DÉPARTEMENT MATHÉMATIQUES

STATISTIQUE - I - Exercices Divers

La consultation des documents et l'usage des calculettes et des portables est interdit.
Seules sont autorisées deux pages format A4, manuscrites ,(non photocopiées).

Connaissances requises

- **Probabilité , Variables Aléatoires, Fonction de densité, Fonction de répartition, Moments,...** (voir Polycopié : **Annexe A**).
- **lois usuelles continues et discrètes** (voir Polycopié ch.3)

Thèmes

- **Lecture des tables (lois discrètes et continues)**
- **Convergence des variables aléatoires**
- **Inégalités de Markov, Bienaymé-Tchebychev**
- **Lois des grands nombre**
- **T.C.L.** (Théorème Central Limite)
- **Approximation des lois avec ou sans correction de continuité**
- **Échantillonnage**
- **Estimateurs et leurs propriétés**
- **Estimateur du Maximum de Vraisemblance**

Exercice à faire par tous les élèves et surtout par les élèves qui ont des difficultés **Exercice 1.** - On se propose à jouer au jeu suivant : On mise 7 F et ensuite on lance deux dés équilibrés. La somme gagnée est égale au total obtenu sur les dés.

1. Construire l'espace probabilisé (Ω, A, P) pour cette expérience.
   (a) Définir une v.a. X qui représente le gain.
   (b) Donner sa loi de probabilité et sa fonction de répartition.
   (c) On change la règle du jeu : La somme gagnée à chaque partie est égale à $(X-7)^2$. Calculer la loi de probabilité et la fonction de repartition pour cette *v.a.*.
2. (a) Calculer le gain moyen (espéré) par jeu dans le cas (1a).
   (b) Calculer le gain moyen par jeu pour le cas (1b).
3. On suppose que plusieurs personnes jouent ce jeu et que le gain moyen Z par joueur et par jour est de 120fr et d'un écart-type 10fr. On considère que dans une journée il y a 70 personnes qui jouent.
   (a) Quel est le pourcentage de personnes qui gagnent plus de 150fr par jour ?
   (b) Quel est le pourcentage des personnes dont le gain diffère du gain moyen 20fr ou plus ?
   (c) Le responsable du casino a déclaré que le samedi dernier le gain moyen s'élevait à 105fr par personne. Peut-on dire qu'il s'agissait d'une journée particulière avec un risque de 8% ?
   (d) Quelle est la probabilité pour que le gain moyen de de cette journée dépasse 125 fr ?
   (e) On suppose que le Lundi 12/09 il y a eu 60 personnes et le Mardi on en a compté 70. Quelle est la probabilté pour que les moyennes de deux journées diffèrent de 12 fr ou plus ?
4. Nous supposons que l'on ne connaît pas la moyenne de cette loi et on a choisi pour l'estimer la moyenne expérimentale (moyenne arithmétique de chaque jour). On suppose que la variance du gain journalier est égale à 1unité. Quel doit-être le nombre de personnes qui jouent dans une journée pour que l'erreur de notre estimation soit inférieure de 0.1 unité avec une probabilité égale à 0.95 ?

Partiels des l'années ?

I Exercice 1 .- Soit $X_1, X_2, ..., X_8$, 8 variables aléatoires de Poisson indépendantes de moyenne $\lambda = 2$.

1. Utilisez l'inégalité de Markov pour obtenir une borne pour : $P(X_1 + X_2 + ... + X_8 \geq 32)$
2. Qu'obtenez-vous en utilisant l'inégalité de Bienaymé-Chebyshev ?

Corrigé :

1. $Z = X_1 + X_2 + ... + X_8$ est une variable aléatoire positive de moyenne 2*8=16, donc l'inégalité de Markov donne : $P(Z \geq 32) \leq E(Z)/32 = 1/2$
2. Si X est une variable aléatoire de Poisson de paramètre λ, elle vérifie : $E(X) = \lambda$ et $Var(X) = \lambda$ Ainsi, $X_1, X_2, ..., X_8$ sont de paramètre 2 et, donc, de variance 2. Puis : $Var(Z) = Var(X_1) + ... + Var(X_8) = 16$ On peut maintenant appliquer l'inégalité de Bienaymé-Chebyshev : $P(Z \geq 32) = P(Z - EZ \geq 16) \leq P(|Z - EZ| \geq 16) \leq Var(Z)/16^2 = 1/16$.

Exercice 2 .-

1. Énoncer le théorème central limite. Indiquer avec précision les hypothèses nécessaires.
2. Une pièce équilibrée est lancée 500 fois. Quelle est la probabilité pour que le nombre de faces obtenu diffère de 250 d'au plus 10 ? (utiliser la calculatrice).

Corrigé :

1. Soient $X_1, X_2, ..., X_8$ (voir poly, TD et cours) Soit X le nombre de faces obtenu. X est une somme de 500 variables aléatoires indépendantes de Bernoulli de paramètre $\frac{1}{2}$. X est donc une v.a. binomiale de paramètres $n = 500$ et $p = \frac{1}{2}$, son espérance et sa variance valent respectivement $m = E[X] = np = 500\frac{1}{2} = 250$ et $\sigma^2 = Var[X] = np(1-p) = 500\frac{1}{2}\frac{1}{2} = 125$ Y étant une variable gaussienne centrée réduite, on a ensuite :
2. sans la correction de continuité $P(250 - 10 \leq X \leq 250 + 10) = P(\frac{-10}{\sigma} \leq Y \leq \frac{10}{\sigma}) = 0.6289066305...$
3. avec la correction de continuité $P(250 - 10 \leq X \leq 250 + 10) = P(\frac{-10.5}{\sigma} \leq Y \leq \frac{10.5}{\sigma}) = 0.6523455198...$ La probabilité exacte (loi binomiale) $P(240 \leq X \leq 260) = 0.6523358207$. Conclusion !

Exercice 3 .-Un représentant se présente dans les 1000 appartements d'une cité. La probabilité pour que ce représentant place un contrat dans un appartement quelconque est estimée à 0,04. Ces événements étant supposés indépendants, quelle est la probabilité pour qu'il place plus de 30 contrats ?
Corrigé : Soit X le nombre de contrats placés par le représentant. X suit une loi binomiale de paramètre $n = 1000$ et $p = 0,04$. La probabilité $P(X > 30)$ est donnée par . . . (loi binomiale, loi exacte)

La somme de l'expression précédente étant un peu longue, on peut, grâce au théorème de Moivre-Laplace, approcher la distribution de la variable binomiale X par celle d'une variable normale Y de moyenne $np = 40$ et de variance $np(1-p) = 38.4$; on a alors : $P(X > 30) \approx P(Y > 30) = P(\frac{Y-40}{\sqrt{38.4}} > \frac{30-40}{\sqrt{38.4}} \approx 0.947$ où la valeur de la probabilité recherchée a été obtenue en utilisant une table, grâce au fait que $Z = (Y - 40)/\sqrt{38.4}$ est une variable normale standard.

Remarques : Bien que $np = 40$, on peut aussi essayer d'approcher cette binomiale par une loi de Poisson, dans ce cas, on obtient :$P(X > 30) \approx 0.938$.

Remarquons par ailleurs que notre Théorème Central Limite ne permet pas d'évaluer a priori (en fonction de la taille n de l'échantillon) la qualité de l'approximation gaussienne effectuée. **Exercice 4** .- Un professeur sait par expérience que la note obtenue par un étudiant se présentant à un examen final est une variable aléatoire X d'espérance 75.

