

Important Distributions

The nine most important distributions we have discussed so far are listed below, each with its PMF/PDF. Each of these distributions is important because it has a natural, useful story, so understanding these stories (and recognizing equivalent stories) is crucial. It is also important to know how these distributions are related to each other. For example, $\text{Bern}(p)$ is the same as $\text{Bin}(1, p)$, and $\text{Bin}(n, p)$ is approximately $\text{Pois}(\lambda)$ if n is large, p is small, and $\lambda = np$ is moderate.

Name	Param.	PMF or PDF
Bernoulli	p	$P(X = 1) = p, P(X = 0) = 1 - p$
Binomial	n, p	$\binom{n}{k} p^k (1 - p)^{n-k}$, for $k \in \{0, 1, \dots, n\}$
Geometric	p	$(1 - p)^k p$, for $k \in \{0, 1, 2, \dots\}$
Negative Binomial	r, p	$\binom{r+n-1}{r-1} p^r (1 - p)^n$, $n \in \{0, 1, 2, \dots\}$
Hypergeometric	w, b, n	$\frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}}$, for $k \in \{0, 1, \dots, n\}$
Poisson	λ	$\frac{e^{-\lambda} \lambda^k}{k!}$, for $k \in \{0, 1, 2, \dots\}$
Exponential	λ	$\lambda e^{-\lambda x}$ for $x > 0$
Uniform	$a < b$	$\frac{1}{b-a}$, for $x \in (a, b)$
Normal	μ, σ^2	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Some Useful Formulas

De Morgan's Laws

$$(A_1 \cup A_2 \cup \dots \cup A_n)^c = A_1^c \cap A_2^c \cap \dots \cap A_n^c$$

$$(A_1 \cap A_2 \cap \dots \cap A_n)^c = A_1^c \cup A_2^c \cup \dots \cup A_n^c$$

Complements

$$P(A^c) = 1 - P(A)$$

Unions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i), \text{ if the } A_i \text{ are disjoint}$$

$$P(A_1 \cup A_2 \cup \dots \cup A_n) \leq \sum_{i=1}^n P(A_i)$$

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{k=1}^n \left((-1)^{k+1} \sum_{i_1 < i_2 < \dots < i_k} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) \right)$$

(Inclusion-Exclusion)

Intersections

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \cdots P(A_n|A_1, \dots, A_{n-1})$$

Law of Total Probability

If E_1, E_2, \dots, E_n are a partition of the sample space S (i.e., they are disjoint and their union is all of S) and $P(E_j) \neq 0$ for all j , then

$$P(B) = \sum_{j=1}^n P(B|E_j)P(E_j)$$

PROBABILITY AND STATISTICS (MATH-235)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Final exam - questions

Date: 27th of January, 2025

Time: 9:15–12:15

Name: _____

SCIPER: _____

INSTRUCTIONS TO CANDIDATES

- To obtain the maximum number of points you should be clear about your reasoning and present your arguments explicitly. You have **3 hours** to complete the exam.
- All that can be used for this exam is a pen and the cheat sheet that **we provide for you**. No external notes, books, summaries, formula collections or calculators are allowed. All questions should be answered.
- The finest enumerated item in each question will be marked on a scale of 0 – 2 points, indicating an incorrect, partially correct and completely correct answer respectively (half-points are not given). **The exam has 8 questions with a total of 84 points.**
- **Write the answer to every question in the other booklet (final exam - answers)**. Scrap paper will be provided for rough work, but only answers written in the booklet will be marked.
- At the end of the exam, you will have to return everything : the booklet with the questions, the booklet with your answers, and the scrap paper.
- When the exercise directs you to “**Simplify!**” provide the simplest fraction, a decimal, or the simplest form possible using the function or parameter in terms of which your answer must be given. If you are explicitly instructed with “**Do not simplify!**”, it is acceptable to leave your answer as a non-simplified fraction, integral, sum, or another more complex form, depending on the context.
- Please write your name and SCIPER number on every sheet you use.
- You can write your answers in English or French, but please commit to the same language throughout the exam.

