Data Science Theory and Practices

Deviations and Fundamental Tail Probabilities

Joseph Chuang-Chieh Lin
Dept. CSIE, Tamkang University

17th May 2021

Markov's Inequality

- Let X be a random variable that assumes only non-negative values. Then, for all $a > 0$,

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

Andrei Andreyevich Markov (Wikipedia)
1856–1922

Markov's Inequality

- Let X be a random variable that assumes only non-negative values. Then, for all $a > 0$,

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

- *Proof*:

Let $I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$. Since $X \geq 0$, $I \leq \frac{X}{a}$

$$\Pr[X \geq a] = \Pr[I = 1] = \mathbf{E}[I] \leq \mathbf{E}\left[\frac{X}{a}\right] = \frac{\mathbf{E}[X]}{a}.$$

Andrei Andreyevich Markov (Wikipedia)
1856–1922

Example: 75% heads in fair coin flips

- What's the probability of obtaining > $3n/4$ heads in a sequence of n fair coin flips?

Example: 75% heads in fair coin flips

- What's the probability of obtaining > 3*n*/4 heads in a sequence of *n* fair coin flips?

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

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

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

- ✓ Applying Markov's inequality,

Example: 75% heads in fair coin flips

- What's the probability of obtaining > 3*n*/4 heads in a sequence of *n* fair coin flips?

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

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

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

  - Applying Markov's inequality,

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

Scenarios of applying Markov's inequality

- Markov's inequality gives the best tail bound when:
  - All we know is the expectation of the random variable
  - The random variable is non-negative.

The kth moment

- Definition. The kth of a random variable X is $\mathbf{E}[X^k]$.
  - So, the expectation is the *first moment* of X.

The variance

- Definition. The variance of a random variable X is defined as

$$\mathbf{Var}[X] = \mathbf{E}[(X - \boldsymbol{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

The variance

- Definition. The variance of a random variable X is

$$\mathbf{Var}[X] = \mathbf{E}[(X - \boldsymbol{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

- Definition. The standard deviation of a random variable X is

$$\sigma[X] = (\mathbf{Var}[X])^{1/2}.$$

- Definition. The covariance of two random variables X and Y is

$$\mathbf{Cov}(X, Y) = \mathbf{E}[(X - \boldsymbol{E}[X])(Y - \boldsymbol{E}[Y])].$$

Variance of the sum of two random variables

- Theorem. For any two random variables X and Y,

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\,\mathbf{Cov}(X, Y).$$

Variance of the sum of two random variables

- <u>Theorem</u>. For any two random variables X and Y,

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\,\mathbf{Cov}(X, Y).$$

Not as nice as the linearity of expectation!

Variance of the sum of two random variables

- <u>Theorem</u>. For any two random variables X and Y,

$$\textbf{Var}[X+Y] = \textbf{Var}[X] + \textbf{Var}[Y] + 2\,\textbf{Cov}(X, Y).$$

- *Proof*:

$$\begin{aligned}
\textbf{Var}[X+Y] &= \mathbf{E}[(X+Y-\mathbf{E}[X+Y])^2] \\
&= \mathbf{E}[(X+Y-\mathbf{E}[X]-\mathbf{E}[Y])^2] \\
&= \mathbf{E}[((X-\mathbf{E}[X])+(Y-\mathbf{E}[Y]))^2] \\
&= \mathbf{E}[(X-\mathbf{E}[X])^2+(Y-\mathbf{E}[Y])^2+2(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])] \\
&= \mathbf{E}[(X-\mathbf{E}[X])^2]+\mathbf{E}[(Y-\mathbf{E}[Y])^2]+2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])] \\
&= \textbf{Var}[X]+\textbf{Var}[Y]+2\textbf{Cov}(X,Y).
\end{aligned}$$

Expectation of product of two random variables

- Theorem. If X and Y are two **independent** random variables, then

$$\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

Expectation of product of two random variables

- Theorem. If X and Y are two **independent** random variables, then

$$\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

Proof:

$$\begin{aligned}
\mathbf{E}[X \cdot Y] &= \sum_i \sum_j (i \cdot j) \cdot \Pr[(X = i) \cap (Y = j)] \\
&= \sum_i \sum_j (i \cdot j) \cdot \Pr[X = i] \cdot \Pr[Y = j] \\
&= \left(\sum_i i \cdot \Pr[X = i] \right) \cdot \left(\sum_j \cdot \Pr[Y = j] \right) \\
&= \mathbf{E}[X] \cdot \mathbf{E}[Y].
\end{aligned}$$