1. Donner une borne supérieure à la probabilité pour que la note de test d'un étudiant dépasse 85.
   Dans les questions qui suivent, on suppose de plus que X a pour variance $\sigma^2 = 25$.
2. Donner une borne supérieure à la probabilité pour que la note de test d'un étudiant dépasse 85.
3. Le professeur est satisfait si la moyenne de classe obtenue à l'issue des corrections est comprise entre 70 et 80. Combien d'étudiants doivent se présenter à cet examen pour que la probabilité de l'événement E : "le professeur est satisfait" dépasse 0,9 ? Ne pas utiliser le théorème central limite.
4. Est-il pertinent de faire appel au théorème central limite pour répondre à la question c) ?

Corrigé :

1. Comme X est une variable aléatoire positive : $P(X > 85) \leq E[X]/85 = 15/17 \cong 0.88$.
2. Remarquons tout de suite que la connaissance du 2 ème moment de X permet d'améliorer la majoration précédente : $P(X > 85) = P(X^2 > 85^2) \leq E[X^2]/85^2 = (\sigma^2 + \frac{E[X^2]}{85^2}) = \frac{(25+75^2)}{85^2} = \frac{226}{289} \cong 0.78$.
   On a aussi : $P(65 \leq X \leq 85) = P(|X - E[X]| \leq 10) = 1 - P(|X - E[X]| > 10) \geq 1 - \frac{\sigma^2}{10^2} = \frac{3}{4}$
3. Puisque les notes de chacun des n étudiants sont des variables indépendantes, leur moyenne arithmétique $Mn = \frac{(X1+X2+...+Xn)}{n}$ est encore d'espérance 75, et sa variance vaut $\frac{\sigma^2}{n} = \frac{25}{n}$ on a ensuite : $P(|Mn - E[Mn]| \leq 5) = 1 - P(|Mn - E[Mn]| > 5) \geq 1 - \frac{25}{(5^2 n)} = 1 - \frac{1}{n}, \frac{3}{4}$ avec 10 étudiants notre probabilité vaut au moins 0,9 .
4. Voyons s'il est pertinent d'utiliser le TCL : $P(|Mn - E[Mn]| \leq 5) \cong P(-n^{\frac{1}{2}} \leq Z \leq n^{\frac{1}{2}})$ et Z est une variable normale standard. Donc on a : $P(-n^{\frac{1}{2}} \leq Z \leq n^{\frac{1}{2}})$ pour $n = 1$, 0.6826894920 ; $n = 2$, 0.8427007929 ; $n = 3$, 0.9167354835 ; $n4$, 0.9544997360 ; $n = 5$, 0.9746526813
   Une application aveugle du T.C.L. donne donc ici pour nombre minimum d'étudiants : $n = 3$, un résultat franchement différent ! On constate que dans cet exemple simple l'utilisation du T.C.L. n'est pas appropriée puisque l'échelle des données du problème ($n \approx 10$) n'est pas assez grande.

II Exercice 1 .- On considère le nombre de passages de véhicules en un point d'une autoroute reliant deux villes A et B durant un intervalle d'une minute. On suppose que dans le sens de A à B, ce nombre X est une variable de Poisson de paramètre 17.8; dans le sens de B à A, le nombre Y de passages est une variable de Poisson de paramètre 7.1.

1. Quelle est la distribution de nombre total de passages durant une minute ?
2. Donner une approximation de la probabilité pour que le nombre total de passages enregistré durant une heure soit strictement plus grand que 1500. (On fera l'hypothèse d'indépendance qui s'impose et on effectuera une correction de continuité.)

Corrigé :

1. par la stabilité de la distribution de Poisson la distribution de nombre total de passages durant une minute est une distribution est une distribution de Poisson de paramètre $17.8 + 7.1 = 24.9$
2. Soit T le nombre total de passages enregistré durant une heure. T est une somme de 60 variables aléatoires indépendantes de Poisson de paramètre $\lambda = 24.9$. Comme la distribution de Poisson est stable X suit une loi de Poisson de paramètre $\lambda_{60} = 60 * 24.9 = 1494$. pour une bonne approximation de la probabilité cherchée on va appliquer le théorème central limite.
   L'espérance et la variance de T valent respectivement $m = E[T] = \lambda_{60} = 60 * 24.9 = 1494$ et $\sigma^2 := Var[T] = \lambda_{60} = 1494$.
   D'après le théorème central limite (T - m) /s peut être approché par une variable aléatoire gaussienne centrée réduite. Alors on pose $Z = \frac{(T-m)}{\sigma}$ avec $Z \sim N(0,1)$. Pour la probabilité cherchée on trouve $P(X > 1500) = P(X \geq 1500.5) = P(\frac{(X-1494)}{\sqrt{1494}} \geq \frac{(1500.5-1494)}{\sqrt{1494}} \simeq P(Z \geq \frac{(1500.5-1494)}{\sqrt{1494}} \simeq 0.4332263648$

Exercice 2 .- Soit X une variable aléatoire dont la distribution a pour densité $f_X(x) = \frac{x}{2}$ si $0 < x < 2$ et est nulle sinon.

1. Donner l'espérance et la variance de X
2. On considère N variables aléatoires $X_1, ..., X_N$ i.i.d. avec densité f_X. Donner l'espérance et la variance de la variable aléatoire $Z_N = \frac{(X_1+...+X_N)}{N}$
3. A partir de quelle valeur de N pouvez-vous affirmer que $P(|Z - \frac{4}{3}| > 10^{-3}) \leq 0.01$

Ind :

1. $E[X] = \frac{4}{3}$; $E[X^2] = 2$ et donc $Var[X] = \frac{2}{9}$
2. $E[Z_N] = \frac{4}{3}$ et $Var[Z_N] = \frac{Var[X]}{N} = \frac{2}{9N}$
3. On peut appliquer l'inégalité de Bienaymé-Chebyshev, : $P(|Z_N - \frac{4}{3}| > 0.001) \leq \frac{2*10^6}{9N}$ si $N > 22222222$.

Exercice 3 .- Des études faites en 1998 ont montré que le coût d'un étudiant par année scolaire suivait une loi Normale de moyenne 30000frs.

On a fait une enquête sur 450 étudiants et on a calculé une moyenne de 32500 frs et un écart-type de 4000frs. Peut-on dire avec un risque de 15% que les dépenses se comportent toujours comme en 1998 ?

Remarques :

- Problème d'échantillonnage.
- Ne pas confondre μ avec $\overline{x}$ (une réalisation de $\overline{X}$).
- La moyenne empirique est $\overline{x}$, l'écart - type expérimentale s
- On doit construire un intervalle pour $\overline{X}$ et ensuite vérifier à partir de l'échantillon si les dépenses se comportent toujours comme en 1998.
- $X\mathcal{N}(\mu, \sigma^2)$, $\mu = 30000, \overline{x} = 32500, s = 4000, n = 450, \alpha = 0.15, \sigma^2$ inconnu. Loi de Student.

Réponse :

On doit avoir : $P(a \leq \frac{\overline{X}-\mu}{s}\sqrt{n} \leq b) = 0.85$; $\mu - t_{0.075}\frac{s}{\sqrt{n}} \leq \overline{X}\mu + t_{0.075}\frac{s}{\sqrt{n}}$ Le point $t_{0.075}$ se trouve dans la table de Student, (on peut utiliser aussi la point de la loi normale $\nu = 449$ étant grand. On trouve l'intervalle $I = [29728; 30271]$ On vérifie ensuite si $\overline{x} \in I$

Exercice 4 .- La résistance à la rupture d'un fil synthétique de canne à pêche suit une *loi normale* de moyenne égale à 30kg et d'écart-type égal à 4kg. Il modifie son processus de fabrication pour gagner du temps et de l'argent. Pour contrôler on prélève un échantillon de 25 pièces dans la nouvelle production et on mesure une moyenne égale à 28kg et un écart-type égal à 4.8kg.

