EPFL

Probability and Statistics for SIC
2018, Spring semester

---

# Probability and Statistics: Exam

22 June 2018

---

**Duration:** The exam will start at 12:15 and end at 15:15.

---

Family name:

First name:

SCIPER number:

| Exercise | Points | Indicative marks |
|---|---|---|
| 1 | | /22 points |
| 2 | | /9 points |
| 3 | | /14 points |
| 4 | | /13 points |
| Total: | | /58 points |

**PROTOCOL**:

- Personal documents may not be brought into the exam.
- A simple calculator may be used but use of other electronic devices, including mobile phones, is forbidden. You may not share a calculator.
- Do NOT unstaple pages.
- French versions of the questions are provided for your convenience. The English version is the reference in case of discrepancies between the two versions.
- Answers may be given in English or in French.
- Explain your reasoning! An unjustified answer will be treated as incorrect.
- Write your answers in the exam scripts. If you need more space or scratch paper, use the blank pages at the end of the script, or ask for another blank page and staple it into the script.
- The assistants will reply to questions only if there is a typo. If you find a question unclear, explain how you understand it when giving your solution.

**Exercise 1. (a)** I draw a sample of two balls at random without replacement from a bag containing two red balls and three white balls. Give the sample space for this random experiment. Are its elements equiprobable? What is the probability that both balls in the sample are red, given that at least one is red?

**(b)** If $\Pr(X > x) = 1/x^2$, for $x > 1$, find the probability density function of $Y = 1/X$.

**(c)** Find the median of the Weibull distribution, $F(x) = 1 - \exp\{-(x/\mu)^\alpha\}$, for $x > 0$ and $\alpha, \mu > 0$.

**(d)** If $X$ is normally distributed with mean 2 and $\Pr(X > 6) = 0.1$, find $\Pr(X < 0)$.

**(e)** If $X_1, X_2 \overset{\text{iid}}{\sim} N(\mu, 1)$, find the maximum likelihood estimator of $\mu$.

**(f)** If $X_1, X_2, X_3$ are independent exponential variables with parameters $\lambda_1, \lambda_2, \lambda_3$, find the distribution of $\min(X_1, X_2, X_3)$.

**(g)** The joint probability mass function of random variables $(X, Y)$ is given by the following table, in which $c$ is a positive constant:

| | $x$ | | |
|---|---|---|---|
| $y$ | 1 | 3 | 5 |
| 2 | $c$ | $2c$ | $3c$ |
| 4 | $3c$ | $2c$ | $c$ |

Find $\mathrm{E}(X)$ and $\mathrm{E}(X \mid Y = 4)$. Are $X$ and $Y$ independent?

**(h)** If $X$ is uniformly distributed on interval $(-1, 1)$, find its moment-generating function.

**(i)** Explain the terms *null hypothesis*, *test statistic* and *P-value*. A hypothesis test is performed and has P-value 0.001. What do you conclude?

**(j)** If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poiss}(\lambda)$ and $I(\cdot)$ is the indicator function, find the mean and variance of

$$T = n^{-1} \sum_{j=1}^{n} I(X_j = 0).$$

How would the mean of $T$ change if $X_1, \ldots, X_n$ were dependent?

**Français**:

**(a)** Je tire un échantillon aléatoire de deux boules sans remplacement d'un sac contenant deux boules rouges et trois boules blanches. Donnez l'espace d'échantillon pour cette expérience aléatoire. Ses éléments sont-ils equiprobables? Quelle est la probabilité que les deux boules soient rouges, étant donné qu'au moins une des boules est rouge?

**(b)** Si $\Pr(X > x) = 1/x^2$, for $x > 1$, trouvez la fonction de densité de $Y = 1/X$.

**(c)** Trouvez la médiane de la distribution de Weibull, $F(x) = 1 - \exp\{-(x/\mu)^\alpha\}$, pour $x > 0$ et $\alpha, \mu > 0$.

**(d)** Si $X$ suit la loi normale avec espérance 2 et $\Pr(X > 6) = 0.1$, trouvez $\Pr(X < 0)$.

**(e)** Si $X_1, X_2 \overset{\text{iid}}{\sim} N(\mu, 1)$, trouvez l'estimateur du maximum de vraisemblance de $\mu$.

**(f)** Si $X_1, X_2, X_3$ sont des variables exponentielles indépendantes avec paramètres $\lambda_1, \lambda_2, \lambda_3$, trouvez la fonction de répartition de $\min(X_1, X_2, X_3)$.

