CS145: Probability & Computing

Lecture 8: Expected value, Markov inequality , Variance, Chebyshev inequality

Figure credits:
*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*
*Pitman, **Probability**, 1999*

CS145: Lecture 7 Outline

- Expectation and Payoff
- Markov’s Inequality
- Variance
- Chebyshev's Inequality

Probability Expectation and Payoff

Consider the following game:

You pay \$C and receive \$1 if a (fair) dice role gives 6, and 0 otherwise.

For what values of C would you play the game?

A rational person will play for any $C \leq 1/6$.

Probability Expectation and Payoff

- The *expectation* or *expected value* of a discrete random variable is:

$$E[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

Consider the following game:

You pay \$C and receive \$i if a fair dice role gives i.

For what values of C will you pay the game?

A rational person will play for any C$\leq E[X] = 3.5$.

Expectation is not Enough

A rational person considers more than just the expectation

Which of these games do you prefer:

1. You pay \$1 and receive \$2 with probability ½
2. You pay \$1 and receive \$1,000 with probability $\frac{1}{1{,}000}$
3. You pay \$5 and receive \$1,000,000 with probability $\frac{1}{1{,}000{,}000}$

Expectation is not Enough

Which job would you prefer?

1. A job that pays \$150,000 a year
2. A job that pays \$100,000 a year plus a bonus of \$100,000 with probability ½
3. A job that pays \$70,000 with equity option of \$1,000,000 with probability 0.01

The deviation from the expectation captures the risk

[Variance, standard deviation, value at risk, large deviation bounds…]

CS145: Lecture 7 Outline

- Expectation and Payoff
- Markov's Inequality, Variance, Chebyshev's Inequality

The Most Basic Deviation Bound

Theorem

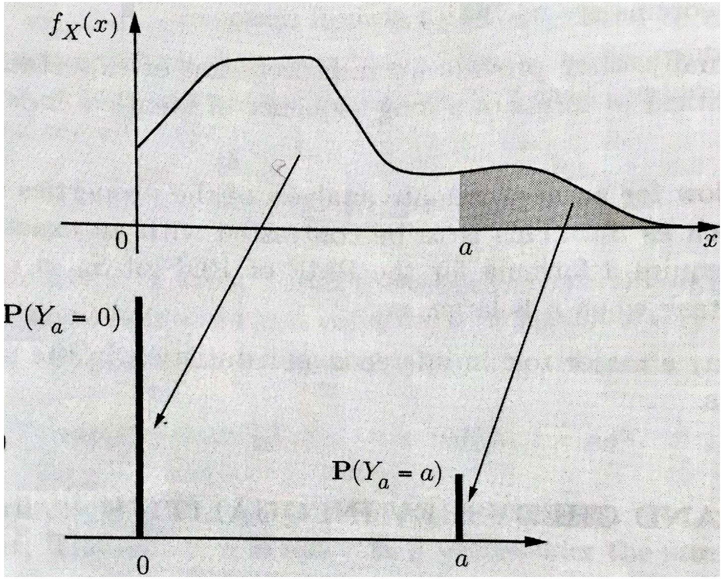
[Markov Inequality] *For any non-negative random variable, and for all* $a > 0$,

$$Pr(X \geq a) \leq \frac{E[X]}{a}.$$

Fix some constant *a>0*, and define

$$Y_a = \begin{cases} 0, & \text{if } X < a, \\ a, & \text{if } X \geq a. \end{cases}$$

$$aP(X \geq a) = E[Y_a] \leq E[X]$$



Markov's Inequality

Theorem

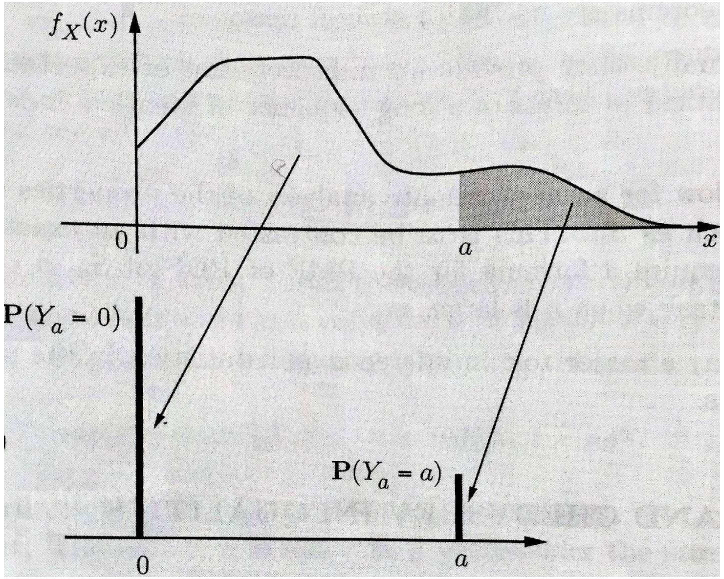
[Markov Inequality] *For any non-negative random variable, and for all* $a > 0$,

