1 Overview

In previous lectures, we introduced some basic probability, the Chernoff bound, the coupon collector problem, and game tree evaluation.

In this lecture, we will introduce concentration inequalities.

2 Coupon Collector Problem

Draw numbers (coupons) independently from $[n] = \{1, 2, \ldots, n\}$. How long does it take to see all of the numbers?

Suppose $T_i$ is the number of draws to get the $i$-th new number. Let $T = \sum_i T_i$.

**Fact 1.** The $T_i$’s are independent of each other.

**Fact 2.** $T_i$ follows geometric distribution with success probability, $p = \frac{n+1-i}{n}$.

**Fact 3.** If $X \sim \text{Geometric}(p)$,

\[
E[X] = p \cdot 1 + (1 - p) \cdot (E[X|X \geq 2])
= p + (1 - p)(1 + E[X])
\]

\[
\Rightarrow E[X] = \frac{1}{p}
\]

\[
E[X^2] = p \cdot 1^2 + (1 - p) \cdot E[X^2|X \geq 2]
= p + (1 - p)E[(X + 1)^2]
= p + (1 - p)(E[X^2] + 2E[X] + 1)
= p + (1 - p)E[X^2] + 2(1 - p)/p + (1 - p)
\]

\[
\Rightarrow E[X^2] = \frac{2 - p}{p^2}
\]

\[
Var(X) = E[(X - E(X))^2] = E[X^2] - (E[X])^2
\]

\[
\Rightarrow Var(X) = \frac{1 - p}{p^2} \leq \frac{1}{p^2}
\]
Therefore in the Coupon Collector Problem,

\[ E[T] = \sum_{i=1}^{n} E[T_i] = \sum_{i=1}^{n} \frac{n}{n+1-i} = nH_n \approx n \log n \]

\[ \text{Var}[T_i] \leq \frac{1}{p_i^2} = \left( \frac{n}{n+1-i} \right)^2 \]

\[ \Rightarrow \text{Var}[T] = \sum_{i=1}^{n} \text{Var}[T_i] \leq n^2 \left( \sum_{i=1}^{n} \frac{1}{i^2} \right) \leq n^2 \cdot \frac{\pi^2}{6} = O(n^2) \]

3 Concentration Inequalities

\[ \forall i, \ Pr[T_i \geq 1 + \alpha_i] \leq \left( 1 - \frac{n+1-i}{n} \right)^\alpha \]

Assume \( \delta \) is some failure probability. Setting \( \alpha_i = \left( \frac{n}{n+1-i} \right) \log \frac{n}{\delta} \) and because \((1 - \frac{1}{x})^x < \frac{1}{e}\), we have

\[ \forall i, \ Pr[T_i \geq 1 + \alpha_i] \leq \frac{\delta}{n} \]

Definition 4. Union Bound

\[ Pr[X_1 \cup X_2 \cup \ldots \cup X_n] \leq \sum_i \Pr[X_i] \]

Using a union bound, we have

\[ Pr \left[ T \geq n + n \log n \log \frac{n}{\delta} \right] \]

\[ = Pr \left[ \sum_i T_i \geq n + \sum_i \alpha_i \right] \]

\[ = Pr[T_1 \geq 1 + \alpha_1 \cup \ldots \cup T_n \geq 1 + \alpha_n] \]

\[ \leq \sum_i Pr[T_i \geq 1 + \alpha_i] \]

\[ \leq \delta \]

Definition 5. With High Probability (w.h.p.)

\[ X \leq O(y) \text{ w.h.p. } \Leftrightarrow \forall c_2, \exists c_1, \text{ s.t. } Pr[X \leq c_1 y] \leq n^{-c_2} \]

