Probabilités et statistiques

Durée : $2^{\text{h}}45$

Barème indicatif : I sur 2 points, II sur 4, III sur 3, IV sur 2 et V sur 9.

Exercice I

Lors de la mise au point du test de dépistage d'une maladie $\mathcal{M}$, les analyses ont montré que sur un échantillon de 100 personnes atteintes de $\mathcal{M}$, le test est positif 96 fois tandis que pour 100 personnes en bonne santé, le test est négatif 97 fois. On estime par ailleurs que 1% de la population est atteint de $\mathcal{M}$. Quelle est la probabilité qu'une personne dont le test serait positif soit réellement malade?

Exercice II

1) Soient A et B deux événements incompatibles (exclusifs). A quelle condition ces événements sont-ils indépendants?

2) On reprend l'expérience de BUFFON en remplaçant l'aiguille par une pièce de monnaie de rayon r, que l'on lance aléatoirement sur le "parquet" de droites parallèles équidistantes de d. Quelle est la probabilité pour que la pièce intersecte l'une des droites?

3) On distribue au hasard n bonbons à n enfants.

- **a)** Quelle est la probabilité que chaque enfant ait un bonbon exactement?
- **b)** Quelle est la probabilité qu'un enfant au moins n'ait pas de bonbon?
- **c)** Facultatif car plus difficile : quelle est la probabilité qu'un enfant exactement n'ait pas de bonbon?

Exercice III

Calculer le paramètre d'asymétrie ("*skewness*") et d'aplatissement ("*kurtosis*") d'une loi de probabilité uniforme dans l'intervalle $[a, b]$ (avec $b > a$).

Exercice IV

Soient X et Y deux variables aléatoires indépendantes ; X suit une loi binomiale $\mathcal{B}(n, p)$ de paramètres $n = 2$, $p = 1/2$ alors que Y est uniforme sur l'intervalle $[0, 1]$. Calculer la loi de probabilité de la variable aléatoire $Z = X + Y$.

./..

Exercice V : Méthode de Box-Muller

On souhaite engendrer un couple (x_1, x_2) de variables aléatoires gaussiennes indépendantes, de densité de probabilité

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2}{2} - \frac{x_2^2}{2}\right). \qquad (\mathcal{G})$$

1) On introduit les coordonnées polaires (r, θ) : $x_1 = r\cos\theta$ et $x_2 = r\sin\theta$. Quelle est la densité de probabilité jointe $q(r, \theta)$ pour le couple (r, θ) ?

2) Pour la suite, on fait le choix de prendre r et θ indépendantes. Quelles sont alors les expressions des densités de probabilité $p_R(r)$ et $p_\Theta(\theta)$?

3) Vérifier que les densités $p_R(r)$ et $p_\Theta(\theta)$ sont convenablement normalisées.

4) Comment peut-on engendrer θ suivant la distribution p_Θ voulue ?

5) Même question pour la distance à l'origine r.

6) Justifier la procédure suivante pour créer x_1 et x_2 indépendantes et distribuées suivant la loi $(\mathcal{G})$:

- échantillonner y_1 et y_2 uniformément dans $[0, 1]$
- calculer $$x_1 = \sqrt{-2\ln(y_1)}\,\cos(2\pi y_2)$$ $$x_2 = \sqrt{-2\ln(y_1)}\,\sin(2\pi y_2).$$

7) Quel peut-être l'intérêt informatique de cette méthode ?

8) Généraliser la procédure au cas où l'on souhaite obtenir x_1 et x_2 gaussiennes de moyennes $(\langle x_1\rangle$ et $\langle x_2\rangle)$ et d'écart type $(\sigma_1$ et $\sigma_2)$ quelconques.

9) Proposer succinctement une méthode permettant d'engendrer un couple de variables gaussiennes corrélées.

◇

Probabilités et statistiques

Durée : 3^{h}

Barème indicatif : exercice I sur 3 points, II sur 3, III sur 3 et problème sur 11.

Exercice I

Dans une population, 20% des individus sont atteints d'une maladie. On constitue un échantillon de 100 personnes choisies au hasard dans cette population. Quelle est la probabilité que le nombre de malades soit supérieur à 22 ? De même, si l'échantillon contient 1000 personnes, quelle est la probabilité que le nombre de malades soit supérieur à 220 ?

Exercice II

On prélève un échantillon de N composants dans une production industrielle, afin de déterminer la proportion p de composants défectueux. Quel doit être l'effectif N de l'échantillon pour que la probabilité d'une erreur sur p inférieure à 0.05 soit d'au moins 80% ?

Exercice III

Sur un échantillon de 100 personnes sondées avant une élection, 52% se déclarent favorables à un candidat. Quelle est la probabilité que ce candidat soit élu ?

Problème : Distribution du produit de nombres aléatoires

1) Préliminaires

On vous propose le jeu suivant : si un dé à six faces, non truqué, tombe sur les valeurs 5 ou 6, votre mise initiale de 1 € est multipliée par 6. Lorsque le dé tombe sur les valeurs 1, 2, 3 ou 4, votre mise est divisée par 3. La somme obtenue est alors remise en jeu pour une seconde partie, et ainsi de suite.

a) Dressez l'inventaire des situations possibles après une partie. Quelle est la probabilité d'avoir augmenté sa mise (situation de gain) après une partie ?

b) De même, quelle est la probabilité de gain après 2 parties ? Après 3 parties ? Il peut être utile de recenser les cas possibles et les probabilités associées.

c) Pensez-vous avoir intérêt à jouer un grand nombre de parties ? La justification rigoureuse de votre choix fait l'objet de la suite du problème.