1. Peut-il prétendre que la qualité reste toujours la même ?
2. Quelle est la probabilité d'avoir une moyenne pour son échantillon inférieure ou égale à 27kg ?
3. Quelle est la valeur s_0^2 (de S^2) qui a 5 chances sur cent d'être dépassée ?

Exercice 5.- Une certaine machine usine des pièces.

1. D'une façon générale, elle produit 3% de pièces mauvaises. Un client reçoit une caisse de 500 pièces, en provenance directe de la machine et il trouve 25 pièces défectueues. Peut-t-on dire avec une probabilité de 1% que la machine a besoin d'un réglage ?
2. On s'intéresse maintenant au diamètre des pièces fabriquées. Le réglage de la machine était fait pour produire avec une moyenne de 5cm et un écart-type de 0.5cm. Suite à un contrôle de 25 pièces on a calculé une moyenne de 4.7cm et un écart-type de 0.8cm. Peut-on accepter que la machine est conforme au réglage effectué avec un risque $\alpha = 8\%$?

Exercice 6.- Le fabricant d'un médicament breveté affirmait que 90% de patients guérissaient en 8 heures. On a administré le médicament à 200 personnes et on a constaté que le médicament en a guéri 160 . Déterminer si l'affirmation du fabricant est légitime avec un risque de 2% . **Exercice 7.-** Une machine automatique fabrique des pièces dont la longueur est une v.a. qui suit une loi $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma = 1,5$ et $\mu = 60cm$.

On prélève régulièrement des échantillons de taille n .

Pour contrôler la variance de l'échantillon , on étudie les valeurs que prend la quantité :

$$S_n^2 = \sum_i (x_i - \overline{x})^2$$

1. Quelle est la loi de probabilité de la v.a. $\frac{S_n^2}{\sigma^2}$?
2. Dans quel intervalle $(0, a)$ doit se trouver S_n^2 avec une probabilité de 95% ?

Application numérique : On a pris un échantillon de taille 9 et on a calculé $s_n^2 = 14,4$. Conclusion ? **Exercice 8.-** On désire tester la résistance d'un matériel. Le test indique une résistance moyenne de 7750 *kilogrammes* avec un écart-type de 145 *kilogrammes*, sur 6 articles testés. Le fabricant affirme que la résistance moyenne est de 8000 *kilogrammes*. Peut-on maintenir les affirmations du fabricant au seuil de 96% ?

EXAMEN DE STATISTIQUE

Exercice 1.- Soient $X_i, i = 1, \ldots, 100$ des variables aléatoires uniformes sur l'intervalle $]0, 1[$. On cherche à évaluer approximativement $P\left\{\sum_{i=1}^{100} X_i > 60\right\}$.

N.B. : $E(X_i) = \frac{1}{2}; V(X_i = \frac{1}{12}$

Exercice 2.- Le nombre d'articles fabriqués dans une usine pendant une semaine est une variable aléatoire de moyenne 500.

1. Que peut-on dire de la probabilité que pendant une semaine particulière la production est d'au moins 1000 articles ?
2. Si la variance de la production hebdomadaire est connue et égale à 100, que peut-on dire de la probabilité que la production pendant une semaine partuculière soit comprise entre 400 et 600 ?

Exercice 3.- On prélève 25 pièces dans une production industrielle. Une étude préalable a montré que le diamètre de ces pièces suivait une loi gaussienne $\mathcal{N}(10, 4)$.

Donner une fourchette dans laquelle on doit trouver :
- la moyenne de cette échantillon avec une probabilité 0.90
- l'écart-type ?

Exercice 4.- Une tréfilerie fabrique des câbles métalliques conçus pour résister à des lourdes charges. Leur résistance à la rupture est en moyenne égale à 3 tonnes, avec un écart-type égal à 200 kg. Un contrôle de qualité de la fabrication est effectué sur un échantillon de 100 câbles.

1. En supposant que la production est conforme aux normes de résistance, quelle est la probabilité que la résistance moyenne des câbles de l'échantillon soit comprise entre 2.96 tonnes et 3.02 tonnes ?
2. Quelle est la probabilité qu'elle soit inférieure à 2.9 tonnes ?

Exercice 5.- On considère une urne contenant deux boules blanches et quatre boules bleues. On y effectue n tirages avec remise. À chaque tirage $i = 1, \ldots, n$ on associe la variable aléatoire

$$X_i = \begin{cases} 1, & \text{si la boule tirée est blanche} \\ 0, & \text{sinon} \end{cases}$$

On définit la variable aléatoire $\overline{X}_n = \frac{1}{n}\sum_{n=1}^{n} X_i$.

1. En utilisant l'inégalité de Bienaymé-Tchebychev, montrer que $\overline{X}_n$ converge vers une variable aléatoire que l'on précisera.
2. Déterminer le nombre minimum n_0 de tirages nécesaires pour que $P\left\{\left|\overline{X}_n - \frac{1}{3}\right| \geq 0.2\right\} \leq 0.01$ en utilisant l'inégalité de Bienaymé-Tchebychev.
3. En utilisant le théorème central limite, déterminer une autre valeur n_0 répondant à la question précédente. Comparer.

Exercice 6.- Chaque année une personne donnée effectue deux fois par jour, cinq jours par semaine et pendant 46 semaines, un trajet en voiture dont la durée est une variable aléatoire X qui suit une loi d'espérance 45mn et d'écart-type 10 mn.

On suppose que les durées des trajets sont mutuellement indépendantes.

Quelle est la probabilité pour que cette personne passe au moins 350 h dans sa voiture au cours de l'année ?

ex1

$P(\sum_{i=1}^{100} X_i > 60)$; on pose $S_n = \sum_{i=1}^{100} X_i$ et on applique T.C.L.

$E(S_n) = 100E(X_i) = 50$; $V(S_n) = 100V(X_i) = 100 * \frac{1}{12} = 100 * 0.08$.

$\Rightarrow \sigma_{S_n} \simeq 10 * 0.3$ et $P\left(\frac{S_n - E(S_n)}{\sigma_{S_n}} \geq \frac{60 - E(S_n)}{\sigma_{S_n}}\right) =$

$= P\left(T \geq \frac{60-50}{3}\right) = P\left(T \geq \frac{10}{3}\right) = P(T \geq 3,3) = 0.0^3483$

$= 0.000483$.

ex2 $\mu = 500$

1. Inégalité Markov.

$P(X \geq 1000) \leq \frac{E(X)}{1000} = \frac{500}{1000} = \frac{1}{2}$.

2. B.T inégalité $P(|X - E(X)| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$ (I) $\qquad \sigma^2 = 100$

$P(|X-500| \geq 100) \leq \frac{100}{100^2} = \frac{1}{100}$

car : $P(400 \leq X \leq 600) = P(-100 \leq X - 500 \leq 100)$

$= P(|X-500| \leq 100) = 1 - P(|X-500| \geq 100)$ et

donc $P(|X-500| \leq 100) \leq 0,99$. (voir (I))

c.a.d la probabilité est au moins égale à 0,99

ex3 Pb échantillonnage (voir TD et cours).

