STOR 435.001 Lecture 7

Random Variables - I

Shankar Bhamidi

UNC Chapel Hill

Example 1a: Suppose that our experiment consists of tossing 3 fair coins. Let Y denote the number of heads that appear. So:

☼

Example 1b: Three balls are to be randomly selected without replacement from an urn containing 20 balls numbered 1 through 20. Let X be the largest ball selected. So:

☼

Definition

A random variable is a real-valued function defined on the sample space. Think of this as a "measurement" of some property of a random experiment.

Typical notation: X, Y, Z, etc.

Intuition/Rationale behind random variables

Might have a super complicated sample space (e.g. state of the universe next Monday at 9:00 a.m. Our brain too tiny to comprehend all the probabilities associated with this complicated space. However we could be interested in a small aspect ("measurement") of this sample space: e.g. price of Netflix stock next Monday at 9:00 a.m. This "measurement" is our random variable.)

Characterizing a random variable: indicate what values it takes and with what probabilities.[1]

Example 1a: Suppose that our experiment consists of tossing 3 fair coins. Let Y denote the number of heads that appear.

☼

Example 1b: Three balls are to be randomly selected without replacement from an urn containing 20 balls numbered 1 through 20. Let X be the largest ball selected.

☼

[1] Note that the values of a random variable can be regarded as outcomes in a new probability space, with probabilities assigned.

Random variables

Types of random variables

Two main types of random variables **Discrete** [mainly in Chapter 3] and **Continuous** [mainly in Chapter 5]. Treatment and analysis differ slightly.

1. For discrete, use counting arguments/summations etc.
2. For continuous, use Calculus etc.

Probability mass function (p.m.f.) [Characterizing a random variable 1]

$$p(a) = P(X = a), \qquad a \in (-\infty, \infty)$$

Discrete random variables

X assumes one of its values $x_1, x_2, \ldots$:

$$p(x_i) \geq 0, \quad i = 1, 2, \ldots, \quad p(x) = 0 \quad \text{for other } x.$$

Note: p.m.f. characterizes discrete X. Moreover: $\sum_{i=1}^{\infty} p(x_i) = 1$. Graphically:

☼

Characterizing a random variable 2:

Through the **cumulative distribution function** (c.d.f.):

$$F(x) = P(X \leq x), \quad -\infty < x < \infty.$$

Question: Why called random variables?

Discrete random variable: takes at most a countable number of possible values.

Note: the random variables in Examples 1a, 1b are discrete.

Example 1a: Suppose that our experiment consists of tossing 3 fair coins. Let Y denote the number of heads that appear. What are p.m.f. and d.f. of Y?

☼

Example: You are told that the cdf of a random variable X is given by

$$F_X(x) = \begin{cases} 0 & \text{for } x < .5 \\ \frac{1}{4} & \text{for } .5 \leq x < 1 \\ \frac{3}{8} & \text{for } 1 \leq x < 1.75 \\ \frac{1}{2} & \text{for } 1.75 \leq x < 3 \\ 1 & \text{for } 3 \leq x < \infty. \end{cases}$$

1. Draw this function.
2. Find the values this random variable takes and the corresponding probabilities.

Solution: Drawing the function

Solution: Finding the values the function takes

Definition

If X is a discrete random variable taking values x_i with probability $p(x_i)$, its expected value (or mean) is defined as

$$EX = \sum_i x_i p(x_i).$$

Motivation:

$$\begin{aligned}
EX &= \text{long run average of values of } X \text{ in } n \text{ repeated experiments} \\
&= \lim_n \frac{1}{n} \sum_{i=1}^{n} X^{(i)} \\
&= \lim_n \sum_k x_k \frac{\text{number of } x_k \text{ in } n \text{ repeated experiments}}{n} \\
&= \sum_k x_k p(x_k).
\end{aligned}$$

Example 3a: Find EX, where X is the outcome when we roll a fair six-faced die.

☼

Note: 3.5 is **not** one of the values the dice can take. So don't get confused with " expected value".

Example 3b: We say that I is an indicator variable for the event A if

$$I = \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{if } A^c \text{ occurs.} \end{cases}$$

Find EI.

