CSE 312

Foundations of Computing II

Lecture 14: Markov, Chebyshev / Conditional Expectation

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Stefano Tessaro
tessaro@cs.washington.edu

Announcements

- Practice Midterms are online.
  - New references online.
- Q&A session on Monday, Oct 28, in class.
- My office hours
  - Moved to Today 3-4pm + Wednesday, 1-3pm
- Next week on Wednesday: Data compression
  - How much can we compress data? (And how does this relate to probability theory?)

Expectation and Variance

So far: Given a random variable X, find $\mathbb{E}(X)$ and $\mathrm{Var}(X)$.

Today: Given $\mathbb{E}(X)$ and $\mathrm{Var}(X)$, can we “bound” X?

Markov's Inequality

Andrei Andreyevich Markov

1856-1922

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 takes into account the expectation.

However: Extremely powerful when combined with clever tricks.

Markov's Inequality – Proof

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

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

$$\mathbb{E}(X) = \sum_{x} x \cdot \mathbb{P}(X = x)$$

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

≥ 0 because $x \geq 0$ whenever $\mathbb{P}(X = x) \geq 0$ (takes only non-negative values)

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

$$\geq \sum_{x \geq t} t \cdot \mathbb{P}(X = x) \quad = t \cdot \mathbb{P}(X \geq t)$$

Follows by re-arranging terms

…

Example – Geometric Random Variable

Let X be geometric RV with parameter p

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

"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: $\mathbb{P}(X \geq 2/p) \leq \frac{\mathbb{E}(X)}{2/p} = \frac{1}{p} \cdot \frac{p}{2} = \frac{1}{2}$ **Can we do better?**

Chebyshev's Inequality

Pafnuty Chebyshev

1821-1894

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

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

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

Definition of Variance

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

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

Markov's inequality ($Z^2 \geq 0$)

Example – Geometric Random Variable

Let X be geometric RV *with* parameter p

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

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

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

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

Not better, unless $p > 1/2$ ☹

Chebyshev's Inequality – Repeated Experiments

"How many times does Alice need to flip a biased coin <u>until she sees heads n</u> times, if heads occurs with probability p?

X = # of flips until n times "heads"

X_i = # of flips between $(i-1)$-st and i-th "heads"

$$X = \sum_i X_i$$

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

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_i X_i\right) = \sum_i \mathbb{E}(X_i) = \frac{n}{p} \qquad \mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i) = \frac{n(1-p)}{p^2}$$

Chebyshev's Inequality – Coin Flips

"How many times does Alice need to flip a biased coin <u>*until she sees heads n*</u> *times, if heads occurs with probability p?*

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_i X_i\right) = \sum_i \mathbb{E}(X_i) = \frac{n}{p} \qquad \mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i) = \frac{n(1-p)}{p^2}$$

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

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

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

Goes to zero as $n \to \infty$ ☺

Something different …

Example – Number of Samples is a Random Variable

Consider the following.

- We roll a die twice, and sum the outcomes.
  - Result is random variable $N \in \{2, \ldots, 12\}$
- Then, we flip N independent coins, each with heads with probability p
  - X_i is Bernoulli indicator variable, 1 if i-th coin is heads.

Q: What is the expectation of $X_1 + \cdots + X_N$?

<u>Problem:</u> Linearity cannot be used directly, because number of terms is a random variable itself! We cannot just write $N \cdot p$

Conditional Expectation

Definition. The **expectation** of X conditioned on event $\mathcal{A}$ is

$$\mathbb{E}(X|\mathcal{A}) = \sum_x x \cdot \mathbb{P}(X = x|\mathcal{A})$$

Theorem. (Law of total expectation). If $\mathcal{A}_1, \dots, \mathcal{A}_n$ partition the sample space Ω, then

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(\mathcal{A}_i) \cdot \mathbb{E}(X|\mathcal{A}_i)$$

Law of Total Expectation – Proof

Theorem. (Law of total expectation). If $\mathcal{A}_1, \dots, \mathcal{A}_n$ partition the sample space Ω, then

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(\mathcal{A}_i) \cdot \mathbb{E}(X|\mathcal{A}_i)$$

$$\mathbb{E}(X) = \sum_{x} \mathbb{P}(X = x) \cdot x$$

$$= \sum_{x} \left(\sum_{i=1}^{n} \mathbb{P}(X = x|\mathcal{A}_i) \cdot \mathbb{P}(\mathcal{A}_i) \right) \cdot x$$

Law of total probability

$$= \sum_{i=1}^{n} \mathbb{P}(\mathcal{A}_i) \sum_{x} (\mathbb{P}(X = x|\mathcal{A}_i) \cdot x)$$

$$= \sum_{i=1}^{n} \mathbb{P}(\mathcal{A}_i) \cdot \mathbb{E}(X|\mathcal{A}_i)$$

Example – Number of Samples is a Random Variable

- Sum two fair dice rolls $N \in \{2, \ldots, 12\}$
- N independent Bernoulli $X_1, \ldots, X_N$ with parameter p

$$\mathbb{E}(X_1 + \cdots + X_N) = \sum_{i=2}^{12} \mathbb{P}(N = i) \cdot \mathbb{E}(X_1 + \cdots + X_N | N = i)$$

$$= \sum_{i=2}^{12} \mathbb{P}(N = i) \cdot \mathbb{E}(X_1 + \cdots + X_i)$$

$$= \sum_{i=2}^{12} \mathbb{P}(N = i) \cdot ip \;\; = p \sum_{i=2}^{12} \mathbb{P}(N = i) \cdot i \;\; = p\mathbb{E}(N) = 7p$$