Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture # 9

Discrete RVs – Part II

Last Time

Random Variable (RV)

$X : S \mapsto \mathbb{R}$

Probability Mass Function (pmf)

$p(x) = P(X = x)$

Cumulative Distribution Function (CDF)

$F(x_0) = P(X \leq x_0)$

Expected Value (aka Expectation/Mean)

$$\mu = E[X] = \sum_{\text{all } x} x \cdot p(x)$$

Today

More About Expected Value

- Random Variables and Parameters
- Expectation of Functions
- Variance & Standard Deviation
- Variance of Bernoulli RV

Binomial Random Variable

Random Variables and Parameters

Notation: $X \sim \text{Bernoulli}(p)$

Means X is a Bernoulli RV with $P(X = 1) = p$ and $P(X = 0) = 1 - p$. The tilde is read "distributes as."

Parameter

Any constant that appears in the definition of a RV, here p.

Important

Use RVs to model populations $\implies$ this definition of parameter corresponds to the one from earlier in the semester: a "feature of the population (e.g. mean)."

Constants Versus Random Variables

This is a crucial distinction that students sometimes miss:

Random Variables

- Suppose X is a RV – the values it takes on are random
- A function $g(X)$ of a RV is itself a RV as we'll learn today.

Constants

- $\mu = E[X]$ is a constant (you should convince yourself of this)
- Realizations x are constants. What is random is *which* realization the RV takes on.
- Parameters are constants (e.g. p for Bernoulli RV)
- Sample size n is a constant

The St. Petersburg Game

How Much Would You Pay?

How much would you be willing to pay for the right to play the following game?

> *Imagine a fair coin. The coin is tossed once. If it falls heads, you receive a prize of \$2 and the game stops. If not, it is tossed again. If it falls heads on the second toss, you get \$4 and the game stops. If not, it is tossed again. If it falls heads on the third toss, you get \$8 and the game stops, and so on. The game stops after the first head is thrown. If the first head is thrown on the x^{th} toss, the prize is \$$2^x$*

X = Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) =$$

X = Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	1/2	1
2	4	1/4	1
3	8	1/8	1

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) =$$

X = Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	2^n	$1/2^n$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) =$$

X = Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	2^n	$1/2^n$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) = 1 + 1 + 1 + \dots$$

X = Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	2^n	$1/2^n$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) = 1 + 1 + 1 + \ldots = \infty$$

Functions of Random Variables are Themselves Random Variables

Example: Function of Bernoulli RV

$$\boxed{\text{Let } Y = e^X \text{ where } X \sim \text{Bernoulli}(p)}$$

Support of Y

Example: Function of Bernoulli RV

$$\boxed{\text{Let } Y = e^X \text{ where } X \sim \text{Bernoulli}(p)}$$

Support of Y

$\{e^0, e^1\} = \{1, e\}$

Probability Mass Function for Y

Example: Function of Bernoulli RV

$$\boxed{\text{Let } Y = e^X \text{ where } X \sim \text{Bernoulli}(p)}$$

Support of Y

$\{e^0, e^1\} = \{1, e\}$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} p & y = e \\ 1 - p & y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Expectation: Function of Bernoulli RV

$\boxed{\text{Let } Y = e^X \text{ where } X \sim \text{Bernoulli}(p)}$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} p & y = e \\ 1 - p & y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Expectation of $Y = e^X$

Expectation: Function of Bernoulli RV

$$\boxed{\text{Let } Y = e^X \text{ where } X \sim \text{Bernoulli}(p)}$$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} p & y = e \\ 1 - p & y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Expectation of $Y = e^X$

$$\sum_{y \in \{1, e\}} y \cdot p_Y(y) =$$

Expectation: Function of Bernoulli RV

$\boxed{\text{Let } Y = e^X \text{ where } X \sim \text{Bernoulli}(p)}$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} p & y = e \\ 1 - p & y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Expectation of $Y = e^X$

$$\sum_{y \in \{1, e\}} y \cdot p_Y(y) = (1 - p) \cdot 1 + p \cdot e = 1 + p(e - 1)$$

Expectation: Function of Bernoulli RV

$$\boxed{\text{Let } Y = e^X \text{ where } X \sim \text{Bernoulli}(p)}$$

Expectation of the Function

$$\sum_{y \in \{1,e\}} y \cdot p_Y(y) = (1-p) \cdot 1 + p \cdot e = 1 + p(e-1)$$

Function of the Expectation

$$e^{E[X]} = e^p$$

$$E[g(X)] \neq g(E[X])$$

(Expected value of Function $\neq$ Function of Expected Value)

Expectation of a Function of a Discrete RV

Let X be a random variable and g be a function. Then:

$$\boxed{E[g(X)] = \sum_{\text{all } x} g(x)p(x)}$$

This is how we proceeded in the St. Petersburg Game Example

Your Turn: Calculate $E[X^2]$

X has support $\{-1, 0, 1\}$, $p(-1) = p(0) = p(1) = 1/3$.

