CSE 312
Foundations of Computing II

Lecture 14: Markov, Chebyshev / Conditional Expectation

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 $E(X)$ and $\text{Var}(X)$.

Today: Given $E(X)$ and $\text{Var}(X)$, can we “bound” $X$?
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 takes into account the expectation.

However: Extremely powerful when combined with clever tricks.
Markov’s Inequality – Proof

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

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

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

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

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

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

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

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

\[ \text{follows by re-arranging terms} \ldots \]
Example – Geometric Random Variable

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

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

$$\mathbb{E}(X) = \frac{1}{p}$$

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

Can we do better?
Chebyshev’s Inequality

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

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

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

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

$|Z| \geq t$ 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 $p > 1/2$ ☹️
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 \( p \)?

\( X = \# \) of flips until \( n \) times “heads”

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

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

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

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

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

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

**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 \( n \to \infty \). 😊
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 $X_1 + \cdots + X_N$?

Problem: Linearity cannot be used directly, because number of terms is a random variable itself! We cannot just write $N \cdot p$
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 $\mathcal{A}_1, \ldots, \mathcal{A}_n$ partition the sample space $\Omega$, 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

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

\[ = \sum_{x} \left( \sum_{i=1}^{n} P(X = x | A_i) \cdot P(A_i) \right) \cdot x \]

\[ = \sum_{i=1}^{n} P(A_i) \sum_{x} (P(X = x | A_i) \cdot x) \]

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

**Theorem. (Law of total expectation).** If \( A_1, \ldots, A_n \) partition the sample space \( \Omega \), then

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

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

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

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