\( T = O(n \log^2 n) \) with high probability.
3.1 Markov’s Inequality

For a non-negative random variable $T$ and any non-negative $\alpha$,

$$E[T] \geq Pr[T \geq \alpha] \cdot \alpha$$

$$\Rightarrow Pr[T \geq \alpha] \leq \frac{E[T]}{\alpha}$$

In the Coupon Collector Problem,

$$\alpha = \frac{E[T]}{\delta} = \frac{nH_n}{\delta}$$

$$\Rightarrow Pr\left[T \geq \frac{nH_n}{\delta}\right] \leq \delta$$

3.2 Chebyshev’s Inequality

For a random variable, $X$, let $\mu = E[X]$ denote the expectation and $\sigma^2 = Var[X]$ denote the variance. Starting from Markov’s Inequality, we find

$$Pr[(X - \mu)^2 \geq \alpha^2] \leq \frac{E[(X - \mu)^2]}{\alpha^2} = \frac{\sigma^2}{\alpha^2}$$

Setting $\alpha \rightarrow \alpha \sigma$

$$Pr[(X - \mu)^2 \geq \alpha^2 \sigma^2] \leq \frac{1}{\alpha^2}$$

Taking the square root, we find

$$Pr[X \geq \mu + \alpha \sigma] \leq \frac{1}{\alpha^2}$$

$$Pr[X \leq \mu - \alpha \sigma] \leq \frac{1}{\alpha^2}$$

Using this result in the Coupon Collector Problem, gives us

$$Pr[T \geq nH_n + \frac{1}{\sqrt{\delta}} O(n)] \leq \delta$$

Setting $\delta = \frac{1}{\log^2 n}$

$$Pr[T \geq nH_n + O(n \log n)] \leq O\left(\frac{1}{\log^2 n}\right)$$

Most of the time, the typical deviation is $O(\sigma)$.

$$Pr[|x - \mu| \leq O(\sigma)] \approx 1 - \delta$$
3.3 Moment Method

If $f$ is non-negative, by Markov’s inequality,

$$Pr[f(X - \mu) \geq f(\alpha)] \leq \frac{E[f(X - \mu)]}{f(\alpha)}$$

For $f$ increasing,

$$Pr[X - \mu \geq \alpha] \leq \frac{E[f(X - \mu)]}{f(\alpha)}$$

Set $f = |t|^k$,

$$Pr[|X - \mu|^k \geq |\alpha|^k] \leq \frac{E[|x - \mu|^k]}{|\alpha|^k}$$

For one side,

$$Pr[X \geq \mu + \alpha] \leq \frac{E[|x - \mu|^k]}{|\alpha|^k}$$

Setting $\delta = \frac{E[|x - \mu|^k]}{|\alpha|^k}$, we have

$$Pr \left[ X \leq \mu + E[|x - \mu|^k]^{1/k} \cdot \left( \frac{1}{\delta} \right)^{1/k} \right] \geq 1 - \delta$$

If we consider $X \sim N(0, \sigma^2)$, we know

$$E[|x|^k] \approx (k\sigma^2)^{k/2} \forall k > 0$$

which means

$$Pr \left[ X \geq \mu + O \left( \sqrt{k} \cdot \sigma \cdot \left( \frac{1}{\delta} \right)^{1/k} \right) \right] \leq \delta$$

Setting $k = \log \frac{1}{\delta}$, we get

$$Pr \left[ X \geq \mu + O \left( \sqrt{\log \frac{1}{\delta}} \right) \right] \leq \delta$$

3.4 Moment Generating Function

**Definition 6.** The moment generating function, parameterized by $\lambda$, is defined as

$$MGF_X(\lambda) = E[e^{\lambda(X-\mu)}]$$

Assume $X$ is centered ($E[X] = 0$).

$$e^{\lambda x} = 1 + \lambda x + \frac{(\lambda x)^2}{2} + \frac{(\lambda x)^3}{3!} + \ldots + \frac{(\lambda x)^k}{k!}$$