Mark question 1 (TOT: 14 points):

Mark question 2 (TOT: 8 points):

Mark question 3 (TOT: 8 points):

Mark question 4 (TOT: 12 points):

Mark question 5 (TOT: 14 points):

Mark question 6 (TOT: 6 points):

Mark question 7 (TOT: 8 points):

Mark question 8 (TOT: 14 points):

VARIABLE DEFINITIONS

All the random variables have finite mean and variance.

Question 1.

Write the most appropriate of \leq , \geq , $=$ or $?$ in the blank for each part (where “?” means that no relation holds in general.). No justification is needed for full points. In this exercise, you do not get partial points, for each subquestion you can get 2 or 0 points.

In this exercise, suppose A and B are independent events. The positive random variables X and Y are not necessarily independent, but they are identically distributed.

- (a) You toss two fair coins, Coin C_1 and C_2 twice and four times, respectively.

$$P(\text{At least 1 Head out of the 2 tosses with } C_1) \text{ ______}$$

$$P(\text{At least 2 Heads out of the 4 tosses with } C_2) \text{ ______}$$

(b) $P((A \cup B)^c) \text{ ______ } P(A^c)P(B^c)$

(c) $P(A|B) \text{ ______ } P(B|B)$

(d) $E((X + Y)^2) \text{ ______ } 2E(X^2)$

(e) $E(E(\log(Y)|X)) \text{ ______ } \log(E(Y))$

(f) $E\left(\frac{X}{Y}\right) \text{ ______ } 1$

(g) $P(X > 3) \text{ ______ } \frac{1}{20}E(X^3)$

Question 2.

A company designed two tests to identify SARS-CoV-infected individuals, Test A and Test B. An individual who has the disease (event D) or is disease-free (event D^c) is tested. Suppose that 5% of the population has the disease. The tests can either be positive (events T_A and T_B) or negative (events T_A^c and T_B^c). Suppose that Test A has a sensitivity and specificity of 90% and Test B has a sensitivity and specificity of 50%, that is

$$P(T_A|D) = P(T_A^c|D^c) = 0.9$$

$$P(T_B|D) = P(T_B^c|D^c) = 0.5.$$

Given whether an individual is infected or not, the test results are conditionally independent.

- (a) An individual sampled at random is tested with Test A and returns as positive. Conditional on the test result, what is the probability that this individual has the disease? Do not simplify!
- (b) A second individual independent of the first one is tested with Test B and returns as negative. Conditional on the test result, what is the probability that the second individual is disease-free? Simplify!
- (c) Would multiple negative results of Test B increase the conditional probability of the disease-free event, given these results? Justify!
- (d) Are T_A and T_B independent(unconditionally)? It is sufficient to justify with words, a numerical computation is not necessarily required.

Question 3.

Let X and Y have joint probability density function (PDF) $f_{X,Y}(x,y) = c(x+y)$, for $0 < x < y < 1$, and $f_{X,Y}(x,y) = 0$ otherwise.

- (a) Find c to make this a valid joint PDF.
- (b) Find the marginal PDFs of X and Y .
- (c) Are X and Y independent?
- (d) Find the conditional PDF of Y given $X = x$.

Question 4.

Suppose a factory produces a random number of gadgets, N , where $N \sim \text{Pois}(\lambda)$. Each gadget is independently operational with probability p and is faulty with probability $q = 1 - p$. Let X be the number of operational gadgets and Y the number of faulty gadgets, so $X + Y = N$.

- (a) Find the conditional distribution functions $X|N = 1$ and $Y|N = 1$, that is, find $P(X = k|N = 1)$ and $P(Y = k|N = 1)$ for all k .
- (b) Find the conditional distribution functions $X|N = n$ and $Y|N = n$, that is, find $P(X = k|N = n)$ and $P(Y = k|N = n)$ for all k .
- (c) Is $P(X = x, Y = y|N = x + y)$ equal to $P(X = x|N = x + y)$? Justify!
- (d) What is the joint PMF of X and Y ? **Hint:** $e^z = e^{zc+z(1-c)}$
- (e) What are the marginal PMFs of X and Y ?
- (f) Are X and Y independent? Are X and Y conditionally independent given $N = n$? Justify!

Question 5.