2) On considère un ensemble de N variables aléatoires indépendantes strictement positives $x_1, x_2 \ldots, x_N$. Toutes ces variables ont même loi de probabilité, et l'on suppose l'existence de $\langle \ln x_1 \rangle$ et de $\langle (\ln x_1)^2 \rangle$. On s'intéresse à la distribution du produit

$$p = \prod_{i=1}^{N} x_i, \quad \text{et l'on note} \quad \begin{cases} m = \langle \ln x_1 \rangle = \langle \ln x_2 \rangle = \ldots = \langle \ln x_N \rangle \\ \sigma^2 = \langle (\ln x_1)^2 \rangle - \langle \ln x_1 \rangle^2 \end{cases}$$

a) Quelle est la valeur moyenne $\langle p \rangle$?

b) Donner l'expression de l'espérance et de la variance de la variable aléatoire $s = (\ln p)/N$.

c) Lorsque N devient grand, quelle est l'expression de la densité de probabilité de s, que l'on notera $g(s)$?

d) Soit $\widetilde{p}$ la variable aléatoire définie par $\widetilde{p} = p^{1/N}$. Quelle est la relation entre la densité de probabilité de $\widetilde{p}$, notée $f(\widetilde{p})$, et $g(s)$?

e) Donner l'expression de la densité $f(\widetilde{p})$ lorsque N devient grand.

f) On note $\widetilde{p}^*$ la valeur de $\widetilde{p}$ la plus probable dans la limite $N \to \infty$. Montrer que

$$\widetilde{p}^* = e^{\langle \ln x \rangle}.$$

g) Lorsque N croît, est-il possible que $\langle p \rangle$ tende vers l'infini tout en ayant $\widetilde{p}^* < 1$?

h) Donner l'expression de $\langle p \rangle$ et $\widetilde{p}^*$ lorsque les différentes variables x_i n'ont pas toutes la même densité de probabilité.

3) Comment s'interprètent ces résultats dans le contexte du jeu introduit dans la partie préliminaire ? Quel est le gain moyen après N tirages ? Comparer avec le gain le plus probable que l'on calculera. Finalement, doit-on accepter de jouer ? On pourra considérer ici que p^*, la valeur la plus probable de p, se comporte pour les grandes valeurs de N comme $(\widetilde{p}^*)^N$.

4) On peut reprendre le problème du jeu de dés sans faire appel au théorème de la limite centrale.

a) Quelles sont les différentes valeurs possibles pour le gain après N parties ? Préciser les probabilités associées.

b) En déduire la valeur du gain le plus probable lorsque N devient grand. Retrouve-t-on les résultats de la question **3)** ?

◇

Probabilités et statistiques

Durée : $2^{\text{h}}45$

Barème indicatif : I sur 2 points, II sur 2, III sur 3, IV sur 2 et V sur 5 et VI sur 6.

Exercice I

On a décelé dans un élevage de moutons que 30% des animaux sont atteints d'une maladie M. Si un mouton n'est pas atteint, il a 9 chances sur 10 de ne pas réagir à un test T. Dans le cas contraire, un mouton malade a 80% de chances d'être positif au test. Quelle est la probabilité qu'un mouton choisi au hasard et présentant un test positif soit malade?

Exercice II

Des études sismologiques ont montré que 20 tremblements de terre avaient eu lieu dans une région donnée en 10 ans. En supposant que le nombre de tremblements sur une période arbitraire suit une loi de Poisson, calculer la probabilité qu'aucun tremblement ne se produise en 2 ans. Quelle est la probabilité d'observer plus de deux tremblements en une année?

Exercice III Le coup du parapluie

On cherche un parapluie qui peut se trouver dans une maison de 7 pièces avec une probabilité p. On a exploré en vain les 6 premières pièces. Quelle est la probabilité que le parapluie se trouve dans la dernière? Comparer ce résultat à la probabilité *a priori* que le parapluie soit dans la première pièce. On supposera qu'il n'y a pas de pièce dans laquelle le parapluie est plus susceptible de se trouver que dans les autres.

Exercice IV

Si X est une variable aléatoire suivant une loi de Poisson de paramètre λ, quelle est l'espérance de la variable 1/(1+X)?

Exercice V

Le dispositif de chauffage d'une maison est constitué d'un élément de base et d'un chauffage d'appoint. On dira que l'on se trouve dans l'état 1 si seul le chauffage de base fonctionne, dans l'état 2 sinon. Si l'on se trouve dans l'état 1 un jour donné, on se trouve dans le même état le lendemain avec une probabilité 1/2. Lorsque l'on est dans l'état 2, la maison est encore chaude le lendemain et l'on passe à l'état 1 avec la probabilité 3/4.

1) Sachant que l'on se trouve dans l'état 1 un dimanche, trouver la probabilité d'être dans le même état le dimanche suivant. Comparer à la probabilité d'être resté toute la semaine dans l'état 1.

