$.
Sous l'hypothèse H_0, la loi de probabilité de $\overline{x_1} - \overline{x_2}$ est normale de moyenne $m_1 - m_2 = 0$ et d'écart-type $\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}} = \sigma.\sqrt{\dfrac{2}{n}}$.

L'estimateur non biaisé de σ^2 est $s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$ où s_1 et s_2 sont les écarts-type non corrigés des échantillons.

Calcul de $s_1^2 = 0{,}1820$ et $s_2^2 = 0{,}0228$, d'où $s^2 = 0{,}1138$ et $s = 0{,}3373$. ($\overline{x}_1 = 31{,}903$ et $\overline{x}_2 = 31{,}589$).

Comme les échantillons sont petits, la variable $\dfrac{|\overline{x_1} - \overline{x_2}|}{s\sqrt{2/n}}$ suit une loi de Student-Fisher à $\upsilon = n_1 + n_2 - 2 = 18$ degrés de liberté. Au risque $\alpha = 0{,}02$ nous obtenons dans la table $t_{0,99} = 2{,}55$.

La règle de décision est

ne pas rejeter $H_0(m_1 = m_2)$ si $|t_{obs}| < 2{,}55$,
rejeter $H_0(\text{donc } m_1 \neq m_2)$ si $|t_{obs}| \geq 2{,}55$.

Or $t_{obs} = \dfrac{\overline{x_1} - \overline{x_2}}{s\sqrt{2/n}} = \dfrac{31{,}903 - 31{,}589}{0{,}3373\sqrt{2/10}} = 2{,}08$

donc $|t_{obs}| < 2{,}55$ et on en déduit que les observations ne permettent pas d'affirmer que les deux échantillons proviennent de deux productions de moyennes différentes.

Exercice 2 -

1ère partie

On a observé les durées de vie en heures d'un transistor d'un type bien précis sur un très grand nombre de transistors venant du même fabricant.

Durée obs	Eff.
]1700;1800]	4
]1800;1900]	44
]1900;1950]	40
]1950;2000]	62
]2000;2050]	58
]2050;2100]	46
]2100;2200]	38
]2200;2300]	8

1. Réalisez le polygone des fréquences relatives non cumulées et le polygone des fréquences relatives cumulées
2. Montrez et évaluez sur ces graphiques la proportion des transistors dont la durée de vie est comprise entre 1850 et 2100.
3. Calculez la durée de vie atteinte ou dépassée par 20% des transistors.

4. L'aspect du polygone des fréquences non cumulées vous fait sans doute penser à la densité de probabilités d'une loi connue. Laquelle ?
5. Réalisez un test statistique au seuil de signification de 5% afin de confirmer ou non votre impression.

2ème partie (indépendante)

Lors de la fabrication d'une poudre, la concentration en détergent doit suivre une loi normale de moyenne 40 (unités) et d'écart-type 2,5 (unités).

Dans le contrôle de fabrication, il importe de pouvoir détecter une augmentation de 2 (unités) pour la moyenne. Si on accepte un risque de 1ère espèce de 0,05 et un de 2ème espèce de 0,10, quelle taille minimale doit-on donner aux échantillons prélevés régulièrement.

Solution

1ère partie

Classes	Eff.	fréq	fréq cum
]1700;1800]	4	0,0133	0,0133
]1800;1900]	44	0,1467	0,1600
]1900;1950]	40	0,1333	0,2933
]1950;2000]	62	0,2067	0,5000
]2000;2050]	58	0,1933	0,6933
]2050;2100]	46	0,1533	0,8466
]2100;2200]	38	0,1267	0,9733
]2200;2300]	8	0,0267	1,0000
	300		