We can use parameter $\lambda$ to adjust the weights on each term. When $\lambda$ is larger, more weight is on higher order terms.
From the derivation of the Moment Method, setting $f(x) = e^{\lambda x}$,

$$Pr[X \geq \mu + \alpha] \leq \frac{MGF_X(\lambda)}{e^{\lambda \alpha}}$$

**Fact 7.** If $X \sim N(0, \sigma^2)$, $MGF_X(\lambda) = E[e^{\lambda X}] \leq e^{\lambda^2 \sigma^2}$, $\forall \lambda \in \mathbb{R}$

Using this,

$$Pr[X \geq \mu + \alpha] \leq \frac{MGF_X(\lambda)}{e^{\lambda \alpha}} = e^{\lambda^2 \sigma^2 - \lambda \alpha} = e^{\frac{1}{2} (\lambda^2 - \frac{\alpha}{\sigma^2})^2 - \frac{\alpha^2}{2\sigma^2}}$$

Set $\lambda = \frac{\alpha}{\sigma^2}$, we get

$$Pr[X \geq \mu + \alpha] \leq e^{-\frac{\alpha^2}{2\sigma^2}}$$

If $\delta = e^{-\frac{\alpha^2}{2\sigma^2}}$, we have

$$\alpha = \sigma \sqrt{2\log \frac{1}{\delta}}$$

Note that this is the same $O\left(\sqrt{\log \frac{1}{\delta}}\right)$ bound as we found in the method of moments, except that now we know the constant.

### 3.5 Subgaussian Variables

**Claim 8.** The following three statements are equivalent if we only care up to a constant for $\sigma$ (i.e. $\forall i, j \in \{1, 2, 3\}, \sigma_i = \theta(\sigma_j)$)

1. $X$ is subgaussian with parameter $\sigma$, i.e. $\forall \lambda \in \mathbb{R}$, $MGF_X(\lambda) \leq e^{\frac{\lambda^2 \sigma^2}{2}}$ (1)
2. $Pr[X \geq \mu + t] \leq e^{-\frac{t^2}{2\sigma^2}}$ (2)
3. $E[|x|^k]^{1/k} \leq O\left(\sigma_3 \sqrt{k}\right)$ (3)

**Fact 9.** The sum of subgaussian variables are subgaussian.

$$X = X_1 + \ldots + X_n$$

$$MGF_X(\lambda) = E\left[e^{\lambda X}\right] = E\left[e^{\lambda(\sum_i x_i)}\right] = E\left[\prod_i e^{\lambda x_i}\right]$$

$$= \prod_i E\left[e^{\lambda x_i}\right] \quad \text{(by independence)}$$

$$= \prod_i MGF_{X_i}(\lambda)$$

$$\leq \prod_i e^{\lambda^2 \sigma^2_i / 2} \quad \text{(by subgaussian)}$$

$$= e^{\frac{1}{2} (\sum_i \sigma^2_i)}$$

5
This implies $X$ is subgaussian with parameter $\sqrt{\sum_i \sigma_i^2}$.

**Fact 10.** If $X \in [0, 1]$, then $X$ is subgaussian with $\sigma = 1/2$ by Hoeffding’s Lemma.¹

Let $X = \sum_i X_i$ where $X_i \in [0, 1]$. $X$ is subgaussian with $\sigma = \sqrt{n}/2$. Plug this into (2), we have

$$Pr[x \geq \mu + \alpha] \leq e^{-\frac{2\alpha^2}{n}}$$

which is exactly the Chernoff bound.

4 Next Class

In the Coupon Collector Problem, we had

$$Pr[T_n \geq \alpha] \leq \left(1 - \frac{1}{n}\right)^\alpha \approx e^{-\alpha}$$

This is not of the form $e^{-\alpha^2}$ so we cannot use the subgaussian results. We will relax the subgaussian requirement to subexponential and subgamma. This will lead to Bernstein’s inequality.

$$MGF_X(\lambda) = E[e^{\lambda X}] \leq e^{\frac{\lambda^2 \sigma^2}{2}} \forall |\lambda| \leq B$$

**References**


¹https://en.wikipedia.org/wiki/Hoeffding%27s_inequality