STAT 305: Chapter 5

Part II

Amin Shirazi

Course page:
ashirazist.github.io/stat305.github.io

Discrete Random Variables

Meaning, Use, and Common Distributions

General Info

Reminder: RVs

General Info About Discrete RVs

Reminder: What is a Random Variable?

Random Variables, we have already defined, take real-numbered ($\mathbb{R}$) values based on outcomes of a random experiment.

- If we know the outcome, we know the value of the random variable (so that isn't the random part).
- However, before we perform the experiment we do not know the outcome - we can only make statements about what the outcome is likely to be (i.e., we make "probabilistic" statements).
- In the same way, we do not know the value of the random variable before the experiment, but we can make probability statements about what value the RV might take.

General Info

Reminder: RVs

Discrete?

Terms & Notation

Common Terms and Notation for Discrete RVs

Of course, we can't introduce a *sort of* new concept without introducing a whole lot of new terminology.

We use capital letters to refer to discrete random variables: X, Y, Z, ...

We use lower case letters to refer to values the discrete RVs can take: x, x_1, y, z, ...

While we can use $P(X = x)$ to refer to the probability that the discrete random variable takes the value x, we usually use what we call the **probability function**:

- For a discrete random variable X, the probability function $f(x)$ takes the value $P(X = x)$
- In otherwords, we just write $f(x)$ instead of $P(X = x)$.

General Info

Reminder: RVs

Discrete?

Terms & Notation

Common Terms and Notation for Discrete RVs

We also have another function related to the probabilities, called the **cumulative probability function**.

For a discrete random variable X taking values $x_1, x_2, \ldots$ the CDF or **cumulative probability function** of X, $F(x)$, is defined as

$$F(x) = \sum_{z \leq x} f(z)$$

Which in other words means that for any value x,

$$f(x) = P(X = x)$$

and

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

Common Terms and Notation for Discrete RVs (cont)

The values that X can take and the probabilities attached to those values are called the **probability distribution** of X (since we are talking about how the total probability 1 gets spread out on (or distributed to) the values that X can take).

Example

Suppose that the we roll a die and let T be the number of dots facing up. Define the probability distribution of T. Find $f(3)$ and $F(6)$.

General Info

Reminder: RVs

Discrete?

Terms & Notation

Example: [Torque]

Let $Z =$ the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

Calculate the following probabilities:

- $P(Z \leq 14)$
- $P(Z > 16)$
- $P(Z \text{ is even})$
- $P(Z \in \{15, 16, 18\})$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Example: [Torque]

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

- $P(Z \leq 14)$

- $P(Z > 16)$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Example: [Torque]

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

- $P(Z \text{ is even})$

- $P(Z \in \{15, 16, 18\})$

General Info

Reminder: RVs

Discrete?

Terms & Notation

More on CDF

> The *cumulative probability distribution (cdf)* for a random variable X is a function $F(x)$ that for each number x gives the probability that X takes that value or a smaller one, $F(x) = P[X \leq x]$.

Since (for discrete distributions) probabilities are calculated by summing values of $f(x)$,

$$F(x) = P[X \leq x] = \sum_{y \leq x} f(y)$$

More on CDF

Properties of a mathematically valid cumulative distribution function:

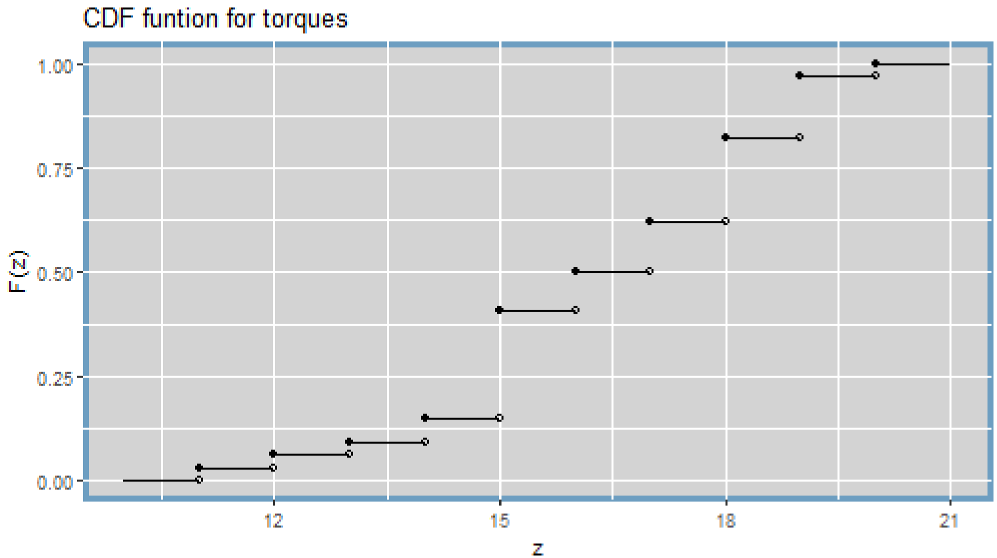
- $F(x) \geq 0$ for all real numbers x
- $F(x)$ is monotonically **increasing**
- $F(x)$ is right continuous
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$
    - This means that $0 \leq F(x) \leq 1$ for **any CDF**

> In the discrete cases, the graph of $F(x)$ will be n stair-step graph with jumps at possible values of our random variable and height equal to the probabilities associated with those values

More on CDF

Example: [Torque] Let $Z =$ the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03



General Info

Reminder: RVs

Discrete?

Terms & Notation

More on CDF

Calculate the following probabilities using the **cdf only**:

- $F(10.7)$
- $P(Z \leq 15.5)$
- $P(12.1 < Z \leq 14)$
- $P(15 \leq Z < 18)$

More on CDF

One more example

Say we have a random variable Q with pmf:

q	f(q)
1	0.34
2	0.10
3	0.22
7	0.34

Draw the CDF

General Info

Reminder: RVs

Discrete?

Terms & Notation

Mean of a Discrete Random Variable

For a discrete random varable, X, which can take values $x_1, x_2, \ldots$ we define **the mean of X** (also known as **the expected value of X**) as:

$$E(X) = \sum_{i=1}^{n} x_i \cdot f(x_i)$$

We often use the symbol μ instead of $E(X)$.

Also, just to be confusing, you will often see EX instead of $E(X)$. Use context clues.

Example:

Suppose that the we roll a die and let T be the number of dots facing up. Find the expected value of T.

General
Info

Reminder:
RVs

Discrete?

Terms &
Notation

Variance of a Discrete Random Variable

For a discrete random varable, X, which can take values $x_1, x_2, \ldots$ and has mean μ we define **the variance of X** as:

$$Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot f(x_i)$$

There are other usefule ways to write this, most importantly:

$$Var(X) = \sum_{i=1}^{n} x_i^2 \cdot f(x_i) - \mu^2$$

which is the same as

$$\mathrm{Var} X = \sum_x (x - \mathrm{E}X)^2 f(x) = \sum_x x^2 f(x) - (\mathrm{E}X)^2.$$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Variance of a Discrete Random Variable

Example:

Suppose that the we roll a die and let T be the number of dots facing up. What is the variance of T?

Variance of a Discrete Random Variable

Example

Say we have a random variable Q with pmf:

q	f(q)
1	0.34
2	0.10
3	0.22
7	0.34

Find the variance and standard deviation

General Info

Reminder: RVs

Discrete?

Terms & Notation

Summary

Discrete Random Variables

- Discrete RVs are RVs that will take one of a countable set of values.
- When working with a discrete random variable, it is common to need or use the RV's
  - probability distribution: the values the RV can take and their probabilities
  - probability function: a function where $f(x) = P(X = x)$
  - cumulative probability function: a function where $F(x) = P(X \leq x)$.
  - mean: a value for X defined by $EX = \sum_x x \cdot f(x)$
  - variance: a value for X defined by $VarX = \sum_x (x - \mu)^2 \cdot f(x)$

Common Distributions

Working with Off The Shelf Random Variables

General Info

Common Distributions

Background

Common Distributions

Why Are Some Distributions Worth Naming?