Moyenne = 2001,5

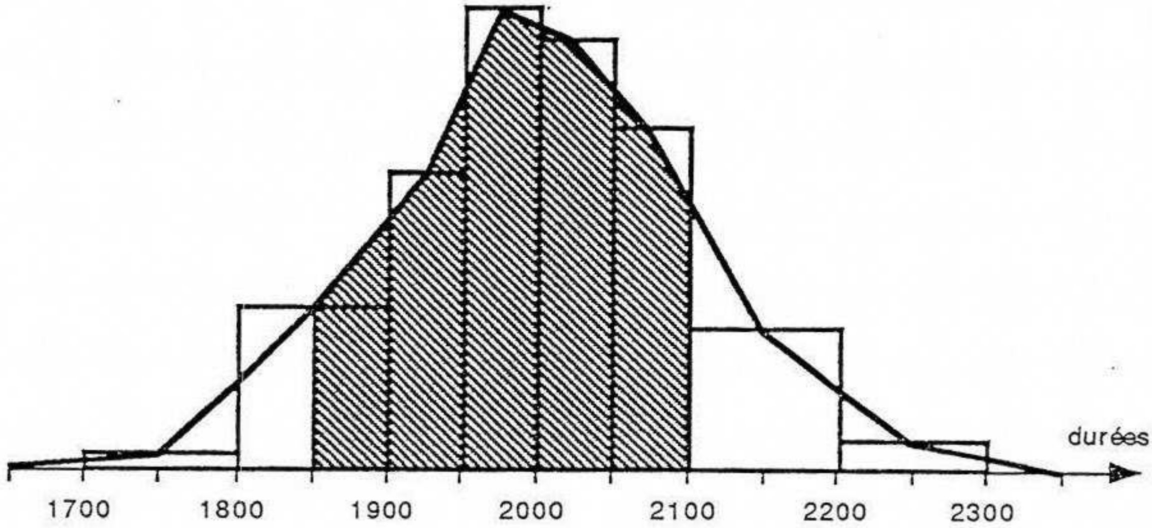
Ecart type = 102,5

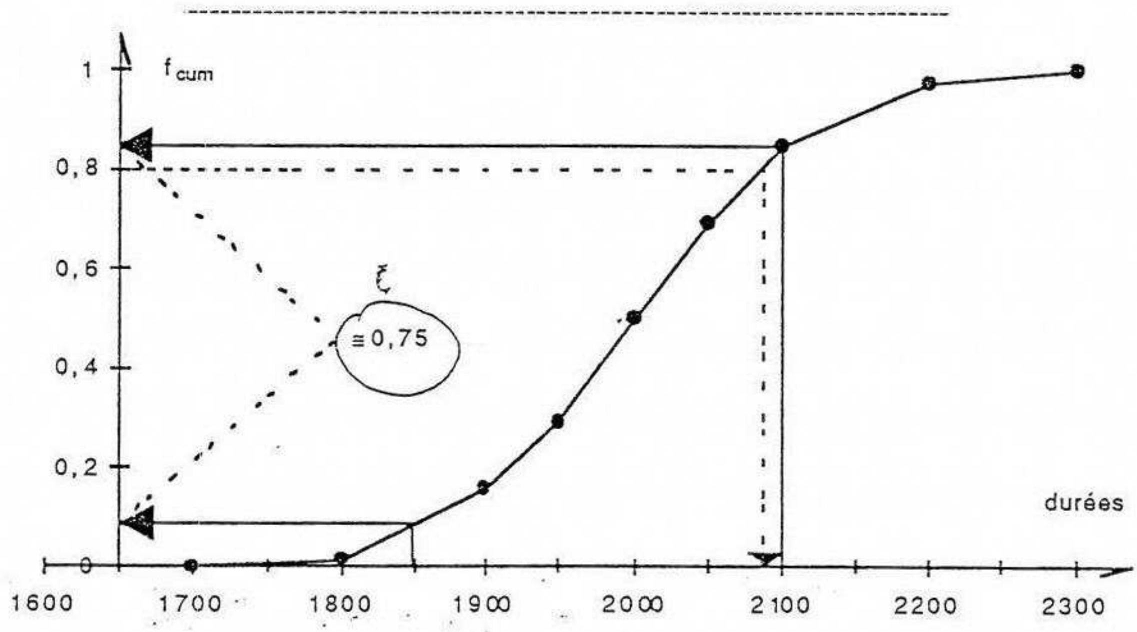
1) et 2). Graphique des fréquences : hauteurs :

$\frac{k}{100}.0,0133; \frac{k}{100}.0,1467; \frac{k}{50}.0,1333; \frac{k}{50}.0,2067; \frac{k}{50}.0,1933; \frac{k}{50}.0,1533; \frac{k}{100}.0,1267$ et $\frac{k}{100}.0,0267$

La plus grande hauteur est $\frac{k}{50}.0,2067 = 0,004134.k \Rightarrow$ choix $k = 2000cm$

$\Rightarrow$ hauteurs : 0,27; 2,93; 5,33; 8,27; 7,73; 6,13; 2,53; 0,53





3) ↑ de 0,6933 à 0,8466 ⇒ → 2050 à 2100 = +50
↑ de 1 ⇒
↑ de 0,6933 à 0,8000 ⇒ + 34,8

0,6933 0,8 0,8466
2050 2100
(x = 34,8) x

80% des transistors ont une durée d'au plus $\boxed{\text{2084,8 heures}}$

4) La loi Normale

5)