2) Montrer que si l'on se trouve un jour dans l'état 1 avec la probabilité 3/5, alors, pour tous les jours suivants, la probabilité d'être dans cet état est également 3/5.

3) Etablir que quel que soit l'état du dispositif un jour donné, le système atteint au bout de quelques semaines un état stationnaire que l'on caractérisera.

Exercice VI Introduction à la détection

Un signal codé en binaire est envoyé par un émetteur E à destination d'un récepteur R au moyen de deux canaux *parallèles* de transmission, $\mathcal{C}_1$ et $\mathcal{C}_2$. Avec le signal considéré, la probabilité d'émission d'un 0 est $P(0) = 0.3$ tandis que $P(1) = 0.7$. Les canaux sont indépendants l'un de l'autre et entachés d'erreur en réception. Pour $\mathcal{C}_1$, la probabilité de mal transmettre le symbole émis (c'est-à-dire un 0 à la place d'un 1 et vice versa) est $q_1 = 10^{-7}$. Pour $\mathcal{C}_2$, la probabilité d'erreur est $q_2 = 2.10^{-7}$. Il se peut donc que le récepteur reçoive deux informations contradictoires de la part des canaux "concurrents" $\mathcal{C}_1$ et $\mathcal{C}_2$, auquel cas il doit prendre une décision pour fournir un symbole au destinataire.

1) Quelle est la probabilité $P\left(\begin{smallmatrix}0\\1\end{smallmatrix}\,|\,1\,\text{émis}\right)$ que $\mathcal{C}_1$ transmette le bit 0 et $\mathcal{C}_2$ le bit 1 sachant que c'est un 1 qui a été émis ? Donner de même l'expression de $P\left(\begin{smallmatrix}1\\0\end{smallmatrix}\,|\,1\,\text{émis}\right)$.

2) Calculer la probabilité $P\left(0\,\text{émis}\,|\,\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ qu'un 0 ait été émis sachant que $\mathcal{C}_1$ transmet un 0 et $\mathcal{C}_2$ un 1. Même question avec $P\left(0\,\text{émis}\,|\,\begin{smallmatrix}1\\0\end{smallmatrix}\right)$.

3) Lorsque les canaux $\mathcal{C}_1$ et $\mathcal{C}_2$ sont en désaccord, le récepteur adopte une règle du maximum de vraisemblance, et choisit le symbole qui a le plus de chances d'avoir été émis. Préciser les choix effectués par R en cas de conflit entre $\mathcal{C}_1$ et $\mathcal{C}_2$.

4) Calculer la probabilité d'erreur globale de ce dispositif. Comparer à la probabilité d'erreur de chaque canal individuellement. Conclusion ?

◇

Probabilités et statistiques

Durée : 3^{h}

Barème indicatif : Exercice sur 3 points, Problème I sur 9 et Problème II sur 8.

Exercice

Dans une population donnée, on note p la probabilité de naissance d'une fille.

1) Quelle est la loi suivie par f, le nombre de naissances de filles parmi n naissances ?

2) Par quelle loi limite peut-on l'approcher lorsque n devient grand ?

3) Sur 100 naissances recensées, 52 ont concerné des filles. Est-il raisonnable d'admettre que les naissances de garçons et de filles sont équiprobables ?

4) Même question avec 520 naissances de filles pour un total de 1000, puis 5200 sur un total de 10000 ?

Problème I : fragmentation d'agrégats

Un corps de masse M peut subir une collision et se fragmenter en deux morceaux de masses respectives xM et $(1-x)M$, où x est une variable aléatoire réelle uniformément répartie entre 0 et 1. Chaque entité ainsi formée peut également subir une collision, et se désagréger elle-même. Un fragment ayant subi n collisions possède alors une fraction $x_1 x_2 ... x_n$ de la masse initiale, où les x_i sont des variables aléatoires indépendantes, toutes équidistribuées dans l'intervalle [0,1]. On s'intéresse ainsi à la densité de probabilité du produit

$$p_n = \prod_{i=1}^{n} x_i \ ,$$

relié à la masse d'une entité résultant de n collisions, et l'on introduit les variables auxiliaires $y_i = -\ln x_i$.

1) Quelle est la densité de probabilité $f(y)$ de la variable $y = -\ln x$?

2) Quelle est la densité de probabilité de $s_2 = y_1 + y_2$? Vérifier que l'expression obtenue est convenablement normalisée.

3) On note $s_n = y_1 + y_2 + \ldots + y_n$. Montrer que g, la densité de probabilité de s_n se met sous la forme

$$g(s) = \alpha_n \, s^{n-1} \, e^{-s}.$$

Donner l'expression du coefficient α_n.

4) En déduire la densité de probabilité h du produit p_n.

5) Le résultat obtenu à la question 3 est-il compatible avec le théorème de la limite centrale? On pourra utiliser la formule de STIRLING :

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \qquad \text{lorsque} \qquad n \to \infty.$$

Problème II : estimation du paramètre d'une loi exponentielle

Soient X et Y deux variables aléatoires indépendantes et θ un paramètre positif. La variable X suit une loi exponentielle de densité

$$f(x) = \frac{1}{\theta}\exp(-x/\theta) \qquad \text{pour} \qquad x \in [0, \infty[.$$

La variable aléatoire Y est quant à elle équidistribuée sur le segment $[0, 2\pi]$. On pose

$$U = \sqrt{X}\cos(Y) \qquad \text{et} \qquad V = \sqrt{X}\sin(Y).$$

1) Calculer l'espérance et la variance de Y.