Even though you may create a random variable in a unique scenario, the way that it's probability distribution behaves (mathematically) may have a lot in common with other random variables in other scenarios. For instance,

> I roll a die until I see a 6 appear and then stop. I call X the number of times I have to roll the die in total.
>
> I flip a coin until I see heads appear and then stop. I call Y the number of times I have to flip the coin in total.
>
> I apply for home loans until I get accepted and then I stop. I call Z the number of times I have to apply for a loan in total.

Why Are Some Distributions Worth Naming? (cont)

In each ot the above cases, we count the number of times we have to do some action until we see some specific result. The only thing that really changes from the random variables perspective is the likelyhood that we see the specific result each time we try.

Mathematically, that's not a lot of difference. And if we can really understand the probability behavior of one of these scenarios then we can move our understanding to the different scenario pretty easily.

By recognizing the commonality between these scenarios, we have been able to identify many random variables that behave very similarly. We describe the similarity in the way the random variables behave by saying that they have a common/shared distribution.

We study the most useful ones by themselves.

The Bernoulli Distribution

The Bernoulli Distribution

Origin: A random experiment is performed that results in one of two possible outcomes: success or failure. The probability of a successful outcome is p.

Definition: X takes the value 1 if the outcome is a success. X takes the value 0 if the outcome is a failure.

probability function:

$$f(x) = \begin{cases} p & x = 1, \\ 1-p & x = 0, \\ 0 & o.w. \end{cases}$$

which can also be written as

$$f(x) = \begin{cases} p^x(1-p)^{1-x} & x = 0, 1 \\ 0 & o.w. \end{cases}$$

Bernoulli Distribution

Expected Value and Variance

The Bernoulli Distribution

Expected value: $E(X) = p$

General Info

Common Distributions

Background

Bernoulli

The Bernoulli Distribution

Variance: $Var(X) = (1 - p) \cdot p$

The Bernoulli Distribution

A few useful notes:

- In order to say that " X has a bernoulli distribution with success probability p " we write $X \sim Bernoulli(p)$
- Trials which results in which the only possible outcomes are "success" or "failure" are called **Bernoulli Trials**
- The value p is the Bernoulli distribution's **parameter**. We don't treat parameters like random values - they are fixed, related to the real process we are studying.
- "Success" does not mean something we would perceive as "good" has happened. It just means the outcome we were watching for was the outcome we got.
- Please note: we have two outcomes, but the probability for each outcome is **not** the same (duh!).

The Binomial Distribution

The Binomial Distribution

Origin: A series of n independent random experiments, or trials, are performed. Each trial results in one of two possible outcomes: successful or failure. The probability of a successful outcome, p, is the same across all trials.

Definition: For n trials, X is the number of trials with a successful outcome. X can take values $0, 1, \ldots, n$.

probability function:

With $0 < p < 1$,

$$f(x) = \begin{cases} \dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & x = 0, 1, \ldots, n \\ 0 & o.w. \end{cases}$$

`

where $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$ and $0! = 1$.

The Binomial Distribution

Example [10 component machine]

Suppose you have a machine with 10 independent components in series. The machine only works if all the components work. Each component succeeds with probability $p = 0.95$ and fails with probability $1 - p = 0.05$.

Let Y be the number of components that succeed in a given run of the machine. Then

$$Y \sim \text{Binomial}(n = 10, p = 0.95)$$

Question: what is the probability of the machine working properly?

The Binomial Distribution

Example [10 component machine]

$$Y \sim \text{Binomial}(n = 10, p = 0.95)$$

What if I arrange these 10 components in parallel? This machine succeeds if at least 1 of the components succeeds.

What is the probability that the new machine succeeds?

Binomial Distribution

Expected Value and Variance

The Binomial Distribution

Expected value: $E(X) = n \cdot p$

The Binomial Distribution

Variance: $Var(X) = n \cdot (1 - p) \cdot p$

The Binomial Distribution

Example [10 component machine]

Calculate the expected number of components to succeed and the variance.

