CMPSCI 240: Reasoning about Uncertainty

Lecture 10: Variance and Tail Bounds

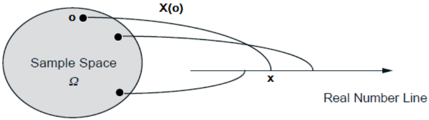
Andrew McGregor

University of Massachusetts

Last Compiled: March 2, 2017

Outline

Random Variables



- Formally, a random variable X is a mapping from Ω to $\mathbb{R}$.
- Given a random variable $X : \Omega \to \mathbb{R}$, we can define events $\{X = k\}$ for each value of $k \in \mathbb{R}$ and the probability of these events is the *probability mass function*.

$$P(X = k) = \sum_{o \in \Omega \text{ such that } X(o)=k} P(o)$$

- It's often sufficient to just consider the probability mass function rather than the probabilities of every event.

Expected Value

- The expected value $E[X]$ is a probability-weighted average of the possible values of the random variable X:

$$E[X] = \sum_k k\,P(X = k)$$

- E.g., if $P(X = 2) = 1/2$, $P(X = 3) = 1/4$, and $P(X = 5) = 1/4$,

$$E[X] = 2 \times \frac{1}{2} + 3 \times \frac{1}{4} + 5 \times \frac{1}{4} = 3$$

- If X is a random variables and $f : \mathbb{R} \to \mathbb{R}$ then $Y = f(X)$ is another random variable with $E(Y) = \sum_k f(k)P(X = k)$.

Variance

- Variance measures how far we expect a random variable to be from its average. The variance of X is defined as the expectation of $Y = (X - E[X])^2$:

$$\mathrm{var}(X) = E[(X - E[X])^2] = \sum_k (k - E[X])^2 \cdot P(X = k)$$

- E.g., if $P(X = 2) = 1/2$, $P(X = 3) = 1/4$, and $P(X = 5) = 1/4$, then since $E[X] = 3$ the variance is

$$\mathrm{var}[X] = (2 - 3)^2 \times \frac{1}{2} + (3 - 3)^2 \times \frac{1}{4} + (5 - 3)^2 \times \frac{1}{4} = 1.5$$

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

$$\mathrm{std}(X) = \sqrt{\mathrm{var}(X)}\ .$$

Example 1

- Consider a random variable X where

$$P(X = 2) = 1/2 \qquad P(X = 3) = 1/4 \qquad P(X = 5) = 1/4$$

- The expected value is:

$$E[X] = \frac{1}{2} \times 2 + \frac{1}{4} \times 3 + \frac{1}{4} \times 5 = 3$$

- The variance is:

$$\text{var}[X] = \frac{1}{2} \times (2 - 3)^2 + \frac{1}{4} \times (3 - 3)^2 + \frac{1}{4} \times (5 - 3)^2 = 1.5$$

Example 2

- Consider a random variable X where

$$P(X = -1) = 1/2 \qquad P(X = 7) = 1/2$$

- The expected value is:

$$E[X] = \frac{1}{2} \times (-1) + \frac{1}{2} \times 7 = 3$$

- The variance is:

$$\text{var}[X] = \frac{1}{2} \times (-1 - 3)^2 + \frac{1}{2} \times (7 - 3)^2 = 16$$

Variance Value: Clicker Question

- The variance var$[X]$ of X is defined as:

$$\text{var}(X) = E[(X - E[X])^2] = \sum_k (k - E[X])^2 \cdot P(X = k)$$

- If X maps to $\{1, 2, 6\}$ and

$$P(X = 1) = 1/3 \quad , \quad P(X = 2) = 1/3 \quad , \quad P(X = 6) = 1/3$$

then is the variance:

$$A)\ 3 \qquad B)\ 4 \qquad C)\ 4.66\ldots \quad D)\ 8.33\ldots \quad E)\ 9\ldots$$

- Answer is

$$\text{var}[X] = (1 - 3)^2 \times 1/3 + (2 - 3)^2 \times 1/3 + (6 - 3)^2 \times 1/3 = 4.66\ldots$$

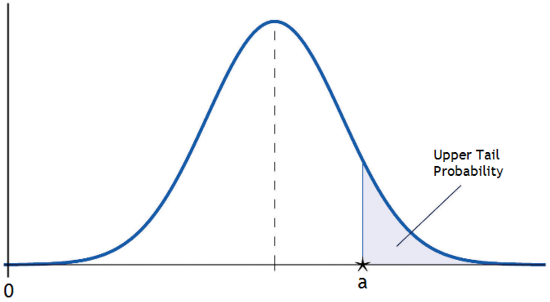
Variance of Standard Random Variables

- **Bernoulli:** $\text{var}[X] = p(1-p)$
- **Binomial:** $\text{var}[X] = np(1-p)$
- **Geometric:** $\text{var}[X] = \frac{1-p}{p^2}$
- **Uniform:** $\text{var}[X] = \frac{(b-a+1)^2-1}{12}$
- **Poisson:** $\text{var}[X] = \lambda$

Outline

Example: Snowfall in Amherst

- Consider the probability density for the total snowfall in Amherst this year. The probability that the total snowfall exceeds some value a is called the **upper tail probability**.



