Definition. Let $X$ and $Y$ be discrete random variables. The Joint PMF of $X$ and $Y$ is

$$p_{X,Y}(a, b) = \Pr(X = a, Y = b)$$

Definition. The joint range of $p_{X,Y}$ is

$$\Omega(X, Y) = \{(c, d) : p_{X,Y}(c, d) > 0\} \subseteq \Omega(X) \times \Omega(Y)$$

Note that

$$\sum_{(s, t) \in \Omega(X, Y)} p_{X,Y}(s, t) = 1$$
Law of Total Expectation

**Law of Total Expectation (event version).** Let $X$ be a random variable and let events $A_1, \ldots, A_n$ partition the sample space. Then,

$$E[X] = \sum_{i=1}^{n} E[X|A_i] \Pr(A_i)$$

**Law of Total Expectation (random variable version).** Let $X$ be a random variable and $Y$ be a discrete random variable. Then,

$$E[X] = \sum_{y \in \Omega(Y)} E[X|Y = y] \Pr(Y = y)$$
Example: Computer Failures

Suppose your computer operates in a sequence of steps, and that at each step $i$ your computer will fail with probability $p$ (independently of other steps). Let $X$ be the number of steps it takes your computer to fail. What is $E[X]$?
Agenda

• Markov’s Inequality
• Chebyshev’s Inequality
Tail Bounds (Idea)

Bounding the probability a random variable is far from its mean. Usually statements of the form:

\[
\Pr(X \geq a) \leq b \\
\Pr(|X - E[X]| \geq a) \leq b
\]

Useful tool when
• An approximation that is easy to compute is sufficient
• The process is too complex to analyze exactly
Markov’s Inequality

**Theorem.** Let $X$ be a random variable taking only non-negative values. Then, for any $t > 0$, 

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$ 

Incredibly simplistic – only requires that the random variable is non-negative and only needs you to know expectation. You don’t need to know anything else about the distribution of $X$. 


Markov’s Inequality – Proof

\[ \mathbb{E}(X) = \sum_x x \cdot \mathbb{P}(X = x) \]

\[ = \sum_{x \geq t} x \cdot \mathbb{P}(X = x) + \sum_{x < t} x \cdot \mathbb{P}(X = x) \]

\[ \geq \sum_{x \geq t} x \cdot \mathbb{P}(X = x) \]

\[ \geq \sum_{x \geq t} t \cdot \mathbb{P}(X = x) = t \cdot \mathbb{P}(X \geq t) \]

\[ \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}. \]

\[ \geq 0 \text{ because } x \geq 0 \]

\[ \text{whenever } \mathbb{P}(X = x) \geq 0 \text{ (takes only non-negative values)} \]

\[ \text{Follows by re-arranging terms} \]

…
Example – Geometric Random Variable

Let $X$ be geometric RV with parameter $p$

\[ P(X = i) = (1 - p)^{i-1}p \]
\[ \mathbb{E}(X) = \frac{1}{p} \]

“How many times does Alice need to flip a biased coin until she sees heads, if heads occurs with probability $p$?

What is the probability that $X \geq 2\mathbb{E}(X) = 2/p$?

Markov’s inequality: \[ P(X \geq 2/p) \leq \frac{\mathbb{E}(X)}{2/p} = \frac{1}{p} \cdot \frac{p}{2} = \frac{1}{2} \]

Can we do better?
Example

Suppose that the average number of ads you will see on a website is 25. Give an upper bound on the probability of seeing a website with 75 or more ads.

Poll: pollev.com/hunter312
a. $0 \leq p < 0.25$
b. $0.25 \leq p < 0.5$
c. $0.5 \leq p < 0.75$
d. $0.75 \leq p$
e. Unable to compute
Example

Suppose that the average number of ads you will see on a website is 25. Give an upper bound on the probability of seeing a website with 20 or more ads.