Let $X \sim \text{Unif}(-1, 1)$.

(a) Define $Y = X^2 - \frac{1}{3}$.

(i) Find the PDF of Y .

(ii) Find $E(Y)$.

(iii) Find $\text{Cov}(X, Y)$.

(iv) Are X and Y independent? Justify!

(b) Define $W = \frac{X+1}{2}$.

(i) What is the distribution of W ?

(ii) As a function of W create Z with CDF $F(z) = e^z/(1 + e^z)$.

(iii) Find $E(Z)$.

Question 6.

Suppose X is a non-negative random variable.

(a) Argue that $X = XI(X > 0)$, where $I(X > 0)$ is the indicator random variable of $X > 0$.

(b) Show that $P(X > 0) \geq \frac{(E(X))^2}{E(X^2)}$.

Hint: Use the Cauchy-Schwarz inequality.

(c) Suppose a student is presented with 5 multiple choice questions with 4 possible answers to each of them. They answer all questions completely at random and independently from each other. Using the inequality above, give an upper bound for the probability that the student made no mistakes. Simplify!

Question 7.

Let $Y = 4e^{3X}$, with $X \sim \text{Exp}(7)$.

(a) Find the mean of Y . Simplify!

(b) Find the variance of Y . Simplify!

(c) Suppose Y_1, \dots, Y_n are i.i.d. with the same distribution as Y . What is the approximate distribution of the sample mean $\bar{Y}_n = \frac{\sum_{i=1}^n Y_i}{n}$ when n is large? Simplify, and specify all parameters!

(d) You receive a random sample y_1, \dots, y_7 from the distribution of Y . Suppose that the mean (μ) of Y is unknown; however, the variance (σ^2) is known, as calculated in part (b).

We find that

$$\bar{y}_7 = \frac{\sum_{i=1}^7 y_i}{7} = 6.5 .$$

Suppose that 7 is "large" in the sense of part (c).

Construct a 95% confidence interval for μ . Simplify!

Hint: You can take $\Phi^{-1}(0.975) \approx 2$

Question 8.

A total of 200 teams compete in a marathon relay. Each team consists of two members, and each member runs a half marathon; that is, 400 runners complete the half marathon distance. The runners' completion times are assumed to be independent and identically distributed (iid) with a continuous cumulative distribution function (CDF) F .

- (a) Five-digit starting numbers are to be created using the digits 1, 2, 3, 4, 5, 6, 7, 8. If no digit can be repeated, and the first, third, and fifth digits of the numbers must be odd, how many distinct starting numbers can be formed? Simplify!
- (b) The runners with 10 fastest half-marathon times receive a prize. What is the probability that *exactly* one team has both of its members included among the 10 runners with the fastest half-marathon times? Do not simplify!
- (c) Outside the 400 marathon relay participants, Alice also completes the half marathon. Her running time is independent and identically distributed according to the same CDF F . For the rest of this question, let A represent the number of runners from the 400 relay participants who finish the half marathon faster than Alice, and let T represent Alice's finishing time.
 - (i) Find the conditional distribution of A , given that $T = t$.
 - (ii) Find $E(A|T)$.
 - (iii) Find $E(A)$. Simplify!
 - (iv) Find $\text{Var}(A|T)$.
 - (v) Find $\text{Var}(A)$. Simplify!

Unofficial French Translation

Question 1.

Pour chaque partie, écrivez dans l'espace le symbole le plus appropriée parmi \leq , \geq , $=$ ou $?$ (où "?" signifie qu'il n'y a pas de relation en général). Aucune justification n'est nécessaire pour obtenir tout les points. Dans cet exercice, vous n'obtenez pas de points partiels, pour chaque sous-question vous pouvez obtenir 2 ou 0 points.

Dans cet exercice, supposons que A et B sont des événements indépendants, tandis que les variables aléatoires strictement positives X et Y ne sont pas nécessairement indépendantes, mais identiquement distribuées.