2) Même question pour X.

3) Donner l'expression de $h(u, v)$, densité de probabilité jointe du couple (U, V).

4) En déduire les densités marginales de U et V.

5) Les variables U et V sont-elles indépendantes?

6) Soit $(X_1, X_2, \ldots, X_n)$ une famille de variables aléatoires indépendantes et de même loi que X. On pose $S_n = X_1 + X_2 + \ldots + X_n$.

a) En exprimant chaque X_i à l'aide des variables U_i et V_i associées, montrer qu'il existe un préfacteur $\lambda(\theta)$ tel que $\lambda(\theta)S_n$ suive une loi du χ^2 dont on précisera le nombre de degrés de liberté.

b) Proposer une approximation valable pour les grands n pour cette loi du χ^2.

c) En supposant l'approximation précédente licite, proposer un intervalle de confiance à 90% pour le paramètre θ, connaissant $\overline{X}_n = S_n/n$, mesuré sur un échantillon donné de taille n.

◇

Université de PARIS SUD
NFI Optronique $4^{\text{ième}}$ année

Examen d'Octobre 2002
Cours E. Trizac

Probabilités et statistiques

Durée : $2^{\text{h}}30$

Calculatrice interdite; réflexion autorisée
Barème indicatif : I sur 5 points, II sur 2, III sur 7 et IV sur 6.

Exercice I

Soit Y une variable aléatoire continue uniformément distribuée entre 0 et 5. On considère l'équation du second degré en x :

$$4x^2 - 4xY + Y + 2 = 0.$$

1) A quelle condition sur Y les deux racines de cette équation sont-elles réelles?

2) Quelle est alors la probabilité que ces deux racines soient positives?

3) Même question avec les racines du polynôme $4x^2 + 4xY + Y + 2 = 0$.

Exercice II

On considère n molécules p-fois atomiques. A haute température, ces molécules se dissocient en atomes, et l'on prélève n atomes au hasard (avec $n < p$). Calculer la probabilité pour que l'échantillon prélevé soit constitué d'atomes ayant appartenu initialement à la même molécule.

Exercice III

Soit X une variable aléatoire positive décrivant un processus exponentiel, c'est-à-dire associée à une densité de probabilité :

$$f(x) = \lambda \exp(-\lambda x); \quad x \in [0, +\infty[.$$

1) Calculer la probabilité $P(X < x_2 \text{ et } X > x_1)$ que X prenne des valeurs comprises entre x_1 et x_2 (on suppose $x_1 < x_2$).

2) En déduire la probabilité conditionnelle que X soit inférieur à x_2 sachant que par ailleurs $X > x_1$.

3) Justifier alors l'assertion affirmant qu'un processus exponentiel est "sans mémoire".

4) On s'intéresse désormais à la proposition réciproque.

a) On considère une variable aléatoire Y positive vérifiant la relation

$$P(Y < s + t \,|\, Y > t) = P(Y < s), \qquad \text{pour tout } s \text{ et } t \text{ positifs.}$$

Montrer que $P(Y > s + t) = P(Y > s)P(Y > t)$.

b) On note $G(y) = P(Y > y)$. Calculer $G(0)$. Etablir que $\dfrac{dG}{dy} = \alpha G(y)$ où α est un paramètre dont on donnera l'expression.

c) En déduire la densité de probabilité de Y et conclure.

Exercice IV

Afin de réaliser une couche métallique mince, on évapore un métal dans le vide, et l'on recueille un à un les atomes sur une plaque de quartz refroidie. La probabilité d'impact de chacun d'entre eux est *a priori* uniforme sur la plaque, dont l'aire est $S = 1\,(\mathrm{cm})^2$. Lorsqu'un atome heurte la surface de la plaque, il s'y fixe irréversiblement et occupe une aire a^2 où $a \simeq 1\,\text{Å}$.

Par souci de simplicité, on considère que la surface de quartz est constituée d'un réseau régulier de sites sur lesquels les atomes peuvent se condenser. On suppose que le nombre de ces sites est S/a^2, ce qui signifie que si chaque site est occupé par un atome condensé et un seul, on obtient une mono-couche uniforme de métal sur la plaque.

1) Quelle est la probabilité qu'un atome se condensant sur la surface a de recouvrir un site de la plaque choisi au préalable?

2) On forme un film métallique dont l'épaisseur moyenne est de 3 couches. Quelle fraction de l'aire de la plaque n'est recouverte par aucun atome?

3) De même, quelle fraction de la surface totale est-elle recouverte par exactement 3 atomes? Les résultats précédents dépendent-ils des valeurs particulières choisies pour S et a?

◇

Probabilités et statistiques

Durée : $2^{\text{h}}30$

Calculatrice interdite
Barème indicatif : chaque exercice sur 4 points.

Exercice I

Soient X et Y deux variables aléatoires indépendantes de POISSON, de paramètres respectifs λ et μ. On s'intéresse à la variable somme $Z = X + Y$.

