B490: Probability Basics

Instructor: Qin Zhang

1 Expectation and Variance

If random variable $X$ takes discrete values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$, then $E[X] = \sum_{i=1}^{\infty} x_i p_i$, and $\text{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$.

**Linearity of Expectation.** $E[X + Y] = E[X] + E[Y]$. Holds even $X$ and $Y$ are *not* independent.

For variance, $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ holds only when $X, Y$ are independent.

The following equality is often useful. If $X$ takes values in $\{0, 1, 2, \ldots\}$, then

$$E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i].$$

2 Markov

$$E[X] \geq \Pr[X \geq a] \cdot a + \Pr[X \leq a] \cdot 0.$$

3 Chebyshev

$$\Pr[|X - E[X]| \geq a] = \Pr[(X - E[X])^2 \geq a^2] \leq \frac{E[(X - E[X])^2]}{a^2} = \frac{\text{Var}[X]}{a^2}$$

4 Coupon Collector

4.1 Using Markov

$X_i$ : # boxes bought when you have exactly $i - 1$ different coupons, and after buying these boxes, you have $i$ different coupons.

Let $X = \sum_{i \in \mathbb{N}} X_i$ : # boxes one bought until at least one of every type of coupon is obtained.

Let $p_i$ be the probability of obtaining a new coupon when we have exactly $i - 1$ different coupons. $p_i = 1 - \frac{i-1}{n}$. Then $E[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$. Therefore

$$E[X] = \sum_{i \in [n]} E[X_i] = n \left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right) = nH_n \approx n \ln n$$
By Markov, \( \Pr[X \geq 2n \ln n] \leq 1/2. \)

### 4.2 Using Chebyshev.

First, note that \( X_i \)'s are independent, hence \( \text{Var}[X] = \sum_{i \in [n]} \text{Var}[X_i] \).

\( X_i \) is a geometric random variable with parameters \( p = \frac{n - i + 1}{n} \), then

\[
E[X_i] = \sum_{j=1}^{\infty} \Pr[X_i \geq j] = \sum_{j=1}^{\infty} (1 - p)^{j-1} = \frac{1}{p},
\]

and

\[
E[X_i^2] = \sum_{j=1}^{\infty} j \cdot \Pr[X_i = j] = \sum_{j=1}^{\infty} p(1 - p)^{j-1}j^2 = \frac{2 - p}{p^2}.
\]

Thus

\[
\text{Var}[X_i] = E[X_i^2] - (E[X_i])^2 = \frac{2 - p}{p^2} - \frac{1}{p^2} < \frac{1}{p^2}.
\]

Therefore,

\[
\text{Var}[X] = \sum_{i \in [n]} \text{Var}[X_i] \leq \sum_{i \in [n]} \left( \frac{n}{n - i + 1} \right)^2 = n^2 \sum_{i \in [n]} \frac{1}{i^2} \leq n^2 \cdot \frac{\pi^2}{6}.
\]

And

\[
\Pr[|X - n \ln n| \geq n \ln n] \leq \frac{n^2 \pi^2/6}{(n \ln n)^2} = \Theta \left( \frac{1}{\ln^2 n} \right).
\]

**Using the Union Bound.** The probability that the \( i \)-th coupon is NOT obtained after \( n \ln n + cn \) steps is

\[
\left( 1 - \frac{1}{n} \right)^{n(\ln n + c)} < e^{-(\ln n + c)} = \frac{1}{e^c \cdot n}.
\]

Setting \( c = \log n \), the probability is at most \( 1/n \). Then use the union bound.