$n = 25 \quad X_i \sim N(\mu, \sigma^2) \quad \mu = 10$ et $\sigma^2 = 4$; $\sigma = 2$.

a) Fourchette pour $\bar{X}$: $P(a \leq \bar{X} \leq b) =$

$= P\left(\underbrace{\frac{a-\mu}{\sigma}\sqrt{n}}_{-t} \leq \frac{\bar{X}-\mu}{\sigma}\sqrt{n} \leq \underbrace{\frac{b-\mu}{\sigma}\sqrt{n}}_{t}\right) = 0,90$

Donc $F(t) - F(-t) = 0,90$ et $2F(t) = 1,90$; $F(t) = 0,95$

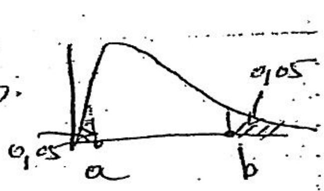
donc $t \simeq 1,65$. On a $\frac{a-\mu}{\sigma}\sqrt{n} = -1,65 \Rightarrow a = 9,34$

et $\frac{b-\mu}{\sigma}\sqrt{n} = 1,65 \Rightarrow b = 10,66$.

b) Fourchette pour S^2 (règle de décision) $S^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$

$\frac{nS^2}{\sigma^2} \sim \chi^2(\nu)$: $n = 25$ donc $\nu = 24$

$P(a \leq S^2 \leq b) = P\left(\frac{an}{\sigma^2} \leq \frac{nS^2}{\sigma^2} \leq \frac{nb}{\sigma^2}\right) = 0,90$.



donc $P\left(\frac{nS^2}{\sigma^2} \leq \frac{an}{\sigma^2}\right) = P\left(\frac{nS^2}{\sigma^2} \geq \frac{nb}{\sigma^2}\right) = 0,05$

Par les tables : $P\left(\frac{nS^2}{\sigma^2} \geq \frac{nb}{\sigma^2}\right) = 0,05 \qquad \nu = 24$; $\chi^2(\nu)$

$t_1 = 36,4 \Rightarrow \frac{n.b}{\sigma^2} = 36,4 \Rightarrow b = \frac{36,4 \cdot \sigma^2}{n}$

et $P\left(\frac{nS^2}{\sigma^2} \leq \frac{na}{\sigma^2}\right) = 1 - P\left(\frac{nS^2}{\sigma^2} \geq \frac{na}{\sigma^2}\right) = 0,05$; $P\left(\frac{nS^2}{\sigma^2} \geq \frac{na}{\sigma^2}\right) = 0.95$

$\frac{na}{\sigma^2} = t_2$ et $t_2 \simeq 13.85$ et $a = \frac{13.85 \cdot \sigma^2}{n}$.

Donc la fourchette pour S^2 : $a \le S^2 \le b$ et

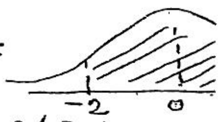
$\sqrt{a} \le S \le \sqrt{b}$ donc $\frac{2}{5}\sqrt{13.85} \le S \le \frac{2}{5}\sqrt{36.4}$

(ou bien $1.48 \le S \le 2.41$)

ex4 : $\mu = 3000$ kg $\quad \sigma = 200 \quad n = 100$.

Pb d'échantillonnage : (voir exercice précédent)

1. $P\left(\frac{2960-\mu}{\sigma}\sqrt{n} \le \frac{\bar{X}-\mu}{\sigma}\sqrt{n} \le \frac{3020-3000}{\sigma}\sqrt{n}\right) =$



$= P\left(-2 \le \frac{\bar{X}-\mu}{\sigma}\sqrt{n} \le 1\right) = P\left(\frac{\bar{X}-\mu}{\sigma}\sqrt{n} \le 1\right) - P\left(\frac{\bar{X}-\mu}{\sigma}\sqrt{n} \le\right.$

$= F(1) - F(-2) = F(1) - (1 - F(2)) = F(2) + F(1)$

$= 0,8413 - 1 + 0,9772 = 0,8185$

2. $P(\bar{X} \le 2900) = P\left(\frac{\bar{X}-\mu}{\sigma}\sqrt{n} \le \frac{2900-3000}{\sigma}\sqrt{n}\right) = F(-5) \simeq 0$

dans la table on trouve cette proba inférieure à

0.000000287

ex5 Proba d'une boule blanche $p = \frac{2}{6} = \frac{1}{3}$ donc $\mu = p$

1. $\bar{X} = \frac{1}{n}\sum X_i$: $E(\bar{X}) = \mu = \frac{1}{3}$ et $V(\bar{X}) = \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{n} = \frac{2}{9n}$, $\sigma = \frac{\sqrt{2}}{3}$

donc B.T. inégalité $P\left(|\bar{X} - \frac{1}{3}| \geq \varepsilon\right) \le \frac{\sigma^2}{\varepsilon^2} = \frac{2}{9n\varepsilon^2} \underset{n \to +\infty}{\longrightarrow} 0$

donc $\bar{X} \underset{P}{\rightarrow} \frac{1}{3}$ (voir déf cours : poly et TD)

2. $P\left(|\bar{X} - \frac{1}{3}| \geq 0.02\right) \le \frac{2}{9n} \cdot \frac{1}{0.02^2} = 0.01 \Rightarrow n = \frac{1000000}{18} = 55.5$

3. On veut calculer n en utilisant T.C.L.

(il faut justifier son utilisation).

$P\left(|\bar{X} - \frac{1}{3}| \geq 0.02\right) = 1 - P\left(|\bar{X} - \frac{1}{3}| \le 0.02\right)$

$= 1 - P\left(-0.02 \le \bar{X} - \frac{1}{3} \le 0.02\right) = 0.01$

$\Rightarrow P\left(\frac{-0.02}{\sigma}\sqrt{n} \le \frac{\bar{X} - \frac{1}{3}}{\sigma}\sqrt{n} \le \frac{0.02}{\sigma}\sqrt{n}\right) = 0,99. \quad \sigma = \frac{\sqrt{2}}{\sqrt{n}} \cdot \frac{1}{3}$

et $P\left(-0.06\sqrt{\frac{n}{2}} \le T \le 0.06\frac{\sqrt{n}}{\sqrt{2}}\right) = 0.99$.

et $F\left(0.06\sqrt{\frac{n}{2}}\right) = \frac{1.99}{2} = 0,995$ et $t = 2.60$ (2.58

$\Rightarrow$ $n = 3756$, valeur très inférieur

EISTI - DÉPARTEMENT MATHÉMATIQUES

EXAMEN DE STATISTIQUE – I

08.12.2001 - DURÉE 1h30

Exercice 1 .- On lance une pièce de monnaie une seule fois avec θ la probabilité d'obtenir pile.

1. Proposer un estimateur pour θ
2. On propose $Z_1 = X$ et $Z_2 = \frac{1}{2}$ deux estimateurs pour θ.

   Lequel de deux estimateurs est préferable ?

Exercice 2 .- La résistance à la rupture d'un fil synthétique de canne à pêche suit une *loi normale* de moyenne égale à 30kg et d'écart-type égal à 4kg. Il modifie son processus de fabrication pour gagner du temps et de l'argent. Pour contrôler on prélève un échantillon de 25 pièces dans la nouvelle production et on mesure une moyenne égale à 28kg et un écart-type égal à 4.8kg.

1. Peut-il prétendre que la qualité reste toujours la même ?
2. Quelle est la probabilité d'avoir une moyenne pour son échantillon inférieure ou égale à 27kg ?
3. Quelle est la valeur s_0^2 (de S^2) qui a 5 chances sur cent d'être dépassée ?

Exercice 3 .- La durée de vie d'une pièce d'équipement est une v.a. qui suit une loi exponentielle (avec densité $\theta e^{-\theta x} \quad x \geq 0 \quad , \quad \theta > 0$).

On dispose de deux ateliers de fabrication.

- Dans le premier atelier on a fait cinq observations et on a mesuré :
  $x_1 = 0.9 \quad , \quad x_2 = 1.7 \quad , \quad x_3 = 0.4 \quad , \quad x_4 = 0.3 \quad , \quad x_5 = 2.4$
- Dans le deuxième atelier on a fait quatre observations et on a mesuré :
  $y_1 = 0.8 \quad , \quad y_2 = 1.4 \quad , \quad y_3 = 1.2 \quad , \quad y_4 = 0.5$

1. Calculer la quantité de l'information contenue dans chaque échantillon.
2. Donner la quantité de l'information contenue à l'ensemble des deux échantillons (de préférence sans calcul).
3. Déterminer l'E.M.V. pour l'ensemble de deux expériences.
4. Contrôler les propriétés de cet estimateur.
5. Proposer un estimateur pour l'espérance de X. Faites vos commentaires.

