EPFL
Probabilités et Statistique pour Informatique et Communications
2012–2013, Semestre d'été

Probabilités et statistique : Examen

24 juin 2013

Durée : L'examen commence à 16:15 et se termine à 19:15.

Nom :

Prénom :

No. SCIPER :

Exercice	Points	Barême indicatif
1		/16 points
2		/7 points
3		/6 points
4		/7 points
Total :		/36 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 ils ne vont pas vous répondre. S'il vous semble qu'une question n'est pas claire, alors expliquer dans votre solution comment vous la comprenez.

Exercice 1. (a) Le Daily Mail du 13 octobre 2010 rapporte que la famille Allali (Angleterre) vient juste d'avoir son troisième enfant et que, fait surprenant, ses trois enfants sont tous nés le même jour (un 7 octobre, en 2005, 2007 et 2010). Le journaliste affirme qu'il y a moins d'une chance sur 48 millions que trois enfants d'une famille (ni jumeaux, ni triplets) naissent un même jour. Que pensez-vous de cette affirmation ?

(b) On admet que 5% des hommes et 0.25% des femmes sont daltoniens. On sélectionne une personne daltonienne au hasard parmis un échantillon composé de deux fois plus de femmes que d'hommes. Quelle est la probabilité qu'il s'agisse d'un homme ?

(c) On rappelle que la fonction génératrice des moments d'une variable aléatoire X est donnée par $M_X(t) = \mathrm{E}(e^{tX})$. Calculer M_X lorsque X suit une loi de Poisson de paramètre λ et en déduire l'espérance et la variance de X.

(d) Soit X une variable exponentielle de paramètre λ et $Y = -\log(X)$. Calculer la fonction de densité de Y.

(e) Soit X_1 et X_2 deux variables indépendantes exponentielles de paramètres respectifs λ_1 et λ_2. Calculer la loi de $Y = \min(X_1, X_2)$.

(f) La mise à jour d'un logiciel nécessite l'installation de 85 fichiers, installés séquentiellement. Le temps d'installation est aléatoire, mais prend en moyenne 15 secondes par fichier, avec une déviation standard de 4 secondes. Quelle est la probabilité que le logiciel soit mis à jour en moins de 20 minutes ?

(g) Soit $X_1, \ldots, X_n$ un échantillon aléatoire d'une variable binomiale X avec $N = 10$ et p inconnu, i.e., $X \sim \mathcal{B}(10, p)$. Calculer l'estimateur du maximum de vraisemblance de p. Est-il biaisé ?

(h) Un constructeur de téléphones annonce que l'autonomie d'un de ses smartphones est en moyenne de 24 heures. On teste 20 de ces smartphones pour trouver une autonomie moyenne $\bar{x} = 21$ heures et un écart-type $s = 5$ heures. Au seuil 95%, rejette-on l'hypothèse que le constructeur a raison ?

Exercice 2. Un programme informatique est composé de trois sections : une section d'initialisation A, une section de calcul B composée de 100 cycles de calcul identiques et indépendents, et une section finale C. On a enregistré les temps de calcul du programme pour chacune des trois sections. Le tableau suivant résume les temps de calcul obtenus.

Section	Temps moyen (ms)	Déviation standard (ms)
Initialisation A	5.5	2.5
Calcul B (composé de 100 cycles)	3.4	2.6
Finalisation C	4.5	1.3

Tous les temps sont indépendants exceptés ceux des sections A et C dont la corrélation vaut 0.2.

(a) Calculer $\mathrm{cov}(A, C)$.

(b) Calculer la moyenne et la variance du temps d'exécution total T du programme.

(c) Les temps des sections A et C sont normalement distribués. Bien que le temps de calcul d'un cycle de la section B ne soit pas normalement distribué, expliquer pourquoi la distribution du temps total de la section B (qui comporte 100 cycles) peut-être approximée par une loi normale et préciser les paramètres de cette distribution.