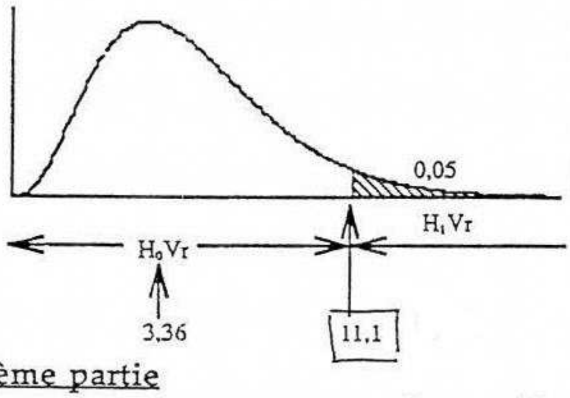
$k = 8$

classer	Probas Normales	Eff ATT	
$]-\infty;1700]$	$\Pr[X \le 1700] = \Pr[Z \le -2,94] = 0,0016$	0,48	7,32
$]1700;1800]$	$\Pr[1700 < X \le 1800] = \Pr[-2,94 < Z \le -1,97] = 0,0228$	6,84	
$]1800;1900]$	$= \Pr[-1,97 < Z \le -0,99] = 0,1367$	41,01	
$]1900;1950]$	$= \Pr[-0,99 < Z \le -0,50] = 0,1474$	44,22	
$]1950;2000]$	$= \Pr[-0,50 < Z \le -0,01] = 0,1875$	56,25	
$]2000;2050]$	$= \Pr[-0,01 < Z \le 0,47] = 0,1848$	55,44	
$]2050;2100]$	$= \Pr[0,47 < Z \le 0,96] = 0,1507$	45,21	
$]2100;2200]$	$= \Pr[0,96 < Z \le 1,94] = 0,1423$	42,69	
$]2200;2300]$	$= \Pr[1,94 < Z \le 2,91] = 0,0244$	7,32	7,86
$]2300;+\infty[$	$= \Pr[2,91 < Z] = 0,0018$	0,54	

Il a fallu estimer deux paramètres $ddl = k-1-a = 8-1-2 = 5$

$$\chi^2 obs = \frac{(4-7,32)^2}{7,32} + \frac{(44-41,01)^2}{41,01} + \cdots + \frac{(8-7,86)^2}{7,86} = 3,36$$

$\left.\begin{array}{l} H_0 : \text{La loi N convient} \\ H_1 : \text{La loi N ne convient PAS} \end{array}\right\}$ test χ^2 UNILAT. DROIT.

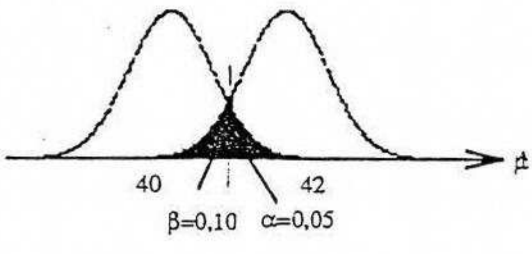


Conclusion

H_0 est vraie. La loi N convient
Risque β de concl. fausse.

2ème partie

$H_0 : \mu = 40 \qquad H_1 : \mu = 42$



On a
$$40 + z_{0,95}.ET\left(\hat{\mu}\middle|N(40;2,5)\right) = 42 + z_{0,10}.ET\left(\hat{\mu}\middle|N(42;2,5)\right)$$
$$40 + 1,645.\frac{2,5}{\sqrt{n}}.1 = 42 + (-1,28).\frac{2,5}{\sqrt{n}}.1$$
$$7,3125/\sqrt{n} = 2$$
$$\sqrt{n} = 3,65625$$
$$n = 13,37 \Rightarrow n = 14$$

Exercice 3 -

Soit une v.a. X prenant ses valeurs dans **N** suivant la loi

$$\Pr(X = x) = \frac{n^x}{(1+n)^{x+1}} \quad x \in \mathbf{N}$$

n désignant un nombre quelconque positif.

Soient X_1, X_2 , 2 v.a indépendantes de même loi que X .

Soit $Z = \inf(X_1, X_2)$.

Quelle est la loi de Z (suggestion: calculer la fonction de répartition de X) ?
Calculez $E(Z), Var(Z)$ en utilisant la méthode des fonctions génératrices .

Solution

Calculons la fonction de répartition de X . Si $k \in N$ on a

$$F_X(k) = \Pr(X \le k) = \sum_{t=0}^{k} \Pr(X = t) = \sum_{t=0}^{k} \frac{n^t}{(1+n)^{t+1}}$$
$$= \sum_{t=0}^{k} \frac{1}{1+n}\left(\frac{n}{1+n}\right)^t = \frac{1}{1+n}\left(\frac{1-\left(\frac{n}{1+n}\right)^{k+1}}{1-\frac{n}{1+n}}\right)$$
$$= 1 - \left(\frac{n}{n+1}\right)^{k+1}$$