8/

DÉPARTEMENT MATHÉMATIQUES

EXAMEN DE STATISTIQUE I
08/12/2001
Quelques Remarques et Indications 09.01.2002

Exercice 1 . Remarques :

- La loi suivie est la loi de Bernoulli.
- On ne connaît pas le paramètre de la loi $\theta = p$. Attention : Si on pose $p = \frac{1}{2}$, on n'a pas besoin de l'estimer.
- Si $a = cte$, $E(a) = a$ et $Var(a) = 0$
- Z_1 est sans biais et Z_2 est biaisé. On doit alors comparer les risques quadratiques : $R_\theta(T) = V(T) + b^2$

1. Soit X la v.a. associée à cette expérience. $X = 0$ si face $X = 1$ si pile. Pour estimer p on peut proposer $Z = \frac{X}{n}$, $n = 1$.
2. On considère $Z_1 = X$; $Z_2 = \frac{1}{2}$. Pour comparer Z_1 et Z_2 on examine :
   - Le biais. $E(Z_1) = p = \theta$ et $E(Z_2) = \frac{1}{2}$, $b(Z_2) = \frac{1}{2} - \theta$. Z_1 sans biais et Z_2 biaisé.
   - Le risque quadratique : $R_\theta(Z) = V(Z) + b^2$. $R_\theta(Z_1) = \theta(1-\theta)$ et $R_\theta(Z_2) = \left(\frac{1}{2} - \theta\right)^2$.

Exercice 2 . Remarques :

- Ne pas confondre μ avec $\overline{x}$ (une réalisation de $\overline{X}$). Les paramètres de la loi $\mu = E(X) = 30$ et $\sigma = 4$ sont connus, on utilise alors la loi Normale.
- La moyenne empirique est $\overline{x} = 28$, l'écart - type expérimentale $s = 4.8$ et l'effectif de l'échantillon est $n = 25$.
- Pour établir que la qualité reste toujours la même, il doit contrôler $\overline{x}$ et s^2.
- Il y a une seule expérience de $n = 25$ épreuves.
- Il n'est pas raisonnable d'essayer de faire un intervalle pour μ ni de calculer $P(\mu \leq 27)$, ($\mu = 30$ valeur connue)

1. On doit établir une règle (un intervalle pour $\overline{X}$ (la moyenne expérimentale) et une règle pour S^2 : $P(a \leq \overline{X} \leq b) = 1 - \alpha$, (α quelconque) et $P(a \leq s^2 \leq b) = 1 - \alpha$. Si $\overline{X}$ se trouve dans $[\mu - t_\alpha \frac{\sigma}{\sqrt{n}}; \mu + t_\alpha \frac{\sigma}{\sqrt{n}}]$ et $\overline{S^2}$ dans $[a; b]$, il peut prétendre que la qualité reste la même avec le risque α.
2. Il suffit de calculer $P(\overline{X} \leq 27)$
3. Il suffit de calculer s_0^2 t.q $P(s^2 > s_0^2) = 0.05$.

Exercice 3 . Remarques :

- La forme de la loi exponentielle est donnée dans l'énoncé ! ! ! (Ce n'est pas la loi de Poisson)
- L'information $I(\theta) = E\left(\frac{\partial \ln L(\theta; X)}{\partial \theta}\right)^2$ ou formule équivalente : $I(\theta) = -E\left(\frac{\partial^2 \ln L(\theta; X)}{\partial \theta^2}\right)$ (La deuxième formule donne le résultat plus rapidement)
- $I(\theta)$ est toujours positive. $I_n(\theta) = nI(\theta)$
- On doit calculer l'espérance des fonctions qui entrent dans la formule précédente.
- $I(\theta)$ ne dépend pas de l'échantillon ! !
- Pour l'E.M.V de l'ensemble de deux expériences on a $5 + 4$ épreuves.

1. $I(\theta) = E\left(\frac{\partial \ln L(\theta; X)}{\partial \theta}\right)^2 = E\left(\frac{1}{\theta} - X\right)^2 = \frac{1}{\theta^2} + E\left(X^2\right) - \frac{2}{\theta} E(X) = \frac{1}{\theta}^2$ $I_1(\theta) = nI(\theta) = 5I(\theta)$, $I_2(\theta) = nI(\theta) = 4I(\theta)$
2. $I_{1,2}(\theta) = I_1(\theta) + I_2(\theta)$
3. On calcule l'E.M.V. à partir de : $\frac{\partial \ln L(x_1, \ldots, x_1, y_1, \ldots y_4 \theta)}{\partial \theta} = \frac{9}{\theta} - \left(\sum_{i=1}^{5} x_i + \sum_{i=1}^{4} y_i\right) = 0$ (en suivant l'algorithme proposé dans les TD et le poly)
   On obtient : $\widehat{\theta} = \frac{9}{\sum_{i=1}^{5} x_i + \sum_{i=1}^{4} y_i}$
4. Calcul de l'$E(\widehat{\theta}), Var(\widehat{\theta})$ e.t.c.
5. $\mu = E(X) = \frac{1}{\theta}$ On peut proposer $\widehat{\mu} = \left(1/\widehat{\theta}\right) = \frac{\sum_{i=1}^{5} x_i + \sum_{i=1}^{4} y_i}{9}$ qui sera l'E.M.V de μ (voir propriété de l'E.M.V.)

9/

EISTI - DÉPARTEMENT MATHÉMATIQUES

EXAMEN DE STATISTIQUE I

Exercice 1.- On souhaite estimer la proportion p de pièces défectueuses fabriquées par une machine. On prélève des échantillons toutes les semaines.

Soit X le nombre de pièces défectueuses parmi les n pièces examinées.

1. Proposer un estimateur T_1 pour estimer p.
2. On considère maintenant un estimateur $T_2 = \dfrac{X+1}{n}$ pour p. Choisissez entre T_1 et T_2. Commenter.

Exercice 2.- Un article paru dans une revue médicale analyse les effets de fortes doses de vitamine C sur les rhumes. Dans l'expérience, 407 volontaires atteints de rhumes prenaient de fortes doses de vitamine C et guérissaient en moyenne après 5.25 jours avec un écart-type de 6.0 jours. La valeur importante de l'écart-type résulte de la spécificité du traitement.

1. Construire un intervalle à 90% de la moyenne du temps de guérison. (de l'échantillon
2. Dans cette expérience la moyenne du groupe qui n'avait pas pris de vitamine C, était de 6.02 jours. Commenter cette expérience.
3. Quel est l'intervalle qui contient la moyenne de guérison dans le 98% de cas ?
4. On a fait une deuxième expérience avec 30 personnes atteintes de rhumes et on a trouvé un écart-type de 5.30. Quelle est la probabilité pour que l'écart-type du temps de guérison dépasse 6.20 jours ?

Exercice 3.- On suppose que le nombre de pannes d'un ordinateur sur un trimestre suit une loi de Poisson de paramètre λ. On souhaite estimer λ.

1. Proposer un estimateur pour λ.
2. On a observé quatre ordinateurs pendant un trimestre et on a constaté 4 pannes pour le premier, 5 , 3 et 6 pannes respectivement pour les autres. Calculer l'E.M.V. en se basant sur cette expérience.
3. On considère maintenant un échantillon de 150 ordinateurs. Pouvez-vous appliquer le $T.C.L.$ pour construire un $I.D.C.$ pour λ avec $\alpha = 15\%$?

EXAMEN DE STATISTIQUE I

Quelques Remarques et Indications

06.12.2000

Exercice 1 .

