Série 6 - loi normale, intervalles de confiance, variance d'échantillon, distribution du Student

Thèmes: distribution normale et intervalle de confiance, taille d'échantillon lors de la loi normale et variance connue, statistique t lors de variance calculée par la variance d'échantillon, distribution de Student, écarts de moyenne, stabilité de la loi exponentielle, distribution exponentielle et χ^2, fonction génératrice des moments.

Exercice 1

On suppose dans cet exercice que la distribution des erreurs suit une loi normale $\mathscr{N}(\mu, \sigma^2)$. Dans une entreprise, 16 personnes sont affairées à des tâches répétitives et fastidieuses sujettes à erreurs. Le nombre d'erreurs moyen par jour est de $\mu = 5.6$ avec un écart-type de $\sigma = 3.6$. On demande

- de construire un intervalle de confiance pour la moyenne d'échantillon du nombre d'erreurs,
- d'établir une limite supérieur de 90 % pour le nombre d'erreurs.

Indication, solution partielle: [3.836; 7.364], 6.761. Pour la dernière valeur utiliser une seule queue.

Exercice 2

Supposons à des fins de simplification, que l'on utilise la règle (avec la notation $s_{\bar{x}} \triangleq x/\sqrt{n}$)

$$\bar{x} \pm 2s_{\bar{x}}$$

quelle que soit la taille n de l'échantillon, pour calculer des limites de confiance à 95 %.

- pour quelle taille d'échantillon est-ce que la probabilité attachée à notre limite est autour des 90 % ?
- Avec la règle $\bar{x} \pm 2.6s_{\bar{x}}$ pour les limites à 99 %, pour quelle taille d'échantillon est-ce que l'on a une probabilité de moins de 97.5 ?
- sous les mêmes hypothèses, on demande la taille d'échantillon pour une probabilité de moins de 95 % ?

Remarque: Pour des petits échantillon, s n'est pas un bon estimateur de σ et il faut agrandir les intervalles de confiance pour permettre à s de se retrouver très loin de σ.

Indication, solution partielle: On a introduit la notation $s_{\bar{x}} \triangleq x/\sqrt{n}$ pour décrire l'écart type standard. Les réponses sont $n \leq 5$, $n \leq 11$, $n \leq 5$.

Exercice 3

Un fabriquant aimerait déterminer la valeur moyenne de rupture μ d'une corde à la précision du kilogramme, que nous interprétons comme un intervalle — de confiance à 95 % — de longueur 2 [kg] pour μ. Si 10 mesures préliminaires ont été effectuées pour déterminer $\sum_i (x_i - \bar{x})^2 = 80$, combien de mesures additionnelles sont-elles nécessaire ?

Indication, solution partielle: Admettre que n est grand dans un premier temps et que la charge de rupture est distribuée selon une loi normale. Utiliser s comme estimée de σ.

Exercice 4

Six automobiles de marque différente ont été utilisées pour tester deux marques différentes A et B de pneumatiques. Chaque voiture a été équipée de la marque A et conduite sur une route sinueuse et difficile (identique pour chaque voiture) jusqu'à ce que les pneus ne soient plus opérationnels. Ensuite, l'opération est répétée mais avec les pneus de la marque B. Les durées de fonctionnement suivantes ont été récoltées :

voiture	1	2	3	4	5	6
marque A	18	23	16	27	19	17
marque B	15	22	16	21	15	16

- Tester si les données sont consistantes avec l'hypothèse que la durée de vie moyenne est identique pour les deux marques A et B.
- On constate en observant les données que pour la marque A, toutes les voitures ont roulé un peu plus que pour la marque B. Suggérer une modification de l'expérience pour éventuellement améliorer la procédure.

Indication, solution partielle: Admettre que les différences sont normales $\mathscr{N}(\mu, \sigma^2)$, et calculer le niveau de signification en utilisant la table de Student.

Exercice 5

27 mesures de rendement ont été récoltées dans un processus industriel, avec les résultats suivants

processus 1	$n = 11$	$\bar{x} = 6.23$	$s_1^2 = 3.79$
processus 2	$n = 16$	$\bar{y} = 12.74$	$s_2^2 = 4.17$

En admettant que les rendements sont distribués selon des lois normales de même variance, déterminer un intervalle de confiance à 95 % pour les moyennes μ_1 et μ_2, ainsi que pour la différence des moyennes $\mu_1 - \mu_2$.

