Discrete Random Variables Notes

Lt Col Ken Horton Lt Col Kris Pruitt Professor Bradley Warner

14 September, 2020

Objectives

1) Define and use properly in context all new terminology.
2) Given a discrete random variable, obtain the pmf and cdf, and use them to obtain probabilities of events.
3) Simulate random variable for a discrete distribution.
4) Find the moments of a discrete random variable.
5) Find the expected value of a linear transformation of a random variable.

Random variables

We have already discussed random experiments. We have also discussed S, the sample space for an experiment. A random variable essentially maps the events in the sample space to the real number line. For a formal definition: A random variable X is a function $X : S \to \mathbb{R}$ that assigns exactly one number to each outcome in an experiment.

> *Example*:
> Suppose you flip a coin three times. The sample space, S, of this experiment is

$$S = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}, \text{THH}, \text{THT}, \text{TTH}, \text{TTT}\}$$

Let the random variable X be the number of heads in three coin flips. Whenever introduced to a new random variable, you should take a moment to think about what possible values can X take? When tossing a coin 3 times, we can get no heads, one head, two heads or three heads. The random variable X assigns each outcome in our experiment to one of these values. Visually:

$$S = \{\underbrace{\text{HHH}}_{X=3}, \underbrace{\text{HHT}}_{X=2}, \underbrace{\text{HTH}}_{X=2}, \underbrace{\text{HTT}}_{X=1}, \underbrace{\text{THH}}_{X=2}, \underbrace{\text{THT}}_{X=1}, \underbrace{\text{TTH}}_{X=1}, \underbrace{\text{TTT}}_{X=0}\}$$

The sample space of X, the support, is the list of numerical values that X can take.

$$S_X = \{0, 1, 2, 3\}$$

Because the sample space of X is a countable list of numbers, we consider X to be a *discrete* random variable (more on that later).

How does this help?

Sticking with our example, we can now frame a problem of interest in the context of our random variable X. For example, suppose we wanted to know the probability of at least two heads. Without our random variable, we have to write this as:

$$\mathrm{P}(\text{at least two heads}) = \mathrm{P}(\{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}\})$$

In the context of our random variable, this simply becomes $\mathrm{P}(X \geq 2)$. It may not seem important in a case like this, but imagine if we were flipping a coin 50 times and wanted to know the probability of obtaining at least 30 heads. It would be unfeasible to write out all possible ways to obtain at least 30 heads. It is much easier to write $\mathrm{P}(X \geq 30)$ and explore the distribution of X (more on that in Lesson 6).

Essentially, a random variable often helps us reduce a complex random experiment to a simple variable that is easy to characterize.

Discrete vs Continuous

A *discrete* random variable has a sample space that consists of a countable set of values. X in our example above is a discrete random variable. Note that "countable" does not necessarily mean "finite". For example, a random variable with a Poisson distribution (a topic for a later lesson) has a sample space of $\{0, 1, 2, ...\}$. This sample space stretches to infinity, but it is considered *countably* infinite, and thus the random variable would be considered discrete.

A *continuous* random variable has a sample space that is a continuous interval. For example, let Y be the random variable corresponding to the height of a randomly selected individual. Y is a continuous random variable because a person could measure 68.1 inches, 68.2 inches, or perhaps any value in between. Note that when we measure height, our precision is limited by our measuring device, so we are technically "discretizing" height. However, even in these cases, we typically consider height to be a continuous random variable.

A *mixed* random variable is exactly what it sounds like. It has a sample space that is both discrete and continuous. How could such a thing occur? Consider an experiment where a person rolls a standard six-sided die. If it lands on anything other than one, the result of the die roll is recorded. If it lands on one, the person spins a wheel, and the angle in degrees of the resulting spin, divided by 360, is recorded. If our random variable Z is the number that is recorded in this experiment, the sample space of Z is $[0, 1] \cup \{2, 3, 4, 5, 6\}$. We will not be spending much time on mixed random variables. However they do occur in practice, consider the job of analyzing bomb data error. If the bomb hits within a certain radius, the error is 0. Otherwise it is measured in a radial direction. This data is mixed.

Discrete distribution functions

Once we have defined a random variable, we need a way to describe its behavior and we will use probabilities for this purpose.