1. Pour estimer p on peut proposer $T_1 = \frac{X}{n}$.
2. On considère $T_2 = \frac{X+1}{n}$. Pour comparer T_1 et T_2 on examine :
   (a) Le biais. $E(T_1) = p$ et $E(T_2) = p + \frac{1}{n}$. T_1 sans biais et T_2 biaisé.
   (b) Le risque quadratique : $R_\theta(T) = V(T) + b^2$. $R_\theta(T_1) = \frac{p(1-p)}{n}$ et $R_\theta(T_2) = \frac{p(1-p)}{n} + \frac{1}{n^2}$ (puisque $V(\frac{1}{n}) = 0$, $V(\frac{X+1}{n}) = V(\frac{X}{n}) + V(\frac{1}{n}) = V(\frac{X}{n})$).

Exercice 2 .

1. Ne pas confondre μ avec $\overline{x}$. La moyenne de la loi est $E(X) = \mu$ inconnue. La moyenne empirique est $\overline{x} = 5.25$, la variance expérimentale $s^2 = 6.0^2$ et l'effectif de l'échantillon est $n = 407$. Par contre σ de la population inconnu. On doit expliquer si on prend $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ ou bien $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
2. On demande un intervalle pour la moyenne (espérance) μ de la loi. Il n'est pas raisonnable d'essayer de faire un intervalle pour $\overline{x} = 5.25$ (valeur connue). Dans ce cas l'I.D.C. est de la forme : $\overline{X} - t\frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + t\frac{s}{\sqrt{n}}$ σ étant inconnu on travaille avec Student. On construit un I.D.C. seulement pour le(s) paramètre(s) inconnus d'une loi $(\mu, \sigma^2, ..)$. Ne pas confondre $\overline{X}$ et $\overline{x}$. $\overline{x} = 5.25$, est une réalisation de la v.a. $\overline{X}$.
3. La moyenne expérimentale du groupe qui n'avait pas pris la vitamine est $\overline{x} = 6.02$. On constate qu'elle se trouve à l'extérieur de cet intervalle. On peut conclure que la vitamine C a influencé le temps de guérison.
4. Même question avec (1).
5. Ici $n = 30$ et $s = 5.30$. $P(\sigma > 6.20)$; $P(\sigma^2 > 6.20^2)$; $P(\frac{ns^2}{\sigma^2} < \frac{ns^2}{6.20^2}) = 0.98$;

Exercice 3 .

1. $E(X) = V(X) = \lambda$; On peut proposer $\overline{X}$ ou S^2 (sous ses deux formes).
2. Il y a $n = 4$ épreuves. On forme la fonction de vraisemblance $L(x_1, ..., x_4)$ et on calcule l'E.M.V. : $\widehat{\lambda} = \sum_{i=1}^{n}\frac{X_i}{4} = 4.5$.
3. On peut appliquer le $T.C.L.$ puisque $n = 150$. $S_n = \sum_{i=1}^{n} X_i$; $\frac{S_n - E(S_n)}{\sigma_{s_n}} \sim \mathcal{N}(0,1)$;

   $\frac{\sum_{i=1}^{n} X_i - n\lambda}{\sqrt{n\lambda}} \sim \mathcal{N}(0,1)$; $V(S_n) = n\lambda$, puisque $\sigma = \sqrt{\lambda}$.

   $\frac{\sum_{i=1}^{n}\frac{X_i}{n} - \lambda}{\sqrt{\frac{\lambda}{n}}} \sim \mathcal{N}(0,1)$; $\frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \sim \mathcal{N}(0,1)$

   L'I.D.C. pour λ sera :

   $\overline{x} - t\sqrt{\frac{\lambda}{n}} \leq \lambda \leq \overline{x} + t\sqrt{\frac{\lambda}{n}}$. On ne connaît pas λ, on remplace λ par son estimateur $\widehat{\lambda} = \overline{x} = 4.5$, (voir TD sur I.D.C. d'une proportion), et on obtient :

   $\overline{x} - t\sqrt{\frac{\widehat{\lambda}}{n}} \leq \lambda \leq \overline{x} + t\sqrt{\frac{\widehat{\lambda}}{n}}$.

11/

EISTI - DÉPARTEMENT MATHÉMATIQUES

EXAMEN DE STATISTIQUE I

29 novembre 2002 –

DURÉE 2h00

Exercice 1 .- Des études faites en 1998 ont montré que le coût d'un étudiant par année scolaire suivait une loi Normale de moyenne 30000frs. En 2000 on a fait une noiuvelle enquête 450 étudiants et on a calculé une moyenne de 32500 frs et un écart-type de 4000frs.

Peut-on dire avec un risque de 15% que les dépenses se comportent toujours comme en 1998 ?

Exercice 2 .- Les durées de fonctionnement d'une photocopieuse entre deux pannes successives sont des variables aléatoires indépendantes $T_1, T_2, \ldots, T_n$ distribuées identiquement selon une loi gamma dont la fonction de densité est :

$$f(y) = \begin{cases} \frac{y}{\theta^2} e^{-\frac{y}{\theta}}, & \text{si } y > 0 \\ 0, & \text{sinon} \end{cases}$$

avec $\theta > 0$ est un paramètre inconnu, à estimer.

N.B. : Pour les paramètres de la loi on a posé : $\alpha = 2$; $\Gamma(\alpha) = 1$; $\beta > 0$ où $\theta = \beta$

1. Pouvez-vous proposer un estimateur pour θ ? Justifier votre réponse.
2. On observe les n premiers temps $t_1, t_2, \ldots, t_n$ de fonctionnement entre deux pannes successives ; estimer le paramètre θ à partir de $t_1, t_2, \ldots, t_n$ en utilisant la méthode du Maximum de Vraisemblance (on appellera $\widehat{\theta}_n$ l'estimateur obtenu).
   (a) Donner le biais b_n et le risque quadratique R_n de l'estimateur $\widehat{\theta}_n$ de θ.
   (b) $\widehat{\theta}_n$ est-il efficace ?
   (c) Démontrer que $\widehat{\theta}_n$ converge en probabilité vers θ.
   (d) Quelle est la loi asymptotique de $\widehat{\theta}_n$?
3. On a fait une deuxième série d'observations des m temps $x_1, x_2, \ldots, x_m$ de fonctionnement entre deux pannes successives.
   (a) Calculer l'estimateur du maximum de vraisemblance de θ en tenant compte de deux séries d'observations.
   (b) On souhaite estimer la quantité $\theta^2 + 3\theta$. Pouvez-vous proposer un estimateur pour cette quantité ? Justifier.

Application numérique :

$n = 5;\ t_1 = 2;\ t_2 = 2.5;\ t_3 = 1.8;\ t_4 = 2.2;\ t_5 = 1.9$

$m = 4;\ x_1 = 2.3;\ x_2 = 2.5;\ x_3 = 1.8;\ x_4 = 2.$

Exercice 3 .- La probabilité d'avoir un garçon à l'échelon national est de 0.48.

1. Dans une maternité sur 500 naissances, nous avons observé 230 naissances de garçons. Peut-on dire que ce résultat est conforme à la moyenne nationale avec un risque de 12% ?
2. Dans une autre maternité on a annoncé une fourchette de $47 - 50\%$ pour les naissances des garçons. Déterminer avec un risque de 5% quel est le nombre de naissances pris en compte pour le calcul de cette fourchette.

EXAMEN DE STATISTIQUE I - 29/11/2002 - corrigé
Quelques Remarques et Indications

Exercice 1 . Remarques :
- Problème d'échantillonnage.
- Ne pas confondre μ avec $\overline{x}$ (une réalisation de $\overline{X}$).
- La moyenne empirique est $\overline{x}$, l'écart - type expérimentale s
- On doit construire un intervalle pour $\overline{X}$ et ensuite vérifier à partir de l'échantillon si les dépenses se comportent toujours comme en 1998.
- $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu = 30000, \overline{x} = 32500, s = 4000, n = 450, \alpha = 0.15, \sigma^2$ inconnu. Loi de Student.