(d) Calculer la proportion des cas pour lesquels le programme prend

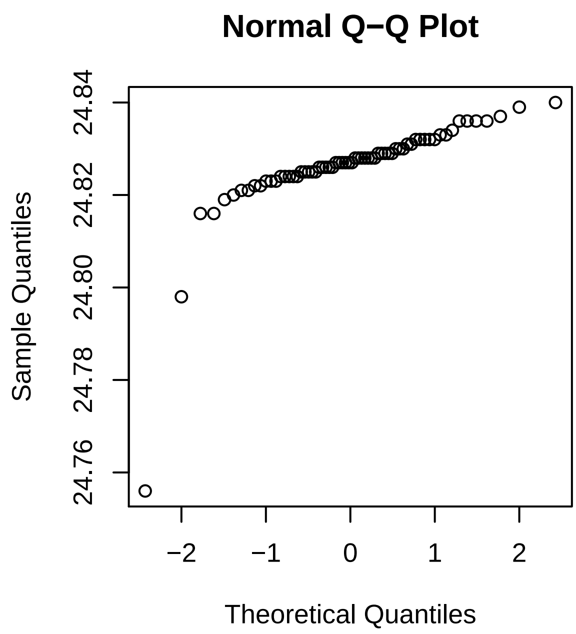
(i) moins de 10 ms,

(ii) plus de 20 ms.

Exercice 3. En 1882, Simon Newcomb construit une expérience afin de déterminer la vitesse la lumière ; plus précisement, il calcule le temps mis par la lumière pour parcourir 7442 mètres au niveau de la mer. L'expérience fut répétée 66 fois, avec les résultats suivants (en 10^{-6} sec) :

24.828	24.826	24.833	24.824	24.834	24.756	24.827	24.816	24.840	24.798
24.829	24.822	24.824	24.821	24.825	24.830	24.823	24.829	24.831	24.819
24.824	24.820	24.836	24.832	24.836	24.828	24.825	24.821	24.828	24.829
24.837	24.825	24.828	24.826	24.830	24.832	24.836	24.826	24.830	24.822
24.836	24.823	24.827	24.827	24.828	24.827	24.831	24.827	24.826	24.833
24.826	24.832	24.832	24.824	24.839	24.828	24.824	24.825	24.832	24.825
24.829	24.827	24.828	24.829	24.816	24.823.				

(a) Un QQ-plot normal de ces données est présenté ci-dessous. Commenter ce graphique. Que représente t-il ? Quelles conclusions peut-on en tirer ?



(b) Les données précédentes ont une moyenne $\bar{x} = 24.8262$ et une déviation standard $s = 0.0107$. Donner un intervalle de confiance à 99% pour la vitesse de la lumière estimée à partir de ces données en précisant les hypothèses que vous faites.

(c) L'estimation actuelle de la vitesse de la lumière dans ces mêmes conditions est de 24.8330×10^{-6} sec. L'estimation basée sur l'expérience de Newcomb concorde-t-elle avec cette valeur ?

(d) Comment l'analyse précédente pourrait-elle être améliorée ? Comment pensez-vous que les résultats précédents en seraient affectés ?

Exercice 4. Soit $X_1, \ldots, X_n$ des variables indépendantes issu d'une loi de Poisson de paramètre λ inconnu. Nous allons comparer deux approches pour estimer λ.

(a) Dans un premier temps nous considérons une approche fréquentiste.

- **(i)** Calculer l'estimateur du maximum de vraisemblance de λ.
- **(ii)** Calculer l'information espérée et donner la loi limite de l'estimateur du maximum de vraisemblance.

(b) Nous considérons maintenant une approche Bayésienne. Nous choisissons une loi à priori Gamma(α,β) pour λ,

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda > 0,$$

d'espérance α/β et de variance α/β^2, avec $\alpha, \beta > 0$.

- **(i)** Calculer la loi à postériori de λ, sachant $X_1, \ldots, X_n$.
- **(ii)** Discuter l'approche Bayésienne par rapport au maximum de vraisemblance lorsque $n \to \infty$.

Rappel : X suit une loi de Poisson de paramètre $\lambda > 0$ si $\Pr(X = x) = \lambda^x e^{-\lambda}/x!$, $x = 0, 1, 2, \ldots$.