

EPFL

Probabilités et Statistique pour Informatique et Communications  
2013–2014, Semestre d'été

---

## Probabilités et statistique : Examen

18 juin 2014

---

**Durée :** L'examen commence à 12:15 et se termine à 15:15.

---

Nom :

Prénom :

No. SCIPER :

Exercice	Points	Barême indicatif
1		/16 points
2		/8 points
3		/6 points
4		/10 points
Total :		/40 points

### REMARQUES:

- Aucun document personnel n'est autorisé.
- Les calculatrices simples sont permises. Il est interdit de s'échanger les calculatrices.
- Justifiez vos réponses ! Une réponse non justifiée sera considérée comme fausse.
- Merci d'écrire vos réponses directement sur les feuilles d'examens. Si vous manquez de place, utilisez les feuilles vides se trouvant à la fin de l'examen ou demandez une nouvelle feuille et agrafez-la.
- Vous avez le droit de questionner un assistant seulement si vous trouvez une faute de frappe. Sinon il ne va pas vous répondre. S'il vous semble qu'une question n'est pas claire, alors expliquez dans votre solution comment vous la comprenez.

- Exercice 1.** (a) Pour gagner le jackpot à l'EuroMillions, il faut avoir choisi correctement cinq nombres parmi 50 et deux étoiles parmi 11. Si toutes les combinaisons de nombres et d'étoiles ont la même chance d'être tirées, quelle est la probabilité de gagner le jackpot en un seul essai ? Si le jackpot s'élève à 150 millions de francs et que participer coûte 3 francs, quel est votre gain espéré ? (Vous pouvez ignorer la possibilité de gagner tout autre prix.)
- (b) Pour une Américaine de 40 ans, la probabilité de développer un cancer du sein est de 0.01. La probabilité qu'une mammographie donne un résultat positif pour une femme atteinte de cancer du sein est de 0.8, mais une mammographie est positive avec probabilité 0.1 pour une femme saine. Quelle proportion des femmes déclarées positives a-t-elle développé un cancer du sein ? Si, étant donné l'état de santé d'une femme, des mammographies distinctes donnent des résultats indépendants, et qu'une femme a obtenu deux mammographies positives, quelle est la probabilité qu'elle ait développé un cancer ?
- (c) Trouver la médiane d'une variable aléatoire dont la fonction de répartition est  $F(x) = 1 - (\beta/x)^\alpha$  pour  $x > \beta$  et  $\alpha, \beta > 0$ .
- (d) Si  $X \sim U(0, 1)$ , trouver la fonction de densité de  $Y = 1/X^2$ .
- (e) On rappelle que la fonction génératrice des moments d'une variable aléatoire  $X$  est  $M_X(t) = E(e^{tX})$ , définie pour tout  $t \in \mathbb{R}$  pour lequel  $M_X(t)$  est finie. Si  $X_1, X_2 \stackrel{\text{iid}}{\sim} \exp(\lambda)$ , trouver la fonction génératrice des moments de  $Z = X_1 - X_2$ .
- (f) Si  $X$  est de moyenne 1 et de variance 4 et  $Y$  de moyenne 2 et de variance 9, et si leur corrélation vaut  $-0.5$ , trouver la variance de  $4 + 2X - 5Y$ .
- (g) Si  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\lambda)$  sont indépendantes de  $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \exp(1/\lambda)$ , trouver l'estimateur du maximum de vraisemblance pour  $\lambda$ .
- (h) Qu'entend-on par "le résultat d'un test d'hypothèse est significatif au niveau 4%" ?

