dénombrement et probabilités élémentaires

1. Le responsable de l'entretien d'un immeuble doit remplacer deux lampes dans un bureau. Il descend au sous-sol chercher deux nouvelles lampes dans une boîte de 24 dont deux sont défectueuses, puis il remonte les tester. Quelle est la probabilité qu'il doive redescendre ?

corrigé succint :

Le responsable devra redescendre si et seulement si il choisit au moins une lampe défectueuse. On peut calculer plus facilement la probabilité que les deux lampes choisies fonctionnent : le nombre de choix est $\binom{22}{2} = \dfrac{22 \times 21}{2} = 231$, alors que le nombre de choix total est $\binom{24}{2} = \dfrac{24 \times 23}{2} = 276$.

La probabilité qu'à le responsable de devoir redescendre est donc de $1 - \dfrac{231}{276} = \dfrac{45}{276}$, soit approximativement 0.163.

2. Deux empaqueteuses fonctionnent indépendemment. Pour une journée donnée, la probabilité que l'empaqueteuse 1 soit en panne est de 0,025, et de 0,02 pour l'empaqueteuse 2. Quelle est la probabilité pour qu'aucune ne fonctionne ? Pour que les deux fonctionnent simultanément ?

corrigé succint :

Soit A l'événement "l'empaqueteuse 1 tombe en panne" et B l'événement "l'empaqueteuse 2 tombe en panne". On a $p(A) = 0.025$ et $p(B) = 0.02$.

(a) L'événement "aucune empaqueteuse ne fonctionne" est $A \cap B$. A et B sont indépendants, donc $p(A \cap B) = p(A) \times p(B) = 0.0005$.

(b) L'événement "les deux empaqueteuses fonctionnent simultanément" est $\bar{A} \cap \bar{B}$, et donc $p(\bar{A} \cap \bar{B}) = p(\bar{A})p(\bar{B}) = (1 - p(A))(1 - p(B)) = 0.975 \times 0.98 = 0.9555$

3. Lors d'un jeu télévisé, on propose à un candidat trois enveloppes identiques ; une seule d'entre-elles contient un chèque. Le candidat commence par désigner une des enveloppes, puis le présentateur ouvre alors une autre enveloppe, qui est vide. Le candidat peut alors choisir d'ouvrir l'enveloppe qu'il avait initialement désignée, ou bien ouvrir la troisième enveloppe. Quel est son intérêt ?

corrigé succint :

(objet d'un jeu télévisé américain, ce « problème » avait dans les années 50 passionné le grand public)

Initialement, le candidat a une chance sur trois de gagner. Cela reste évidemment le cas s'il applique comme stratégie de ne rien changer.

S'il applique la stratégie de changer systématiquement d'enveloppe, il transforme les gains en pertes et les pertes en gain, et par conséquent sa probabilité de gagner est de $2/3$.

Ainsi, le candidat a intérêt à ouvrir systématiquement la troisième enveloppe.

4. Dans une boîte il y a 12 pièces : une vis et un écrou de diamètre d_1, une vis et un écrou de diamètre d_2, . . ., une vis et un écrou de diamètre d_6.
On choisit deux pièces au hasard. Calculer la probabilité :
(a) De pouvoir reconstituer un boulon.
(b) Qu'il s'agisse d'un vis et d'un écrou.

corrigé succint :

(a) Il y a donc $\binom{12}{2} = 66$ tirages possibles.
   i. 6 tirages sont constitués de pièces de même diamètre, donc la probabilité est de $6/66 = 1/11$.
   ii. Choisir un tirage constitué d'un vis et d'un écrou revient à choisir une vis parmi 6 et un écrou parmi 6. La probabilité cherchée est ainsi $\dfrac{\binom{6}{1} \times \binom{6}{1}}{\binom{12}{2}} = 6/11$.

5. Durant un jeu avec paquet de 32 cartes, chacun des deux joueurs reçoit 5 cartes : 2 connues de tous et 3 connues de lui seul.
1) Quelle est la probabilité pour qu'un joueur ait un carré d'as ?
2) Vous avez vu que votre adversaire a reçu deux as ; de votre côté vous n'en avez aucun. Quelle est la probabilité que votre adversaire ait un carré d'as en main ?

corrigé succint :

1) Il y a $\binom{32}{5}$ mains possibles. Parmi celles-ci, 28 contiennent un carré d'as (28 choix possibles pour l'unique carte qui n'est pas un as). Donc la probabilité est $\dfrac{28}{\binom{32}{5}} = 1/7192$ soit environ $1,39 \times 10^{-4}$ soit 0.013%.