$$Pr(X \geq a) \leq \frac{E[X]}{a}.$$

Fix some constant *a>0*, and define

$$Y_a = \begin{cases} 0, & \text{if } X < a, \\ a, & \text{if } X \geq a. \end{cases}$$

- No such inequality would hold if *X* could take negative values. Why?
- If *a* < *E[X]*, Markov's inequality is vacuous, but no better bound is possible. Why?



Examples

We flip 100 fair coins. X is the number of heads. $E[X] = 50$

$$Prob(X \geq 75) \leq \frac{50}{75} = \frac{2}{3}.$$

We flip a fair coin till the first head. Y is the numnber of flips.

$$E[Y] = \sum_{i \geq 1} Prob(Y \geq i) = \sum_{i \geq 1} \frac{1}{2^{i-1}} = 2$$

$$Prob(Y \geq 4) \leq \frac{2}{4} = \frac{1}{2}$$

Variance

- **Reminder.** The *expectation* or *expected value* of a random variable

$$E[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

- The *variance* is the *expected squared deviation* of a random variable from its mean (the following definitions are equivalent):

$$\mathrm{Var}[X] = E[(X - E[X])^2] = \sum_{x \in \mathcal{X}} (x - E[X])^2 p_X(x)$$

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \left[\sum_{x \in \mathcal{X}} x^2 p_X(x)\right] - \left[\sum_{x \in \mathcal{X}} x p_X(x)\right]^2$$

- The *standard deviation* is the square root of the variance:

$$\sigma_X = \mathrm{Std}[X] = \sqrt{\mathrm{Var}[X]}$$

Bernoulli Distribution

- A *Bernoulli* or *indicator* random variable *X* has one parameter *p*:

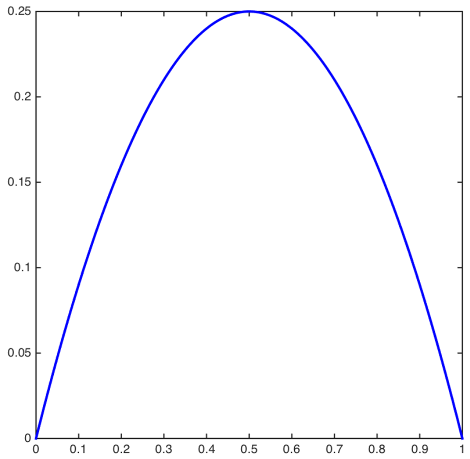
$$p_X(1) = p, \qquad p_X(0) = 1 - p, \qquad \mathcal{X} = \{0, 1\}$$

- For an indicator variable, *expected values are the probabilities:*

$$E[X] = p$$

- Variance of Bernoulli distribution:

$$\mathrm{Var}[X] = E\left[(X - p)^2\right] = p(1 - p)$$



- Fair coin (*p=0.5*) has largest variance
- Coins that always come up heads (*p=1.0*), or always come up tails (*p=0.0*), have variance 0

Sums of Independent Variables

- If $Z=X+Y$ and random variables X and Y are independent, we have

$$E[Z] = E[X] + E[Y] \qquad \mathrm{Var}[Z] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

For any variables X, Y. *Only for independent X, Y.*

- Interpretation: Adding independent variables increases variance

$$\mathrm{Var}[Z] \geq \mathrm{Var}[X] \quad \text{and} \quad \mathrm{Var}[Z] \geq \mathrm{Var}[Y]$$

Examples

Let X and Y be two Bernoulli r.v. such that $P(X=Y)=1$, $Z=X+Y$

$P_{XY}=$

	0	1
0	$(1-p)^2$	$p(1-p)$
1	$P(1-p)$	p^2

$$Var[Z] = E[(Z-2p)^2]$$
$$= p^2(2-2p)^2 + 2(1-p)p(1-2p)^2 + (-2P)^2(1-P)^2$$
$$= p(1\text{-}p)[4p(1\text{-}p)+2(1-2p)^2 + 4p(1-p)]$$
$$= 2p(1-P) = Var[X] + Var[Y]$$

