

EPFL

Probabilités et Statistique pour Informatique et Communications
2014–2015, Semestre d'été

Probabilités et statistique : Examen

3 juillet 2015

Durée : L'examen commence à 8:15 et se termine à 11:15.

Nom :

Prénom :

No. SCIPER :

Exercice	Points	Barème indicatif
1		/16 points
2		/8 points
3		/8 points
4		/6 points
Total :		/38 points

REMARQUES:

- Aucun document personnel n'est autorisé.
- L'utilisation d'une calculatrice simple est permise mais l'utilisation de tout autre appareil électronique y compris les téléphones portables est interdite. Il est interdit de s'échanger les calculatrices.
- Ne dégrafez pas les feuilles d'examen.
- Justifiez vos réponses ! Une réponse non justifiée sera considérée comme fausse.
- Les réponses peuvent être écrites soit en français soit en anglais.
- Merci d'écrire vos réponses directement et lisiblement 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) Notons X le nombre de faces visibles lorsque trois pièces de monnaie équilibrées sont lancées indépendamment. Donnez un espace de probabilité qui correspond à cet expérience aléatoire et utilisez-le pour obtenir $\Pr(X = 2 \mid X \geq 2)$.
- (b) Au jeu du craps, deux dés à six faces équilibrés sont lancés et leurs scores sont additionnés. Le joueur perd si la somme est égale à 2, 3 ou 12, il gagne si la somme est égale à 7 ou 11 et il relance les dés si la somme est égale à n'importe quel autre nombre. Quelle est la distribution du nombre total de lancés X jusqu'à ce que le jeu s'arrête (i.e., le lanceur gagne ou perd)? Si on observe $X = 3$, quelle est la probabilité que le joueur ait perdu?
- (c) Je lance une pièce de monnaie équilibrée donnant le score $Y = 0$ si le résultat est pile et le score $Y = 1$ sinon. Si $Y = 0$ je lance un dé à six faces équilibré donnant le résultat X ; si $Y = 1$ alors $X = 0$. Trouvez l'espérance de $X + Y$.
- (d) Si une variable aléatoire continue X a une fonction de densité x^{-2} pour $x \geq 1$, trouvez la fonction de répartition de $Y = 1/X$.
- (e) Pour rappel, la fonction génératrice des moments d'une variable aléatoire X est $M_X(t) = E(e^{tX})$, définie pour tout $t \in \mathbb{R}$ tel que $M_X(t)$ est fini. Si X_1, X_2 sont deux variables aléatoires de Poisson indépendantes et de paramètres θ_1, θ_2 , trouvez la fonction génératrice des moments de $Z = X_1 + X_2$. Quelle est la distribution de Z ?
- (f) Les données $(x_1, y_1), \dots, (x_n, y_n)$ ont un coefficient de corrélation empirique r . Pour chacun des cas de figure suivants, esquissez une configuration possible des données : (i) $r = 1$, (ii) $r = -0.5$, (iii) $r = 0$.
- (g) Si $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} U(0, \theta)$, trouvez l'estimateur du maximum de vraisemblance de θ .
- (h) Qu'entend-on par (L, U) est un intervalle de confiance à $(1 - \alpha) \times 100\%$ pour un paramètre θ ?

English :

- (a) Let X denote the number of heads seen when three fair coins are thrown independently. Write down a probability space that represents this random experiment, and use it to obtain $\Pr(X = 2 \mid X \geq 2)$.
- (b) In the game of craps, two fair six-sided dice are rolled and their scores added. If the sum of the scores equals 2, 3, or 12 the roller loses, if this sum equals 7 or 11 the roller wins, and if this sum equals any other number, the dice are rolled again. What is the distribution of the total number of rolls X until the game ends (i.e., the roller wins or loses)? If we observe $X = 3$, what is the probability that the roller lost?
- (c) I toss a fair coin, giving a score $Y = 0$ if the coin shows tails and a score $Y = 1$ otherwise. If $Y = 0$, then I throw a fair six-sided die, giving result X , whereas if $Y = 1$ then $X = 0$. Find the expected value of $X + Y$.
- (d) If a continuous random variable X has probability density function x^{-2} , for $x \geq 1$, find the cumulative distribution function of $Y = 1/X$.
- (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 are independent Poisson variables with parameters θ_1, θ_2 , find the moment-generating function of $Z = X_1 + X_2$. What is the distribution of Z ?
- (f) Data $(x_1, y_1), \dots, (x_n, y_n)$ have empirical correlation r . For each of the following values of r , sketch one possible data configuration : (i) $r = 1$, (ii) $r = -0.5$, (iii) $r = 0$.
- (g) If $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} U(0, \theta)$, find the maximum likelihood estimator of θ .
- (h) What is meant by saying that (L, U) is a $(1 - \alpha) \times 100\%$ confidence interval for a parameter θ ?

