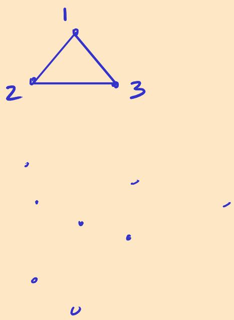


$$\frac{m}{\binom{n}{2}} \cdot \frac{m-1}{\binom{n}{2}-1} \cdot \frac{m-1}{\binom{n}{2}-2} \cdot \binom{n}{3}$$

$$\binom{n}{2} = K$$

$$\binom{n}{3}\frac{\binom{K-3}{m-3}}{\binom{K}{m}}$$

$$\frac{(K-3)!}{(m-3)!\,\cancel{(K-m)!}} \cdot \frac{m!\,\cancel{(K-m)!}}{K!} = \frac{m(m-1)(m-2)}{K(K-1)(K-2)}$$

11

Random Variables II, Variance, LLN, Concentration Inequalities

Computer Science Mentors 70

Nov 16 - Nov 20
Fall 2020

Prepared by: Aishani Sil, Austin Lei, Agnibho Roy, Debayan Bandyopadhyay, Abinav Routhu

I Random Variables II

1. **E[*levator*]**
   A building has n floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each gets off at a uniformly random one of the n floors (independently of everybody else).

   (a) What is the expected value of the last floor someone gets off at? For example, if people leave at floors 1, 2, and 5, Floor 5 is the last floor.

   L the last floor that someone gets off at. $\mathbb{E}[L] = \sum_{i=1}^{n} i \cdot \Pr[L=i]$

   $\Pr[L=i] = \Pr[$everyone gets off at floor $\le i \cap$ someone gets off at floor $i] = \Pr[L\le i] - \Pr[L \le i-1] = \left(\frac{i}{n}\right)^m - \left(\frac{i-1}{n}\right)^m$

   $\sum_{i=1}^{n} i \cdot \Pr[L=i] = \sum_{i=1}^{n} i \cdot \left[\left(\frac{i}{n}\right)^m - \left(\frac{i-1}{n}\right)^m\right] = n - \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^m = \mathbb{E}[L]$

   $\underbrace{\_\ \_\ \_\ \_\ \_\ \_\ \_\ \_\ \_\ \_}_{i}$\ _\ _\ _\ _\ _

   (b) What is the expected number of times the elevator stops (not including the ground floor)? $=F$

   $F_i = \mathbb{1}\{$elevator stopped at floor $i\} = \begin{cases} 1 & \text{if elevator stopped at floor } i \\ 0 & \text{o/w} \end{cases}$

   $\mathbb{E}[F_i] = 1 \cdot \Pr[F_i = 1] + 0 \cdot \Pr[F_i = 0] = \Pr[F_i = 1] = 1 - \Pr[$no one gets off at floor $i] = 1 - \left(\frac{n-1}{n}\right)^m$

   $F = \sum_{i=1}^{n} F_i$, $\mathbb{E}[F] = \mathbb{E}\left[\sum_{i=1}^{n} F_i\right] = \sum_{i=1}^{n} \mathbb{E}[F_i] = n \cdot \left(1 - \left(\frac{n-1}{n}\right)^m\right)$

2. **Chaotic Santa**
   (Fall '17 Disc) A delivery guy is out delivering n packages to n customers, where $n \in \mathbb{N}$, $n > 1$. Not only does he hand a random package to each customer, he opens the package before delivering it with probability $\frac{1}{2}$.

   (a) What is the expected number of customers that get their own package unopened? X

   $X = \sum_{i=1}^{n} X_i$ $\quad X_i = \mathbb{1}\{$customer i gets package i unopened$\}$

   $\mathbb{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2n}$

   $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n \cdot \frac{1}{2n} = \frac{1}{2}$

   (b) What is the variance for the random variable above? $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X]$ $\quad (1/2)^2 = 1/4$ $\quad = \frac{3}{4} - \frac{1}{4} = \boxed{\frac{1}{2}}$

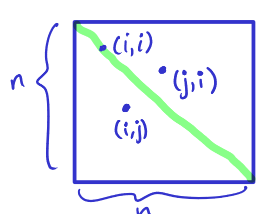
   $\mathbb{E}\left[(X_1 + X_2 + \cdots + X_n)^2\right] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2 \sum_{1 \le i < j \le n} \mathbb{E}[X_i X_j] = \frac{1}{2} + 2\binom{n}{2} \frac{1}{4} \cdot \frac{1}{n(n-1)} = \frac{1}{2} + \cancel{n(n-1)} \cdot \frac{1}{4} \cdot \frac{1}{\cancel{n(n-1)}} = \frac{3}{4}$

