STAT 230 - Probability

Cameron Roopnarine

Last updated: November 18, 2019

Contents

1 Lecture 13 **4**
- 1.1 Summary . . . 4
- 1.2 Negative Binomial Distribution (5.5) . . . 4
- 1.3 Example . . . 4
- 1.4 Example . . . 4
- 1.5 Range and Probability Function of the Negative Binomial Distribution . . . 5
- 1.6 Example . . . 5

2 Lecture 14 **5**
- 2.1 Summary . . . 5
- 2.2 Example . . . 6
- 2.3 Geometric Distribution (5.6) . . . 6
- 2.4 Range and Probability Function of the Geometric Distribution . . . 6

3 Lecture 15 **7**
- 3.1 Summary . . . 7
- 3.2 Example . . . 8
- 3.3 Example . . . 8
- 3.4 Example . . . 8
- 3.5 Example . . . 8
- 3.6 Poisson Distribution from Binomial (5.7) . . . 8
- 3.7 Example . . . 9
- 3.8 Poisson Distribution from Poisson Process (5.8) . . . 9

4 Lecture 16 **9**
- 4.1 Summary . . . 9
- 4.2 Example . . . 10
- 4.3 Combining Other Models with the Poisson Process (5.9) . . . 10
- 4.4 Example (Continued) . . . 10
- 4.5 Summarizing Data on Random Variables (7.1) . . . 11
  - 4.5.1 Definition (Median) . . . 11
  - 4.5.2 Definition (Mode) . . . 11

5 Lecture 17* **11**
- 5.1 Summary . . . 11
- 5.2 Expectation of a Random Variable (7.2) . . . 12
  - 5.2.1 Definition (Expected Value) . . . 13
- 5.3 Example . . . 13
  - 5.3.1 Theorem . . . 13
- 5.4 Example . . . 13
- 5.5 Example . . . 14

6 Lecture 18 **14**
6.1 Means and Variances of Distributions (7.4) . . . 14
6.1.1 Definition (Variance) . . . 14
6.2 Example . . . 14
6.3 Example (Roulette) . . . 15
6.3.1 Definition (Standard Deviation) . . . 15
6.4 Linear Transformations . . . 15

7 Lecture 19* **16**
7.1 Summary . . . 16
7.2 Example . . . 16

8 Lecture 20* **18**
8.1 Summary . . . 18
8.2 Example . . . 19

9 Lecture 21 **19**
9.1 Summary . . . 19
9.2 Example . . . 20
9.2.1 Definition (Probability Density Function) . . . 20
9.3 Example . . . 21
9.3.1 Definition (Percentiles) . . . 22

10 Lecture 22 **22**
10.1 Example . . . 22
10.2 Continuous Uniform Distribution (8.2) . . . 22
10.3 Change of Variables . . . 23
10.4 Example (Change of Variable) . . . 23

11 Lecture 23 **24**
11.1 Example . . . 24
11.2 Exponential Distribution (8.3) . . . 24

12 Lecture 24 **25**
12.1 Memoryless Property . . . 26
12.2 Example . . . 26
12.3 Normal Distribution (8.5) . . . 27
12.4 Empirical rule . . . 28

13 Lecture 25 **28**
13.1 Example . . . 29
13.2 Example . . . 30
13.3 Example . . . 30

14 Lecture 26 **31**
14.1 Example . . . 31
14.2 Basic Terminology and Techniques (9.1) . . . 31
14.3 Example . . . 31

15 Lecture 27 **32**
15.1 Thought Question . . . 33
15.2 Example . . . 33
15.3 Definition (Conditional pf) . . . 34
15.4 Example . . . 34
15.5 Example . . . 34

15.6 Functions of 2 or more random variables 34

16 Lecture 28 **35**
16.1 Thought Question 35
16.2 Sums of rvs 35
16.3 Multinomial Distribution (9.2) 36
16.4 Example 36

1 Lecture 13

1.1 Summary

Today we started with our SWAG on counting cards in Blackjack (more info on Learn), which is a great application of the Hypergeometric distribution (and why, if you were to play with an infinite number of decks, it would be a Binomial.)

Then we defined the Negative Binomial distribution. It is also based on Bernoulli trials, but instead of doing n trials and counting the S's (as with Binomial), we do trials until we obtain k Successes and count how many Failures occurred along the way. You can usually recognize a Negative Binomial situation if we are waiting until something occurs. The range is all non-negative integers, and we found its pf is $f(x) = \binom{x+k-1}{k-1} p^x (1-p)^x$. It's sometimes tricky to tell the difference between Binomial and Negative Binomial, but the key is that the very last trial must be a Success with NB, or else we would have stopped doing trials sooner.

We looked at an example of deriving the distribution of the total number of trials using the Negative Binomial for the number of Fails.