Exercice 2. Supposons que le temps dans une ville est quantifié par le coefficient d'ensoleillement, c'est-à-dire la proportion d'heures d'ensoleillement par rapport à la durée de la journée pour un jour donné. Des études ont montré que pour de nombreuses villes le coefficient d'ensoleillement de deux jours consécutifs X et Y en été ont la fonction de densité conjointe

$$f_{X,Y}(x,y) = c(1 + 4\theta xy), \quad 0 \leq x, y \leq 1, \quad \theta \geq 0,$$

où le paramètre θ varie de ville en ville et la valeur $c \equiv c(\theta)$ dépend de θ .

- (a) Trouvez les fonctions de densité marginales de X et Y .
- (b) Pour quelle(s) valeur(s) de θ les coefficients d'ensoleillement de deux jours consécutifs sont-ils indépendants ?
- (c) Donnez l'espérance du coefficient d'ensoleillement de Londres, où $\theta = 1$.
- (d) Etant donné que le coefficient d'ensoleillement d'aujourd'hui à Londres était $x = 1$, donnez la distribution conditionnelle du coefficient d'ensoleillement de demain à Londres.

English : Suppose that the weather in a city is measured by the sunshine quotient, i.e., the proportion of hours of sunshine to the total hours of daylight on a given day. Research has shown that in many cities, the sunshine quotients on two successive summer days, X and Y , have joint probability density function

$$f_{X,Y}(x,y) = c(1 + 4\theta xy), \quad 0 \leq x, y \leq 1, \quad \theta \geq 0,$$

where the parameter θ varies from city to city and the positive quantity $c \equiv c(\theta)$ depends on θ .

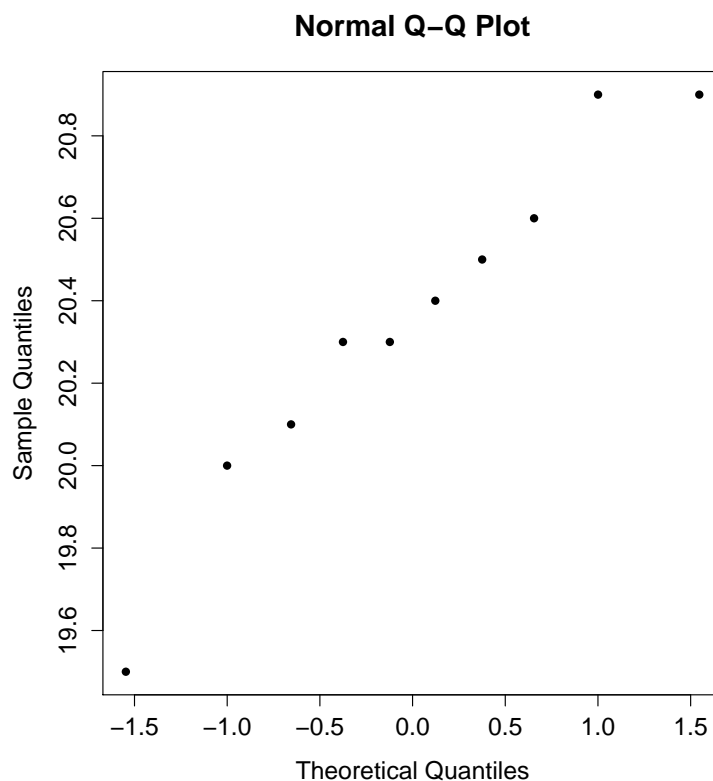
- (a) Find the marginal probability density functions of X and Y .
- (b) For what value(s) of θ are the sunshine quotients on two successive days independent ?
- (c) Give the expected sunshine quotient for London, where $\theta = 1$.
- (d) Given that the sunshine quotient today in London equalled $x = 1$, give the conditional density of the sunshine quotient there tomorrow.

Exercice 3. Maintenant que leur session d'examen est terminée, Marco et Chris préparent chacun leur valise pour partir en vacances en Thaïlande. La compagnie aérienne fait payer une surtaxe aux voyageurs dont le poids du bagage excède 20 kg.