**English :**

- (a) To win the EuroMillions jackpot, you must correctly guess 5 numbers out of 50, and two stars out of 11. If all combinations of numbers and stars are equally likely, what is the probability of winning the jackpot with a single trial? If the jackpot is 150 million francs, and it costs 3 francs to participate, what are your expected winnings? (You may ignore the possibility of winning any other prizes.)
- (b) The probability that an American woman aged 40 has breast cancer is 0.01. The probability that a mammography will give a positive outcome for a woman with breast cancer is 0.8, but a mammography will give a positive outcome for a woman without breast cancer with probability 0.1. What proportion of women with positive mammographies have breast cancer? If different mammographies have independent outcomes, conditional on the state of the woman's health, and a woman has two positive mammographies, what is the probability that she has cancer?
- (c) Find the median of a random variable whose distribution function is  $F(x) = 1 - (\beta/x)^\alpha$  for  $x > \beta$ , where  $\alpha, \beta > 0$ .
- (d) If  $X \sim U(0, 1)$ , find the probability density function of  $Y = 1/X^2$ .
- (e) Recall that the moment-generating function of a random variable  $X$  is  $M_X(t) = E(e^{tX})$ , defined for all  $t \in \mathbb{R}$  such that  $M_X(t)$  is finite. If  $X_1, X_2 \stackrel{\text{iid}}{\sim} \exp(\lambda)$ , find the moment-generating function of  $Z = X_1 - X_2$ .
- (f) If  $X$  has mean 1 and variance 4, and  $Y$  has mean 2 and variance 9, and their correlation is  $-0.5$ , find the variance of  $4 + 2X - 5Y$ .
- (g) If  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\lambda)$  independent of  $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \exp(1/\lambda)$ , find the maximum likelihood estimator of  $\lambda$ .
- (h) What is meant by saying that the result of a statistical hypothesis test is significant at level 4%?

**Exercice 2.** Un site internet de partage de fichiers compte un très grand nombre de fichiers dont la taille  $S$  (gigabytes, GB) et le temps de téléchargement  $T$  (min) ont pour fonction de densité conjointe

$$f_{S,T}(s, t) = \frac{c}{(1+s)^4} \exp\{-t/(1+s)\}, \quad s, t > 0.$$

- (a) Donner la fonction de densité de la taille d'un fichier choisi au hasard. Le temps de téléchargement est-il indépendant de la taille?
- (b) Donner la taille moyenne d'un fichier.
- (c) Donner la fonction de densité conditionnelle du temps de téléchargement pour un fichier de taille 1 GB.
- (d) Quel est le temps espéré de téléchargement d'un fichier de taille 2 GB?
- (e) Donner le temps espéré de téléchargement pour des fichiers du site.

**Note :** Si  $n = 1, 2, \dots$  et  $\lambda > 0$ , alors  $\int_0^\infty \lambda^n x^{n-1} e^{-\lambda x} dx = (n-1)!$ .

**English :** A file-sharing website has a very large number of files, whose sizes  $S$  (GB) and download times  $T$  (min) have joint probability density function

$$f_{S,T}(s, t) = \frac{c}{(1+s)^4} \exp\{-t/(1+s)\}, \quad s, t > 0.$$

- (a) Give the probability density function of the size of a file selected at random. Is download time independent of file size?
- (b) Give the mean file size.
- (c) Give the conditional density function of the download time for a file of size 1 GB.
- (d) What is the expected time to download a file of size 2 GB?
- (e) Give the expected download time for files from the site.

**Note :** If  $n = 1, 2, \dots$  and  $\lambda > 0$ , then  $\int_0^\infty \lambda^n x^{n-1} e^{-\lambda x} dx = (n-1)!$ .

