

École Polytechnique Fédérale de Lausanne

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

Instructor : Anthony Davison

27 janvier 2021

Nom et prénom : _____

SCIPER : _____

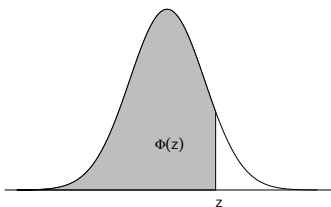
Cet examen contient 7 pages (y compris la page de couverture) et 4 questions. Bonne chance!

Distribution de points

Question	Points	Score
1	20	
2	10	
3	10	
4	10	
Total:	50	

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) 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 équiprobables? Quelle est la probabilité que les deux boules soient rouges, étant donné qu'au moins une des boules est rouge?
- (b) Si $\Pr(X > x) = 1/x^2$, pour $x > 1$, trouvez la fonction de densité de $Y = 1/X$.
- (c) Trouvez la médiane de $Y = \exp(X)$, où $X \sim U(a, b)$ pour $a < b$.
- (d) Si X suit la loi normale avec espérance 2 et $\Pr(X > 6) = 0.1$, trouvez $\Pr(X < 0)$.
- (e) La fonction de masse conjointe des variables aléatoires (X, Y) est donnée dans ce tableau :

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

Trouvez $E(X)$ et $E(X | Y = 4)$. X et Y sont-elles indépendantes?

- (f) 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 $M_X(t) < \infty$. Si $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\lambda)$ et $S = \sum_{j=1}^n X_j$, trouvez $M_S(t)$.
- (g) 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?
- (h) Soient $X_1, \dots, X_n$ un échantillon aléatoire avec fonction de répartition $F(x) = 1 - (4/x)^\alpha$, pour $x > 4$ et $\alpha > 0$. Trouver l'estimateur de maximum de vraisemblance de α .
- (i) Soient $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Pois}(\lambda)$ et $I(\cdot)$ 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 $X_1, \dots, X_n$ étaient dépendantes?

English :

- (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 $Y = \exp(X)$, where $X \sim U(a, b)$ for $a < b$.
- (d) If X is normally distributed with mean 2 and $\Pr(X > 6) = 0.1$, find $\Pr(X < 0)$.
- (e) The joint probability mass function of random variables (X, Y) is given by the following table :

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

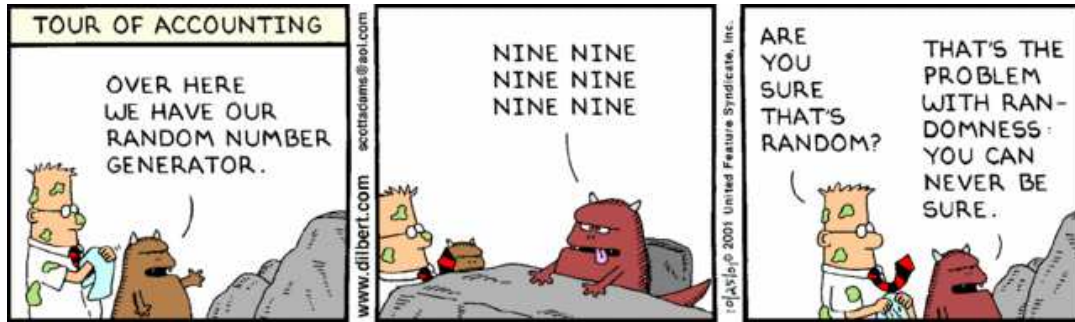
Find $E(X)$ and $E(X | Y = 4)$. Are X and Y independent?

- (f) 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 $M_X(t) < \infty$. If $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\lambda)$ and $S = \sum_{j=1}^n X_j$, find $M_S(t)$.

- (g) 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?
- (h) Let $X_1, \dots, X_n$ be a random sample with cumulative distribution function $F(x) = 1 - (4/x)^\alpha$, for $x > 4$ and $\alpha > 0$. Find the maximum likelihood estimator of α .
- (i) If $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Pois}(\lambda)$ 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 $X_1, \dots, X_n$ were dependent?



2. (10 points) (a) Les comptables utilisent deux générateurs de chiffres aléatoires. Le bon, choisi avec probabilité $1 - p$, génère des entiers uniformément distribués sur $\{0, \dots, 9\}$. Le mauvais, choisi avec probabilité p , génère toujours des 9. Dilbert ne sait pas quel générateur a été choisi, mais il observe une suite de 9. Combien de 9 successifs doit-il observer pour avoir une probabilité d'au moins q que la suite provienne du mauvais générateur ?
- (b) Dilbert a enfin identifié le bon générateur de la partie (a), qui lui livre une suite de variables indépendantes $X_1, X_2, \dots$. De plus il a une source indépendante de variables $U_1, U_2, \dots, \stackrel{\text{iid}}{\sim} U(0, 1)$.
Démontrer que l'algorithme suivant génère $R \sim B(9, a)$, où $a \in (0, 1)$:
1. Générer X du bon générateur et $U \sim U(0, 1)$.
 2. Si $U \leq \binom{9}{X} a^X (1 - a)^{9-X}$, alors mettre $R = X$ et terminer.
 3. Sinon, rejeter X et U , et revenir à l'étape 1.
- (c) Montrer que le nombre d'itérations N de l'algorithme de (b) avant de terminer suit une loi géométrique et trouver son espérance.

