Midterm exam (March, 27th)

Problem 1 (Basics)**.** Give a concise (but justified) answer to the following questions. When using a theorem from the course, state it clearly.

1. We flip a fair coin (i.e., heads with probability $\frac{1}{2}$ and tails with probability $\frac{1}{2}$) 6 times independently. What is the probability that the number of tails is exactly equal to the number of heads we obtain? `Sol:` $\binom{6}{3}2^{-6}$

2. Let X_1 and X_2 be independent and uniformly distributed on $\{1, 2, 3\}$.
Compute $\mathbf{E}\{\max(X_1, X_2) + \min(X_1, X_2)\}$.

3. Let X be a binomial random variable with parameters n and $\frac{1}{2}$. For all $k \in \{0, \dots, n\}$, compute $\mathbf{P}\{X = k | X \text{ is even}\}$.

4. The following version of Chernoff's bound has a missing assumption. Show that this statement is not true and identify the missing assumption.

   Let $X_1, \dots, X_n$ be independent random variables with $\mathbf{E}\{\sum_{i=1}^{n} X_i\} = \mu$. Then for any $\epsilon > 0$
$$\mathbf{P}\left\{\sum_{i=1}^{n} X_i \geq (1+\epsilon)\mu\right\} \leq \exp\left(-\mu\frac{\epsilon^2}{2+\epsilon}\right)$$

5. Let $t \in]0, \infty[$ and let X_n be a binomial random variable with parameters n and $\frac{t}{n}$. For fixed values of t and $k \in \mathbb{N}$, compute $\lim_{n\to\infty} \mathbf{P}\{X_n = k\}$.

6. Suppose X and Y are independent random variables with $\mathbf{E}\{X\} = \mathbf{E}\{Y\} = 0$. Compute the fourth moment of $X + Y$, i.e., $\mathbf{E}\{(X + Y)^4\}$ as a function of the moments of X and Y. `Sol:` $\mathbf{E}\{X^4\} + 6\mathbf{E}\{X^2\}\mathbf{E}\{Y^2\} + \mathbf{E}\{Y^4\}$.

7. We construct a random bipartite graph on vertex set $\{1, \dots, 2n\}$ as follows. We add edges of the form $\{i, j\}$ one by one independently by choosing $i \in \{1, \dots, n\}$ and $j \in \{n+1, \dots, 2n\}$ uniformly at random. Let X be the number of edges added when we first have $\deg(u) \geq 1$ for all vertices $u \in \{1, \dots, n\}$. Compute $\mathbf{E}\{X\}$.

8. Let $\Omega = \{1, 2, 3, 4\}$. Describe explicitly the σ-algebra generated by $\{\{1\}, \{2\}\}$.

Problem 2 (Las Vegas and Monte Carlo algorithms)**.** Let $f : \{0, 1\}^* \to \{0, 1\}^*$ be a function we would like to compute.

1. Suppose you have a Las Vegas algorithm $\mathcal{A}$ for f, i.e., the algorithm $\mathcal{A}$ on input x always outputs $f(x)$ but the running time is random. We only know that the expected running time is upper bounded by $T(n)$, where $n = |x|$. Design an algorithm with running time upper bound deterministically by $O(T(n))$ but that may output the wrong answer with probability at most $\frac{1}{10}$.

2. Now suppose you have another algorithm $\mathcal{A}$ that on input x outputs $f(x)$ with probability at least $p > 0$ in time bounded by $T_1(n)$ with $n = |x|$. Suppose in addition that given x and $\mathcal{A}(x)$, it is possible to deterministically check whether $\mathcal{A}(x) = f(x)$ in time $T_2(n)$. Describe a Las Vegas algorithm for computing f with expected running time $O(\frac{T_1(n)+T_2(n)}{p})$.

Problem 3. Let $\vec{X}$ be a random vector in $\mathbb{R}^d$, i.e., $\vec{X} = (X_1, \ldots, X_d)$ where each X_i is a random variable. We naturally define $\mathbf{E}\left\{\vec{X}\right\} = (\mathbf{E}\{X_1\}, \mathbf{E}\{X_2\}, \ldots, \mathbf{E}\{X_d\}) \in \mathbb{R}^d$. We write $\vec{\mu} = \mathbf{E}\left\{\vec{X}\right\} \in \mathbb{R}^d$.

