

École Polytechnique Fédérale de Lausanne

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

Instructor : Anthony Davison

29 janvier 2024

Nom et prénom : _____

SCIPER : _____

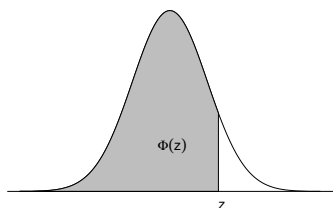
Cet examen contient 7 pages (y compris la page de couverture) et 4 questions. Bonne chance !

Distribution de points

Question	Points	Score
1	20	
2	8	
3	8	
4	10	
Total:	46	

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. (20 points) (a) On jette six fois de suite un dé équilibré à six faces, et on note les nombres apparaissant sur la face du haut. Donnez l'espace de probabilité de cette expérience aléatoire. Est-il plus probable que les faces du haut des jets montrent le même nombre, ou que toutes ces faces aient un nombre différent ?
- (b) Si X et Y sont deux variables aléatoires indépendantes géométriques de paramètre p , calculez $\Pr(Y = X)$.
- (c) Si $X \sim \exp(\lambda)$, calculez la fonction de densité de $Y = X^2$.
- (d) Si $Z_1, Z_2 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ et a, b, c, d, e, f sont des constantes, calculez la corrélation de $X = a + bZ_1 + cZ_2$ et $Y = d + eZ_1 + fZ_2$.
- (e) Si la densité conjointe de variables aléatoires continues U, V est $f(u, v) = cu^2v$ ($0 < u < 2, 0 < v < 1$) pour un certain $c > 0$, et $f(u, v) = 0$ sinon, calculez la densité conditionnelle de V sachant que $U = 1$.
- (f) Si X a une distribution continue avec p quantile x_p , trouvez $\Pr(X > x_p)$.
- (g) Que veut-on dire par la phrase 'la médiane empirique est plus robuste que la moyenne empirique' ? Comment quantifier ces notions ?
- (h) Si, pour n grand, $\bar{X}$ a approximativement une distribution normale de moyenne $\mu \neq 0$ et variance σ^2/n , déterminez la distribution approximative de $1/\bar{X}$.
- (i) Si $I_1, \dots, I_{2n}$ sont des variables aléatoires indépendantes de Bernoulli avec probabilité p , montrez que

$$T = n^{-1} \sum_{j=1}^n I_{2j-1}(1 - I_{2j})$$

est un estimateur non biaisé de $p(1-p)$, et déterminez son erreur moyenne quadratique.

English :

- (a) A fair six-sided die is thrown six times in succession, and the numbers shown on the top faces are recorded. Write down a probability space that represents this random experiment. Is it more likely that all the top faces show the same number, or that these numbers are all different ?
- (b) If X and Y are two independent geometric random variables with parameter p , find $\Pr(Y = X)$.
- (c) If $X \sim \exp(\lambda)$, find the probability density function of $Y = X^2$.
- (d) If $Z_1, Z_2 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ and a, b, c, d, e, f are constants, find the correlation of $X = a + bZ_1 + cZ_2$ and $Y = d + eZ_1 + fZ_2$.
- (e) If the joint density of continuous random variables U, V is $f(u, v) = cu^2v$ ($0 < u < 2, 0 < v < 1$) for some $c > 0$, and $f(u, v) = 0$ elsewhere, find the conditional density of V given that $U = 1$.
- (f) If X has a continuous distribution with p quantile x_p , give $\Pr(X > x_p)$.
- (g) In what sense is the sample median is more robust than the sample average ? How can this be quantified ?
- (h) If, for large n , $\bar{X}$ has an approximate normal distribution with mean $\mu \neq 0$ and variance σ^2/n , give an approximate distribution for $1/\bar{X}$.
- (i) If $I_1, \dots, I_{2n}$ are independent Bernoulli variables with probability p , show that

$$T = n^{-1} \sum_{j=1}^n I_{2j-1}(1 - I_{2j})$$

is an unbiased estimator of $p(1-p)$, and give its mean square error.

2. (8 points) Une transaction de carte de crédit est frauduleuse avec probabilité $p = 0.01$. Lorsque la transaction est légitime, son montant, X (milliers de dollars), peut être modélisé par une densité standard uniforme sur l'intervalle $(0, 1)$, tandis que si la transaction est frauduleuse, X a la densité

$$f(x) = \begin{cases} \frac{9}{(1+9x)^2}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