Indication, solution partielle: Comme $n_1 \neq n_2$, déterminer une moyenne pondérée σ^2 à partir de σ_1^2 et σ_2^2 pour l'intervalle de confiance pour μ_1 et pour μ_2. En ce qui concerne l'intervalle de confiance pour $\mu_1 - \mu_2$, utiliser le résultat de la dernière slide du cours qui donne avec $\phi \triangleq \mu_1 - \mu_2$

$$\hat{\phi} = \bar{X}_1 - \bar{X}_2 \sim \mathscr{N}\left(\phi, \frac{1}{n}\sigma_1^2 + \frac{1}{m}\sigma_2^2\right)$$

Une explication intuitive de cette formule est donnée par le fait que les estimateurs de chacune des moyenne $\bar{X}_1$ et $\bar{X}_2$ ont chacun un étalement respectif correspondant aux variances respectives de $\frac{1}{n}\sigma_1^2$ et $\frac{1}{m}\sigma_2^2$. Lorsque on fait la différence $\bar{X}_1 - \bar{X}_2$ ces deux estimateurs peuvent donner des signes qui additionne les écarts. L'étalement de l'estimateur $\hat{\phi}$ correspond donc à l'addition des variances de chacun des estimateur $\bar{X}_1$ et $\bar{X}_2$.

$$6.23 \pm 1.25 \qquad 12.74 \pm 1.03 \qquad -6.51 \pm 1.62$$

Exercice 6

Cet exercice concerne la stabilité de la distribution du χ^2 lorsque celle-ci est convoluée. Elle conserve son type, elle demeure χ^2 mais les degrés de libertés augmentent.

On demande de démontrer à l'aide des fonctions génératrices des moments que lorsque on additionne plusieurs variables aléatoires distribuées selon une loi du χ^2, le résultat est une variable aléatoire qui obéit à une loi en χ^2. De manière plus précise, soit X_1, X_2, ..., X_n, des variables aléatoires distribuées selon $\chi^2_{(\nu_1)}$, $\chi^2_{(\nu_2)}$, ..., $\chi^2_{(\nu_n)}$ alors

$$Z = \sum_{k=1}^{n} X_i \sim \chi^2_{(\sum_i \nu_i)}$$

Utiliser la relation entre les fonctions génératrices des moments et la convolution.

On sait que la densité de probabilité d'une somme de variables aléatoires et la convolution des densités de probabilité de chacune des variables aléatoires. Ensuite utiliser la correspondance entre la fonction génératrice de la densité du χ^2 et observer l'élégante association. Rappel de la définition, de la propriété des convolutions et de la fonction génératrice des moments $M_X(u)$. Soit X une variable aléatoire de densité de probabilité $f(x)$, on a

$$M_X(u) \triangleq \mathbb{E}[e^{uX}] = \int_{-\infty}^{+\infty} e^{sx} f(x)\, dx$$

Propriétés: $Z = X + Y$ avec h, f et g resp. pour les densités de probabilité

$$h = f * g \leftrightarrow M_Z(u) = M_X(u)\, M_Y(u)$$

Et pour un $\chi^2_{(\nu)}$ on a

$$\frac{1}{2^{\nu/2}\Gamma(\nu/2)} \cdot x^{\frac{\nu}{2}-1} \cdot e^{\frac{x}{2}} \leftrightarrow (1-2u)^{-\nu/2}$$

Exercice 7

- Montrer que si X est une variable aléatoire exponentielle de moyenne θ alors $\frac{2X}{\theta}$ est distribué selon une loi $\chi^2_{(2)}$, i.e.

$$X \sim \text{Exp}\left(\frac{1}{\theta}\right) \Rightarrow \frac{2X}{\theta} \sim \chi^2_{(2)}$$

- Définir

$$T = X_1 + X_2 + \ldots + X_n$$

avec X_i des variables aléatoires indépendantes de distribution exponentielle de même moyenne θ. Montrer que

$$\frac{2T}{\theta} \sim \chi^2_{(2n)}$$

- La durée de vie d'un composant électronique est distribuée selon une loi exponentielle de durée moyenne de 100 heures. 15 composants sont commandés et 1 de ces composants est mis en service. Dès que celui-ci cesse de fonctionner, il est remplacé par un autre composant du lot. Quelle est la probabilité que l'ensemble conduit à une durée de fonctionnement d'au moins 2000 heures ?