- **Upper Tail Probabilities:** $P(X \geq a) = \sum_{k \in \mathbb{R}: k \geq a} P(X = k)$
- If X just takes integer values and a is an integer then

$$P(X \geq a) = P(X = a) + P(X = a+1) + P(X = a+2) + P(X = a+3) + \ldots$$

Probability and Risk

- Certain classes of reasoning problems, particularly risk assessment, revolve around the probability and cost of extreme events.
- Upper tail probabilities tell you how likely extreme events are, e.g., probability snowfall exceeds twice the expected snowfall.
- One of the main problems in risk assessment is that the probability distributions of the random variables involved are not exactly known and can be quite complex but we can still make some deductions using just the expectation and variance.

Outline

Known Expectation

- Suppose we have a discrete non-negative random variable X.
- **Question:** What can we say about $P(X \geq a)$ if we don't know the PMF but do know the expected value $E[X]$?
- **Markov's Inequality:** $P(X \geq a) \leq \frac{E[X]}{a}$
- Let's prove it ...

Proof of Markov Inequality

$$
\begin{aligned}
E[X] &= \sum_{k} kP(X = k) \\
&= \sum_{k<a} kP(X = k) + \sum_{k \geq a} kP(X = k) \\
&\geq \sum_{k<a} 0P(X = k) + \sum_{k \geq a} aP(X = k) \\
&= a \sum_{k \geq a} P(X = k) \\
&= aP(X \geq a)
\end{aligned}
$$

1. And therefore $P(X \geq a) \leq E[X]/a$ as required.

Example: Snowfall

- **Question:** Suppose the average total snowfall in Amherst over the past hundred years has been 5 feet per winter. Find an upper bound on the probability that there is 20 or more feet of snow this winter.
- **Answer:** Let X be the total snowfall.
  - The average snowfall is 5 feet so assume $E[X] = 5$.
  - The probability we are interested in is $P(X \geq 20)$.
  - Using Markov's Inequality, $P(X \geq 20) \leq 5/20 = 0.25$.

Example: Dice Rolls

- **Question:** You roll a pair of fair six sided dice repeatedly until the first pair of sixes. Find an upper bound on the probability that it takes you 100 or more rolls.
- **Clicker Question:** The Markov bound implies that the probability

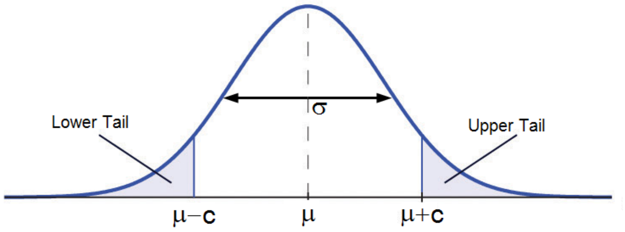
$$A)\ 0.01 \qquad B)\ 0.1 \qquad C)\ 0.36 \quad D)\ 0.63 \quad E)\ 0.66$$

- **Answer:** Let X be the number of rolls up to and including the first pair of sixes.
  - The PMF of X is geometric with $p = 1/36$
  - We thus have $E[X] = 1/p = 36$.
  - The probability we are interested in is $P(X \geq 100)$.
  - Using Markov's Inequality, $P(X \geq 100) \leq 36/100 = 0.36$.

Outline

Double Sided Risk

- Suppose you run a website hosting company. Every month you have to decide how much bandwidth to buy. If you buy more than you need you'll lose money. If you buy too little, you won't have enough to serve your customers' sites and they'll leave.
- Let μ denote the expected value and σ be the standard deviation.



- $P(|X-\mu| \geq c) = P(X \leq \mu - c) + P(X \geq \mu + c)$
- Can we estimate $P(|X-\mu| \geq c)$ using $E(X)$ and $var(X)$?

Chebyshev's Inequality

- Chebyshev's Inequality: For any random variable X and tail threshold $c > 0$:
$$P(|X - \mu| \geq c) \leq \frac{var[X]}{c^2}$$
- Intuition: if random variable X has a small variance, then the probability that the value of X is far from its mean is small
- Note that X does not need to be nonnegative.

Chebyshev Example: Bandwidth Usage

- Suppose your website hosting company has an average bandwidth usage of 15TB per month with a standard deviation of 2TB per month. How can we bound the probability that the company will see a month where the actual usage deviates from mean usage by more than 5TB?

- We can use Chebyshev's Inequality as follows:
  - Let X be the bandwidth usage.
  - $\mu = 15$, $c = 5$ and $\sigma = 2$
  - We have $P(|X - 15| \geq 5) \leq \frac{2^2}{5^2} = \frac{4}{25}$

Outline

Multiple Random Variables

- Given a sample space Ω and probability rule, it is possible to define multiple random variables.
- Linearity of Expectation: If X, Y are random variables then

$$E[X+Y] = E[X] + E[Y]$$

  and this extends to more than two random variables.
- Independence: We say X and Y are independent if for all i and j,

$$P(X = i \text{ and } Y = j) = P(X = i)P(Y = j)$$

- Linearity of Variance: If X, Y are *independent* random variables then

$$var[X+Y] = var[X] + var[Y]$$

  and this extends to more than two random variables.