Your Turn: Calculate $E[X^2]$

X has support $\{-1, 0, 1\}$, $p(-1) = p(0) = p(1) = 1/3$.

$$E[X^2] = \sum_{\text{all } x} x^2 p(x) = \sum_{x \in \{-1,0,1\}} x^2 p(x)$$

Your Turn: Calculate $E[X^2]$

X has support $\{-1, 0, 1\}$, $p(-1) = p(0) = p(1) = 1/3$.

$$\begin{aligned} E[X^2] &= \sum_{\text{all } x} x^2 p(x) = \sum_{x \in \{-1,0,1\}} x^2 p(x) \\ &= (-1)^2 \cdot (1/3) + (0)^2 \cdot (1/3) + (1)^2 \cdot (1/3) \end{aligned}$$

Your Turn: Calculate $E[X^2]$

X has support $\{-1, 0, 1\}$, $p(-1) = p(0) = p(1) = 1/3$.

$$
\begin{aligned}
E[X^2] &= \sum_{\text{all } x} x^2 p(x) = \sum_{x \in \{-1,0,1\}} x^2 p(x) \\
&= (-1)^2 \cdot (1/3) + (0)^2 \cdot (1/3) + (1)^2 \cdot (1/3) \\
&= 1/3 + 1/3 \\
&= 2/3 \approx 0.67
\end{aligned}
$$

Linearity of Expectation

Holds for Continuous RVs as well, but proof is different.

Let X be a RV and a, b be constants. Then:

$$E[a + bX] = a + bE[X]$$

This is one of the most important facts in the course: the special case in which $E[g(X)] = g(E[X])$ is $g = a + bX$.

Proof: Linearity of Expectation For Discrete RV

$$E[a + bX] =$$

Proof: Linearity of Expectation For Discrete RV

$$
\begin{aligned}
E[a+bX] &= \sum_{\text{all } x} (a+bx)p(x) \\
&=
\end{aligned}
$$

Proof: Linearity of Expectation For Discrete RV

$$
\begin{aligned}
E[a + bX] &= \sum_{\text{all } x} (a + bx)p(x) \\
&= \sum_{\text{all } x} p(x) \cdot a + \sum_{\text{all } x} p(x) \cdot bx \\
&=
\end{aligned}
$$

Proof: Linearity of Expectation For Discrete RV

$$
\begin{aligned}
E[a+bX] &= \sum_{\text{all } x} (a+bx)p(x) \\
&= \sum_{\text{all } x} p(x)\cdot a + \sum_{\text{all } x} p(x)\cdot bx \\
&= a\sum_{\text{all } x} p(x) + b\sum_{\text{all } x} x\cdot p(x) \\
&=
\end{aligned}
$$

Proof: Linearity of Expectation For Discrete RV

$$
\begin{aligned}
E[a + bX] &= \sum_{\text{all } x} (a + bx)p(x) \\
&= \sum_{\text{all } x} p(x) \cdot a + \sum_{\text{all } x} p(x) \cdot bx \\
&= a \sum_{\text{all } x} p(x) + b \sum_{\text{all } x} x \cdot p(x) \\
&= a + bE[X]
\end{aligned}
$$

Variance and Standard Deviation of a RV

The Defs are the same for continuous RVs, but the method of calculating will differ.

Variance (Var)

$$\sigma^2 = Var(X) = E\left[(X-\mu)^2\right] = E\left[(X - E[X])^2\right]$$

Standard Deviation (SD)

$$\sigma = \sqrt{\sigma^2} = SD(X)$$

Key Point

Variance and std. dev. are *expectations of functions of a RV*

It follows that:

1. Variance and SD are constants
2. To derive facts about them you can use the facts you know about expected value

How To Calculate Variance for Discrete RV?

Remember: it's just a function of X!

$$\text{Recall that } \mu = E[X] = \sum_{\text{all } x} x p(x)$$

$$Var(X) = E\left[(X-\mu)^2\right] = \sum_{\text{all } x} (x-\mu)^2 p(x)$$

Shortcut Formula For Variance

This is *not* the definition, it's a shortcut for doing calculations:

$$Var(X) = E\left[(X-\mu)^2\right] = E[X^2] - (E[X])^2$$

We'll prove this in an upcoming lecture.