1) Quelle est la valeur moyenne de Z et sa variance ?

2) En considérant la loi de POISSON comme la limite d'une loi binomiale, préciser la loi de probabilité de Z.

3) Calculer explicitement la probabilité $P(Z = z)$ que Z prenne une valeur z donnée, et retrouver la conclusion de la question précédente.

4) Peut-on déduire des résultats ci-dessus la loi de probabilité de la variable $2X$? Quelle est cette loi ?

Exercice II

Un sondage, effectué sur un échantillon de 900 individus, indique que 49% des personnes interrogées ont une opinion favorable concernant un homme politique.

1) Donner un intervalle de confiance à 98% pour la proportion de la population favorable au candidat.

2) On suppose désormais que cette proportion est exactement 49%. Un nouveau sondage, réalisé un mois plus tard, donne 50% d'opinions favorables au politicien. Quelle est la probabilité que sa cote de popularité soit *en baisse* ? Même question si le second sondage donne 51% d'avis positifs. Méditer…

Exercice III

On suppose que l'on dispose d'un générateur de nombres pseudo-aléatoires y uniformément distribués entre 0 et 1.

1) Proposer une procédure –directement utilisable dans un programme informatique– qui permette de créer une variable aléatoire Lorentzienne x de densité de probabilité

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

Quel est le domaine de définition de x ?

2) Généraliser la méthode au cas de la densité

$$f_\alpha(t) = \frac{\beta}{\pi} \frac{1}{\alpha^2 + t^2},$$

où α est un paramètre donné. Quelle doit être la valeur de β ? Comment peut-on interpréter le coefficient α ? Peut-on considérer qu'il s'agit de l'écart-type de la distribution ?

Exercice IV

Soient x et y deux variables aléatoires indépendantes continues, uniformément distribuées sur le segment $[0, 1]$. On définit deux nouvelles variables $p = xy$ et $r = x/y$.

1) Calculer les densités de probabilité de p et r. Les valeurs moyennes de r et p sont-elles bien définies ?

2) Calculer la densité de probabilité jointe du couple (r, p). Ces deux variables sont-elles indépendantes ?

Exercice V

On suppose que X suit une loi de Poisson $\mathcal{P}(\lambda)$. Dans quel cas peut-on approcher cette loi par une Gaussienne ? Quelle est alors sa moyenne et son écart-type ? En déduire la formule de Stirling

$$n! \sim \gamma\sqrt{n}\left(\frac{n}{e}\right)^n$$

valable lorsque n devient grand. On précisera la valeur du coefficient numérique γ.

◇

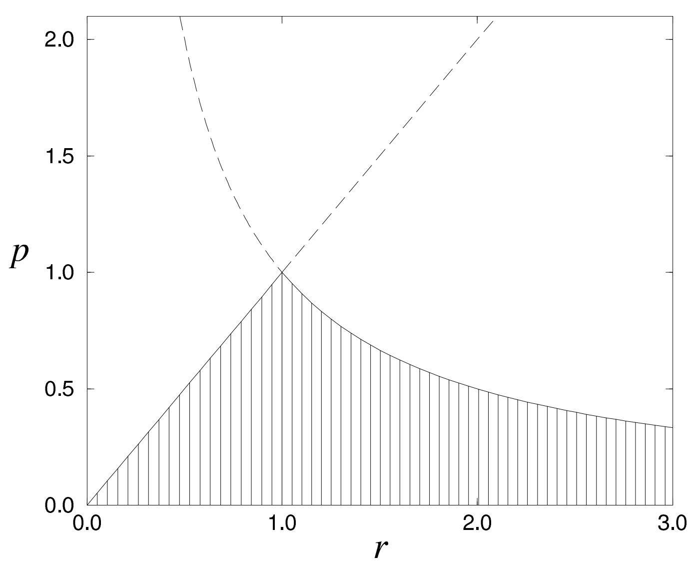


Fig. 1 – *Un graphe d'ornement*

Probabilités et statistiques

Durée : 2^{h}

Barème indicatif : exercice I sur 3 points, II sur 8 points et III sur 9.

Exercice I

Dans une région à risque, on a observé 50 tremblements de terre en un siècle. Quelle est la probabilité qu'un séisme exactement se produise durant une période de 2 ans ? Même question avec un séisme ou plus. Combien de temps faudrait-il attendre pour observer un tremblement avec une probabilité supérieure à 0.99 ?

Exercice II

Trois composants électroniques sont insérés dans un montage. Ces composants sont identiques, et leurs durées de vie T_1, T_2, T_3 sont des quantités aléatoires que l'on suppose de densité exponentielle :

$$f(t) = \tau^{-1} e^{-t/\tau}, \qquad t \in [0, \infty].$$

Les composants sont indépendants ; la défaillance de l'un d'entre eux est sans conséquence sur le fonctionnement des autres.

1) Quelle est la durée de vie moyenne d'un composant ?

2) Les trois composants sont associés en série. La durée de vie T_s du circuit ainsi formé est par conséquent $T_s = \min(T_1, T_2, T_3)$. Comparer les événements "$T_s > t$" et "$T_1 > t$ et $T_2 > t$ et $T_3 > t$". En déduire

$$P(T_s > t) = \exp(-t/\tau_s),$$

où τ_s est un temps caractéristique dont on donnera l'expression.