**Exercice 3.** La sandwicherie de mon quartier ne vend que de la soupe et des sandwiches.

- (a) Le patron prépare 40 litres de soupe quotidiennement. Chaque client choisit l'une des deux contenances de bol indépendamment et remplit celui-ci d'une quantité aléatoire de soupe. La quantité de soupe versée dans un petit bol est d'espérance 300ml et d'écart-type 30ml et la quantité versée dans un grand bol est d'espérance 600ml et d'écart-type 60ml. Si 50 clients choisissent un petit bol et 40 un grand, donner la distribution de la quantité totale de soupe consommée. Quelle est la probabilité qu'il n'y ait pas suffisamment de soupe aujourd'hui?
- (b) A l'ouverture de la sandwicherie, 100 sandwiches sont prêts à la vente, et après  $t$  heures, chacun a été vendu indépendamment avec probabilité  $1 - e^{-t}$ . Quelle est la probabilité qu'il n'y ait plus de sandwich si j'arrive 4 heures après l'ouverture?

**English :** My local sandwich bar serves only soup and sandwiches.

- (a) The chef makes 40 litres of soup each day. Customers each independently choose one of two sizes of soup bowl, and fill it with a random amount of soup. The amount of soup put into a small bowl has expectation 300 ml and standard deviation of 30 ml, and the amount put into a large bowl has expectation 600 ml and standard deviation 60 ml. If 50 clients choose small bowls and 40 clients choose large bowls, give the distribution of the total amount of soup consumed. What is the probability that there will not be enough soup today?
- (b) When the bar opens, there are 100 sandwiches for sale, and after  $t$  hours each has been sold independently with probability  $1 - e^{-t}$ . What is the probability that no sandwiches are left, if I arrive 4 hours after opening time?

**Exercice 4.** Le nombre quotidien de spams sur mon compte EPFL Mailcleaner a évolué ainsi pendant les 20 derniers jours :

13 9 10 12 5 7 10 19 14 9 7 7 26 19 22 19 11 5 14 20

Un modèle simple consiste à considérer ces nombres comme des réalisations de variables aléatoires de Poisson indépendantes, avec fonction de masse  $f(x) = \theta^x e^{-\theta} / x!$ , pour  $x = 0, 1, \dots$ , et  $\theta > 0$ .

- (a) Etablir la fonction de vraisemblance pour ce modèle. Trouver l'estimateur du maximum de vraisemblance  $\hat{\theta}$  pour  $\theta$  ainsi que l'information observée.
- (b) Donner un intervalle de confiance à 95% pour  $\theta$ . Comment faut-il l'interpréter ?
- (c) Plus tôt dans l'année on avait établi  $\theta = 15$ . Est-ce toujours plausible étant donné ces observations ?
- (d) On peut montrer que si  $X$  suit une distribution de Poisson de moyenne  $\theta$ , alors  $2(X^{1/2} - \theta^{1/2}) \overset{\sim}{\sim} \mathcal{N}(0, 1)$ . La figure ci-dessous montre un qq-plot normal des quantités correspondantes pour les observations ci-dessus.
  - (i) Expliquer ce qu'est un qq-plot normal et quelle est son utilité.
  - (ii) Qu'en déduire dans ce cas-ci ?

**English :** The numbers of spam emails in my EPFL Mailcleaner account for the last 20 days are

13 9 10 12 5 7 10 19 14 9 7 7 26 19 22 19 11 5 14 20

A simple model is that these numbers are realisations of independent Poisson random variables, each with probability mass function  $f(x) = \theta^x e^{-\theta} / x!$  for  $x = 0, 1, \dots$ , and  $\theta > 0$ .

- (a) Write down the likelihood under this model. Find the maximum likelihood estimate  $\hat{\theta}$  of  $\theta$ , and the observed information.
- (b) Give a 95% confidence interval for  $\theta$ . How should it be interpreted ?
- (c) Earlier this year it was said that  $\theta = 15$ . Is this now plausible, based on these data ?
- (d) It can be shown that if  $X$  has the Poisson distribution with mean  $\theta$ , then  $2(X^{1/2} - \theta^{1/2}) \overset{\sim}{\sim} \mathcal{N}(0, 1)$ . The figure below shows a normal QQ-plot of the corresponding quantities for the data above.
  - (i) Explain what such a plot is, and why it is useful.
  - (ii) What do you deduce in this case ?