English :

- a) The accounting department has two random number generators. The good one, chosen with probability $1 - p$, generates integers uniformly distributed on $\{0, \dots, 9\}$. The bad one, chosen with probability p , always generates 9s. Dilbert is unsure which generator has been chosen, but observes the series 9, 9, 9, ... How many successive 9s must he see to have a probability of at least q that the series comes from the bad generator ?
- b) Dilbert has finally identified the good generator from part (a), which gives him independent variables $X_1, X_2, \dots$. He also has an independent source of uniform variables $U_1, U_2, \dots \stackrel{\text{iid}}{\sim} U(0, 1)$. Show that the following algorithm generates $R \sim B(9, a)$, where $a \in (0, 1)$:
1. Generate X from the good generator and $U \sim U(0, 1)$.
 2. If $U \leq \binom{9}{X} a^X (1 - a)^{9-X}$, then set $R = X$ and terminate.
 3. Otherwise reject X and U and go back to step 1.
- (c) Show that the number of iterations N needed for the algorithm in (b) to terminate has a geometric distribution and find its expectation.

3. (10 points) Deux téléchargements successifs par une application durent (en minutes) $X_1 \sim \mathcal{N}(8, 3^2)$ et $X_2 \sim \mathcal{N}(16, 4^2)$. Les temps de téléchargement sont indépendants.
- (a) Quelle est la distribution de la durée totale de téléchargement $T = X_1 + X_2$?
 - (b) Quelle est la probabilité que T excède 30 minutes ?
 - (c) Etant donné $X_1 = 10$, donnez la probabilité que T excède 30 minutes.
 - (d) Etant donné que T était de 30 minutes, quelle est la probabilité que X_1 était inférieur à 7 minutes ?

Rappel : si

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim \mathcal{N}_2 \left\{ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} \right\},$$

alors $W_1 \mid W_2 = w \sim \mathcal{N}\{\mu_1 + \omega_{12}\omega_{22}^{-1}(w - \mu_2), \omega_{11} - \omega_{12}^2/\omega_{22}\}$.

English : Two successive software downloads take times (minutes) $X_1 \sim \mathcal{N}(8, 3^2)$ and $X_2 \sim \mathcal{N}(16, 4^2)$. The download times are independent.

- (a) What is the distribution of the total download time $T = X_1 + X_2$?
- (b) What is the probability that T exceeds 30 minutes ?
- (c) Given that $X_1 = 10$, find the probability that T exceeds 30 minutes.
- (d) Given that T was 30 minutes, what is the probability that X_1 was less than 7 minutes ?

Reminder : if

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim \mathcal{N}_2 \left\{ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} \right\},$$

then $W_1 \mid W_2 = w \sim \mathcal{N}\{\mu_1 + \omega_{12}\omega_{22}^{-1}(w - \mu_2), \omega_{11} - \omega_{12}^2/\omega_{22}\}$.

4. (10 points) Des résumés des résultats de deux groupes d'élèves à un examen se trouvent au tableau et aux diagrammes ci-dessous.
- Expliquez les éléments du diagramme de gauche de la figure 1. Qu'en déduisez-vous ?
 - Expliquez les éléments du diagramme de droite de la figure 1. Qu'en déduisez-vous ? En le dessinant sur le diagramme, indiquez ce à quoi le diagramme ressemblerait pour le groupe 2.
 - En supposant que les tailles des groupes sont assez grandes pour que les erreurs d'estimation des déviations standards puissent être négligées, donnez la variance de la différence des moyennes des groupes. Fournissez un intervalle de confiance de niveau 95% pour l'espérance de cette différence, et expliquez soigneusement comment votre intervalle doit être interprété.
 - Listez les hypothèses faites en (c) et discutez de leur plausibilité.

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

- Explain the elements of the left-hand panel of Figure 1. What do you infer from it ?
- 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.
- 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.
- Explain what assumptions were made in (c) and discuss their plausibility.

TABLE 1 – Résumé des résultats de deux groupes d'élèves à un examen

	Taille du groupe, n	Moyenne, $\bar{x}$	Déviations standard, s
Groupe 1	100	3.66	1.06
Groupe 2	80	4.05	0.75

FIGURE 1 – Graphiques sommaires pour les résultats