Variance of Bernoulli RV – via the Shortcut Formula

Step 1 – $E[X]$

$$\mu = E[X] = \sum_{x \in \{0,1\}} p(x) \cdot x = (1-p) \cdot 0 + p \cdot 1 = p$$

Variance of Bernoulli RV – via the Shortcut Formula

Step 1 – $E[X]$

$$\mu = E[X] = \sum_{x \in \{0,1\}} p(x) \cdot x = (1-p) \cdot 0 + p \cdot 1 = p$$

Step 2 – $E[X^2]$

$$E[X^2] = \sum_{x \in \{0,1\}} x^2 p(x) = 0^2(1-p) + 1^2 p = p$$

Variance of Bernoulli RV – via the Shortcut Formula

Step 1 – $E[X]$

$$\mu = E[X] = \sum_{x \in \{0,1\}} p(x) \cdot x = (1-p) \cdot 0 + p \cdot 1 = p$$

Step 2 – $E[X^2]$

$$E[X^2] = \sum_{x \in \{0,1\}} x^2 p(x) = 0^2(1-p) + 1^2 p = p$$

Step 3 – Combine with Shortcut Formula

$$\sigma^2 = Var[X] = E[X^2] - (E[X])^2 = \; p - p^2 = p(1-p)$$

Variance of Bernoulli RV – Without Shortcut

You will fill in the missing steps on Problem Set 5.

$$
\begin{aligned}
\sigma^2 &= Var(X) = \sum_{x \in \{0,1\}} (x - \mu)^2 p(x) \\
&= \sum_{x \in \{0,1\}} (x - p)^2 p(x) \\
&\vdots \\
&= p(1 - p)
\end{aligned}
$$

Variance of a Linear Function

Suppose X is a random variable with $Var(X) = \sigma^2$ and a, b are constants. What is $Var(a + bX)$?

(a) σ^2

(b) $a + \sigma^2$

(c) $b\sigma^2$

(d) $a + b\sigma^2$

(e) $b^2\sigma^2$

Variance and SD are *NOT* Linear

$$
\begin{aligned}
Var(a + bX) &= b^2\sigma^2 \\
SD(a + bX) &= b\sigma
\end{aligned}
$$

These should look familiar from the related results for sample variance and std. dev. that you worked out on an earlier problem set.

Variance of a Linear Transformation

$$Var(a + bX) = E\left[\{(a + bX) - E(a + bX)\}^2\right]$$

Variance of a Linear Transformation

$$
\begin{aligned}
Var(a + bX) &= E\left[\{(a + bX) - E(a + bX)\}^2\right] \\
&= E\left[\{(a + bX) - (a + bE[X])\}^2\right]
\end{aligned}
$$

Variance of a Linear Transformation

$$\begin{aligned} Var(a+bX) &= E\left[\{(a+bX)-E(a+bX)\}^2\right] \\ &= E\left[\{(a+bX)-(a+bE[X])\}^2\right] \\ &= E\left[(bX-bE[X])^2\right] \end{aligned}$$

Variance of a Linear Transformation

$$
\begin{aligned}
Var(a + bX) &= E\left[\{(a + bX) - E(a + bX)\}^2\right] \\
&= E\left[\{(a + bX) - (a + bE[X])\}^2\right] \\
&= E\left[(bX - bE[X])^2\right] \\
&= E[b^2(X - E[X])^2]
\end{aligned}
$$

Variance of a Linear Transformation

$$\begin{aligned}
Var(a+bX) &= E\left[\{(a+bX)-E(a+bX)\}^2\right] \\
&= E\left[\{(a+bX)-(a+bE[X])\}^2\right] \\
&= E\left[(bX-bE[X])^2\right] \\
&= E[b^2(X-E[X])^2] \\
&= b^2E[(X-E[X])^2]
\end{aligned}$$

Variance of a Linear Transformation

$$
\begin{aligned}
Var(a+bX) &= E\left[\{(a+bX) - E(a+bX)\}^2\right] \\
&= E\left[\{(a+bX) - (a+bE[X])\}^2\right] \\
&= E\left[(bX - bE[X])^2\right] \\
&= E[b^2(X - E[X])^2] \\
&= b^2 E[(X - E[X])^2] \\
&= b^2 Var(X) = b^2\sigma^2
\end{aligned}
$$

The key point here is that variance is defined in terms of expectation and expectation is linear.