For this problem we assume that $\vec{X} \in [0,1]^d$. Our objective is to estimate $\vec{\mu}$ to some fixed precision ϵ. For this we have access to independent samples $\vec{X}^1, \ldots, \vec{X}^n$ all having the same distribution as $\vec{X}$.

1. Propose an estimator $\vec{Y}$ for $\vec{\mu}$ as a function of $\vec{X}^1, \ldots, \vec{X}^n$. Your estimator should have the property that $\mathbf{E}\left\{\vec{Y}\right\} = \vec{\mu}$.

2. As a function of ϵ and n, find an upper bound on $\mathbf{P}\left\{\|\vec{Y} - \vec{\mu}\|_\infty \geq \epsilon\right\}$ where for a vector $\vec{v} = (v_1, \ldots, v_d)$, we define $\|\vec{v}\|_\infty = \max_{i\in\{1,\ldots,d\}} |v_i|$.

Problem 4 (Dominating sets)**.** Let $G = (V, E)$ be an undirected graph with $|V| = n$. A dominating set of the graph is a subset $S \subseteq V$ such that for any vertex $v \in V$, v is connected to an element in S, i.e., there exists $s \in S$ such that $\{s, v\} \in E$.

The objective of this problem is to explore the size of the smallest dominating set of a graph with minimum degree $\geq \mu n$, where $\mu \in [0, 1]$ is a constant.

1. Compute the size of the smallest dominating set for the complete graph K_n on n vertices and the complete bipartite graph $K_{n,n}$ between two sets of n vertices each. `Sol: 1 and 2`

2. Naturally, we expect that if the minimum degree of the graph is large, then there is a small dominating set. Prove that this intuition is false: for any $\mu \in]0, 1[$ and k, there exists an integer n and a graph G on n vertices with minimum degree at least μn but with no dominating set of size $\leq k$.

   Hint: you may consider a random graph $\mathcal{G}_{n,p}$ for a well-chosen p. `Sol:` $\mathbf{P}\{S\text{dominating}\} = (1-(1-p)^k)^{n-k}$. `So` $\mathbf{P}\{\exists S, \text{dominating}\} \leq \binom{n}{k}(1-(1-p)^k)^{n-k}$ `which goes to 0 as` $n \to \infty$. `Then pick` $p \in]\mu, 1[$ `will do the job, with Chernoff.`

3. Let us call a *distance-two dominating set* a subset $S \subseteq V$ such that for any $v \in V$, there exists a vertex $s \in S$ such that v is at distance at most 2 from s. We will show that for any $\mu \in]0, 1[$, there exists a k only depending on μ such that for every integer n and every graph G on n vertices with minimum degree μn, G has a distance-two dominating set of size $\leq k$.

   In order to prove this result, we fix $\mu \in]0, 1[$ and we consider an integer n and a graph $G = (V, E)$ on n vertices, with minimum degree $\geq \mu n$.

   Choosing an appropriate value for k (which depends on μ but not on n), prove that G has a distance-two dominating set of size $\leq k$.

   Hint: you can select randomly a set S of size at most k and consider the random variable that corresponds to the number of vertices in V that are connected to (at least) one vertex of S.

   `Sol: Take` k `so that` $1-(1-\mu)^k \geq 1-\mu$ `and then take an` S `such that` $N > (1-\mu)n$. `But then for any` $v \in V$ `it has to have at least one neighbour in this set of size` N.

French version

Problem 1 (Basiques)**.** Donner une réponse concise (mais justifiée) aux questions suivantes. Si vous utilisez un théorème du cours, énoncez-le clairement.

1. On lance une pièce équilibrée (pile ou face avec probabilité $\frac{1}{2}$) 6 fois de manière indépendante. Quelle est la probabilité que l'on obtienne autant de piles que de faces? `Sol:` $\binom{6}{3}2^{-6}$

2. Soit X_1 et X_2 indépendentes et uniformément distribuées sur $\{1, 2, 3\}$.
   Calculer $\mathbf{E}\{\max(X_1, X_2) + \min(X_1, X_2)\}$.