X_i	X_i^2	$\Pr$
0	0	$1 - 1/2n$
1	1	$1/2n$

   $\mathbb{E}[X_i^2] = \mathbb{E}[X_i] = \frac{1}{2n}$

$X_j \backslash X_i$	0	1
0	0	0
1	0	1

   $\mathbb{E}[X_i X_j] = 1 \cdot \Pr[X_i = 1 \cap X_j = 1] + 0 \cdot \ldots$
   $= \Pr[X_i = 1] \Pr[X_j = 1 \mid X_i = 1]$
   $= \frac{1}{2n} \cdot \frac{1}{n-1} \cdot \frac{1}{2} = \frac{1}{4} \cdot \frac{1}{n(n-1)}$

   $X_i X_j = X_j X_i$

   $n^2 - n = n(n-1)$ $\quad \frac{n^2 - n}{2} = \binom{n}{2}$

   $(X_1 + X_2 + \cdots + X_n)^2 = (X_1 + X_2 + \cdots + X_n)(X_1 + X_2 + \cdots + X_n)$
   $= \sum_{i,j} X_i X_j = \sum_{i=1}^{n} X_i^2 + \sum_{i \ne j} X_i X_j = \sum_{i=1}^{n} X_i^2 + 2 \sum_{1 \le i < j \le n} X_i X_j$

II Variance

3. **Introduction to Variance**
   Let us say that we are dealing with a biased die and we want to know how often my roll varies from turn to turn. Consider the

following 6-length tuple assigning probabilities to each of the 6 rolls on the die, whose distribution we name X:

$$(p_1, p_2, p_3, p_4, p_5, p_6)$$

(a) Find the mean and variance for the die with distribution $(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$

(b) Find the mean and variance for the die with distribution $(\frac{1}{3}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{3})$

(c) Give an intuitive explanation for the difference between the results in part(a) and part (b)

4. **Money Balls**

We are drawing dollar bills out of a bag consisting of R red balls and B blue balls. We get 1 dollar for every blue ball, but whenever we draw a red ball, the game stops and we retrieve the amount of money we have made so far.

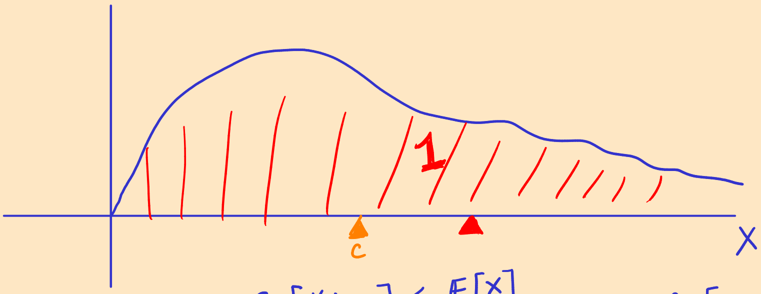
(a) Find the expected value of the amount of money made from this game.

(b) Find the variance of the amount of money made in this game.

Markov

$X \geq 0$

$\mathbb{E}[X] < \infty$



$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c}$$

linear

$$\Rightarrow c\Pr[X \geq c] \leq \mathbb{E}[X]$$

$$\begin{aligned}\mathbb{E}[X] &= \Pr[X<c]\cdot\mathbb{E}[X \mid X<c] + \Pr[X\geq c]\cdot\mathbb{E}[X \mid X\geq c] \\ &\geq \Pr[X\geq c]\cdot\mathbb{E}[X \mid X\geq c] \\ &\geq \Pr[X\geq c]\cdot\mathbb{E}[c] \\ &= \Pr[X\geq c]\cdot c\end{aligned}$$

$c\Pr[X=c] + (c+1)\Pr[X=c+1] + \cdots$

$$\Pr[X=x \mid X\geq c] = \frac{\Pr[X=x \cap X\geq c]}{\Pr[X\geq c]}$$

Chebyshev

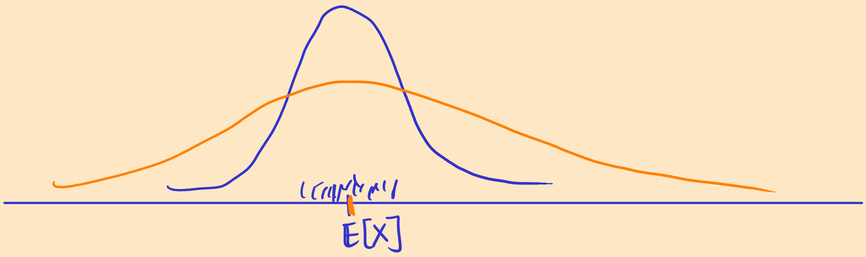
X

$\mathbb{E}[X] < \infty$

$\mathrm{Var}(X) < \infty$

$$\Pr[|X-\mathbb{E}[X]| \geq c] \leq \frac{\mathrm{Var}(X)}{c^2}$$

quadratic



$$\Pr[(X-\mathbb{E}[X])^2 \geq c^2] \leq \frac{\mathbb{E}[(X-\mathbb{E}[X])^2]}{c^2} = \frac{\mathrm{Var}(X)}{c^2}$$

Intuition: RV usually take on vals close to $\mathbb{E}[X]$

Quantification: By how much?