3) Quelle est la durée de vie moyenne $\langle T_s \rangle$ du circuit série ?

4) On associe les 3 composants en parallèle, si bien que la durée de vie du circuit est $T_p = \max(T_1, T_2, T_3)$. En vous inspirant des questions précédentes, calculer $P(T_p < t)$. Quelle est la densité de probabilité g_p du temps de vie T_p ?

5) Quelle est la durée de vie moyenne $\langle T_p \rangle$ du circuit parallèle ?

6) Reprendre l'ensemble des questions précédentes dans le cas d'une fonction f quelconque et d'un nombre arbitraire de composants.

Exercice III Le paradoxe des temps d'attente

Des bus passent à un endroit donné suivant un processus de Poisson, de cadence c. Le nombre moyen de bus passant durant un intervalle de temps t est ainsi ct, et le temps moyen entre deux bus consécutifs est $1/c$. Une personne arrive à un instant T donné. Pour calculer son temps moyen d'attente $\langle t_a \rangle$, deux raisonnements semblent possible :

- t_a doit être une variable de Poisson de cadence c, si bien que $\langle t_a \rangle = c^{-1}$.

- l'instant d'arrivée T est *a priori* uniformément distribué dans l'intervalle de temps correspondant au passage entre deux bus consécutifs (intervalle de durée s_T, dont on peut naïvement penser que la valeur moyenne est c^{-1}). Le temps d'attente moyen est la moitié de cette durée, d'où $\langle t_a \rangle = c^{-1}/2$.

Le but de cet exercice est de lever cette apparente contradiction. Nous allons montrer que le temps moyen entre le dernier bus passant avant T, et le premier passant après T, est $\langle s_T \rangle = 2c^{-1}$ et pas $1/c$. Dans ces conditions, le second raisonnement conduit à $\langle t_a \rangle = 2c^{-1}/2 = c^{-1}$, et le problème disparaît.

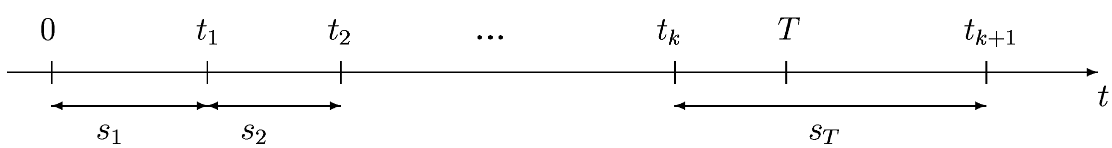
Le processus démarre à $t = 0$. On note t_i, le temps de passage du $i^{\text{ème}}$ bus et s_i le temps écoulé entre le $(i-1)^{\text{ème}}$ et le $i^{\text{ème}}$ bus : $s_i = t_i - t_{i-1}$. Les variables aléatoires s_i sont toutes distribuées suivant la même loi exponentielle

$$f(s) = c\,\exp(-cs),$$

et on admet que $t_n = \sum_{i=1}^{n} s_i$ a pour densité de probabilité la fonction g_n définie par

$$g_n(t) \;=\; c\,\frac{(ct)^{n-1}}{(n-1)!}\,\exp(-ct). \tag{1}$$

On appelle s_T la durée de l'intervalle (entre deux bus consécutifs) contenant l'instant T. Dans la figure ci-dessous, on a donc $s_T = s_{k+1}$.



On cherche dans un premier temps h_T, la densité de probabilité de s_T.

1) Soit $x < T$.

a) Quelle est la probabilité qu'aucun bus ne passe dans l'intervalle $[T-x, T]$?

b) Montrer que

$$P(s_T > x) = e^{-cx} + \sum_{k=1}^{\infty} \int_{T-x}^{T} g_k(y) e^{-cx}\, dy.$$

c) Calculer $\sum_{k=1}^{\infty} g_k(y)$.

d) En déduire que pour $0 < x < T$,

$$h_T(x) = c^2 x e^{-cx}. \tag{2}$$

2) On considère maintenant le cas $x > T$. Montrer que

$$P(s_T > x) \;=\; (1 + cT)\, e^{-cx}. \tag{3}$$

3) Lorsque $cT \gg 1$ (formellement $T \to \infty$), la formule (2) est valable pour toutes les valeurs possibles de s_T, et l'on peut « oublier » la relation (3). Dans ces conditions, calculer $\langle s_T \rangle$. Conclusion ?

4) Par définition, le temps d'attente d'un usager de la ligne de bus arrivant à l'instant T est $t_a = s_T - T$. Calculer $P(t_a < x)$, où x est un seuil fixé. En déduire la densité de probabilité de t_a ainsi que sa valeur moyenne.

5) Démontrer la relation (1), admise plus haut. Il est utile de calculer au préalable la fonction de répartition $G_n(\alpha) = P(t_n < \alpha)$.

Probabilités et statistiques

Durée : $2^{\text{h}}15$

Barème indicatif : exercice I sur 4 points, II sur 4 points, III sur 5 points et IV sur 7.

Exercice I

On lance 100 fois un dé à 6 faces et on calcule la somme totale S des faces obtenues à chaque lancer. Quelle est la probabilité que S soit compris entre 350 et 400 ? Donner un intervalle de confiance à 97% pour S.