2) On ne s'intéresse qu'aux 3 cartes inconnues, à choisir parmi $25 = 32 - 5 - 2$ (vos 5 cartes et les 2 as connus). Il y a donc $\binom{25}{3}$ choix possibles, et parmi ceux-ci, 23 contiennent les deux as manquants (choix des 2 as et d'une seule autre carte parmi les 23 qui ne sont pas des as). La probabilité est donc de $\dfrac{23}{\binom{25}{3}}$ soit exactement 1%.

6. On lance deux fois un dé bien équilibré.
(a) Quelle est la probabilité que la somme des points soit strictement supérieure à 10 sachant que le premier résultat est 6 ? Sachant qu'un des résultats est 6 ?
(b) Sachant que la somme des points vaut 6 calculer la probabilité pour que l'un des dés ait donné un 2.

corrigé succint :

(a) Si l'on sait que le premier résultat est 6, on peut prendre pour univers l'ensemble des tirages dont le premier résultat est un 6, de cardinal 6 (on a six tirages possibles en tout, $(6,1),(6,2),\ldots,(6,6)$).

L'événement "la somme est strictement supérieure à 10" est constitué des deux issues $(6,5)$ et $(6,6)$. Ainsi, la probabilité cherchée est $2/6 = 1/3$.

Si on sait que l'un des dés a donné 6, l'univers est constitué de tous les tirages comportant au moins un 6, soit $(6,1),(6,2),\ldots,(6,6),(1,6),\ldots,(5,6)$, de cardinal 11 (attention, $(6,6)$ n'apparaît qu'une seule fois).

Les tirages pour lesquels la somme est strictement supérieure à 10 sont $(6,6)$, $(6,5)$ et $(5,6)$, et la probabilité est donc $3/11$.

(b) On note A l'événement « l'un des dés a donné un 2 » et B l'événement « la somme des points fait 6 ».

Il s'agit de calculer $p(A|B) = p(A \cap B)/p(B)$.

Mais $A \cap B = \{(2,4);(4,2)\}$ est de cardinal 2, et $B = \{(1,5);(5,1);(4,2);(2,4);(3,3)\}$ est de cardinal 5, donc la probabilité cherchée est de $\boxed{2/5.}$

7. Deux machines M_1 et M_2 produisent quotidiennement respectivement 1000 et 3000 pièces. Le taux de pièces défectueuses est de 2% pour M_1 et de 5% pour M_2.
Quelle est la probabilité qu'une pièce défectueuse provienne de M_1 ?

corrigé succint :

On tire une pièce au hasard : on note D l'événement « la pièce est défectueuse », et M_i l'événement « la pièce provient de la machine M_i », $i = 1, 2$.

$$p(D) = \frac{0.02 \times 1000 + 0.05 \times 3000}{1000 + 3000} = \frac{17}{400}, p(M_1) = \frac{1}{4}, p(M_2) = \frac{3}{4}.$$

On sait que $p(D|M_1) = 0.02$, donc $p(D \cap M_1) = 0.02 \times p(M_1) = 0.005$. Et donc

$$p(M_1|D) = \frac{p(D \cap M_1)}{p(D)} = \boxed{\frac{2}{17} \simeq 11.76\%.}$$

variables aléatoires

8. 5% des interrupteurs sortant d'une chaîne de production sont défectueux. On en prend deux au hasard. Soit X la variable aléatoire « nombre d'interrupteurs défectueux dans l'échantillon prélevé ».
Donner la loi de probabilité de X, calculer $E(X)$ et $\mathrm{Var}(X)$.

corrigé succint : Les trois résultats possibles sont :

« deux interrupteurs défectueux », de probabilité $0.05^2 = 0.0025$,

« aucun interrupteur défectueux », de probabilité $0.95^2 = 0.9025$,

« un seul interrupteur défectueux », de probabilité $1 - 0.95^2 - 0.05^2 = 0.095$.

Alors $\boxed{E(X) = 0.0025 \times 2 + 0.095 = 0.1,}$ $E(X^2) = 0.0025 \times 4 + 0.095$, $E(X) = 0.105$, et donc $\boxed{\mathrm{Var}(X) = E(X^2) - E(X)^2 = 0.095.}$

Remarque : on peut aussi voir que X suit une loi binomiale de paramètres $n = 2$ et $p = 0.05$, donc $E(X) = np = 0.1$ et $\mathrm{Var}(X) = np(1-p) = 0.1 \times 0.95 = 0.095$.