$P_{XY}=$

	0	1
0	$1-p$	0
1	0	p

$$Var[Z] = E[(Z-2p)^2] = (2-2p)^2 p + (1-p)(-2p)^2$$
$$= (1-p)p(4(1-p)+4p) = (1-p)p$$
$$= Var[X] = Var[Y]$$

Sums of Independent Variables

- If $Z=X+Y$ and random variables X and Y are independent, we have

$$E[Z] = E[X] + E[Y] \qquad \mathrm{Var}[Z] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

For any variables X, Y. *Only for independent X, Y.*

- The *standard deviation* of a sum of independent variables is then

$$\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2} \qquad \sigma_X = \sqrt{\mathrm{Var}[X]}, \sigma_Y = \sqrt{\mathrm{Var}[Y]}, \sigma_Z = \sqrt{\mathrm{Var}[Z]}$$

- Identity used in proof: If X and Y are *independent* random variables,

$$E[XY] = E[X]E[Y] \qquad \text{if} \qquad p_{XY}(x,y) = p_X(x)p_Y(y)$$

This equality does not hold for general, dependent random variables.

Some Math

Assume that X and Y are independent random variable.

$$
\begin{aligned}
E[XY] &= \sum_x \sum_y xyProb(X = x, Y = y) \\
&= \sum_x \sum_y xyProb(X = x)Prob(Y = y) \\
&= \sum_x xProb(X = x) \left(\sum_y yProb(Y = y) \right) \\
&= E[X]E[Y]
\end{aligned}
$$

More Math

Assume that X and Y are independent random variables,

$$\begin{aligned} Var[X+Y] &= E[(X+Y)^2] - (E[X] + Y[Y])^2 \\ &= E[X^2] + E[Y^2] + 2E[XY] - (E[X])^2 - (E[Y])^2 - 2E[X]E[Y] \\ &= Var[X] + Var[Y] \end{aligned}$$

Binomial Probability Distribution

- Suppose you flip *n* coins with bias *p*, count number of heads
- A *binomial* random variable *X* has parameters *n, p*:

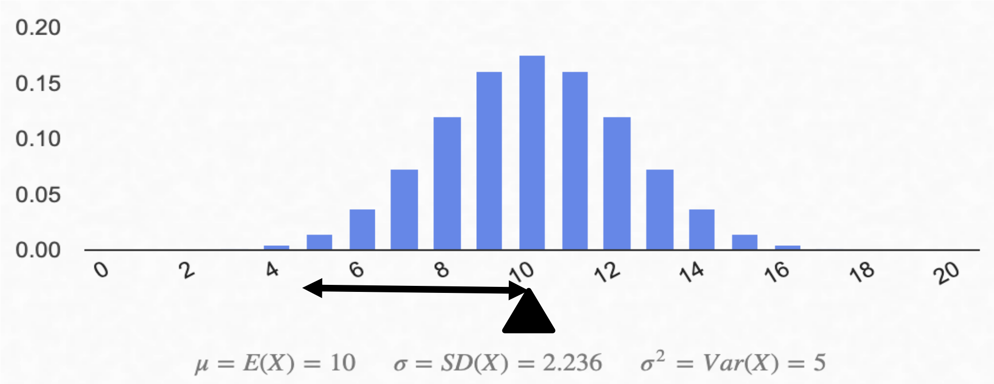
$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \mathcal{X} = \{0, 1, 2, \ldots, n\}$$

X_i is a Bernoulli variable indicating whether toss *i* comes up heads,

because tosses are *independent*:

$$X = \sum_{i=1}^{n} X_i \qquad E[X] = np \qquad \mathrm{Var}[X] = np(1-p)$$

$$n = 20, p = 0.5$$



Binomial Probability Distribution

- Suppose you flip *n* coins with bias *p*, count number of heads
- A *binomial* random variable *X* has parameters *n, p*:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

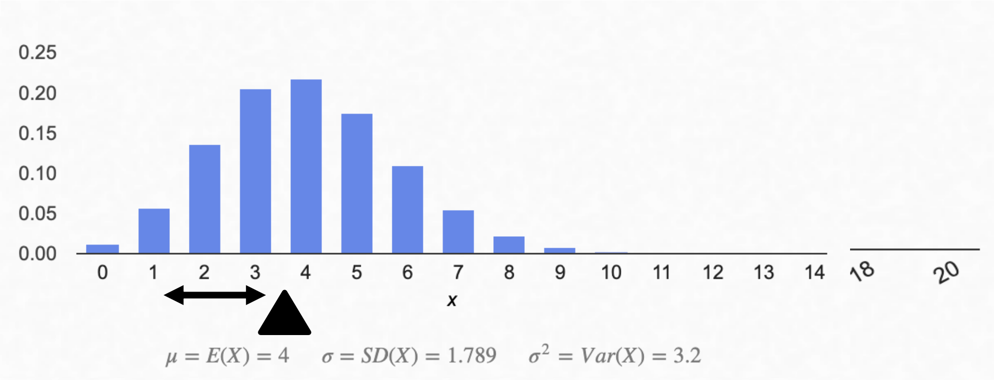
$$\mathcal{X} = \{0, 1, 2, \ldots, n\}$$