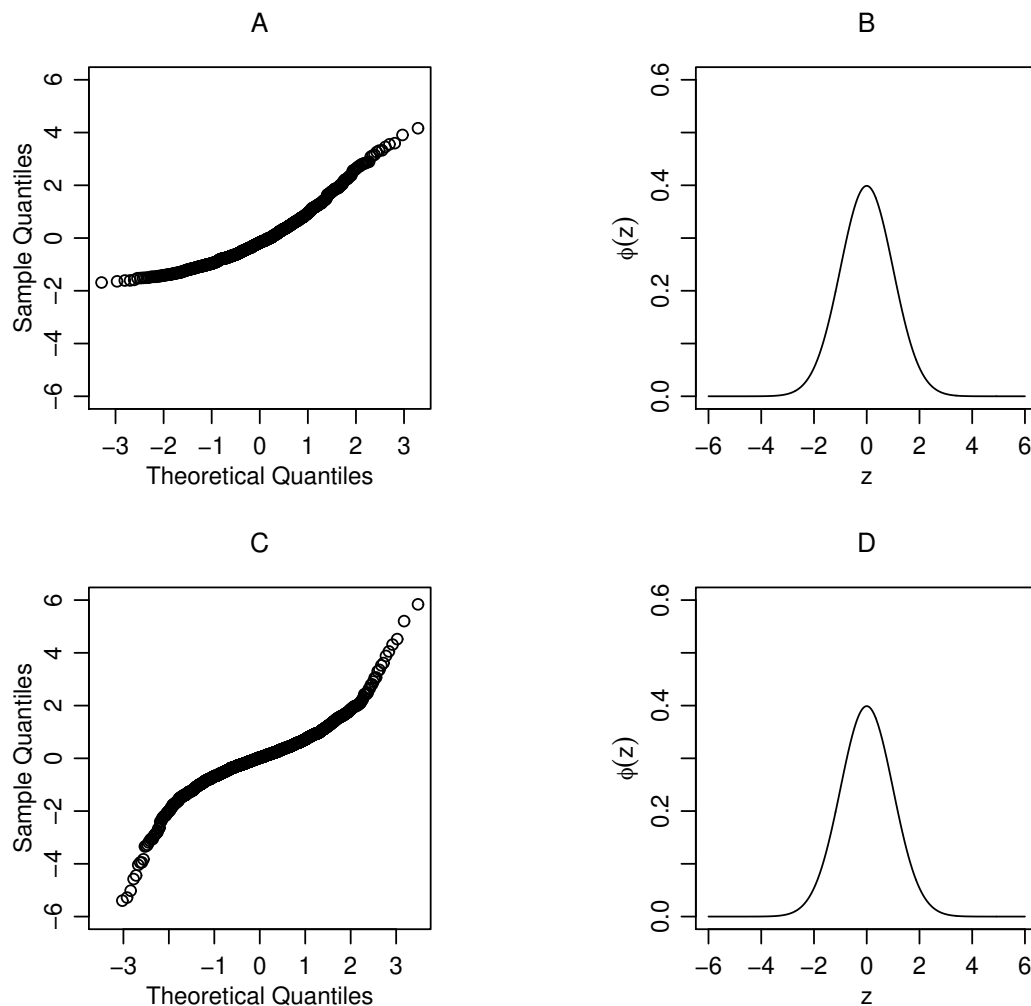
- (a) Toute transaction tentée avec $X > 1$ est automatiquement classée comme frauduleuse et refusée. Quelle est la probabilité de cet événement ?
- (b) Calculez la probabilité qu'une transaction soit frauduleuse, sachant qu'elle est de x milliers de dollars.
- (c) Ma banque doit me rembourser en cas de transaction frauduleuse. Quelle est la perte espérée, pour les transactions qui n'ont pas été refusées ?

English : A credit card transaction is fraudulent with probability $p = 0.01$. When the transaction is legitimate, the amount of money paid, X (thousands of dollars), can be taken to have a standard uniform density on the interval $(0, 1)$, while if the transaction is fraudulent, X has probability density function

$$f(x) = \begin{cases} \frac{9}{(1+9x)^2}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

- (a) Any transaction attempted with $X > 1$ is automatically classified as fraudulent and refused. What is the probability of this event ?
- (b) Find the probability that a transaction is fraudulent, given that it is for x thousand dollars.
- (c) My bank must reimburse me in case of a fraudulent transaction. What is its expected loss, for transactions that have not been refused ?

3. (8 points) (a) Soient $x_1, \dots, x_n$ des observations avec moyenne empirique $\bar{x}$ et variance empirique $s^2 = 0$. Lesquelles des affirmations suivantes sont correctes? Justifier.
- (i) La taille de l'échantillon est trop petite.
 - (ii) Toutes les observations sont égales.
 - (iii) $\bar{x} = 0$.
 - (iv) Les observations sont normalement distribuées.
- (b) Quelles affirmations ci-après s'appliquent à la corrélation empirique r calculée pour données $(x_1, y_1), \dots, (x_n, y_n)$, avec $n \geq 2$? Justifier.
- (i) r dépend des unités dans lesquelles sont exprimées x et y .
 - (ii) r mesure l'association des variables x et y .
 - (iii) r est comprise entre -1 et 1 .
 - (iv) r ne peut jamais être *exactement* zéro.
- (c) Les panneaux A et C du graphique ci-dessous comparent deux échantillons avec les quantiles de la loi normale standard.
- (i) Expliquer comment les données du panneau A diffèrent d'un échantillon normal, et esquisser une densité correspondante sur le panneau B.
 - (ii) Expliquer comment les données du panneau C diffèrent d'un échantillon normal, et esquisser une densité correspondante sur le panneau D.



