# Homework 2

1. (2 + 2 + (2 + 2 + 3 + 4) points) Consider a joint distribution $(\mathbb{X}_1, \dots, \mathbb{X}_n)$. Each marginal distribution $\mathbb{X}_i$ is over the same sample space $\Omega$. Let $\mathbb{U}_\Omega$ be a uniform distribution over $\Omega$.

   **Definition 1** ($t$-wise Independence). *The joint distribution $(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is a $t$-wise independent if for all distinct indices $i_1 < i_2 < \dots < i_t$, where $i_1, i_2 \dots, i_t \in \{1, \dots, n\}$, we have the following distribution*

   $$\left(\mathbb{X}_{i_1}, \mathbb{X}_{i_2}, \dots, \mathbb{X}_{i_t}\right)$$

   *is identical to $\left(\mathbb{U}_\Omega^{(1)}, \dots, \mathbb{U}_\Omega^{(t)}\right)$. (Recall that each $\mathbb{U}_\Omega^{(i)}$ represents a independent uniform distribution over $\Omega$)*

   (a) Let $\Omega = \mathbb{F} = \{0, 1, \dots, n-1\}$ be a field. A sample according to the joint output distribution $(\mathbb{X}_0, \dots, \mathbb{X}_{n-1})$ is defined as follows.

   - Sample $a$ uniformly at random from $\mathbb{F}$, and sample $b$ uniformly (and independently) at random from $\mathbb{F}$.
   - Let $f_{a,b}(x) := a \cdot x + b$, where the $\cdot$ and $+$ are respectively the product and addition operator corresponding to the field $\mathbb{F}$.
   - Generate the sample $(\omega_0, \omega_1 \dots, \omega_{n-1}) = \big(f(0), f(1), \dots, f(n-1)\big)$.

   Prove that this joint distribution is 2-wise independent (i.e., pair-wise independent).

   (b) Prove that the above distribution is *not* 3-wise independent.

   (c) Consider a balls-and-bins problem where $n$ balls are thrown into $n$ bins by first drawing a sample $(\omega_1, \dots, \omega_n)$ from a $t$-wise independent distribution $(\mathbb{X}_1, \dots, \mathbb{X}_n)$. Then, for $i \in \{1, \dots, n\}$, the $i$-th ball is thrown into the $\omega_i$-th bin. We will obtain upper bounds on the max-load for this experiment.

   i. For every $t$ distinct indices $i_1 < \dots < i_t$ and $i_1, \dots, i_t \in \{1, \dots, n\}$, define a random variable $\mathbb{H}_{i_1,\dots,i_t}$ over the sample space $\{0, 1\}$. The random variable $\mathbb{H}_{i_1,\dots,i_t}$ is 1 if the balls $i_1, \dots, i_t$ all fall in the same bin. Intuition: Think of this as a random variable that indicates the collision of balls $i_1, \dots, i_t$.
   Compute $\mathbb{E}\left[\mathbb{H}_{i_1,\dots,i_t}\right]$.

   ii. Define

   $$\mathbb{H} = \sum_{\substack{i_1,\dots,i_t \in \{1,\dots,n\} \\ i_1 < \dots < i_t}} \mathbb{H}_{i_1,\dots,i_t}$$

   Compute $\mathbb{E}[\mathbb{H}]$.

   iii. Let $\mathbb{M}$ denote the random variable for the max-load of this experiment. Prove that,

   $$\mathbb{E}\left[\binom{\mathbb{M}}{t}\right] \leqslant \mathbb{E}[\mathbb{H}]$$

   iv. Show that $\mathbb{E}[\mathbb{M}]$ is (roughly) upper bounded by $n^{1/t}$.

2. (2 + 8 points) Consider the random variable $B(n,p)$. That is, we are counting the number of heads in $n$ independent coin tosses and each coin outputs heads with probability $p$. We will like to show the tightness of the Chernoff Bound.