X_i is a Bernoulli variable indicating whether toss *i* comes up heads,

because tosses are *independent*:

$$X = \sum_{i=1}^{n} X_i \qquad E[X] = np \qquad \mathrm{Var}[X] = np(1-p)$$

$$n = 20, p = 0.2$$



Binomial Probability Distribution

- Suppose you flip *n* coins with bias *p*, count number of heads
- A *binomial* random variable *X* has parameters *n, p*:

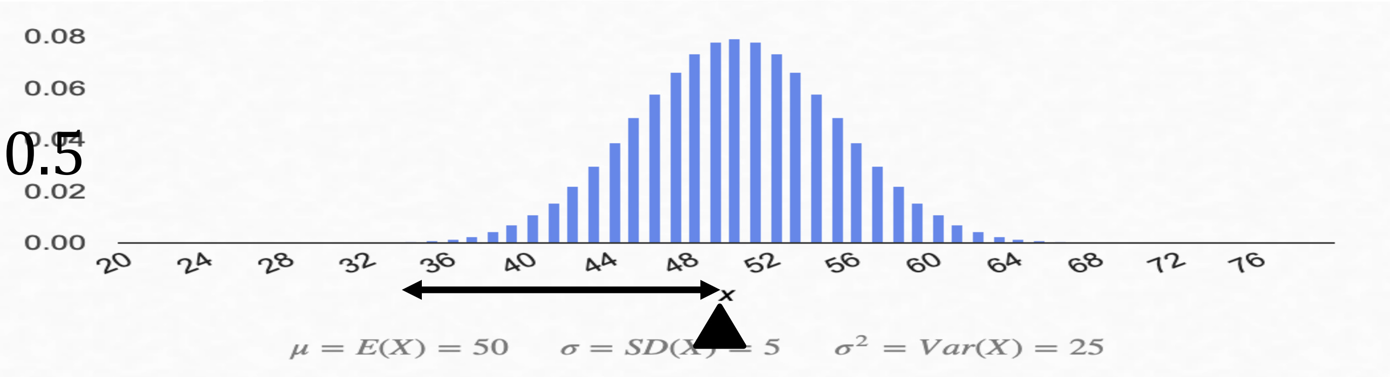
$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \mathcal{X} = \{0, 1, 2, \ldots, n\}$$

X_i is a Bernoulli variable indicating whether toss *i* comes up heads,

because tosses are *independent*:

$$X = \sum_{i=1}^{n} X_i \qquad E[X] = np \qquad \mathrm{Var}[X] = np(1-p)$$

$n = 100, p = 0.5$



Variance: What comes next?

- Expectation and Variance
- Markov’s Inequality
- Variance
- Chebyshev's Inequality

Chebyshev's Inequality

Theorem

For **any** *random variable* X*, and any* $a > 0$*,*

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

Proof.

$$Pr(|X - E[X]| \geq a) = Pr((X - E[X])^2 \geq a^2)$$

By Markov inequality

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

$$= \frac{Var[X]}{a^2}$$

Chebyshev's Inequality

Theorem

For **any** *random variable* X, *and any* $a > 0$,

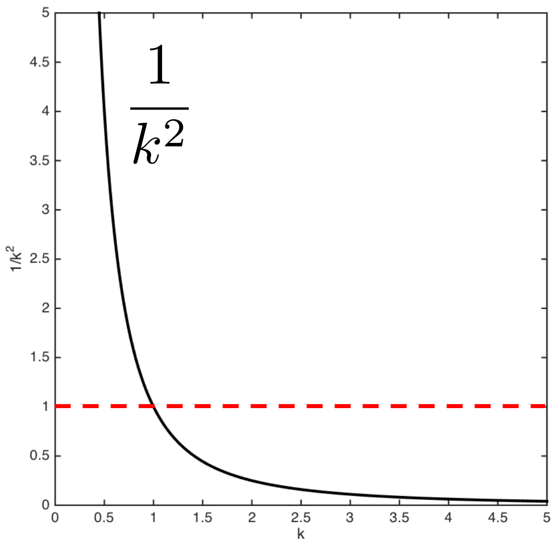
$$Pr(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}.$$