☼

Example

If Y denotes the outcome of an unfair 6-faced die where

$$\mathbb{P}(Y = 1) = 1/3, \quad \mathbb{P}(Y = 2) = 1/9, \quad \mathbb{P}(Y = 3) = 1/18$$

$$\mathbb{P}(Y = 4) = 1/18, \quad \mathbb{P}(Y = 5) = 1/9, \quad \mathbb{P}(Y = 6) = 1/3$$

Calculate $\mathbb{E}(Y)$.

Random variables

If X is a discrete random variable and g is a function, then $g(X)$ is also a discrete random variable. If one wanted to compute the expected value of $g(X)$, one could do this by definition after finding the values it takes and the probabilities.

Example: Suppose X takes values -1, 0 and 1 with probabilities 0.2, 0.5 and 0.3. Let $Y = X^2$. Then, EY is:

☼

Note also that:

☼

Computing expectation of function of random variable

If X is a discrete random variable taking values x_i with probability $p(x_i)$, and g is a function, then

$$Eg(X) = \sum_i g(x_i)p(x_i).$$

Terminology: $EX =$ mean, 1st moment; $EX^n = n$th moment, $n \geq 1$.

Simple consequence: for two constants a, b,

$$E(aX + b) = \sum_i (ax_i + b)p(x_i) = a \sum_i x_i p(x_i) + b \sum_i p(x_i) = aEX + b.$$

“Newspaper vendor” problem

A salesman sees that a new smartphone, the JPhone 6X is about to be released and wonders if he should buy a few and sell them. He can buy these phones at \$100 and sell it to customers who are willing to pay \$150. However the demand is variable. Let X be the demand for the phone from this salesman then this is a random variable with distribution

$$\mathbb{P}(X = k) := \binom{10}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{10-k}, \qquad k = 0, 1, 2, \ldots, 10$$

Approximately how many phones should he purchase so as to maximize his **expected profit**?

☼

Random variables

Mean EX is a measure of center of p.m.f. It says nothing about variation or spread of the values. For example $W = 0$, $Y = \pm 1$ with probability $1/2$, and $Z = \pm 100$ with probability $1/2$, have all mean zero $EW = EY = EZ = 0$ but very different spreads.

⚠

Definition

If X is a random variable with mean μ, its variance is defined as

$$\text{Var}(X) = E(X - \mu)^2.$$

$\sqrt{\text{Var}(X)}$ is called the standard deviation of X.

Note 1: Var$(X) = E(X^2) - (EX)^2$.

Example: X is the outcome when a fair die is rolled.

Note 2: for two numbers a, b, Var$(aX + b) = a^2$Var(X).

Moment generating function

Given a random variable X, the moment generating function **(MGF)** of the random variable is defined as the function

$$M_X(t) := E(e^{tX}), \qquad t \in \mathbb{R}.$$

Why super important

Small note

In some cases might be hard to get explicit formula for the MGF.

Binomial Random variables

Bernoulli experiment: 2 possible outcomes: success (with probability p) and failure (with probability $1 - p$)

Bernoulli random variable X: takes 2 values: 1 (success) with probability p and 0 (failure) with probability $1 - p$.

Example: tossing a fair coin; $H =$ success, $T =$ failure; $p = 1/2$.

Binomial random variable X: the number of successes in n independent Bernoulli trials:

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i = 0, 1, \ldots, n.$$

Notation: $X = B(n, p)$. Bernoulli: $X = B(1, p)$.

Example: tossing a fair coin n times; $X =$ number of H's $= B(n, 1/2)$.

Where does the Binomial pmf come from?

Let us do an example: Suppose you do 3 Bernoulli trials. Probability of success (S) is $1/4$. Let $X =$ number of successes out of 3. What is

1. $P(X = 0)$?
2. $P(X = 1)$?
3. $P(X = 2)$?