9. On donne la distribution conjointe de deux variables discrètes :

$X\|Y$	0	1	2
10	0	0.25	0
12	0.25	0	0.25
14	0	0.25	0

(a) déterminer les distributions marginales, puis $E(X)$ et $E(Y)$.

(b) Déterminer $E(X.Y)$ et la covariance de X et Y.

(c) les deux variables sont-elles indépendantes ?

corrigé succint : Pour trouver les $p(X = i)$ on additionne les coefficients $p(X = i, Y = j)$ pour $j = 0, 1, 2$: $\boxed{p(X = 10) = 0.25, p(X = 12) = 0.5, p(X = 14) = 0.25.}$

De même, $\boxed{p(Y = 0) = 0.25, p(Y = 1) = 0.5, p(Y = 2) = 0.25.}$

$E(X.Y) = 10p(X = 10, Y = 1) + 12p(X = 12, Y = 1) + 14p(X = 14, Y = 1) + 10p(X = 10, Y = 2) + 12p(X = 12, Y = 2) + 14p(X = 14, Y = 2) = 0.25 \times 10 + 0.25 \times 14 + 0.25 \times 24 = 12.$

$E(X) = 12$ et $E(Y) = 1$, et donc $E(X.Y) = E(X).E(Y)$, mais pourtant $\boxed{X \text{ et } Y \text{ ne sont pas indépendantes :}}$ on voit par exemple que $p(X = 10, Y = 0)$ n'est pas égal à $p(X = 10)p(Y = 0)$.

10. X suit une loi normale de paramètres 2 et 4.
Calculer $p(X = 2)$, $p(X > 3)$, $p(-1 \leq X \leq 4)$, $p(-2 \leq X \leq 6)$.

corrigé succint :

Si $Z = \dfrac{X-2}{2}$, Z suit une loi normale centrée réduite.

$p(X = 2) = 0$ car X est continue.

$p(X > 3) = p(Z > 1/2) = 1 - p(Z < 1/2) = 1 - 0.6915 = 0.3085$,

$p(-1 \leq X \leq 4) = p(-3/2 \leq Z \leq 1) = p(Z \leq 1) - p(Z \leq -3/2) = p(Z \leq 1) - (1 - p(Z \leq 3/2)) = p(Z \leq 1) - 1 + p(Z \leq 3/2) = 0.8413 - 1 + 0.9332 = 0.7745$

Pour $p(-2 \leq X \leq 6)$ on peut refaire un calcul analogue au précédent, ou simplement remarquer que c'est la probabilité d'être dans un intervalle centré sur l'espérance et de "rayon" 2 écarts-types : la probabilité est donc 95.4%.

11. Dix boites sont posées dans une pièce. Un objet est caché dans l'une d'entre elles, et un joueur ouvre une boîte au hasard pour trouver l'objet.

    Il effectue une série de 32 essais pour le retrouver (à chaque essai, réussi ou non, l'objet est caché à nouveau hors de la présence du joueur).

    (a) Quelle est la loi de la variable X « nombre de fois où l'objet est trouvé » ?

    (b) Déterminer l'espérance et la variance de X.

    (c) Calculer la probabilité d'avoir 0 succès, 3 succès, moins de 3 succès, plus de 10 succès.

    (d) À partir de combien de succès peut-on considérer le résultat comme extraordinaire (c'est-à-dire qu'un tel nombre de succès n'est obtenu que pour 1% des joueurs) ?

    corrigé succint :

    (a) Il s'agit d'une $\boxed{\text{loi binomiale de paramètres 32 et 0.1}}$ (on répète 32 fois l'opération qui consiste à choisir au hasard une boite, qui a une chance sur 10 de contenir l'objet, et on compte le nombre total de succès).

    (b) Son espérance est $\boxed{E(X) = np = 3.2}$ et sa variance $\boxed{\text{Var}(X) = np(1-p) = 2.88.}$

    (c) $p(X = 0) = 0.034$, $p(X = 3) = 0.234$, $p(X \leq 3) = 0.600$, $p(X \geq 10) = 0.0008$.