Exercice II

Soient X_1 et X_2 deux variables aléatoires gaussiennes indépendantes de même loi $\mathcal{G}(m,\sigma)$. Quelle est la loi de la variable $Y = X_1 - X_2$? Même question dans le cas où les deux variables sont respectivement de lois $\mathcal{G}(m_1,\sigma_1)$ et $\mathcal{G}(m_2,\sigma_2)$.

Exercice III

On souhaite estimer l'écart-type σ d'une variable aléatoire X gaussienne de moyenne m. Pour ce faire, on prélève un échantillon de n réalisations de cette variable, et on calcule la variance empirique S^2. On pourra supposer n assez grand pour faire les développements limités qui s'imposent.

1) Connaissant la valeur de S^2 mesurée sur l'échantillon, donner un intervalle de confiance à 97% pour σ^2. Donner également un intervalle de confiance à 99% pour σ.

2) On désire connaître σ avec une précision de 1% et une confiance de 99%. Quelle est dans ces conditions la valeur minimale de n ?

Exercice IV : La loi de Pareto

Partie A Pour décrire la distribution des richesses, l'économiste V. Pareto a introduit la loi de probabilité suivante, de fonction de répartition

$$\begin{aligned} F(x) &= 1 - \frac{1}{x^\beta} \quad \text{pour} \quad x \geqslant 1 \\ &= 0 \qquad\quad \text{pour} \quad x \leqslant 1, \end{aligned}$$

où β est un nombre positif fixé. On parle de loi de Pareto de paramètre β. La richesse S_n d'une population de n individus est dans ces conditions la somme de n variables aléatoires de Pareto $(X_1, X_2, \ldots, X_n)$:

$$S_n = X_1 + X_2 + ... + X_n.$$

On introduit également M_n, la richesse de la personne la plus fortunée de l'échantillon $(X_1, X_2, \ldots, X_n)$

$$M_n = \max(X_1, X_2, \ldots, X_n).$$

1) Quelle est la densité de probabilité d'une variable de Pareto ?

2) À quelle condition portant sur β peut-on appliquer le théorème central limite pour calculer la loi de probabilité de S_n ?

3) Donner l'expression de la fonction de répartition de M_n.

4) On définit M_n^*, la valeur « typique » de M_n, comme étant celle que M_n a une probabilité 1/2 de dépasser. Montrer que lorsque n devient grand, on a

$$M_n^* \sim \left(\frac{n}{\ln 2}\right)^{1/\beta}.$$

5) Quelle serait la valeur typique de la plus grande valeur parmi n variables aléatoires gaussiennes $\mathcal{G}(0,1)$? On donne

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt \sim 1-\frac{1}{\sqrt{2\pi}}\frac{\exp(-x^2/2)}{x} \qquad \text{pour} \qquad x\to\infty.$$

Partie B On se place désormais dans le cas où $\beta < 1$ pour lequel le théorème central limite ne peut être appliqué. On peut montrer (nous l'admettrons) que pour n grand, l'espérance $E(S_n/M_n) = 1/(1-\beta)$ (le point important est que cette espérance ne diverge pas). Ce résultat est utile pour interpréter les données statistiques relatives à certains types de catastrophes. Pour traiter cette partie du problème, seul le résultat de la question A-4) est nécessaire. La figure 1 correspond à des observations recueillies entre 1965 et 1991, et montre la probabilité $P(X)$ que les inondations ayant eu lieu une année donnée dans le monde aient fait plus de X sans abris. En déduire que le nombre de sans abris consécutifs aux inondations d'une année donnée suit une loi de Pareto, dont on précisera le paramètre.