Distribution functions describe the behavior of random variables. We can use these functions to determine the probability that a random variable takes a value or a range of values. For discrete random variables, there are two distribution functions of interest: the *probability mass function* (pmf) and the *cumulative distribution function* (cdf).

Probability mass function

Let X be a discrete random variable. The probability mass function (pmf) of X, given by $f_X(x)$ is a function that assigns probability to each possible outcome of X.

$$f_X(x) = \mathrm{P}(X = x)$$

Note that the pmf is a *function*. Functions have input and output. The input of a pmf is any real number. The output of a pmf is the probability that the random variable takes the inputted value. The pmf must follow the axioms of probability described in the Probability Rules lesson. Primarily,

1) For all $x \in \mathbb{R}$, $0 \leq f_X(x) \leq 1$.
2) $\sum_x f_X(x) = 1$, where the x in the index of the sum simply denotes that we are summing across the entire domain or support of X.

   Example:
   Recall our example again. You flip a coin three times and let X be the number of heads in those three coin flips. We know that X can only take values 0, 1, 2 or 3. But at what probability does it take these three values? In that example, we had listed out the possible outcomes of the experiment and denoted what value of X corresponds to each outcome.

$$S = \{\underbrace{\text{HHH}}_{X=3}, \underbrace{\text{HHT}}_{X=2}, \underbrace{\text{HTH}}_{X=2}, \underbrace{\text{HTT}}_{X=1}, \underbrace{\text{THH}}_{X=2}, \underbrace{\text{THT}}_{X=1}, \underbrace{\text{TTH}}_{X=1}, \underbrace{\text{TTT}}_{X=0}\}$$

Each of these eight outcomes is equally likely (each with a probability of $\frac{1}{8}$). Thus, building the pmf of X becomes a matter of counting the number of outcomes associated with each possible value of X:

$$f_X(x) = \begin{cases} \frac{1}{8}, & x = 0 \\ \frac{3}{8}, & x = 1 \\ \frac{3}{8}, & x = 2 \\ \frac{1}{8}, & x = 3 \\ 0, & \text{otherwise} \end{cases}$$

Note that this function specifies the probability that X takes any of the four values in the sample space (0, 1, 2, and 3). Also, it specifies that the probability that X takes any other value is 0.

Graphically, the pmf is not terribly interesting. The pmf is 0 at all values of X except for 0, 1, 2 and 3.

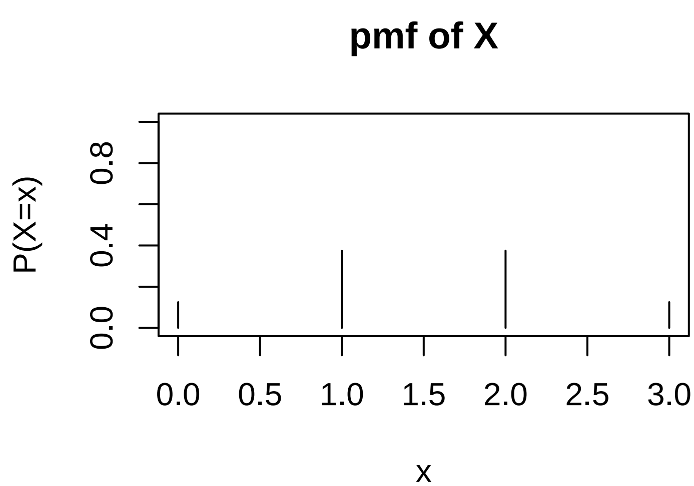

Figure 1: Probability Mass Function of X from Coin Flip Example

Example:
We can use a pmf to answer questions about an experiment. For example, consider the same context. What is the probability that we flip at least one heads? We can write this in the context of X:

$$\mathrm{P}(\text{at least one heads}) = \mathrm{P}(X \geq 1) = \mathrm{P}(X = 1) + \mathrm{P}(X = 2) + \mathrm{P}(X = 3) = \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8}$$

Alternatively, we can recognize that $\mathrm{P}(X \geq 1) = 1 - \mathrm{P}(X = 0) = 1 - \frac{1}{8} = \frac{7}{8}$.