    Comme $n \geq 30$, $p \leq 0.1$ et $np = 3.2 \leq 5$, les conditions énoncées dans le cours sont vérifiées, et on utilisera effectivement l'approximation par une $\boxed{\text{loi de Poisson de paramètre } \lambda = np = 3 \text{ car on n'a pas la table pour la loi de paramètre 3.2.}}$

    En utilisant l'approximation par une loi de Poisson de paramètre 3, on trouve donc :

    — $\boxed{p(X = 0) = 0.050}$

    — $\boxed{p(X = 3) = 0.022}$

    — $\boxed{p(X \leq 3) = 0.647}$

    — $p(X \geq 10) = 1 - p(X < 10) = 1 - p(X \leq 9) = 1 - 0.9989$, $\boxed{p(X \geq 10) = 0.001}$

    — En utilisant la table de la loi de Poisson on constate que la probabilité d'obtenir moins de 7 essais est 0.988, d'obtenir moins de 8 essais est 0.996. Donc la probabilité d'obtenir 9 essais ou plus est inférieure à 1%.

12. Pour se prémunir contre les 10% défections tardives habituellement constatées, une compagnie aérienne pratique la surréservation : elle vend 270 billets pour 250 sièges dans un avion.

    Soit X la variable aléatoire « nombre de personnes ayant réservé qui se présentent pour embarquer ».

    (a) Montrer que X suit une loi binomiale, et que l'approximation par une loi normale est justifiée.

    (b) Quelle est la probabilité qu'exactement 250 personnes se présentent à l'embarquement ?

    Quelle est la probabilité que toute personne ayant réservé et se présentant soit assurée d'un siège ?

    corrigé succint :

    (a) La probabilité qu'une personne ayant acheté un billet se présente à l'embarquement vaut $\boxed{p = 0.9,}$ puisque la compagnie constate 10% de défections.

    Par conséquent, si on répète 270 fois un "tirage au sort" pour déterminer si un passager se présente (avec une probabilité de 0.9), et l'on compte le nombre de passagers qui se présentent : X suit bien une loi binomiale de paramètres $n = 270$ et $p = 0.9$.

    Comme $n > 50$ et et que $np(1-p) = 24.3$, on peut effectivement approcher la loi de X par une $\boxed{\text{loi normale de paramètres } np = 243 \text{ et } np(1-p) = 24.3.}$

    (b) On souhaite déterminer $p(X = 250)$ et $p(X \leq 250)$; si on utilise l'approximation par la loi normale, on calcule en fait $p(249.5 \leq X \leq 250.5)$ et $p(X \leq 250.5)$.

    Pour cela, on se ramène à une loi normale centrée réduite : $p(X \leq 250.5) = p(\frac{X - 243}{\sqrt{24.3}} \leq \frac{7.5}{\sqrt{24.3}}) = p(\frac{X - 243}{\sqrt{24.3}} \leq 1.52) \simeq 0.9357$, de même $p(X \leq 249.5) \simeq 0.905$ Ainsi, $\boxed{\text{la probabilité que toute personne ayant réservé ait un siège est de 93.57\%}}$, et la probabilité qu'exactement 250 personnes se présentent est 0.03.

    (remarque : un calcul exact donne, avec la loi binomiale, $p(X \leq 250) = 94.11\%$; par ailleurs un calcul de $p(X \leq 250)$ (et non 250.5) avec la loi normale donne une moins bonne approximation, 92.22%).

estimation de paramètres

13. Dans un grand lot de pièces circulaires, on a prélevé au hasard 40 pièces dont on vérifie le diamètre. Les mesures (en cm) sont :

4.9	5.0	5.2	4.7	4.8	5.1	4.5	5.2	4.9	4.8
4.9	4.9	4.9	5.3	5.0	4.8	4.8	4.9	5.1	5.3
5.4	4.9	4.9	5.0	4.8	4.8	5.3	4.8	5.1	5.0
5.1	4.8	4.7	5.0	4.9	4.8	4.6	4.7	4.9	4.7

    Estimer par un intervalle de confiance 95% le diamètre moyen.

corrigé succint :

On trouve les estimations ponctuelles $\boxed{\bar{x} = 4.93}$, $\boxed{s = 0.202}$ et $\boxed{s^2 = 0.0406.}$

L'effectif de l'échantillon étant « grand » (supérieur à 30), on peut considérer que $\sqrt{n}\dfrac{\bar{X} - \mu}{s}$ suit une loi normale centrée réduite.