- (a) Vous lancez deux pièces justes, les pièces C_1 et C_2 , deux fois et quatre fois, respectivement.

$$\frac{P(\text{Au moins 1 Pile sur les 2 lancers avec } C_1)}{P(\text{Au moins 2 Piles sur les 4 lancers avec } C_2)} \text{ ______}$$

(b) $P((A \cup B)^c) \text{ ______ } P(A^c)P(B^c)$

(c) $P(A|B) \text{ ______ } P(B|B)$

(d) $E((X + Y)^2) \text{ ______ } 2E(X^2)$

(e) $E(E(\log(Y)|X)) \text{ ______ } \log(E(Y))$

(f) $E\left(\frac{X}{Y}\right) \text{ ______ } 1$

(g) $P(X > 3) \text{ ______ } \frac{1}{20}E(X^3)$

Question 2.

Une entreprise a conçu deux tests pour identifier les personnes infectées par le SARS-CoV, le test A et le test B. Une personne atteinte de la maladie (événement D) ou non (événement D^c) est testée. Supposons que 5% de la population soit atteinte de la maladie. Les tests peuvent être positifs (événements T_A et T_B) ou négatifs (événements T_A^c et T_B^c). Supposons que le test A ait une sensibilité et une spécificité de 90% et que le test B ait une sensibilité et une spécificité de 50%, soit

$$P(T_A|D) = P(T_A^c|D^c) = 0.9$$

$$P(T_B|D) = P(T_B^c|D^c) = 0,5.$$

Qu'un individu soit infecté ou non, les résultats du test sont conditionnellement indépendants.

- (a) Un individu échantillonné au hasard est testé avec le test A et son résultat est positif. Etant donné le résultat du test, quelle est la probabilité que cet individu soit atteint de la maladie? Ne pas simplifier!
- (b) Un deuxième individu indépendant du premier est testé avec le test B et le résultat est négatif. Etant donné le résultat du test, quelle est la probabilité que le deuxième individu soit indemne de la maladie? Simplifiez!

- (c) Compte tenu de ces résultats, plusieurs résultats négatifs au test B augmenteraient-ils la probabilité conditionnelle que l'individu soit indemne de la maladie? Justifiez!
- (d) Est-ce que T_A et T_B sont (inconditionnellement) indépendants? Il suffit de justifier votre réponse avec des mots, un calcul numérique n'est pas nécessaire.

Question 3.

Supposons que la fonction de densité de probabilité (PDF) jointe de X et Y est $f_{X,Y}(x, y) = c(x + y)$, pour $0 < x < y < 1$, et $f_{X,Y}(x, y) = 0$ sinon.

- (a) Trouvez c pour en faire une PDF jointe valide.
- (b) Trouvez les PDF marginales de X et Y .
- (c) Est-ce que X et Y sont indépendants?
- (d) Trouver la PDF conditionnelle de Y étant donné $X = x$.

Question 4.

Supposons qu'une usine produise un nombre aléatoire de gadgets, N , où $N \sim Pois(\lambda)$. Chaque gadget est indépendamment opérationnel avec une probabilité p et est défectueux avec une probabilité $q = 1 - p$. Soit X le nombre de gadgets opérationnels et Y le nombre de gadgets défectueux, tels que $X + Y = N$.

- (a) Trouver les fonctions de distribution conditionnelle $X|N = 1$ et $Y|N = 1$, c'est-à-dire trouver $P(X = k|N = 1)$ et $P(Y = k|N = 1)$ pour tout k .
- (b) Trouver les fonctions de distribution conditionnelle $X|N = n$ et $Y|N = n$, c'est-à-dire trouver $P(X = k|N = n)$ et $P(Y = k|N = n)$ pour tout k .
- (c) Est-ce que $P(X = x, Y = y|N = x + y)$ est égal à $P(X = x|N = x + y)$? Justifiez!
- (d) Quel est la fonction de masse (PMF) conjointe de X et Y ?
Indice: $e^z = e^{zc+z(1-c)}$
- (e) Quel est la PMF marginale de X et Y ?
- (f) Les variables X et Y sont-elles indépendantes? Est-ce que X et Y sont conditionnellement indépendants étant donné $N = n$? Justifiez!

Question 5.

Soit $X \sim Unif(-1, 1)$.