3. Soit X une variable aléatoire binomiale de paramètres n et $\frac{1}{2}$. Pour tout $k \in \{0, \ldots, n\}$, calculer $\mathbf{P}\{X = k | X \text{ est pair}\}$.

4. La version suivante de la borne de Chernoff a une hypothèse manquante. Montrer que cet énoncé est faux et identifier l'hypothèse manquante.

   Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes avec $\mathbf{E}\left\{\sum_{i=1}^n X_i\right\} = \mu$. Alors, pour tout $\epsilon > 0$

   $$\mathbf{P}\left\{\sum_{i=1}^n X_i \geq (1+\epsilon)\mu\right\} \leq \exp\left(-\mu\frac{\epsilon^2}{2+\epsilon}\right)$$

5. Soit $t \in]0, \infty[$ et soit X_n une variable aléatoire binomiale avec pour paramètres n et $\frac{t}{n}$. Pour des valeurs fixées de t et $k \in \mathbb{N}$, calculer $\lim_{n\to\infty} \mathbf{P}\{X_n = k\}$.

6. Soit X et Y des variables aléatoires indépendantes avec $\mathbf{E}\{X\} = \mathbf{E}\{Y\} = 0$. Calculer le quatrième moment de $X + Y$, i.e., $\mathbf{E}\{(X + Y)^4\}$ en l'exprimant en fonction des moments de X et Y. `Sol:` $\mathbf{E}\{X^4\} + 6\mathbf{E}\{X^2\}\mathbf{E}\{Y^2\} + \mathbf{E}\{Y^4\}$.

7. On construit un graphe biparti aléatoire que l'ensemble de sommets $\{1, \ldots, 2n\}$ comme suit. On ajoute des arêtes de la forme $\{i, j\}$ une par une indépendamment en choisissant aléatoirement $i \in \{1, \ldots, n\}$ et $j \in \{n+1, \ldots, 2n\}$ uniformément. Soit X le nombre d'arêtes ajoutées lorsque l'on a pour la première fois $\deg(u) \geq 1$ pour tout sommet $u \in \{1, \ldots, n\}$. Calculer $\mathbf{E}\{X\}$.

8. Soit $\Omega = \{1, 2, 3, 4\}$. Décrire explicitement la σ-algèbre (ou tribu) générée par $\{\{1\}, \{2\}\}$.

Problem 2 (Algorithmes de Las Vegas et Monte Carlo)**.** Soit $f : \{0, 1\}^* \to \{0, 1\}^*$ une fonction que l'on cherche à calculer.

1. Supposons que l'on a un algorithme de Las Vegas $\mathcal{A}$ pour f, i.e., l'algorithme $\mathcal{A}$ sur l'entrée x renvoie toujours $f(x)$ mais le temps d'exécution est aléatoire. On sait seulement que le temps d'exécution est majoré par $T(n)$, où $n = |x|$. Construire un algorithme de temps d'exécution majoré $O(T(n))$ mais qui peut renvoyer une réponse fausse avec probabilité au plus $\frac{1}{10}$.

2. Supposons maintenant que l'on a un autre algorithme $\mathcal{A}$ qui, sur l'entrée x, renvoie $f(x)$ avec probabilité au moins $p > 0$, et dont le temps d'exécution est majoré par $T_1(n)$ où $n = |x|$. Supposons en plus qu'étant donné x et $\mathcal{A}(x)$, il est possible de vérifier de manière déterministe $\mathcal{A}(x) = f(x)$ en temps $T_2(n)$. Décrire un algorithme de Las Vegas pour calculer f avec un temps d'exécution d'espérance $O(\frac{T_1(n)+T_2(n)}{p})$.

Problem 3. Soit $\vec{X}$ un vecteur aléatoire dans $\mathbb{R}^d$, i.e., $\vec{X} = (X_1, \ldots, X_d)$ où chaque X_i est une variable aléatoire. On définit naturellement $\mathbf{E}\left\{\vec{X}\right\} = (\mathbf{E}\{X_1\}, \mathbf{E}\{X_2\}, \ldots, \mathbf{E}\{X_d\}) \in \mathbb{R}^d$. On écrit $\vec{\mu} = \mathbf{E}\left\{\vec{X}\right\} \in \mathbb{R}^d$.