Par conséquent, l'espérance de $\bar{X}$ sera dans 95% des cas dans l'intervalle $[\bar{x} - 1.96s/\sqrt{n}, \bar{x} + 1.96s/\sqrt{n}]$, soit ici dans l'intervalle $\boxed{[4.87, 4.99].}$

(complément hors programme : estimation par intervalle de confiance de la variance : **Si** X **suit une loi normale**, on sait alors que $(n-1)s^2/\sigma^2$ suit une loi du χ^2 à $n-1$ degrés de liberté, donc σ^2 est avec une probabilité de 95% dans l'intervalle $[(n-1)s^2/c_1^2, (n-1)s^2/c_2^2]$ pour $c_1^2 \simeq 59.342$ et $c_2^2 \simeq 24.433$ (lus dans la table du χ^2 à 40 degrés de liberté, car le formulaire ne donne pas la table à 39 degrés de liberté), soit $\boxed{\sigma^2 \in [0.0267, 0.0648].}$ Si on ne sait pas que X suit une loi normale, on ne peut rien dire...)

14. Des essais en laboratoire sur 20 lampes miniatures donnent les durées de vie suivantes, en heures : 451, 412, 412, 375, 407, 454, 375, 393, 355, 364, 414, 413, 345, 432, 392, 329, 439, 381, 451, 413.

    On suppose la durée de vie distribuée normalement.

    Estimer par un intervalle de confiance 95% la durée de vie moyenne.

    corrigé succint : Les estimations ponctuelles de l'espéance, de l'écart-type et de la variance sont respectivement $\boxed{\bar{x} = 400.35}$, $\boxed{s = 36.01}$ et $\boxed{s^2 = 1297.}$

    Alors $\sqrt{n}\dfrac{\bar{X} - \mu}{s}$ suit une loi de Student à 19 degrés de liberté, et donc l'espérance de $\bar{X}$ sera dans 95% des cas dans l'intervalle $[\bar{x} - t(0.95)s/\sqrt{n}, \bar{x} + t(0.95)s/\sqrt{n}]$, et on lit $t(0.95) = 2.093$ dans la table de la loi de Student à 19 degrés de liberté, donc l'intervalle cherché est $\boxed{[383.5, 417.2].}$

    (complément hors-programme : estimation de l'écart-type par intervalle de confiance : X suit une loi normale, donc $(n-1)s^2/\sigma^2$ suit une loi du χ^2 à $n-1$ degré de liberté, donc σ^2 est avec une probabilité de 95% dans l'intervalle $[(n-1)s^2/c_1^2, (n-1)s^2/c_2^2]$ pour $c_1^2 \simeq 32.8523$ et $c_2^2 \simeq 8.9065$ (lus dans la table du χ^2 à 19 degrés de liberté), soit $\sigma^2 \in [750.11, 2766.86]$, et donc l'écart-type a 95% de chances de vérifier $\boxed{\sigma \in [27.39, 52.6].}$)

15. (a) Sur un échantillon de 60 pièces fabriquées par une machine, 12 sont défectueuses. Trouver un intervalle de confiance à 95% de la proportion réelle de pièces défectueuses fabriquées par la machine.

    (b) Même question avec 60 pièces défectueuses sur 300.

    corrigé succint :

    (a) On admet que si la proportion de pièces défectueuses dans la population est p, la proportion F de pièces défectueuses dans un échantillon d'effectif n suit approximativement une loi normale $N(p, \dfrac{p(1-p)}{n})$. Par conséquent, $\dfrac{F-p}{\sqrt{p(1-p)/n}} \sim N(0,1)$, et donc dans 95% des cas, $-1.96 \leq \dfrac{F-p}{\sqrt{p(1-p)/n}} \leq 1.96$.

    Si on approche l'écart-type $\sqrt{p(1-p)/n}$ par son estimation ponctuelle $\sqrt{0.2 * 0.8/40} = 0.0516$, on obtient $p \in [0.2 - 1.96 \times 0.0632, 0.2 + 1.96 \times 0.0632]$, soit l'intervalle $\boxed{[0.099, 0.301].}$

    Cet intervalle n'a pas grand intérêt en pratique : savoir que la proportion de pièces défectueuses se situe entre 9.9% et 30.1% n'est pas un encadrement très utile...

    De plus, on a ici $np(1-p) = 9.6$ inférieur à 18, donc même l'hypothèse de pouvoir faire l'approximation par la loi normale est douteuse : l'échantillon est trop petit, et l'intervalle de confiance 95% réel sans doute encore plus large...

