

École Polytechnique Fédérale de Lausanne

MATH-233– Examen final Probabilités et statistiques (PH)

Instructor : Anthony Davison

24 janvier 2022

Nom et prénom : _____

SCIPER : _____

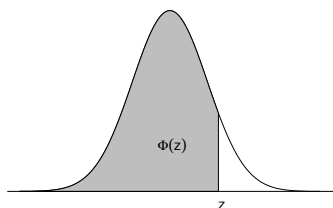
Cet examen contient 6 pages (y compris la page de couverture) et 4 questions. Bonne chance !

Distribution de points

Question	Points	Score
1	19	
2	10	
3	9	
4	9	
Total:	47	

PROTOCOL :

- Aucun document personnel n'est autorisé.
- Les calculatrices simples sont permises. Il est interdit de s'échanger les calculatrices ou d'utiliser d'autres aides électroniques, y compris des natels.
- Justifiez vos réponses ! Une réponse non justifiée sera considérée comme fausse.
- Ecrivez vos réponses sur les feuilles d'examen. Si vous manquez de place, 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 expliquez dans votre solution comment vous la comprenez.
- Des traductions anglaises des questions sont disponibles, mais la version française est celle de référence.
- Vous pouvez donner vos réponses en français ou en anglais.

Distribution normale standard $\Phi(z)$ 

Pour $z < 0$ on utilise la symétrie : $\Pr(Z \leq z) = \Phi(z) = 1 - \Phi(-z)$ pour $z \in \mathbb{R}$.

z	0	1	2	3	4	5	6	7	8	9
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56750	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84850	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92786	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

1. (19 points) (a) Je jette un dé équilibré à six faces trois fois de suite indépendamment et je note les numéros sur leurs faces supérieures. Donner l'ensemble fondamental pour cette expérience aléatoire. Soit X le total de trois numéros notés. Est-il plus probable que $X = 9$ ou que $X = 10$?
- (b) Soient $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\lambda)$, trouver la fonction de densité de $W = \min(X_1^3, \dots, X_n^3)$.
- (c) Si $U \sim U(0, 1)$ trouver la loi de $P = 2 \min(U, 1 - U)$.
- (d) Soient $0 < p < 1$ et x_p le p -quantile de $X \sim \mathcal{N}(3, 4)$, trouver $\Pr(x_{0.1} < X \leq x_{0.95})$.
- (e) Soient $X = a + b_1 Z_1 + c_1 Z_3$ et $Y = b_2 Z_2 + c_2 Z_3$ où $Z_1, Z_2, Z_3 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$, trouver $\text{corr}(X, Y)$.
- (f) Soit X le nombre d'essais jusqu'au premier succès dans une suite d'essais de Bernoulli avec probabilité p de succès. Montrer que $E(X) = p + (1 - p)E(X \mid X > 1)$ et ainsi trouver $E(X)$.
- (g) Soit $X \sim \mathcal{N}(0, \sigma^2)$, trouver $E(|X|)$ et ainsi proposer un estimateur non-biaisé de σ sur la base d'un échantillon aléatoire tiré de cette loi.
- (h) Le taux de reproduction effectif R_e pour le virus Covid-19 indique combien de personnes quelqu'un de contaminé infecte en moyenne. Qu'entend-on par 'l'intervalle de confiance à 95% pour R_e est de (1.11, 1.42)' ?

English :

- (a) I throw a fair six-sided die three times independently and write down the numbers on their top faces. Give the sample space for this experiment. Let X be the total of the three numbers. Is it more likely that $X = 9$ or that $X = 10$?
- (b) If $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\lambda)$, find the density function of $W = \min(X_1^3, \dots, X_n^3)$.
- (c) If $U \sim U(0, 1)$ find the distribution of $P = 2 \min(U, 1 - U)$.
- (d) If $0 < p < 1$ and x_p denotes the p -quantile of $X \sim \mathcal{N}(3, 4)$, find $\Pr(x_{0.1} < X \leq x_{0.95})$.
- (e) If $X = a + b_1 Z_1 + c_1 Z_3$ and $Y = b_2 Z_2 + c_2 Z_3$, where $Z_1, Z_2, Z_3 \stackrel{\text{iid}}{\sim} (0, 1)$, find $\text{corr}(X, Y)$.
- (f) Let X be the number of trials until the first success in a sequence of Bernoulli trials with probability of success p . Show that $E(X) = p + (1 - p)E(X \mid X > 1)$ and thus find $E(X)$.
- (g) If $X \sim \mathcal{N}(0, \sigma^2)$, find $E(|X|)$ and hence propose a non-biased estimator of σ based on a random sample drawn from this distribution.
- (h) The effective reproduction rate R_e for the Covid-19 virus indicates how many people on average are infected by an infected person. What is meant by saying that 'the 95% confidence interval for R_e is (1.11, 1.42)' ?