Binomial Random Variable

What we get if we sum a bunch of indep. Bernoulli RVs

Question

Suppose we flip a fair coin 3 times. What is the probability that we get exactly 2 heads?

Question

Suppose we flip a fair coin 3 times. What is the probability that we get exactly 2 heads?

Answer

Three basic outcomes make up this event: $\{HHT, HTH, THH\}$.

Question

Suppose we flip a fair coin 3 times. What is the probability that we get exactly 2 heads?

Answer

Three basic outcomes make up this event: $\{HHT, HTH, THH\}$. Each of these has probability $1/8 = 1/2 \times 1/2 \times 1/2$ so, since basic outcomes are mutually exclusive we sum to get $3/8 = 0.375$

A More Complicated Example

Question

Suppose we flip an *unfair* coin 3 times, where the probability of heads is $1/3$. What is the probability that we get exactly 2 heads?

Answer

The basic outcomes of the experiment are no longer equally likely, but those with exactly two heads *remain so*

$$\begin{aligned} P(HHT) &= (1/3)^2(1-1/3) = 2/27 \\ P(THH) &= 2/27 \\ P(HTH) &= 2/27 \end{aligned}$$

Summing gives $2/9 \approx 0.22$

Starting to see a pattern?

Suppose we flip an unfair coin *4* times, where the probability of heads is $1/3$. What is the probability that we get exactly 2 heads?

HHTT	TTHH
HTHT	THTH
HTTH	THHT

Six equally likely, mutually exclusive basic outcomes make up this event:

$$\binom{4}{2}(1/3)^2(2/3)^2$$

Binomial Random Variable

Let $X =$ the sum of n independent Bernoulli trials, each with probability of success p. Then we say that: $X \sim \text{Binomial}(n, p)$

Parameters

$p =$ probability of "success," $n = \#$ of trials

Support

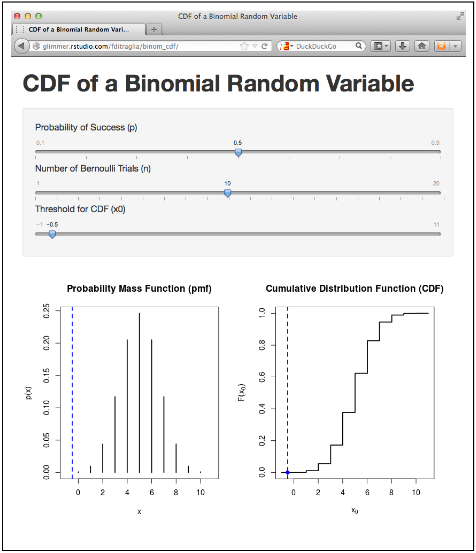
$\{0, 1, 2, \ldots, n\}$

Probability Mass Function (pmf)

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

http://glimmer.rstudio.com/fditraglia/binom_cdf/

Try playing around with all three sliders. If you set the second to 1 you get a Bernoulli.



Don't forget this!

A Binomial Random Variable counts up the *total* number of successes (ones) in n independent Bernoulli trials, each with probability of success p.

We'll learn more about the Binomial RV in the coming lectures...

http://fditraglia.github.com/Econ103Public/Rtutorials/Bernoulli_Binomial.html

Source Code on my Github Page

Getting the Binomial RV from the Bernoulli

www.ditraglia.com/econ103/Bernoulli_Binomial.html

Getting the Binomial RV from the Bernoulli

In class we defined the Binomial(n, p) random variable as the sum of n independent Bernoulli(p) random variables. In other words, the Binomial(n, p) equals the total number of successes (ones) in n independent Bernoulli trials, each with probability of success (one) equal to p. The point of this document is to convince you that this definition actually makes sense and really does lead to the formulas from class.

A complete transcript of the R source code for this document is available on my Github Page

A Function to Simulate Bernoulli Random Variables

R doesn't have a built-in function to simulate Bernoulli Random Variables since it treats them as a Binomial$(n = 1, p)$ random variable. Let's make our own:

```
# This function simulates n independent, identically distributed
Bernoulli
# Random Variables
rbern <- function(n, p) {

    sims <- sample(0:1, size = n, replace = TRUE, prob = c(1 - p, p))
    return(sims)

}
```

The argument `prob = c(1-p, p)` is new. It tells `sample` to draw 0 with probability `1-p` and 1 with probability `p`. This is exactly what we want. Let's test it out:

```
rbern(30, 0.1)
```

```
##  [1] 1 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
```