    (b) De même, avec $n = 300$, on trouve l'intervalle de confiance 95% $\boxed{[0.155, 0.245]}$, large mais quand même plus "intéressant".

    Logiquement, l'augmentation de n aboutit à un encadrement plus précis.

tests d'hypothèses

16. On veut savoir si la résistance moyenne de composants produits dans une usine est 400Ω. On considère que la distribution des résistances est normale, et on mesure pour 16 composants les valeurs 392, 396, 386, 389, 388, 387, 403, 397, 401, 391, 400, 402, 394, 406, 406, 400.

    Peut-on considérer, au seuil de signification $\alpha = 5\%$, que le lot respecte la norme de 400Ω ? Même question avec un seuil de $\alpha = 1\%$.

    corrigé succint : On détermine, à partir de l'échantillon, les estimations ponctuelles des espérance, variance et écart-type de la loi : on trouve $\boxed{\bar{x} = 396.125,\ s = 6.742 \text{ et } s^2 = 45.45.}$

    Si l'on fait l'hypothèse H_0 : "le lot respecte la norme de 400Ω", alors dans 95% des cas la moyenne sur un échantillon d'effectif 16 se trouve dans l'intervalle $[400 - t * 6.742/4, 400 + t * 6.742/4]$, t étant lu dans la table de la loi de Student à 15 degrés de liberté : $t = 2.1314$. Ainsi l'intervalle de confiance 95% pour la résistance est $\boxed{[396.40, 403.59], \text{ et on peut donc, au risque } 5\%, \text{ rejeter l'hypothèse.}}$

Au seuil $\alpha = 1\%$, on a dans l'hypothèse H_0 un intervalle de confiance pour la moyenne $[400 - t * 6.742/4, 400 + t * 6.742/4]$, avec $t = 2.9467$. Ainsi, l'intervalle est $[395.03, 404.97]$. Au risque 1%, on ne rejette pas H_0.

17. Un fabricant se vante de proposer des tubes à essai d'une durée de vie supérieure à 2000h de chauffage. A l'aide d'un échantillon de 100 tubes testés, on estime la durée de vie moyenne à 1975h et l'écart-type à 130h. Peut-on affirmer, au risque 5%, que le fabriquant ment ?

corrigé succint : Il s'agit ici d'un test unilatéral...H_0 est l'hypothèse : "la durée de vie moyenne vérifie $\mu \geq 2000$". On peut supposer, l'effectif de l'échantillon étudié étant grand, que $\sqrt{n}\dfrac{\bar{X} - \mu}{s}$ suit une loi normale centrée réduite.
Si H_0 est vérifiée, on cherche t tel que $p(\mu - ts/\sqrt{n} \leq \bar{X}) = 0.95$, soit $p(-t \leq \sqrt{n}\dfrac{\bar{X} - \mu}{s}) = 0.95$, et donc $1 - F(-t) = F(t) = 0.95$: $t = 1.64$.
Ainsi, dans l'hypothèse H_0, la durée de vie moyenne d'un échantillon d'effectif 100 se trouve, dans 95% des cas, dans l'intervalle $[2000 - 1.64 * 130/10, +\infty[= [1978.68, +\infty[$.
La mesure de 1975h sur l'échantillon n'étant pas dans cet intervalle,

H_0 doit être rejetée : il est probable que le fabriquant mente.

18. Un fabricant de médicaments affirme que les masses d'un composant dans ses comprimés sont réparties selon une loi normale d'espérance 75 mg. Les mesures pour le vérifier étant coûteuses, trois seulement sont réalisées, dont les résultats sont 70, 72 et 74 mg. Peut-on, au risque de 5% de se tromper, contester l'affirmation ?

corrigé succint : Notons X la variable aléatoire correspondante et $\mu = E(X)$: il s'agit donc ici d'effectuer un test bilatéral de l'hypothèse $H_0 : \mu = 75$.
On obtient sur un échantillon de 3 mesures : $n = 3$, $\bar{x} = 72$, $\sigma'^2 = 8/3$ et $s^2 = 8/2 = 4$, donc l'estimation ponctuelle de l'écart-type est $s = 2$.
On sait que $\sqrt{n}\dfrac{\overline{X} - \mu}{s}$ suit une loi de Student à 2 degrés de liberté, donc si $\alpha = 0.05$, $t(\alpha) = 4.3027$.
Ainsi, la moyenne des masses du composant, mesurées sur un échantillon d'effectif 3 sera, dans 95% des cas, dans l'intervalle $[75 - 4.3027 \times s/\sqrt{3}, 75 + 4.3027 \times s/\sqrt{3}] = [70.03, 79.97]$.
La valeur moyenne 72 mesurée sur l'échantillon étant bien dans cet intervalle,

on n'a pas de raisons, au vu de ces mesures, de rejeter H_0.