The Binomial Distribution

A few useful notes:

- In order to say that " X has a binomial distribution with n trials and success probability p" we write $X \sim Binomial(n, p)$
- If $X_1, X_2, \ldots, X_n$ are n independent Bernoulli random variables with the same p then $X = X_1 + X_2 + \ldots + X_n$ is a binomial random variable with n trials and success probability p.
- Again, n and p are referred to as "parameters" for the Binomial distribution. Both are considered fixed.
- Don't focus on the actual way we got the expected value - focus on the trick of trying to get part of your complicated summation to "go away" by turning it into the sum of a probability function.

The Geometric Distribution

The Geometric Distribution

Origin: A series of independent random experiments, or trials, are performed. Each trial results in one of two possible outcomes: successful or failure. The probability of a successful outcome, p, is the same across all trials. The trials are performed until a successful outcome is observed.

Definition: X is the trial upon which the first successful outcome is observed. X can take values $1, 2, \ldots$.

probability function:

With $0 < p < 1$,

$$f(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \ldots \\ 0 & o.w. \end{cases}$$

The Geometric Distribution

Cumulative probability function: $ F(x) = 1 - (1 - p)^x $

Here's how we get that cumulative probability function:

- The probability of a failed trial is $1 - p$.
- The probability the first trial fails is also just $1 - p$.
- The probability that the first two trials both fail is $(1 - p) \cdot (1 - p) = (1 - p)^2$.
- The probability that the first x trials all fail is $(1 - p)^x$.
- This gets us to this math:

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

$$= 1 - P(X > x)$$

$$= 1 - (1 - p)^x$$

The Geometric Distribution

Expected value: $E(X) = \dfrac{1}{p}$

Variance: $Var(X) = \dfrac{1-p}{p^2}$

Common Distributions

Background

Bernoulli

Binomial

Geometric

Example

A shipment of 200 widgets arrives from a new widget distributor. The distributor has claimed that the widgets there is only a 10% defective rate on the widgets.

- How many widgets would you expect to test before finding the first defective widget?

You find your first defective widget while testing the thrid widget.

- What is the probability that a the first defective widget would be found **on** the third test if there are only 10% defective widgets from in the shipment?
- What is the probability that a the first defective widget would be found **by** the third test if there are only 10% defective widgets from in the shipment?
- Is it unusual to find the first defective widget on the third test? What is value of p makes finding the first defective widget **by** the third test the least unusual?

The Poisson Distribution

The Poisson Distribution

Origin: A rare occurance is watched for over a specified interval of time or space.

Definition: X is the number of times the rare occurance is observed.

probability function:

For $\lambda > 0$ $$ f(x) = \begin{cases} \dfrac{e^{-\lambda} \lambda^x}{x!} & x = 0, 1, ... \\ 0 & o.w. \end{cases}$$ This distribution comes from letting(n \rightarrow \infty)and(n \cdot p \rightarrow \lambda)at the same time (in other words, our(n)gets large, but our(p)gets small). For very large(n)and very small(p)`, the binomial probability function is very close to the poisson.

Common Distributions

Background

Bernoulli

Binomial

Geometric

Poisson

The Poisson Distribution

Cumulative probability function: $ F(x) = 1 - (1 - p)^x $

Here's how we get that cumulative probability function:

- The probability of a failed trial is $1 - p$.
- The probability the first trial fails is also just $1 - p$.
- The probability that the first two trials both fail is $(1 - p) \cdot (1 - p) = (1 - p)^2$.
- The probability that the first x trials all fail is $(1 - p)^x$
.
- This gets us to this math:

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

$$= 1 - P(X > x)$$

$$= 1 - (1 - p)^x$$

The Poisson Distribution

Expected value: $E(X) = \lambda$

Variance: $Var(X) = \lambda$

Example

My last slide set contained 912 words and 4 misspellings.

Using this information, describe a possible model for the number of typos in my current slide deck, which has 1,205 words.