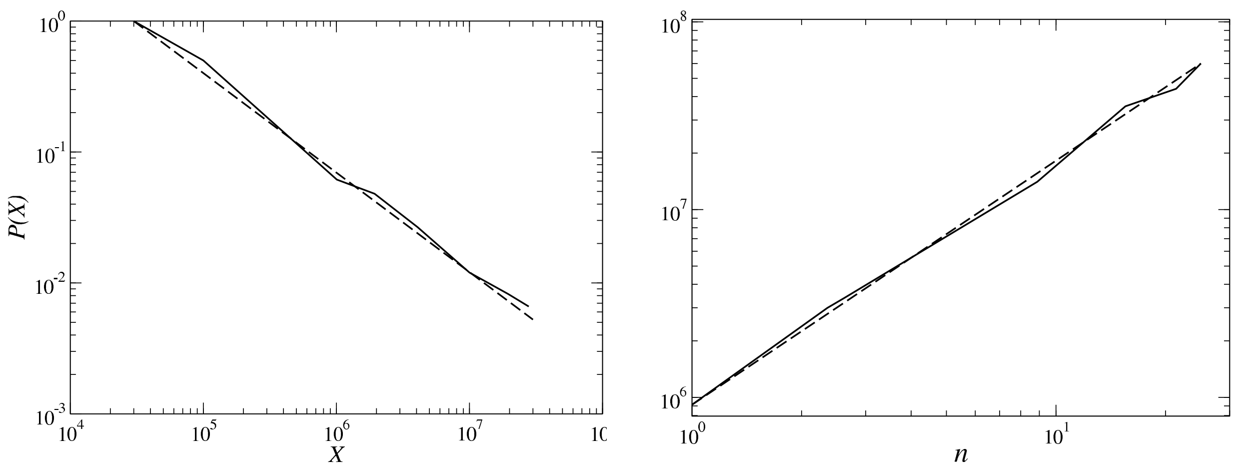


FIG. 1 – **À gauche :** *histogramme de $P(X)$, où $3.10^4 < X < 3.10^7$. Attention, l'échelle est logarithmique pour les abscisses comme pour les ordonnées (tracé dit log-log). La droite en pointillés a une pente -0.76, et montre donc une fonction de la forme $(Ax)^{-0.76}$ où A est un nombre positif sans importance.* **À droite :** *nombre cumulé de sans abris dus aux inondations sur la période 1965-1991. L'année $n = 0$ correspond à 1965 et le tracé est de nouveau log-log. La droite en pointillés représente une croissance proportionnelle à $n^{1.3}$ (plus rapide que linéaire).*

On s'intéresse désormais au nombre cumulé de victimes sur n années : $S_n = X_1+X_2+\ldots+X_n$, où X_i est le nombre de victimes la $i^{\text{ième}}$ année. Quelle loi de croissance peut-on anticiper pour S_n en fonction de n ? Cette dernière loi est-elle compatible

a) avec les données de la figure 1 (graphe de droite) ?

b) avec l'impression *a priori* inquiétante d'une accélération du nombre de victimes dues aux catastrophes naturelles ?

Probabilités et statistique

Durée : 2^{h}

Le hasard sait toujours trouver ceux qui savent s'en servir (R. ROLLAND)
Barème indicatif : I, II et III sur 3 points, IV sur 5, V sur 6 ; points bonus pour le VI.

Exercice I

Dans un élevage, un lapin sur 100 en moyenne est albinos. Quelle est la probabilité pour que dans une population de 100 lapins, un exactement soit albinos ? Même question avec un ou plus.

Exercice II

Une entreprise achète des composants par paquets de 10. Sa technique de contrôle est de n'examiner que 3 des composants tirés au hasard, et de n'accepter le paquet que si aucun des composants testés n'est défectueux. Dans le cas d'école où 30% des paquets contiennent 4 composants à malfaçon, et les 70% restant n'en contiennent qu'un, quelle est la proportion de paquets rejetés ?

Exercice III

Le test de dépistage d'une maladie est tel que 95% des personnes ayant contracté la maladie voient leur test positif, tandis que pour 95% des personnes en bonne santé, le test donne un résultat négatif. Madame XY subit le test de dépistage. Il est positif. Sachant que 1% de la population féminine du même âge et du même milieu que madame XY est atteinte par la maladie, quelle est la probabilité que Mme XY soit réellement malade ?

Exercice IV

On étudie la propagation d'un faisceau d'électrons dans un gaz. L'énergie T d'un électron est une variable aléatoire de moyenne ε et d'écart-type σ. On note L la longueur de pénétration d'un électron du faisceau dans le gaz : lorsque l'électron a une énergie $T = t$, L suit une loi exponentielle de densité de probabilité

$$g(\ell) \;=\; \alpha \exp\left(-\frac{\ell}{\overline{\ell}}\right),$$

où la quantité $\overline{\ell}$, appelée libre parcours moyen, est donnée par la formule empirique $\overline{\ell} = \beta t$ (β est une constante du problème).

1) Relier α aux autres données de l'énoncé.

2) Calculer l'espérance conditionnelle $E(L|T = t)$.

3) Même question avec la variance conditionnelle $V(L|T = t)$.

4) En déduire la moyenne $E(L)$ et la variance $V(L)$.

Exercice V

Un matériau radioactif émet des particules α dans toutes les directions. Le nombre X de particules émises durant l'intervalle de temps T d'une observation suit une loi de Poisson de paramètre λ. Par ailleurs, un observateur ne peut détecter que les particules émises dans sa direction, si bien que chaque particule émise a une probabilité p d'être décelée. Le nombre de particules observées est noté Y.

1) Quelle est la loi de probabilité conditionnelle de la variable aléatoire Y, sachant que le nombre de particules émises est $X = n$?

2) Quelle est la loi jointe du couple (X, Y) ? On rappelle $\displaystyle\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{\lambda}$.

3) Quelle est la loi marginale de Y ?

4) On note $Z = X - Y$ le nombre de particules non détectées. Quelle est la loi de Z ?

5) Quelle est la loi jointe de (Z, Y) ? Conclusion ?

6) Quelle serait la loi du nombre de particules observées si l'on triplait le temps d'observation ($T \to 3T$) ?

Exercice VI : corrélation et causalité

Les deux affirmations suivantes sont extraites d'articles récents parus dans la presse quotidienne "du soir". Quels commentaires vous inspirent-elles ?

1) Les compagnies d'assurance ont constaté que 80% des accidents de voiture se produisaient à moins de 30 km du domicile du conducteur. On en conclut à un relâchement de la vigilance sur les trajets de proximité.

2) Il a été prouvé que l'espérance de vie des personnes pratiquant le jogging à 60 ans était significativement supérieure à celle de la population générale du même âge. Ce qui démontre le bénéfice de cette activité.