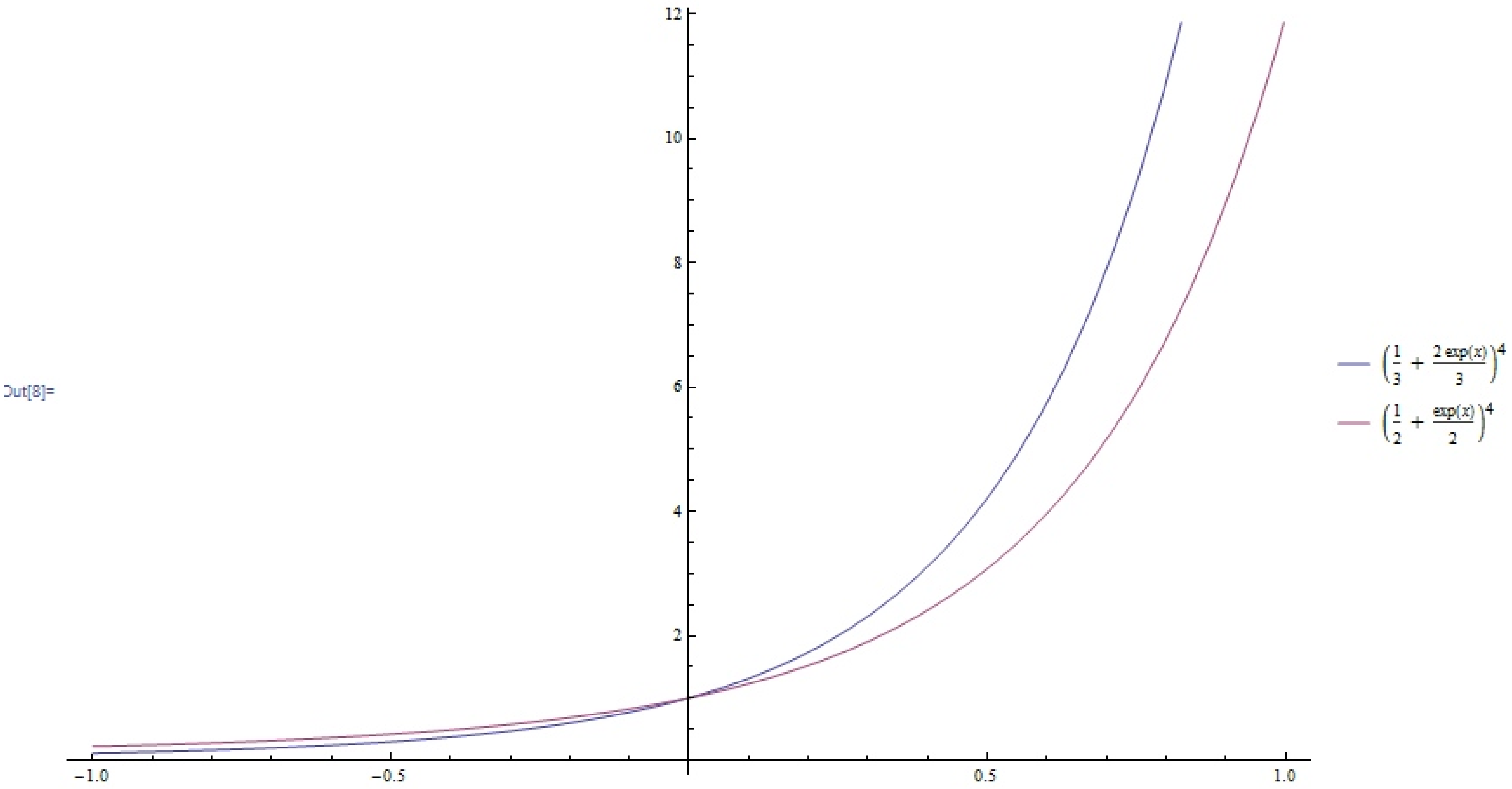
Calculations continued

☼

Some properties: If $X = B(n, p)$, what is the moment generating function of X?

☼

Examples of the MGF for two different Binomial distributions.



Some properties: If $X = B(n, p)$, then $EX = np$:

☼

Note: Book calculates this **without** using MGF, just using properties of combinations $\binom{n}{k}$
Similarly, $EX^2 = (np)^2 + np(1-p)$ and hence

$$\text{Var}(X) = np(1-p).$$

Calculating moments of Binomial without MGF

Calculating $\mathbb{E}(X)$ for Binomial

Calculating $\mathbb{E}(X^2)$ for Binomial

Problem: A big supermarket chain "Motco" sells big boxes of nuts. Each box of nuts actually had 10 small "ready to take to work" pre-packed packets of nuts. It is known that each packet will have one or more raisins with probability .05, independently of any other packet (or any other box). The company offers a money-back guarantee that at most 1 of the 10 packets in the package will have raisins. The guarantee is that the customer can return the entire box if he or she finds more than one packet with raisins in them. Shankar, being addicted to nuts, buys 5 boxes. What is the probability that he will return exactly 1 of these 5 boxes?

☼

Solution continued