- (a) Marco prépare sa valise, puis la pèse en utilisant le pèse-personne de sa salle de bains. Il sait que cette balance n'est pas très fiable, alors il décide de peser sa valise 10 fois, donnant les poids (kg)

20.0 21.0 20.8 20.7 19.6 20.9 20.0 20.7 21.0 20.8

Il obtient ensuite le graphique suivant :



- (i) Expliquer ce qu'est ce graphique et pourquoi il est utile.
(ii) Que déduisez-vous de ce graphique ?
- (b) Effectuez un test d'hypothèse à 95% pour déterminer si la valise de Marco est trop lourde.
- (c) Chris est plus consciencieux. Il utilise sa balance de cuisine pour peser individuellement chaque objet qu'il va emporter y compris sa valise. Le poids total des 49 objets, valise incluse, est de 19.9 kg. Le fabricant stipule que la balance de cuisine a un écart-type de 4 g pour chaque pesée. Effectuez un test d'hypothèse à 99% pour déterminer si Chris devra payer une surtaxe.

English : Now that their exams are over, Marco and Chris are each packing a suitcase before flying to Thailand for a holiday. The airline fines travellers whose suitcases weigh more than 20 kg.

- (a) Marco packs his suitcase, and then weighs it using the bathroom scales. He knows that these scales are very unreliable, so he decides to weigh the suitcase 10 times, resulting in weights (kg)

20.0 21.0 20.8 20.7 19.6 20.9 20.0 20.7 21.0 20.8

He then produces the above plot.

- (i) Explain what such a plot is, and why it is useful.
(ii) What do you deduce in this case ?
- (b) Test at level 95% whether Marco's suitcase is overweight.
- (c) Chris is more careful ; he uses the kitchen scales to weigh individually each item he will take, including his suitcase. The total weight of these 49 objects equals 19.9 kg. The manufacturer claims that the kitchen scales have a standard deviation of 4 g for each item weighed. Test at level 99% whether Chris is likely to be fined.

Exercice 4. Des paquets de données transmis dans un réseau de communication sont corrompus de manière indépendante avec probabilité inconnue θ . Soient x_1, \dots, x_n les valeurs observées de variables indicatrices telles que $x_j = 1$ si le paquet j est corrompu et $x_j = 0$ sinon. Supposons que pour des réseaux similaires θ a la fonction de densité $\pi(\theta) = m(1 - \theta)^{m-1}$, pour $0 < \theta < 1$ et $m > 0$.

- (a) Ecrivons la fonction de densité conjointe des données comme $f(x_1, \dots, x_n | \theta) = \theta^c(1 - \theta)^{n-c}$. Trouvez la fonction de densité a posteriori de θ étant donné x_1, \dots, x_n . Que représente c ?
- (b) Trouvez l'estimateur qui maximise la fonction de densité a posteriori de θ .
- (c) Trouvez la probabilité que le prochain paquet de données qui arrivera sera corrompu étant donné x_1, \dots, x_n .

Note : Pour $a, b > 0$, $\int_0^1 u^{a-1}(1 - u)^{b-1} du = \Gamma(a)\Gamma(b)/\Gamma(a + b)$, où la fonction Gamma $\Gamma(a) = (a - 1)!$ pour $a = 1, 2, \dots$, et $\Gamma(a + 1) = a\Gamma(a)$ pour $a > 0$.

English : Packets sent over a communication network are independently corrupted with unknown probability θ . Let x_1, \dots, x_n be the observed values of indicator variables, with $x_j = 1$ if packet j is corrupted and $x_j = 0$ otherwise, and suppose that for similar networks θ has density function $\pi(\theta) = m(1 - \theta)^{m-1}$, for $0 < \theta < 1$ and some $m > 0$.

- (a) By writing the joint density of the data as $f(x_1, \dots, x_n | \theta) = \theta^c(1 - \theta)^{n-c}$, find the posterior density of θ given x_1, \dots, x_n . What does c represent ?
- (b) Find the maximum a posteriori estimate of θ based on the data.
- (c) Find the probability that the next packet to arrive will be corrupted, conditional on x_1, \dots, x_n .

Note : For $a, b > 0$, $\int_0^1 u^{a-1}(1 - u)^{b-1} du = \Gamma(a)\Gamma(b)/\Gamma(a + b)$, where the Gamma function $\Gamma(a) = (a - 1)!$ for $a = 1, 2, \dots$, and $\Gamma(a + 1) = a\Gamma(a)$ for $a > 0$.