2. (10 points) Un site web de partage de fichiers possède un très grand nombre de fichiers, dont les tailles S (Go) et les temps de téléchargement T (min) ont une fonction de densité conjointe

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

- (a) Trouver la fonction de densité de probabilité de la taille d'un fichier choisi au hasard. Le temps de téléchargement est-il indépendant de la taille du fichier ?
- (b) Trouver la taille moyenne du fichier.
- (c) Trouver la fonction de densité conditionnelle du temps de téléchargement pour un fichier de taille s Go.
- (d) Quel est le temps moyen de téléchargement d'un fichier de taille s Go ?
- (e) Trouver le temps moyen de téléchargement des fichiers du site.

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

3. (9 points) Chaque matin je prends le train pour aller au travail. Ces dix derniers jours, le train est parti avec les retards (en secondes) suivants :

20 0 58 6 16 23 23 0 41 0

- (a) Sous l'hypothèse de normalité, un intervalle de confiance bilatéral de 95% pour le retard moyen est de (5, 32)s. Le chef de gare aimerait dire qu'un départ à l'heure se situe dans les normes. Cela est-il vrai ? Discuter.
- (b) Il y a eu une erreur lors de la saisie du 16 : le retard réel était de 160 secondes et avec ce changement, l'intervalle de confiance devient $(-2, 68)$ s. Avec ce changement, le chef de station est satisfait. Est-ce raisonnable ? Discuter.
- (c) Expliquez les composants de la figure 1. Est-il raisonnable de supposer que les retards sont normalement distribués ? Si non, proposez un meilleur modèle.

English : Every morning I take the train to go to work. In the past ten days, the train has left with the following delays (in seconds) :

20 0 58 6 16 23 23 0 41 0

- (a) Under the assumption of normality, a two-sided 95% confidence interval for the expected delay is (5, 32)s. The station manager would like to say that an on-time departure is within the norm. Is this true ? Discuss.
- (b) There was a mistake when the 16 was entered : the real delay was 160 seconds and with this change the confidence interval becomes $(-2, 68)$ s. With this change, the station manager is satisfied. Is this reasonable ? Discuss.
- (c) Explain the components of Figure 1. Is it reasonable to suppose that the delays are normally distributed ? If not, suggest a better model.

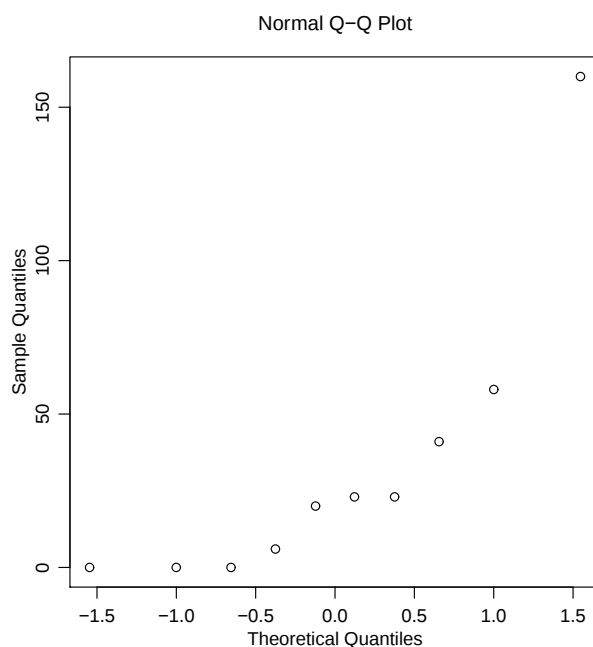


FIGURE 1 – Graphique sommaire pour les retards du train

4. (9 points) Soit X une variable aléatoire de loi binomiale de paramètres n et θ :

$$\Pr(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n, \quad 0 < \theta < 1.$$

- (a) Montrer que l'estimateur de maximum de vraisemblance de θ obtenu à partir d'un échantillon aléatoire $X_1, \dots, X_k$ de cette loi est $\hat{\theta} = (nk)^{-1} \sum_{i=1}^k X_i$.
- (b) 100 étudiants passent un test composé de six questions, auxquelles ils répondent indépendamment. Il faut trois points pour réussir l'examen. Une réponse correcte donne un point et une réponse incorrecte donne zéro point. Pour chaque question, la réponse correcte est trouvée avec la probabilité θ . Les notes de 15 étudiants sont

2 3 2 4 0 5 1 2 3 3 5 2 3 2 3

Donner l'estimation du maximum de vraisemblance de θ basée sur ces données. Trouver un intervalle de confiance pour θ au niveau 0.95.

- (c) Estimer combien des étudiants restants ont réussi l'examen.

English : Let X be a binomial random variable with parameters n and θ :

$$\Pr(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n, \quad 0 < \theta < 1.$$

- (a) Show that the maximum likelihood estimator of θ obtained from a random sample $X_1, \dots, X_k$ from this distribution is $\hat{\theta} = (nk)^{-1} \sum_{i=1}^k X_i$.
- (b) 100 students take a test consisting of six questions, which they answer independently. Three points are required to pass the test. A correct answer gives one point and a wrong answer gives zero points. For each question, the correct answer is found with the probability θ . The scores of 15 students are

2 3 2 4 0 5 1 2 3 3 5 2 3 2 3

Give the maximum likelihood estimate and a 95% confidence interval for θ based on these data.

- (c) Estimate how many of the remaining students passed the exam.