Réponse : On doit avoir : $P(a \leq \frac{\overline{X}-\mu}{s}\sqrt{n} \leq b) = 0.85$; $\mu - t_{0.075}\frac{s}{\sqrt{n}} \leq \overline{X} \leq \mu + t_{0.075}\frac{s}{\sqrt{n}}$ Le point $t_{0.075}$ se trouve dans la table de Student, (on peut utiliser aussi les points de la loi normale ($\nu = 449$ étant grand). On trouve l'intervalle $I = [29728; 30271]$ On vérifie ensuite si $\overline{x} \in I$

Exercice 2 . Remarques :
- La loi Gamma se trouve dans le poly p.38-39. Il ne s'agît pas de la loi exponentielle, ni de la binomiale!!
- L'espérance et la variance de la loi sont données p.39 : $E(Y) = \alpha\beta = 2\theta$; $V(Y) = \alpha\beta^2 = 2\theta^2$ i.e. : $E(T_i) = 2\theta$; $V(T_i) = 2\theta^2$

Réponse :

1. On peut proposer l'estimateur $U = \frac{\overline{T}}{2}$ dont $E(\frac{\overline{T}}{2}) = \theta$
2. E.M.V : Fonction de vraisemblance : $L(t_1, ..., t_n; \theta) = \Pi_{i=1}^{n} \frac{t_i}{\theta^2} e^{\frac{t_i}{\theta}}$ On calcule l'E.M.V. (en suivant l'algorithme proposé dans les TD et le poly) à partir de : $\frac{\partial \ln L(t_1,...,t_n;\theta)}{\partial\theta} = \frac{-2n}{\theta} + \frac{\sum_{i=1}^{n} t_i}{\theta^2} = 0$; $\widehat{\theta_n} = \frac{\sum_{i=1}^{n} t_i}{2n} = \frac{1}{2}\overline{T}$
   (a) Le biais. $E(\widehat{\theta_n}) = \theta$; sans biais. Le risque quadratique : $R(\widehat{\theta_n}) = V(\widehat{\theta_n}) = \frac{\theta^2}{2n}$.
   (b) Efficace : L'information $I(\theta) = E\left(\frac{\partial \ln L(\theta;t)}{\partial\theta}\right)^2$ ou formule équivalente : $I(\theta) = -E\left(\frac{\partial^2 \ln L(\theta;t)}{\partial\theta^2}\right) = \frac{2}{\theta^2}$ (La deuxième formule donne le résultat plus rapidement)
   (c) $I(\theta)$ est toujours positive. $I_n(\theta) = nI(\theta) = \frac{2n}{\theta^2}$. Alors $\widehat{\theta_n}$ est efficace.
   (d) $\widehat{\theta_n} = \frac{\sum_{i=1}^{n} t_i}{2n}$; $E(T_i) = 2\theta$; $V(T_i) = 2\theta^2$ existent, par la loi faible des grands nombres on obtient : $\widehat{\theta}_n = \frac{\sum_{i=1}^{n} t_i}{2n} \to 2\theta, n \to \infty$, en probabilité. Il est facile à démontrer aussi la convergence en m.q. ce qui implique la convergence en probabilité.
   (e) Loi asymptotique : $\widehat{\theta}_n \sim \mathcal{N}(\theta, \frac{\theta^2}{2n})$
3. Pour l'ensemble de deux expériences :
   (a) $L(t_1, ..., t_n, x_1, ..., x_n; \theta) = L(t_1, ..., t_n; \theta)L(x_1, ..., x_n; \theta)$; $lnL(t_1, ..., t_n, x_1, ..., x_n; \theta) = lnL(t_1, ..., t_n; \theta) + lnL(x_1, ..., x_n; \theta)$.En utilisant les résultats déjà obtenus on a : EMV $\theta_{tot} = \frac{1}{2}\frac{\sum_{i=1}^{n} t_i + \sum_{i=1}^{m} x_i}{n+m}$
   (b) EM.V de $\theta^2 + 3\theta$ peut être donné par : $\left(\widehat{\theta}_n\right)^2 + 3\widehat{\theta}_n$ si cette transformation est bijective.

Exercice 3 . Remarques :

1. Il s'agit de la loi de la fréquence voir TD , poly , cours : f_n : $p = 0.48, n = 500, f_{500} = \frac{230}{500} = 0.46$ On doit établir une règle pour f_n et vérifier si 0.46 appartient à cet intervalle. $P(a \leq f_n \leq b) = 0.98$; $p - t_{0.06}\sqrt{\frac{p(1-p)}{n}} \leq f_n \leq p + t_{0.06}\sqrt{\frac{p(1-p)}{n}}$. On calcule $I = [0.445; 0.51]$
2. On veut maintenant : $P(47 \leq f_n \leq 51) = 0.95$ i.e. $P(\frac{a-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{f_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{b-p}{\sqrt{\frac{p(1-p)}{n}}}) = 0.95$; $P(\frac{a-p}{\sqrt{\frac{p(1-p)}{n}}} \leq t_1) = 0.025$ ou $P(\frac{b-p}{\sqrt{\frac{p(1-p)}{n}}} \geq t_2) = 0.025$. Par la table de la loi normale on calcule t_1 ou t_2 et ensuite l'effectif utilisé, on trouve $n = 9558$

DÉPARTEMENT MATHÉMATIQUES

EXAMEN DE STATISTIQUE – I

14.03.2000 - DURÉE 1h30

Exercice 1 .- On considère trois *v.a.* indépendantes qui sont distribuées suivant la loi normale. On suppose que $X_1 \sim \mathcal{N}(\mu_1, 1)$, $X_2 \sim \mathcal{N}(\mu_2, 1)$, $X_3 \sim \mathcal{N}(\mu_1+\mu_2, 1)$. On a réalisé une mesure de chaque *v.a.* et on a obtenu les résultats x_1, x_2, x_3.

Calculer l'*E.M.V.* de μ_1 et μ_2.

Exercice 2 .- La longueur des fibres fabriquées par un atelier suit une loi $N(\mu, \sigma^2)$, les paramètres μ et σ étant inconnus. On a prélevé un échantillon et on a calculé

$$\sum_{i=1}^{9} (x_i - \overline{x})^2 = 45.$$

1. Comparer les intervalles de confiance pour σ^2 avec risque $\alpha = 0.12$:
   (a) bilatéral symétrique ;
   (b) bilatéral dissymétrique . $\alpha_1 = 0,04$ et $\alpha_2 = 0,08$.
   (c) unilatéral à droite, de la forme $]a, +\infty[$.
2. Quelle aurait dû être la taille de l'échantillon, dans le cas d'un intervalle *bilatéral symétrique*, pour que la longueur de cet intervalle soit inférieure à 10 ?

Exercice 3 .- On considère la population des ménages français répartis en deux catégories de communes : urbaines et rurales. Soient p_1 et p_2 les proportions dans chaque catégorie de communes, des ménages possédant une maison. On a réalisé un sondage avec remise, dans chaque catégorie, sur n_1 ménages respectivement.

1. **On suppose $p_1 = p_2 = p$ et on considère l'estimateur $\widehat{p(a)} = a\dfrac{X_1}{n_1} + (1-a)\dfrac{X_2}{n_2}, \ 0 \leq a \leq 1.$**

   Étudier les propriétés de cet estimateur.
2. Pouvez-vous l'améliorer ?