**(g)** La fonction de masse conjointe des variables aléatoires $(X, Y)$ est donnée dans le tableau suivant, où $c$ est une constante positive :

| | $x$ | | |
|---|---|---|---|
| $y$ | 1 | 3 | 5 |
| 2 | $c$ | $2c$ | $3c$ |
| 4 | $3c$ | $2c$ | $c$ |

Trouvez $\mathrm{E}(X)$ et $\mathrm{E}(X \mid Y = 4)$. $X$ et $Y$ sont-elles indépendantes?

**(h)** Si $X$ est uniformément distribué dans l'intervalle $(-1, 1)$, trouvez sa fonction génératrice de moments.

**(i)** Expliquez les termes *hypothèse nulle*, *statistique de test* et *$p$-valeur*. Un test d'hypothèse est exécuté et a une $p$-valeur de 0.001. Qu'en concluez-vous?

**(j)** Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poiss}(\lambda)$ et $I(\cdot)$ est une fonction indicatrice, trouvez l'espérance et la variance de

$$T = n^{-1} \sum_{j=1}^{n} I(X_j = 0).$$

Comment l'espérance de $T$ changerait-elle si $X_1, \ldots, X_n$ étaient dépendantes?

**Exercise 2.** This summer is forecast to be wet. Each day it will rain with probability 0.5, and on rainy days the amounts of rain will be independent with mean 8mm and standard deviation 4mm.

**(a)** What is the probability that it will be dry for the next seven days?

**(b)** Find the expectation and the variance of the rainfall tomorrow.

**(c)** What is the approximate probability that the rainfall over a 30-day period will exceed 16cm?

*Hint: for (b) you may find it helpful to write the rainfall tomorrow as $R = WA$, where $W$ is an indicator variable of tomorrow being wet, $A$ is tomorrow's rainfall amount, and $A$ and $W$ are independent.*

**Français**:
Beaucoup de pluie est annoncée pour l'été. Chaque jour, la probabilité qu'il pleuve est de 0.5 et, les jours de pluie, les quantités de pluie sont indépendantes avec une espérance de 8mm et une deviation standard de 4mm.

**(a)** Quelle est la probabilité qu'il ne pleuve pas pendant les sept prochains jours?

**(b)** Trouvez l'espérance et la variance de la pluie pour demain.

**(c)** Quelle est la probabilité approximative que la hauteur de précipitations dépasse 16cm sur une période de 30 jours?

*Indice: pour (b) vous pourrez trouver utile de dénoter la hauteur de précipitations du lendemain par $R = WA$, où $W$ est une variable indicatrice de la pluie le lendemain, $A$ est la hauteur de précipitations du lendemain, et $A$ et $W$ sont indépendantes.*

**Exercise 3.** Summaries of marks from an exam involving two groups of students are given in Table 1, and the data are plotted in Figure 1.

**(a)** Explain the elements of the left-hand panel of Figure 1. What do you infer from it?

**(b)** Explain the elements of the normal Q-Q plot in the right-hand panel of Figure 1. What do you infer from it? By drawing on this panel, indicate what the Q-Q plot for Group 2 would look like.

**(c)** On the assumption that the group sizes are large enough that any errors in estimating their standard deviations can be neglected, give the variance of the difference of the group averages. Hence give an approximate 95% confidence interval for the mean difference in marks, and explain carefully how your interval should be interpreted.

**(d)** Explain what assumptions were made in (c) and discuss their plausibility.

**Français**: Un résumé des resultats de deux groupes d'élèves à un examen est donné dans le tableau 1 et les données sont représentées dans la figure 1.

**(a)** Expliquez les éléments du diagramme de gauche de la figure 1. Qu'en déduisez-vous?

**(b)** Expliquez les éléments du diagramme Q-Q sur le droite de la figure 1. Qu'en déduisez-vous? En le dessinant sur le diagramme, indiquez ce à quoi le diagramme Q-Q ressemblerait pour groupe 2.