- Another way of parameterizing Chebyshev's inequality:

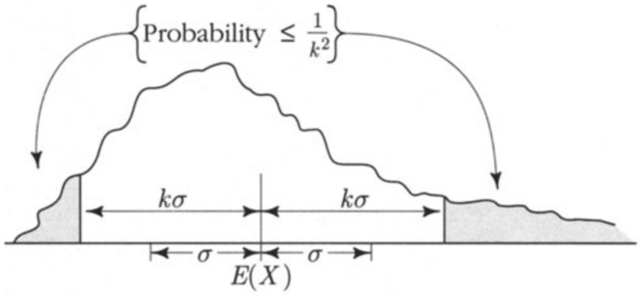
$$\mu = E[X], \qquad \sigma = \sqrt{\mathrm{Var}[X]}$$

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- Chebyshev bound is vacuous (above one) for events less than one standard deviation from the mean. But this could be likely!



Chebyshev's Inequality

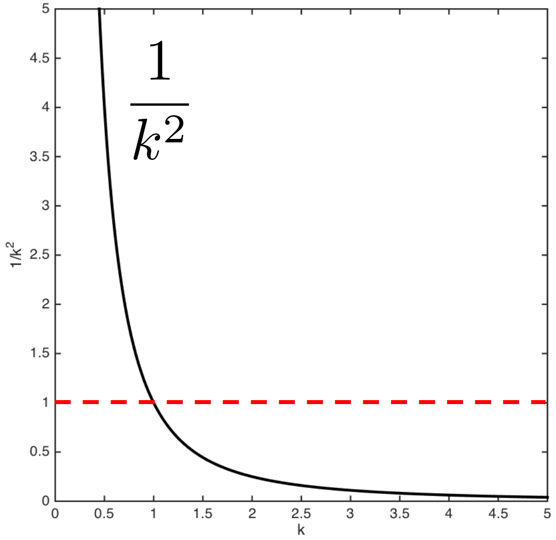


- Another way of parameterizing Chebyshev's inequality:

$$\mu = E[X], \qquad \sigma = \sqrt{\mathrm{Var}[X]}$$

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- Chebyshev bound is vacuous (above one) for events less than one standard deviation from the mean. But this could be likely!



Markov vs. Chebyshev's Inequalities

We flip a fair coin n times.
What is the probability of getting more than $3n/4$ heads?

$X =$ number of heads. $E[X] = n/2$, $Var[X] = n/4$.

Markov's Inequality:

$$Pr\{X \geq \frac{3n}{4}\} \leq \frac{E[X]}{3n/4} \leq \frac{n/2}{3n/4} = \frac{2}{3}$$

Chebyshev's Inequality:

$$Pr\{X \geq \frac{3n}{4}\} \leq Pr\{|X - \frac{n}{2}| \geq \frac{n}{4}\} \leq \frac{Var[X]}{(n/4)^2} = \frac{n/4}{n^2/16} = \frac{4}{n}$$

The Weak Law of Large Numbers

Theorem

Let $x_1, \dots, x_n$ *be independent, identically distributed random variables with finite mean,* $E[x_i] = \mu$*. For any* $\epsilon > 0$

$$Prob\{|\frac{1}{n}\sum_{i=1}^{n} x_i - \mu| \geq \epsilon\} \to 0$$

as $n \to \infty$.

The (Weak) Law of Large Numbers

$X_1, X_2, \ldots$ i.i.d.
finite mean μ and variance σ^2

$$M_n = \frac{X_1 + \cdots + X_n}{n}$$

sample mean or empirical mean

$$E[M_n] = \frac{\mathbf{E}[X_1] + \cdots + \mathbf{E}[X_n]}{n} = \frac{n\mu}{n} = \mu,$$

$$\mathrm{Var}[M_n] = \frac{\mathrm{var}(X_1 + \cdots + X_n)}{n^2} = \frac{\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

- Chebyshev's inequality bounds distance between the true mean and the "empirical" or "sample" mean:

$$\mathbf{P}(|M_n - \mu| \geq \epsilon) \;\leq\; \frac{\mathrm{Var}(M_n)}{\epsilon^2} \;=\; \frac{\sigma^2}{n\epsilon^2}$$

- *The empirical mean converges to the true mean in probability*

$$\lim_{n \to \infty} P(|M_n - \mu| \geq \epsilon) = 0$$

- True even if variance not finite, but proof more challenging.