Cumulative distribution function

Let X be a discrete random variable. The cumulative distribution function (cdf) of X, given by $F_X(x)$ is a function that assigns to each value of X the probability that X takes that value or lower:

$$F_X(x) = \mathrm{P}(X \leq x)$$

Again, note that the cdf is a *function* with an input and output. The input of a cdf is any real number. The output of a cdf is the probability that the random variable takes the inputted value *or less*.

If we know the pmf, we can obtain the cdf:

$$F_X(x) = \mathrm{P}(X \leq x) = \sum_{y \leq x} f_X(y)$$

Like the pmf, the cdf must be between 0 and 1. Also, since the pmf is always non-negative, the cdf must be non-decreasing.

Example:
Obtain and plot the cdf of X of the previous example.

$$F_X(x) = \mathrm{P}(X \leq x) = \begin{cases} 0, & x < 0 \\ \frac{1}{8}, & 0 \leq x < 1 \\ \frac{4}{8}, & 1 \leq x < 2 \\ \frac{7}{8}, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

Visually, the cdf of a discrete random variable has a stairstep appearance. In this example, the cdf takes a value 0 up until $X = 0$, at which point the cdf increases to 1/8. It stays at this value until $X = 1$, and so on. At and beyond $X = 3$, the cdf is equal to 1. See the figure below, Figure 2.

Simulating random variables

We can simulate values from a random variable using the cdf, we will use a similar idea for continuous random variables. Since the range of the cdf is in the interval $[0, 1]$ we will generate a random number in that same interval and then use the inverse function to find the value of the random variable. The pseudo code is:
1) Generate a random number, U.
2) Find the index k such that $\sum_{j=1}^{k-1} x_j \leq U < \sum_{j=1}^{k} x_j$ or $F_x(k-1) \leq U < F_x(k)$.

Example:
Simulate a random variable for the number of heads in flipping a coin three times.

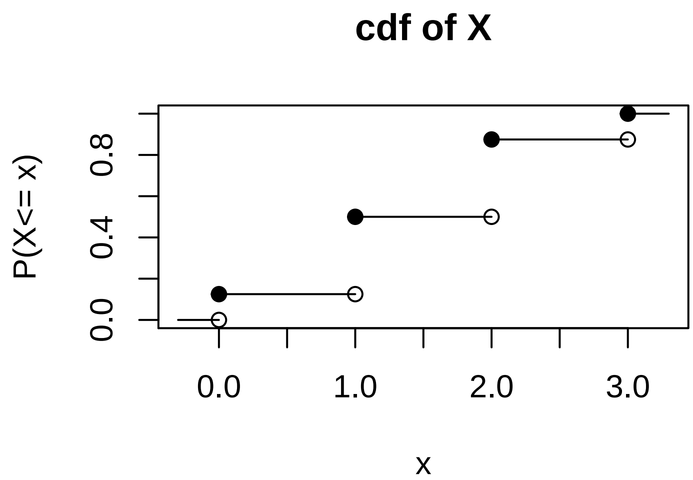

Figure 2: Cumulative Distribution Function of X from Coin Flip Example

First we will create the pmf.

```
pmf <- c(1/8,3/8,3/8,1/8)
values <- c(0,1,2,3)
pmf
```

```
## [1] 0.125 0.375 0.375 0.125
```

We get the cdf from the cumulative sum.

```
cdf <- cumsum(pmf)
cdf
```

```
## [1] 0.125 0.500 0.875 1.000
```

Next, we will generate a random number between 0 and 1.

```
set.seed(1153)
ran_num <- runif(1)
ran_num
```

```
## [1] 0.7381891
```

Finally, we will find the value of the random variable. We will do each step separately first so you can understand the code.

```
ran_num < cdf
```

```
## [1] FALSE FALSE  TRUE  TRUE
```

```
which(ran_num < cdf)
```

```
## [1] 3 4
```

```
which(ran_num < cdf)[1]
```

```
## [1] 3
```

```
values[which(ran_num < cdf)[1]]
```

```
## [1] 2
```

Let's make this a function.

```
simple_rv <- function(values,cdf){
ran_num <- runif(1)
return(values[which(ran_num < cdf)[1]])
}
```

Now let's generate 10000 values from this random variable.

```
results <- do(10000)*simple_rv(values,cdf)
inspect(results)
```

```
##
## quantitative variables:
##           name   class min Q1 median Q3 max   mean       sd     n missing
## ...1 simple_rv numeric   0  1      2  2   3 1.5048 0.860727 10000       0
```

```
tally(~simple_rv,data=results,format="proportion")
```

```
## simple_rv
##      0      1      2      3
## 0.1207 0.3785 0.3761 0.1247
```

Not a bad approximation.