Exercice 4 .- Un fabricant de câbles lance une campagne promotionnelle affirmant que la résistance moyenne de ses câbles est de 4450 kilogrammes. Sur un échantillon de 12 articles testés, on a calculé une résistance moyenne de 4500 kilogrammes avec un écart-type de 14. 5 kilogrammes. Peut-on maintenir les affirmations du fabricant au risque de 4% ?

DÉPARTEMENT MATHÉMATIQUES

EXAMEN DE STATISTIQUE I

Quelques Remarques

14.03.2000

Remarque générale : Dans les annales il y a deux catégories de corrigés :

1. Corrigés faits et distribués par le professeur.
2. Corrigés faits par les élèves ou photocopiés des copies des élèves.

Les seules corrigés garantis sans erreurs sont ceux faits par le professeur concerné.

Par contre dans les corrigés faits par les élèves aussi bien dans les copies il y a plusieurs fautes et inexactitudes.

Exercice 1 .-

1. La fonction de vraisemblance pour l'échantillon (x_1, x_2, x_3) est :

$$L(\theta; x_1, x_2, x_3) = Ke^{-\frac{(x_1-\mu_1)^2}{2\sigma^2}} e^{-\frac{(x_2-\mu_2)^2}{2\sigma^2}} e^{-\frac{(x_1-\mu_1-\mu_2)^2}{2\sigma^2}}$$

$$= Ke^{-(x_1-\mu_1)^2} e^{-\frac{(x_2-\mu_2)^2}{2\sigma^2}} e^{-\frac{(x_1-\mu_1-\mu_2)^2}{2\sigma^2}}$$

Il suffit ensuite de calculer les dérivées partielles par rapport à $\theta_1 = \mu_1$ et $\theta_2 = \mu_2$.

2. On obtient un système de deux équations avec deux inconnues μ_1, μ_2
$\mu_1 = \frac{(x_1-\mu_2+x_3)}{2}$ et $\mu_2 = \frac{(x_2-\mu_1+x_3)}{2}$
La résolution de ce système donne les E.M.V pour μ_1 et μ_2.

Exercice 2 .

1. La courbe de la loi du $\mathcal{X}^2(\nu)$ n'a pas d'axe de symétrie.
2. Il n'y a pas de points négatifs.
3. Pour l'intervalle unilatérale, voir *p*.58 du polycopié.
4. Pour calculer la taille minimale on utilise la formule du polycopié, page 61 . Il suffit d'observer que l'intervalle diminue quand les d.d.l. augmentent. On doit calculer les points t_1, t_2 pour $n = 9, 10, 11, .., \nu = 8, 9, 10, ...$ et recalculer ensuite $n * (s^2/t_2 - s^2/t_1)$.

Exercice 3 .

1. $E(X_1) = n_1p_1$ et $E(X_2) = n_2p_2$. On démontre ensuite que l'estimateur $\widehat{p(a)}$ est sans biais et convergent.
2. Pour l'efficacité, il faut étudier l'expérience de $n = (n_1 + n_2)$. D'après l'énoncé $X_1 \sim B(n_1, p)$ et $X_2 \sim B(n_2, p)$.
On doit d'abord :
   (a) calculer l'information de Fisher pour l'échantillon (X_1, X_2) , $I_2(\vartheta)$, calculer ensuite la BCR.
   (b) en égalisant avec la variance de l'estimateur $\widehat{p(a)}$ on constate que la $V(\widehat{p(a)})$ est égale à la **BCR** pour $a = \frac{n}{n_1+n_2}$.
3. Pour l'améliorer, il suffit de minimiser la fonction de la variance. On constate que $V(\widehat{p(a)})$ est minimale pour $a = \frac{n_1}{n_1+n_2}$, ce qui est cohérent avec le résultat de la question précédente.

Exercice 4 .

1. La moyenne de la loi est $E(X) = 4450kg$.
2. La moyenne empirique est $\overline{x} = 4500$ et la variance expérimentale $s^2 = 145^2$. On doit expliquer si on prend
$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ ou bien $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
3. L'espérance μ de la loi étant connue, il n'est pas raisonnable d'essayer de faire un I.D.C. pour μ.
On construit un I.D.C. seulement pour le(s) paramètre(s) inconnus d'une loi $(\mu, \sigma^2, ..)$.
Pour les quantités expérimentales, on construit tout simplement un intervalle (e.g. pour $\overline{X}$).
4. Ne pas confondre $\overline{X}$ et $\overline{x}$.
$\overline{x} = 4500$, est une réalisation de la v.a. $\overline{X}$.

15

16/

Applications et exercices de révision.
18/11/2002.

ex1 corrigé

1. $\Omega = \{(1,1), \ldots (6,1), \ldots, (6,6)\}$; card $\Omega = 36$. $P(\omega_i) = \frac{1}{36}$
$\forall i = 1, 2, \ldots, 36$.

$(\Omega, \mathcal{A}, P)$: $\mathcal{A} = \{\forall$ les parties de $\Omega\}$.

a) X : gain.

X	2	3	4	5	6	7	8	9	10	11	12
P(X=x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
y	25	16	9	4	1	0	1	4	9	16	25

(I)

$F(x) = P(X \leq x)$; $F(0) = 0$, $F(1) = 0$, $F(2) = \frac{1}{36}, \ldots, F(12) =$

b) $Y = (X-7)^2$:

Y (y_i)	0	1	4	9	16	25
P(Y=y)	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

$F(y) = P(Y \leq y)$. $F_Y(0) = \frac{6}{36}, \ldots, F_Y(25) = 1 \ldots$

2. On a joué n fois ; n_i = nb de fois où on a obtenu le total i. et $n = n_2 + n_3 + \cdots + n_{12}$.

a) le gain moyen : $\frac{2n_2}{n} + \frac{3n_3}{n} + \cdots + 12\frac{n_{12}}{n}$

$\Rightarrow P(X=2) = \frac{n_2}{n}$; $P(X=3) = \frac{n_3}{n}, \ldots$

$\Rightarrow E(X) = \sum x_i p_i = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \cdots \cdot 12 \cdot \frac{1}{36} = 7$

b) on fait pareil .. $E(Y) = \sum_{i=1}^{6} y_i p_i = \sum_{i=1}^{11} (x_i - 7)^2 P(X = x_i$

3. On lance n dés :

a) $Z \sim N(\mu, \sigma^2)$. $\mu = 120$ $\sigma^2 = 10^2$.
$P(Z > 150) = P\left(\frac{Z-120}{10} > \frac{150-120}{10}\right) = 1 - F(3) = 0{,}00$

b) $P(|Z-120| \geq 20) = 1 - P(|Z-120| \leq 20) = 1 - P(-20 \leq Z-120 \leq$
$= 1 - F(2) + F(-2) = \ldots = 0{,}0455$

c) $P(\bar{X} \geq 105)$ ou $P(\bar{X} \leq 105)$ .. centrer réduire et voir la proba - si elle correspond ou non

d) à 5%

$P(\bar{X} \geq 125) = \cdots$ centrer réduire etc.

e) $\bar{X}$: moyenne expérimentale lundi, $\bar{Y}$ = moyenne expér. mardi
$\bar{X} \sim N(\mu_{\bar{X}}, \sigma^2_{\bar{X}})$; $\bar{Y} \sim N(\mu_{\bar{Y}} ; \sigma^2_{\bar{Y}}) \Rightarrow \bar{X} - \bar{Y} \sim N(\mu_{\bar{X}-\bar{Y}}, \sigma^2_{\bar{X}-\bar{Y}}$

$\Rightarrow P(\bar{X} - \bar{Y} \ldots$ etc

4. On cherche n t.q. $P(|\bar{X} - \mu| \leq 0{,}1) = 0{,}95$
$\Rightarrow P(-0{,}1 \leq \bar{X} - \mu \leq 0{,}1) = 0{,}95$; centrer réduire etc
on trouve $n \geq 384{,}18$.