Pour ce problème on suppose que $\vec{X} \in [0,1]^d$. Notre objectif est d'estimer $\vec{\mu}$ avec une précision fixe ϵ. Pour ceci on a accès à des échantillons indépendants $\vec{X}^1, \ldots, \vec{X}^n$ qui ont tous la même distribution $\vec{X}$.

1. Proposer un estimateur $\vec{Y}$ pour $\vec{\mu}$ comme une fonction de $\vec{X}^1, \ldots, \vec{X}^n$. Cet estimateur doit vérifier $\mathbf{E}\left\{\vec{Y}\right\} = \vec{\mu}$.

2. Trouver une borne supérieure de $\mathbf{P}\left\{\|\vec{Y} - \vec{\mu}\|_\infty \geq \epsilon\right\}$ exprimée en fonction de ϵ et n. Pour un vecteur $\vec{v} = (v_1, \ldots, v_d)$, on écrit $\|\vec{v}\|_\infty = \max_{i \in \{1,\ldots,d\}} |v_i|$.

Problem 4 (Ensemble dominants)**.** Soit $G = (V, E)$ un graphe non dirigé avec $|V| = n$. Un ensemble dominant de G est un sous-ensemble $S \subseteq V$ tel que, pour tout sommet $v \in V$, v est connecté à un élément de S, i.e., il existe $s \in S$ tel que $\{s, v\} \in E$.

L'objectif de ce problème est d'étudier la taille du plus petit ensemble dominant d'un graphe de degré minimum $\geq \mu n$, où $\mu \in [0,1]$ est une constante.

1. Calculer la taille du plus petit ensemble dominant du graphe complet K_n sur n sommets et du graphe complet biparti $K_{n,n}$ sur deux ensembles de n sommets chacun. `Sol: 1 and 2`

2. Intuitivement, on s'attend à ce que, si le degré minimum d'un graphe est elevé, alors il existe un ensemble dominant de petite taille. Prouver que cette intuition est fausse: pour tout $\mu \in]0,1[$ et k, il existe un entier n et un graphe G sur n sommets de degré minimum au moins μn sans ensemble dominant de taille $\leq k$.

   Aide: on pourra considérer un graphe aléatoire $\mathcal{G}_{n,p}$ pour un p bien choisi. Sol: $\mathbf{P}\{S\text{dominating}\} = (1-(1-p)^k)^{n-k}$. So $\mathbf{P}\{\exists S, \text{dominating}\} \leq \binom{n}{k}(1-(1-p)^k)^{n-k}$ which goes to 0 as $n \to \infty$. Then pick $p \in]\mu, 1[$ will do the job, with Chernoff.

3. On définit un *ensemble dominant de distance deux* un sous-ensemble $S \subseteq V$ tel que, pour tout $v \in V$, il existe un sommet $s \in S$ tel que v est à distance au plus 2 de s. On montre que, pour tout $\mu \in]0,1[$, il existe un k dépendant seulement de μ tel que pour tout entier n et tout graphe G sur n sommets de degré minimum μn, G a un ensemble dominant de distance deux de taille $\leq k$.

   Pour prouver ce résultat, on fixe $\mu \in]0,1[$ et on considère un entier n ainsi qu'un graphe $G = (V, E)$ sur n sommets, de degré minimum $\geq \mu n$.

   En choisissant une valeur de k appropriée (dépendant de μ mais non de n), montrer que G admet un ensemble dominant de distance deux de taille $\leq k$.

   Aide: vous pouvez choisir aléatoirement un ensemble S de taille au plus k et considérer la variable aléatoire correspondant au nombre de sommets de V connectés à (au moins) un sommet de S.

   Sol: Take k so that $1-(1-\mu)^k \geq 1-\mu$ and then take an S such that $N > (1-\mu)n$. But then for any $v \in V$ it has to have at least one neighbour in this set of size N.