   (a) For $\varepsilon > 0$, prove

   $$\mathbb{P}\left[B(n,p) \geqslant n\,(p+\varepsilon)\right] = \sum_{i \geqslant (p+\varepsilon)n} \binom{n}{i} p^i (1-p)^{n-i}$$

   (b) Recall that, the Chernoff bound proves that

   $$\mathbb{P}\left[B(n,p) \geqslant n\,(p+\varepsilon)\right] \leqslant 2^{-n\mathrm{D}_{\mathrm{KL}}(p+\varepsilon,p)}$$

   Use Stirling's approximation to show that

   $$\mathbb{P}\left[B(n,p) \geqslant n\,(p+\varepsilon)\right] \geqslant \frac{1}{\mathrm{poly}(n)} 2^{-n\mathrm{D}_{\mathrm{KL}}(p+\varepsilon,p)}$$

   This will show that Chernoff bound is essentially tight. For this problem assume that $(p+\varepsilon)n$ is an integer.

3. (2 + 8 points) Let $\mathbb{X}$ be the random variable over the sample space $\{1,2,\dots\}$ such that $\mathbb{P}[\mathbb{X}=i] = 2^{-i}$. Intuitively, $\mathbb{X}$ is identical to the random variable that outputs the number of trials to see the first heads in an infinite sequence of coin tosses, where each coin toss is equally likely to be heads or tails.

   (a) Compute $\mathbb{E}[\mathbb{X}]$.

   (b) Let $(\mathbb{X}^{(1)},\dots,\mathbb{X}^{(n)})$ be $n$ independent samples from the distribution $\mathbb{X}$. Let $\mathbb{S} = \sum_{i=1}^n \mathbb{X}^{(i)}$. Intuitively, $\mathbb{S}$ is the number of coin tosses needed to see $n$ heads in an infinite sequence of coin tosses, where each coin toss is equally likely to be heads or tails.

   Follow a strategy similar to the proof of the Chernoff bound to prove a exponentially decaying upper bound on

   $$\mathbb{P}\left[\mathbb{S} \geqslant \mathbb{E}[\mathbb{S}]\,(1+\varepsilon)\right]$$

4. (10 points) Let $\mathbb{X}$ be the random variable over the sample space $\{0,1,2,\dots\}$ such that $\mathbb{P}[\mathbb{X}=i] = \exp(-\mu)\frac{\mu^i}{i!}$. This is the Poisson distribution with mean $\mu$. Recall that when $m$ balls are thrown into $n$ bins, the probability of a bin getting $i$ balls is *roughly* equal to $\mathbb{P}[\mathbb{X}=i]$.

   Now consider the joint distribution $(\mathbb{X}^{(1)},\dots,\mathbb{X}^{(n)})$. Let $\mathbb{S} = \sum_{i=1}^n \mathbb{X}^{(i)}$. Follow a strategy similar to the proof of the Chernoff bound to prove an upper bound on

   $$\mathbb{P}\left[\mathbb{S} \geqslant \mathbb{E}[\mathbb{S}]\,(1+\varepsilon)\right]$$

5. (2 + 2 + 2 + 9 points) In this problem we will derive the Chernoff-like Concentration bound for the Hypergeometric series.

   Suppose there are $N$ balls in a jar and $pN$ of them are red and rest are black. Our sample is a uniformly random set of $n$ balls from this jar.

   (a) Consider the joint distribution $(\mathbb{X}_1,\dots,\mathbb{X}_n)$ such that $\mathbb{X}_i$ represents the indicator variable for the $i$-th ball being red. What is $\mathbb{E}[X_i]$?

(b) Let $\mathbb{S} = \sum_{i=1}^{n} \mathbb{X}_i$. What is $\mathbb{E}[\mathbb{S}]$?

(c) Prove that the probability that this set of $n$ balls has exactly $k$ red balls is:

$$\frac{\binom{pN}{k}\binom{(1-p)N}{n-k}}{\binom{N}{n}}$$

(d) Define an appropriate filtration and an appropriate martingale sequence with respect to it. Apply Azuma-Hoeffding bound to prove this concentration bound. (Hint: In class we nearly solved this problem!)

6. (Extra Credit) Consider the ball-and-bins experiment where $n$ balls are thrown into $n$ bins uniformly and independently at random. Obtain a concentration bound for the max-load around its mean in this experiment using the Poisson Approximation Theorem.