Calculons la fonction de répartition de Z. Si $k \in N$ on a

$$
\begin{aligned}
F_Z(k) = \Pr(Z \le k) &= \Pr\left(\inf\left(X_1, X_2\right) \le k\right) \\
&= 1 - \Pr\left(\inf\left(X_1, X_2\right) > k\right) \\
&= 1 - \Pr\left(X_1 > k \ \underline{\text{et}}\ X_2 > k\right) \\
&= 1 - \underbrace{\Pr\left(X_1 > k\right)}_{1-F_X(k)}\underbrace{\Pr\left(X_2 > k\right)}_{1-F_X(k)} \qquad (X_1, X_2 \text{independantes}) \\
&= 1 - \left(\frac{n}{n+1}\right)^{2k+2}
\end{aligned}
$$

Calculons la loi de Z

$$\Pr(Z = z) = \Pr(Z \le z) - \Pr(Z \le z-1) \qquad (Z \text{ prenant valeur dans } N)$$

$$= F_Z(z) - F_Z(z-1)$$

$$= \left(1 - \left(\frac{n}{n+1}\right)^{2z+2}\right) - \left(1 - \left(\frac{n}{n+1}\right)^{2z}\right)$$

$$= -\left(\frac{n}{n+1}\right)^{2z+2} + \left(\frac{n}{n+1}\right)^{2z} = \left(\frac{n}{n+1}\right)^{2z}\left(1 - \left(\frac{n}{n+1}\right)^{2}\right)$$

$$= \left(\frac{n}{n+1}\right)^{2z}\left(\frac{n^2 + 2n + 1 - n^2}{(n+1)^2}\right) = \left(\frac{n}{n+1}\right)^{2z}\frac{1+2n}{(n+1)^2}$$

Calculons

$$M_Z(t) = E\left(e^{Zt}\right) = \sum_{z=0}^{\infty} e^{zt}\left(\frac{n}{n+1}\right)^{2z}\frac{1+2n}{(n+1)^2}$$

$$M_Z(t) = \frac{1+2n}{(n+1)^2} \qquad \underbrace{\sum_{z=0}^{\infty}\left(e^t\left(\frac{n}{1+n}\right)^2\right)^z}_{\text{serie geom convergeant conv si } \left|e^t\left(\frac{n}{1+n}\right)^2\right|<1}$$

$$\text{si } e^t < \frac{(1+n)^2}{n^2} \quad \text{si } t < \underbrace{\ln\underbrace{\left(\frac{1+n}{n}\right)^2}_{>1}}_{>0}$$

$$= \frac{1+2n}{(n+1)^2}\frac{1-0}{1-e^t\left(\frac{n}{n+1}\right)^2} = \frac{1+2n}{(n+1)^2\frac{\left((n+1)^2 - e^t n^2\right)}{(n+1)^2}}$$

$$= \frac{1+2n}{(n+1)^2 - e^t n^2}$$

$$\left.\frac{\partial}{\partial t}M_Z(t)\right|_{t=0} = \left.\frac{(1+2n)e^t n^2}{\left((n+1)^2 - e^t n^2\right)^2}\right|_{t=0} = \frac{(1+2n)n^2}{\left((n+1)^2 - n^2\right)^2} = \frac{n^2}{1+2n}$$

$$\Rightarrow \qquad \boxed{E(Z) = \frac{\partial}{\partial t} M_Z(t)\Big]_{t=0} = \frac{n^2}{1+2n}}$$

$$\frac{\partial^2}{\partial t^2} M_Z(t) = \frac{(1+2n)n^2 e^t\left((n+1)^2 - e^t n^2\right)^2 + (1+2n)n^2 e^t 2\left((n+1)^2 - e^t n^2\right)e^t n^2}{\left((n+1)^2 - e^t n^2\right)^4}$$

$$\frac{\partial^2}{\partial t^2} M_Z(t)\Big|_{t=0} = \frac{(1+2n)n^2(1+2n)^2 + (1+2n)n^2 2(1+2n)n^2}{(1+2n)^4}$$

$$= \frac{(1+2n)^2\left((1+2n)n^2 + 2n^4\right)}{(1+2n)^4}$$

$$= \frac{\left((1+2n)n^2 + 2n^4\right)}{(1+2n)^2}$$

$$= \frac{n^2 + 2n^3 + 2n^4}{(1+2n)^2}$$

$$= \mu_2'$$

$$\Rightarrow Var(Z) = \mu_2' - \left(\mu_1'\right)^2$$

$$= \frac{n^2 + 2n^3 + 2n^4}{(1+2n)^2} - \left(\frac{n^2}{1+2n}\right)^2$$

$$= \frac{n^2 + 2n^3 + 2n^4 - n^4}{(1+2n)^2}$$

$$\boxed{Var(Z) = \frac{n^2\left(n^2 + 2n + 1\right)}{(1+2n)^2} = \frac{n^2(n+1)^2}{(1+2n)^2}}$$