- (a) Définir $Y = X^2 - \frac{1}{3}$.
 - (i) Trouvez la fonction de densité de probabilité (PDF) de Y .
 - (ii) Trouvez $E(Y)$.
 - (iii) Trouvez $Cov(X, Y)$.
 - (iv) Les variables X et Y sont-elles indépendantes? Justifiez!

- (b) Définir $W = \frac{X+1}{2}$.
- (i) Quelle est la distribution de W ?
- (ii) En fonction de W , créer Z avec la fonction de coût $F(z) = e^z/(1 + e^z)$.
- (iii) Trouvez $E(Z)$.

Question 6.

Supposons que X soit une variable aléatoire positive ou nulle.

- (a) Argumenter que $X = XI(X > 0)$, où $I(X > 0)$ est la variable aléatoire indicatrice de $X > 0$.
- (b) Montrer que $P(X > 0) \geq \frac{(E(X))^2}{E(X^2)}$.
Indice: Utilisez l'inégalité de Cauchy-Schwarz.
- (c) Supposons qu'un étudiant se voit présenter 5 questions à choix multiples avec 4 réponses possibles pour chacune d'entre elles. Il répond à toutes les questions complètement au hasard et indépendamment les unes des autres. En utilisant l'inégalité ci-dessus, donnez une borne supérieure pour la probabilité que l'étudiant n'ait pas fait d'erreur. Simplifiez!

Question 7.

Soit $Y = 4e^{3X}$, avec $X \sim Exp(7)$.

- (a) Trouver la moyenne de Y . Simplifiez!
- (b) Trouvez la variance de Y . Simplifiez!
- (c) Supposons que Y_1, \dots, Y_n sont i.i.d. avec la même distribution que Y . Quelle est la distribution approximative de la moyenne de l'échantillon $\bar{Y}_n = \frac{\sum_{i=1}^n Y_i}{n}$ lorsque n est grand? Simplifiez et précisez tous les paramètres!
- (d) Vous recevez un échantillon aléatoire y_1, \dots, y_7 de la distribution de Y . Supposons que la moyenne (μ) de Y soit inconnue; cependant, la variance (σ^2) est connue, égale à celle calculée dans la partie (b).

Nous trouvons que

$$\bar{y}_7 = \frac{\sum_{i=1}^7 y_i}{7} = 6.5 .$$

Supposons que 7 soit "grand" au sens de la partie (c).

Construire un intervalle de confiance à 95% pour μ . Simplifiez!

Indice: Vous pouvez prendre $\Phi^{-1}(0.975) \approx 2$.

Question 8.

Un total de 200 équipes participent à un marathon relais. Chaque équipe est composée de deux membres, et chaque membre court un semi-marathon; c'est-à-dire que 400 coureurs parcourent la distance du semi-marathon. Les temps de parcours des coureurs sont supposés

être indépendants et identiquement distribués (iid) avec une fonction de distribution cumulative continue (CDF) F .

- (a) Les nombres de départ à cinq chiffres doivent être créés en utilisant les chiffres 1, 2, 3, 4, 5, 6, 7, 8. Si aucun chiffre ne peut être répété et que les premier, troisième et cinquième chiffres des nombres doivent être impairs, combien de nombres de départ distincts peuvent être formés? Simplifiez!
- (b) Les coureurs ayant réalisé les 10 temps les plus rapides au semi-marathon reçoivent un prix. Quelle est la probabilité que *exactement* une équipe ait ses deux membres parmi les 10 coureurs ayant réalisé les temps les plus rapides au semi-marathon? Ne pas simplifier!
- (c) En dehors des 400 participants au relais marathon, Alice participe également au semi-marathon. Son temps de course est indépendant et identiquement distribué selon la même CDF F . Pour le reste de cette question, notons A le nombre de coureurs parmi les 400 participants au relais qui terminent le semi-marathon plus vite qu'Alice, et notons T le temps d'arrivée d'Alice.
 - (i) Trouvez la distribution conditionnelle de A , étant donné que $T = t$.
 - (ii) Trouvez $E(A|T)$.
 - (iii) Trouvez $E(A)$. Simplifiez!
 - (iv) Trouvez $\text{Var}(A|T)$.
 - (v) Trouvez $\text{Var}(A)$. Simplifiez!