Moments

Distribution functions are excellent characterizations of random variables. The pmf and cdf will tell you exactly how often the random variables takes particular values. However, distribution functions are often a lot of information. Sometimes, we may want to describe a random variable X with a single value or small set of values. For example, we may want to know the average or some measure of center of X. We also may want to know a measure of spread of X. *Moments* are values that summarize random variables with single numbers. Since we are dealing with the population, these moments are population vales and not summary statistics as we used in the first block of material.

Expectation

At this point, we should define the term *expectation*. Let $g(X)$ be some function of a discrete random variable X. The expected value of $g(X)$ is given by:

$$\mathrm{E}(g(X)) = \sum_x g(x) \cdot f_X(x)$$

Mean

The most common moments used to describe random variables are *mean* and *variance*. The mean (often referred to as the expected value of X), is simply the average value of a random variable. It is denoted as μ_X or $\mathrm{E}(X)$. In the discrete case, the mean is found by:

$$\mu_X = \mathrm{E}(X) = \sum_x x \cdot f_X(x)$$

The mean is also known as the first moment of X around the origin. It is a weighted sum with the weight being the probability. If each outcome were equally likely, the expected value would just be the average of the values of the random variable since each wight is the reciprocal of the number of values.

> *Example*:
> Find the expected value (or mean) of X: the number of heads in three flips of a fair coin.
>
> $$\mathrm{E}(X) = \sum_x x \cdot f_X(x) = 0 * \frac{1}{8} + 1 * \frac{3}{8} + 2 * \frac{3}{8} + 3 * \frac{1}{8} = 1.5$$

We are using μ because it is a population parameter.

From our simulation above, we can find the mean as an estimate of the expected value. This is really a statistic since our simulation is data from the population and thus will have variance from sample to sample.

```
mean(~simple_rv,data=results)
```

```
## [1] 1.5048
```

Variance

Variance is a measure of spread of a random variable. The variance of X is denoted as σ_X^2 or $\mathrm{Var}(X)$. It is equivalent to the average squared deviation from the mean:

$$\sigma_X^2 = \mathrm{Var}(X) = \mathrm{E}[(X - \mu_X)^2]$$

In the discrete case, this can be evaluated by:

$$\mathrm{E}[(X - \mu_X)^2] = \sum_x (x - \mu_X)^2 f_X(x)$$

Variance is also known as the second moment of X around the mean.

The square root of $\mathrm{Var}(X)$ is denoted as σ_X, the standard deviation of X. The standard deviation is often reported because it is measured in the same units as X, while the variance is measured in squared units and is thus harder to interpret.

Example:
Find the variance of X: the number of heads in three flips of a fair coin.

$$\text{Var}(X) = \sum_x (x-\mu_X)^2 \cdot f_X(x) = (0-1.5)^2 \times \frac{1}{8} + (1-1.5)^2 \times \frac{3}{8} + (2-1.5)^2 \times \frac{3}{8} + (3-1.5)^2 \times \frac{1}{8}$$

```
(0-1.5)^2*1/8 + (1-1.5)^2*3/8 + (2-1.5)^2*3/8 + (3-1.5)^2*1/8
```

```
## [1] 0.75
```

The variance of X is 0.75.

We can find the variance of the simulation but `R` uses the sample variance and this is the population variance. So we need to multiply by $\frac{n-1}{n}$

```
var(~simple_rv,data=results)*(10000-1)/10000
```

```
## [1] 0.740777
```

Mean and variance of Linear Transformations

Lemma: Let X be a discrete random variable, and let a and b be constants. Then:

$$\text{E}(aX+b) = a\text{E}(X)+b$$

and

$$\text{Var}(aX+b) = a^2\text{Var}(X)$$

The proof of this is left as an Application problem.

File Creation Information

- File creation date: 2020-09-14
- Windows version: Windows 10 x64 (build 18362)
- R version 3.6.3 (2020-02-29)
- `mosaic` package version: 1.7.0
- `tidyverse` package version: 1.3.0