III Concentration Inequalities

5. Introduction to Markov and Chebyshev Bounds

(a) Let X be a nonnegative random variable with $E[X] = 5$. Using Markov's Inequality, determine an upper bound for the probability that $X > 10$.

$$Pr[X>10] \le \frac{E[X]}{c} = \frac{5}{10} = \frac{1}{2}$$

$E[Y] = E[X+10]$
$= E[X]+10$

$X > -10,\ E[X] = 2$ $\quad Pr[X>-5] \le ?$

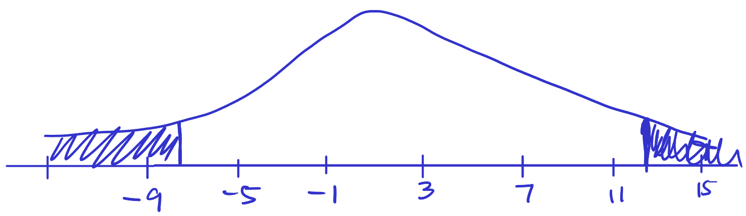
$Y = X+10,\ Y>0,\ E[Y] = 12 \leftarrow Pr[X>-5] = Pr(Y>-5+10] \le \frac{E(Y)}{5} = \frac{12}{5}$

(b) Let X be a random variable with $E[X] = 3$ and $\text{Var}(X) = 4$. Using Chebyshev's Inequality, determine an upper bound for the probability that $X > 13$. $\quad 13 = E[X]+10$

$$Pr[|X-E[X]| \ge c] \le \frac{Var(X)}{c^2}$$

$Pr[X>13] = Pr[X>3+10]$
$= Pr[X-3>10]$
$\le Pr[|X-3|>10] \le \frac{4}{10^2} = \frac{1}{25}$

$|X-E[X]|>c$



6. Strengthening Inequalities

Imagine that we have a biased coin with a probability of getting heads as 0.3. This coin is flipped 10 times.

(a) Find a bound for the probability that the coin lands on heads at least 8 times with Markov's inequality. How good is this bound compared to the true probability? $\quad X \sim Binom(10, 0.3)$

$E[X] = np = 3$

$Pr[X \ge 8] \le \frac{3}{8} = 0.375$

$Pr[X \ge 8] = \binom{10}{8} 0.3^8\, 0.7^2 + \cdots = 0.00159$

(b) Now, find the same bound (coin lands on heads at least 8 times) using Chebyshev's inequality.

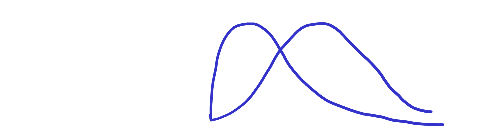
$Var(X) = np(1-p) = 10 \cdot 0.3 \cdot 0.7 = 2.1$

$Pr[X \ge 8] = Pr[X-3 \ge 5] \le Pr[|X-3| \ge 5] \le \frac{2.1}{5^2} = 0.084$

(c) Is this bound better? Can we divide this bound by 2 to get the bound on the tail we are interested in?

Yes if X is from a symm distr.

Binom(n, p) NOT symm $p \ne 0.5$



$\boxed{\text{NO}}$

$X=0$ vs. $X=10$

pip install scipy

from scipy import stats

IV LLN

7. **Introduction to LLN**

Leanne has a weighted coin that shows up heads with probability $\frac{4}{5}$ and tails with probability $\frac{1}{5}$. Leanne flips the coin 100 times, and computes $\overline{X}$, the average number of coins that show up heads.

(a) What is $E[\overline{X}]$?

$\frac{4}{5} \cdot 100 = 80$

Weak LLN

$X_1, \dots, X_n \overset{iid}{\sim} \mathcal{D}$

$E[X_i] < \infty$

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\lim_{n\to\infty} \Pr[|\overline{X} - E[X_i]| > \varepsilon] \to 0 \quad \forall \varepsilon > 0$

(b) What is Var(X)?

$\text{Var}(X) = 100 \cdot \frac{4}{5} \cdot \frac{1}{5} = 16$

(c) Suppose Leanne flips n coins instead of 100. What does the LLN tell us about $\overline{X}$?

$n \to \infty, \quad \overline{X}\,?$

$\overline{X} \xrightarrow{p} 80$

$\cancel{\overline{X} \to 80}$

$\lim_{n\to\infty} \Pr[|\overline{X} - 80| > \varepsilon] \to 0$

$\varepsilon > 0$

$n \to \infty$

$\text{Var}(\overline{X}) \to 0$

Central Limit Thm

$(\overline{X} - E[X_i]) \xrightarrow{d} \mathcal{N}(0, 1/n)$

$\Pr[|\overline{X} - 80| > \varepsilon] \le \frac{0}{\varepsilon^2} = 0$