19. Un laboratoire pharmaceutique désire étudier les effets secondaires potentiels d'un médicament sur le taux de cholestérol des patients. Cent volontaires sont donc choisis pour tester le médicament.

(a) Avant l'expérience, leur taux de cholestérol moyen est de 2.02g.L^{-1}, avec un écart-type de 0.2g.L^{-1}
Le taux de cholestérol moyen dans la population étant de 2g.L^{-1}, vérifier que cet échantillon est représentatif au risque 5%.

(b) Après un mois de traitement, seuls 97 volontaires reviennent faire un test. Leur taux moyen de cholestérol est passé à 2.09g.L^{-1} avec un écart-type d'échantillon de 0.25g.L^{-1}.
La différence est-elle significative au risque 5% ? Au risque 1% ?

corrigé succint :

(a) Soit X_1 la variable aléatoire qui mesure le taux de cholestérol d'un individu ; $E(X_1) = \mu_1 = 2$.
$\overline{X_1}$ est le taux moyen mesuré sur un échantillon de taille $n_1 = 100$.
Alors n_1 étant plus grand que 30, on peut considérer que $\sqrt{n_1}\dfrac{\overline{X_1} - 2}{s_1}$ suit une loi normale, avec $s_1 = 0.2$ estimation ponctuelle de l'écart-type de X_1.
Ainsi, dans 95% des cas le taux moyen observé sur un échantillon sera compris dans $[2 - 1.96 \times 0.2/10, 2 + 1.96 \times 0.2/10] = [1.961, 2.039]$.
Le taux de cholestérol moyen des volontaires étant bien dans cet intervalle, on peut considérer que cet échantillon est représentatif.

(b) Soit X_2 la variable aléatoire mesurant le taux de cholestérol d'un individu après un mois de traitement ; son espérance μ_2 est inconnue. $\overline{X_2}$ est le taux moyen d'un échantillon de taille $n_2 = 97$.

On fait l'hypothèse H_0 : « les taux de cholestérol moyens sont les mêmes avant et après traitement ». Alors $\mu_1 = \mu_2$, et on peut considérer que $\dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \sim N(0, 1)$ (avec $s_1 = 0.2$, $s_2 = 0.25$), et par conséquent on détermine l'intervalle de confiance au risque 5% de $\overline{X_1} - \overline{X_2}$: $[-1.96\sqrt{s_1^2/n_1 + s_2^2/n_2}, 1.96\sqrt{s_1^2/n_1 + s_2^2/n_2}] = [-0.063, 0.063]$.
Comme la différence entre les taux moyens mesurés $2.02 - 2.09 = 0.07$ n'est pas dans cet intervalle, elle est significative, et on rejette H_0 donc

on considère, au risque 5% de se tromper, que le médicament a un effet.

En revanche, l'intervalle de confiance au risque 1% est $[-2.57\sqrt{s_1^2/n_1 + s_2^2/n_2}, 2.57\sqrt{s_1^2/n_1 + s_2^2/n_2}] = [-0.083, 0.083]$, intervalle qui contient la valeur $2.02 - 2.09 = 0.07$, donc la différence n'est pas significative au risque 1%.

20. On relève chaque jour pendant 200 jours le nombre d'atterrissages entre 14h et 15h dans un aéroport :

Nb d'atterrissages	0	1	2	3	4	5	6	7	8	9	10
Nb de jours	11	28	43	47	32	28	7	0	2	1	1

(a) Soit X la variable « nombre d'atterrissages par jour entre 14h et 15h ». Donner les estimations ponctuelles de $E(X)$ et Var(X) et estimer $E(X)$ par un intervalle de confiance 95%. Ces résultats sont-ils compatibles avec une loi de Poisson ? Quel serait son paramètre ?

(b) Tester la validité de ce modèle (test du χ^2 au risque 5%).

(c) Calculez la probabilité d'avoir dans cet aéroport, toujours entre 14h et 15h : 0 atterrissage un jour donné, 1 ou 2 atterrissages un jour donné, 2 atterrissages en tout sur 3 jours quelconques.

corrigé succint :

(a) L'effectif de l'échantillon est $n = 200$.