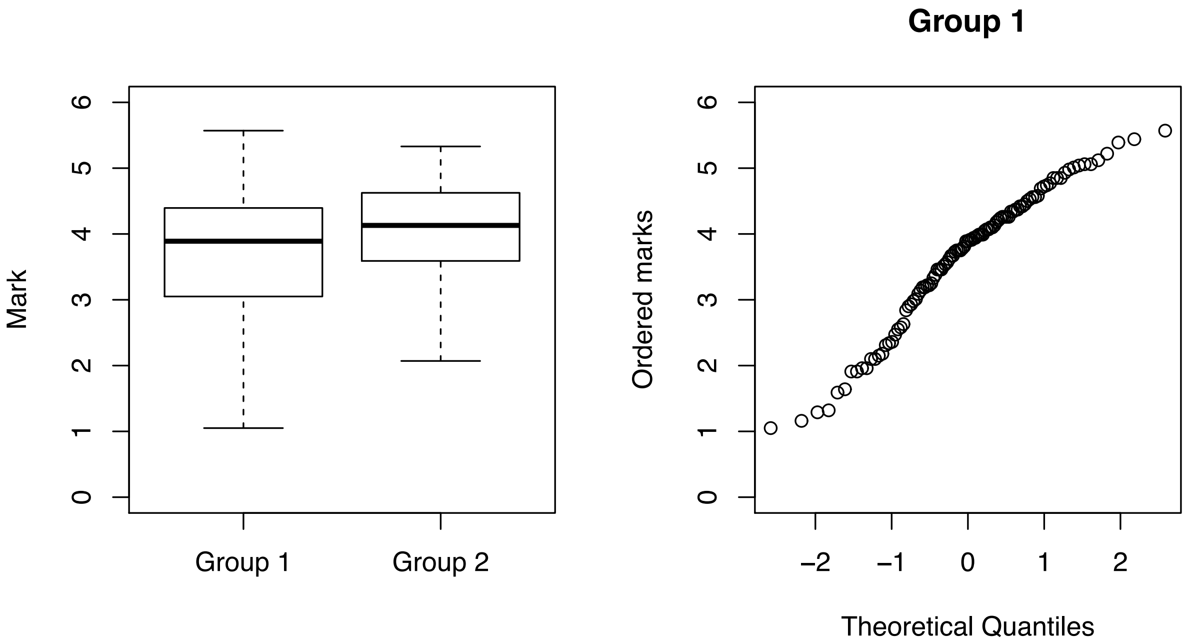
**(c)** En supposant que les tailles des groupes sont assez grandes que les erreurs d'estimation de leur deviation standard puissent être négligées, donnez la variance de la différence de moyennes des deux groupes. Fournissez un intervalle de confiance approximatif de 95% pour la moyenne des différences entre les notes, et expliquez soigneusement comment votre intervalle doit être interprété.

**(d)** Listez les hypothèses faites en (c) et discutez de leur plausibilité.

Table 1: Summary statistics for exam marks for two groups of students

| | Size, $n$ | Average, $\overline{x}$ | Standard deviation, $s$ |
|---|---|---|---|
| Group 1 | 100 | 3.66 | 1.06 |
| Group 2 | 80 | 4.05 | 0.75 |

Figure 1: Plots for data on exam marks.




**Exercise 4.** With the previous operating system my laptop used to crash purely at random, but the average interval between crashes was two weeks. Since a system upgrade a week ago it has crashed three times, and I wonder whether crashes have become more frequent. To assess this, I decide to model the number of crashes each week as a Poisson random variable with mean $\lambda$.

**(a)** Briefly discuss the use of the (i) uniform, (ii) normal, (iii) gamma and (iv) Bernoulli distributions to represent prior information on the average number of crashes per week.

**(b)** If the prior density for $\lambda > 0$ is taken to be $\pi(\lambda) = c\exp(-c\lambda)$ for some $c > 0$, compute the prior expectation of $\lambda$, and hence suggest a suitable value for $c$.

**(c)** Compute the posterior density for $\lambda$, conditional on the number of crashes in the past week.

**(d)** Find the maximum a posteriori (MAP) estimate of $\lambda$. Have crashes become more frequent?

**Français**:
Avec l'ancien système d'exploitation, mon ordinateur plantait aléatoirement, mais l'intervalle moyen entre les problèmes était de deux semaines. Depuis une mise à jour il y a une semaine, il a planté trois fois et je me demande si le problème est devenu plus fréquent. Pour évaluer la situation, je décide de modeliser le nombre de fois où mon ordinateur plante chaque semaine par une variable aléatoire de Poisson avec une moyenne $\lambda$.

**(a)** Expliquez brièvement à quel point les distributions (i) uniforme, (ii) normale, (iii) gamma et (iv) de Bernoulli sont valables pour représenter l'information préalable sur le nombre de fois où mon ordinateur plante chaque semaine.

**(b)** Si la densité à priori pour $\lambda > 0$ est $\pi(\lambda) = c\exp(-c\lambda)$ pour $c > 0$, calculez l'espérance à priori de $\lambda$ et suggérez une valeur appropriée pour $c$.

**(c)** Calculez la densité à posteriori de $\lambda$, conditionnelle au nombre de fois mon ordinateur a planté pendant la semaine précédente.

**(d)** Trouvez l'estimé du maximum à posteriori (MAP) de $\lambda$. Est-ce que les problèmes sont devenus plus fréquents?