Independence → Linearity of Variance

- Corollary. If X and Y are independent random variables, then

$$\mathbf{Cov}(X, Y) = 0$$

and

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y].$$

Independence → Linearity of Variance

- <u>Corollary</u>. If X and Y are independent random variables, then

$$\mathbf{Cov}(X, Y) = 0$$

and

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y].$$

Proof:

$$\begin{aligned}
\mathbf{Cov}[X, Y] &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\
&= \mathbf{E}[X - \mathbf{E}[X]] \cdot \mathbf{E}[Y - \mathbf{E}[Y]] \\
&= 0.
\end{aligned}$$

Remark

- Theorem. For mutually independent random variables X_1, X_2, …, X_n

$$\mathbf{Var}\left[\sum_{i}^{n} X_i\right] = \sum_{i}^{n} \mathbf{Var}[X_i].$$

Example: Variance of a binomial random variable

- **Recall**: a binomial random variable X can be regarded as the sum of n independent Bernoulli trials (Y's), each with success probability p.

Example: Variance of a binomial random variable

- **Recall**: a binomial random variable X can be regarded as the sum of n independent Bernoulli trials (Y's), each with success probability p.

$$\mathbf{Var}[Y] = \mathbf{E}[(Y - \mathbf{E}[Y])^2] = p \cdot (1-p)^2 + (1-p) \cdot (0-p)^2 = p(1-p).$$

Example: Variance of a binomial random variable

- A binomial random variable X can be regarded as the sum of n independent Bernoulli trials (Y's), each with success probability p.

$$\mathbf{Var}[Y] = \mathbf{E}[(Y - \mathbf{E}[Y])^2] = p \cdot (1-p)^2 + (1-p) \cdot (0-p)^2 = p(1-p).$$

- By the theorem in p.19,

$$\mathbf{Var}[X] = n \cdot (p(1-p)) = np(1-p).$$

Chebyshev's Inequality

- A stronger tail bound if you have the expectation and the variance.

- Theorem [Chebyshev's Inequality]. For any $a > 0$,

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

Chebyshev's Inequality

- Theorem [Chebyshev's Inequality]. For any $a > 0$,

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

- *Proof*:

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

Apply Markov's inequality,

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

Chebyshev's Inequality

- Corollary. For any $t > 1$,

$$\Pr[|X - \mathbf{E}[X])| \geq t \cdot \sigma[X]] \leq \frac{1}{t^2}.$$

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

Example: 75% heads in fair coin flips

- $$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \qquad \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

Example: 75% heads in fair coin flips

- $$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

  $$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \qquad \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

- Actually,

  $$\mathbf{E}[X_i^2] = \mathbf{E}[X_i] = \frac{1}{2}.$$

  $$\mathbf{Var}[X_i] = \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

  $$\mathbf{Var}[X] = \mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}[X_i] = \frac{n}{4}.$$

Example: 75% heads in fair coin flips

- $$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \qquad \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

- Actually,

$$\mathbf{E}[X_i^2] = \mathbf{E}[X_i] = \frac{1}{2}.$$

$$\mathbf{Var}[X_i] = \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

$$\mathbf{Var}[X] = \mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}[X_i] = \frac{n}{4}.$$

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

Example: 75% heads in fair coin flips

- $$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{E}[X_i] = \Pr[X_i = 1] = \frac{1}{2} \qquad \mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n}{2}.$$

- Actually,

$$\mathbf{E}[X_i^2] = \mathbf{E}[X_i] = \frac{1}{2}.$$

$$\mathbf{Var}[X_i] = \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

$$\mathbf{Var}[X] = \mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}[X_i] = \frac{n}{4}.$$

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

Example: 75% heads in fair coin flips

- What's the probability of obtaining > $3n/4$ heads in a sequence of n fair coin flips?

$$X_i = \begin{cases} 1 & \text{if the } i\text{th coin flip is head} \\ 0 & \text{otherwise} \end{cases}$$

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

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

- Applying Markov's inequality,

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

Randomized Algorithms, CSIE, TKU, Taiwan 18

Assignment 03

1. Let X be a number chosen uniformly at random from $[1, n]$. Find $\text{Var}[X]$.
2. Suppose that we roll a standard fair die 100 times. Let X be the sum of the numbers that appear over the 100 rolls. Use Chebyshev's inequality to bound

$$\Pr[|X - 350| \geq 50].$$