English : (a) The observations $x_1, \dots, x_n$ have average $\bar{x}$ and sample variance $s^2 = 0$. Which of the following statements are correct? Explain your reasoning. (i) The sample size is too small; (ii) all the observations are equal; (iii) $\bar{x} = 0$; (iv) the observations are normally distributed.

(b) Which of the following statements apply to the empirical correlation r calculated for data $(x_1, y_1), \dots, (x_n, y_n)$, with $n \geq 2$? Explain your reasoning. (i) r depends on the units of measurement of x and y ; (ii) r measures the association between x and y ; (iii) r lies between -1 and 1 ; (iv) r cannot be *exactly* zero.

(c) Panels A and C of the figure compare two samples of data with the quantiles of the standard normal distribution. (i) Explain how the data shown in panel A differ from a normal sample, and sketch the corresponding density function on panel B. (ii) Explain how the data shown in panel C differ from a normal sample, and sketch the corresponding density function on panel D.

4. (10 points) Le tableau 1 montre les nombres de buts marqués par le FC Sion aux six stages de la Coupe Suisse de football depuis la saison 2007–2008.

- (a) Discutez brièvement l'utilisation des distributions (i) hypergéométrique, (ii) binomiale, (iii) exponentielle, et (iv) Poisson pour représenter les nombres de buts.
- (b) En supposant que les nombres de buts peuvent être modélisés par une distribution de Poisson de paramètre λ , déterminez l'estimateur du maximum de vraisemblance de λ et ainsi donner un intervalle de confiance à 95% pour λ .
- (c) Un test d'hypothèse nulle que λ est constant contre l'hypothèse alternative que le nombre moyen de buts marqués dépend du 'round' donne pour p -valeur $p_{\text{obs}} = 3.4 \times 10^{-6}$. Donnez une interprétation de ce nombre. Qu'en concluez-vous?
- (d) Pensez-vous que les analyses en (b) et (c) soient raisonnables?

English : Table 1 shows the numbers of goals scored by FC Sion in the six stages of the Swiss Cup soccer competitions since the 2007–2008 season.

- (a) Briefly discuss the use of the (i) hypergeometric, (ii) binomial, (iii) exponential, and (iv) Poisson distributions to represent football scores.
- (b) On the assumption that the numbers of goals may be modelled by a Poisson distribution with parameter λ , derive the maximum likelihood estimate of λ and use it to give a 95% confidence interval for λ based on the data in Table 1. Discuss.
- (c) A test of the null hypothesis that λ is constant against the alternative hypothesis that the mean number of goals scored depends upon the round gives p -value $p_{\text{obs}} = 3.4 \times 10^{-6}$. Give an interpretation of this number. What do you conclude?
- (d) Do you think the analyses in (b) and (c) seem reasonable?

TABLE 1 – Nombre de buts marqués par le FC Sion lors des tours de la Coupe Suisse de football, 2008–2023. Les espaces indiquent qu'ils ont été battus lors d'un tour précédent. Les matchs qui se terminent en prolongations ou aux tirs au buts sont indiqués par un †. Si le match s'est terminé aux tirs au but, le score donné est le score à la fin des prolongations. En 2009 et en 2011, le FC Sion a remporté la coupe. Il s'est qualifié directement au second tour en 2020 à cause de la pandémie de Covid.

English : Numbers of goals scored by FC Sion in the rounds of the Swiss Cup, 2008–2023. A blank indicates that they were defeated in a previous round. A match with extra time or a penalty shootout is indicated by †; in the case of a penalty shootout, the score given is that at the end of extra time. In 2009 and 2011 FC Sion won the cup, and in 2020 they entered the second round directly due to the Covid pandemic.

Saison	Nombre d'équipes au tour					
	64	32	16	8	4	2
2007–2008	7	2	1			
2008–2009	5	5	1	2	1†	3
2009–2010	1	1†				
2010–2011	4	5	3†	2	2	2
2011–2012	5	2	2	3	0	
2012–2013	1	3	4	2	0	
2013–2014	3	1	0			
2014–2015	3†	1	3	2	1	3
2015–2016	2	2	3†	2†	0	
2016–2017	6	4	5†	5	0†	0
2017–2018	9	1†				
2018–2019	2	1	2†	2†		
2019–2020	10	2	2	2	1	
2020–2021	–	3	2			
2021–2022	3	0				
2022–2023	3†	2	2†	0		
Nombre de matchs	15	16	13	10	8	4
Total nombre de buts au tour	64	35	30	22	5	8