Poll: [pollev.com/hunter312](pollev.com/hunter312)

a. $0 \leq p < 0.25$

b. $0.25 \leq p < 0.5$

c. $0.5 \leq p < 0.75$

d. $0.75 \leq p$

e. Unable to compute
Brain Break
Agenda

• Markov’s Inequality
• Chebyshev’s Inequality
Chebyshev’s Inequality

**Theorem.** Let $X$ be a random variable. Then, for any $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2}.$$ 

**Proof:** Define $Z = X - \mathbb{E}(X)$

$$\mathbb{P}(|Z| \geq t) = \mathbb{P}(Z^2 \geq t^2) \leq \frac{\mathbb{E}(Z^2)}{t^2} = \frac{\text{Var}(X)}{t^2}.$$ 

$|Z| \geq t$ iff $Z^2 \geq t^2$

Definition of Variance

Markov’s inequality ($Z^2 \geq 0$)
Example – Geometric Random Variable

Let $X$ be geometric RV with parameter $p$

$$\mathbb{P}(X = i) = (1 - p)^{i-1}p \quad \mathbb{E}(X) = \frac{1}{p} \quad \text{Var}(X) = \frac{1 - p}{p^2}$$

What is the probability that $X \geq 2\mathbb{E}(X) = 2/p$?

**Markov:** $\mathbb{P}(X \geq 2/p) \leq \frac{\mathbb{E}(X)}{2/p} = \frac{1}{p} \cdot \frac{p}{2} = \frac{1}{2}$

**Chebyshev:** $\mathbb{P}(X \geq 2/p) \leq \mathbb{P}\left(\left|X - \frac{1}{p}\right| \geq \frac{1}{p}\right) \leq \frac{\text{Var}(X)}{1/p^2} = 1 - p$

Not better, unless $p > 1/2$
Example

Suppose that the average number of ads you will see on a website is 25 and the standard deviation of the number of ads is 5. Give an upper bound on the probability of seeing a website with 30 or more ads.

Poll: pollev.com/hunter312
a. $0 \leq p < 0.25$
b. $0.25 \leq p < 0.5$
c. $0.5 \leq p < 0.75$
d. $0.75 \leq p$
e. Unable to compute
Chebyshev’s Inequality – Repeated Experiments

“How many times does Alice need to flip a biased coin until she sees heads $n$ times, if heads occurs with probability $p$?

$X = \# \text{ of flips until } n \text{ times “heads”}$

$X_i = \# \text{ of flips between } (i - 1)-\text{st and } i-\text{th “heads”}$

$X = \sum_i X_i$

Note: $X_1, \ldots, X_n$ are independent and geometric with parameter $p$

$\mathbb{E}(X) = \mathbb{E} \left( \sum_i X_i \right) = \sum_i \mathbb{E}(X_i) = \frac{n}{p}$

$\text{Var}(X) = \sum_i \text{Var}(X_i) = \frac{n(1-p)}{p^2}$
Chebyshev’s Inequality – Coin Flips

“How many times does Alice need to flip a biased coin until she sees heads \( n \) times, if heads occurs with probability \( p \)?

\[
\mathbb{E}(X) = \mathbb{E}\left(\sum_i X_i\right) = \sum_i \mathbb{E}(X_i) = \frac{n}{p} \quad \text{Var}(X) = \sum_i \text{Var}(X_i) = \frac{n(1-p)}{p^2}
\]

What is the probability that \( X \geq 2\mathbb{E}(X) = 2n/p \)?

Markov: \( \mathbb{P}(X \geq 2n/p) \leq \frac{\mathbb{E}(X)}{2n/p} = \frac{n}{p} \cdot \frac{p}{2n} = \frac{1}{2} \)

Chebyshev: \( \mathbb{P}(X \geq 2n/p) \leq \mathbb{P}\left(\left| X - \frac{n}{p} \right| \geq \frac{n}{p} \right) \leq \frac{\text{Var}(X)}{n^2/p^2} = \frac{1-p}{n} \)

Goes to zero as \( n \to \infty \) 😊
Tail Bounds

Useful for approximations of complex systems. How good the approximation is depends on the actual distribution and the context you are using it in.

– Usually loose upper-bounds are okay when designing for worst-case

Generally (but not always) making more assumptions about your random variable leads to a more accurate upper-bounder.