On détermine l'estimation ponctuelle de la moyenne $\bar{x} = 600/200 = 3$, et l'estimation ponctuelle de la variance $s^2 = 596/199 = 2.995$, soit $s = 1.73$.

L'effectif est suffisant pour assimiler la loi de $\sqrt{n}\dfrac{\bar{X}-\mu}{s}$ a une loi normale centrée réduite : l'intervalle de confiance 95% pour la moyenne est donc $[\bar{x} - 1.96\frac{s}{\sqrt{200}}, \bar{x} + 1.96\frac{s}{\sqrt{200}}]$, soit $[2.76, 3.24]$.

(**hors-programme :** On sait que l'intervalle de confiance pour la variance est $[\dfrac{199}{c_1^2}s^2, \dfrac{199}{c_2^2}s^2]$, avec c_1^2 et c_2^2 lus dans la table de la loi du χ^2 à 199 degrés de liberté. En pratique on utilise la table de la loi du χ^2 à 200 degrés de liberté et on lit dans les colonnes $0.025 = 0.05/2$ et $0.975 = 1 - 0.05/2$: $c_1^2 = 241.1$ et $c_2^2 = 162.7$. Ainsi l'intervalle est $[0.825s^2, 1.223s^2] = [2.47, 3.66]$.)

L'espérance et la variance ont (quasiment) la même estimation ponctuelle, égale à trois : les résultats sont compatibles avec le fait que X suive une loi de Poisson de paramètre 3.

(b) On définit l'hypothèse H_0 : « X suit une loi de Poisson de paramètre 3 ».

Pour utiliser un test du χ^2 pour accepter ou refuser H_0, on doit avoir des valeurs (ou classes) d'effectif "théorique" au moins égal à 5. Or ici ce n'est pas le cas : les valeurs 7, 8, 9, 10 ont des effectifs $np(X = 7), np(X = 8), np(X = 9), np(X = 10)$ inférieurs : on doit regrouper en une seule classe les valeurs 7, 8, 9, 10.

Les probabilités des évènements $X = 0$, $X = 1$, ..., $X = 6$, $X \in [7, 10]$ sont déterminées par lecture de la table de la loi de Poisson : $p(X = 0) = 0.0498$, $p(X = 1) = 0.1494$, $p(X = 2) = 0.224$, $p(X = 3) = 0.224$, $p(X = 4) = 0.168$, $p(X = 5) = 0.1008$, $p(X = 6) = 0.0504$, $p(X \geq 7) = 1-p(X \leq 6) = 1-0.9665 = 0.0335$. On obtient en multipliant ces valeurs par l'effectif total 200 les effectifs « théoriques » de chaque classe :

Nb d'atterrissages	0	1	2	3	4	5	6	7+
Effectif mesuré	11	28	43	47	32	28	7	4
Effectif théorique	9.96	29.88	44.8	44.8	33.6	20.16	10.08	6.7

et en en déduit la valeur de $\chi_o^2 = \dfrac{(11-9.96)^2}{9.96} + \dfrac{(28-29.88)^2}{29.88} + \ldots + \dfrac{(4-6.7)^2}{6.7} = 5.56$.

Le critère de décision pour accepter H_0 sera $\chi_o^2 \leq c^2$, avec c^2 lu dans la table du χ^2 à 6 degrés de liberté (on a 8 classes, et le test porte sur une loi de Poisson : le nombre de d.d.l.à considérer est donc 8-2). On lit dans la table, dans la colonne $\alpha = 0.05$ et la ligne 5 : $c^2 = 12.59$.

Ainsi, on accepte H_0.

(c) Si on admet, suite au test du χ^2, que X suit une loi de Poisson de paramètre 3, on lit directement dans la table les valeurs $p(X = 0) = 4.98\%$, $p(1 \leq X \leq 2) = p(X = 1) + p(X = 2) = 37.34\%$.

Si on s'intéresse aux nombre d'atterissages entre 14h et 15h sur trois jours distincts, si on note X_1, X_2 et X_3 les variables donnant le nombre d'atterissage chacun des trois jours, on sait que, X_1, X_2 et X_3 étant indépendantes, $X_1 + X_2 + X_3$ suit une loi de Poisson de paramètre 9. Par conséquent, $p(X_1 + X_2 + X_3 = 2) = 0.5\%$.