Unfortunately there is no nice closed-form expression for the cumulative distribution function $F(x)$ for either the Hypergeometric, Binomial, or Negative Binomial random variables. If you wanted it, you would just have to add up the $f(x)$ values.

Good luck on the midterm tomorrow!

Recall

Binomial approximation to Hypergeometric distribution: If $X \sim \mathrm{Hyp}(N, r, n)$, we can approximate it with $\mathrm{Bin}(n, \frac{r}{N})$ if n is a small proportion of N.

1.2 Negative Binomial Distribution (5.5)

Setup: Bernoulli Trials

- independent
- each trial is a success or fail (S or F)
- $P(\text{success}) = p = \text{constant}$

Suppose we want to get k S's. We do trials until we get k S's and let $X = \#$ of F's. We get

$$X \sim \mathrm{NB}(k, p)$$

in a total of $k + X$ trials.

Binomial	Negative Binomial
know # of trials	unknown # of trials
unknown # of S's	known # of S's
$\binom{n}{x} p^x (1-p)^{n-x}$	$\binom{x+k-1}{k-1} p^k (1-p)^x$

1.3 Example

How many tails until we get the 10th head on a fair coin. $X \sim \mathrm{NB}(10, \frac{1}{2})$

1.4 Example

If courses were independent with probability p of passing and you need 40 courses, then the number of failed courses would be $\mathrm{NB}(40, p)$.

1.5 Range and Probability Function of the Negative Binomial Distribution

range $x \in \{0, 1, \dots\}$ (countably infinite)

$$\begin{aligned} f(x) = P(X = x) &= p(x \text{ F's before } k\text{th S}) \\ &= \binom{x+k-1}{x} p^k (1-p)^x \\ &= \binom{x+k-1}{k-1} p^k (1-p)^x \end{aligned}$$

In a picture:

$$\overbrace{\underbrace{\_\ \_\ \_\ \cdots\ \_\ |}_{(k-1)\text{ S's, } x\text{ F's}}}^{x+(k-1)\text{ Trials}} \underbrace{\text{S}}_{k\text{th S}}$$

1.6 Example

Suppose a startup is looking for 5 investors. They ask investors repeatedly where each independently has a 20% chance of saying yes. Let $X =$ total # of investors that they ask and note that X does not follow a negative binomial distribution. Find $f(x)$ and $f(10)$.

Let $Y =$ # who say no before 5 say yes. $Y \sim \text{NB}(5, 0.2)$, and $X = Y + 5$. So,

$$\begin{aligned} f(x) &= P(X = x) \\ &= P(Y + 5 = x) \\ &= P(Y = x - 5) \\ &= \binom{(x-5)+5-1}{5-1} (0.2)^5 (0.8)^{x-5} \\ &= \binom{x-1}{4} (0.2)^5 (0.8)^{x-5} \qquad \text{for } x = 5, \dots \end{aligned}$$

$$f(10) = \binom{9}{4} (0.2)^5 (0.8)^5$$

note that it's $\binom{9}{4}$ and not $\binom{10}{5}$ because the 10th investor must have said yes.

2 Lecture 14

2.1 Summary

Today we reviewed the Negative Binomial distribution and looked at an example illustrating the difference between Bin and NB.

Then we discussed the Geometric distribution, which is a special case of the Negative Binomial with $k = 1$. We found $f(x) = p(1-p)^x$ for $x = 0, 1, 2, \dots$ (You can also get this same expression by subbing in $k = 1$ to the Negative Binomial pf.) The neat thing about the Geometric distribution is that it actually does have a nice closed-form expression for the cumulative distribution function, $F(x) = 1 - (1-p)^{x+1}$.

Lastly we talked about all the distributions we've seen so far and some clues about how to identify which distribution to use. You will be given the formulas for the pfs on all tests/quizzes, but your task will be to identify which one is the one to use. The more you practice doing this, the better you will get at it, so I encourage you to try to make up problems for yourself or your classmates and see if you can figure out the distribution needed.

To help with this, if something happens to you over the weekend that makes you think hmm.... I wonder what distribution I could use to model this? then please remember it (write it down if you have to, or email it to me, or tweet it to ActSciProf) and we can use it in class on Monday.

2.2 Example

Suppose you send a bit string over a noisy connection with each bit independently having a probability 0.01 of being flipped. What is the probability that

(a) it takes 50 bits to get 5 errors?

(b) a 50 bit message has 5 errors?

(b) Let $Y = \#$ of errors in 50 bits. $Y \sim \text{Bin}(50, 0.01)$.

Then, $P(Y = 5) = \binom{50}{5}(0.01)^5(0.99)^{45}$

(a) Let $X = \#$ of correct bits until 5 errors. $X \sim \text{NB}(5, 0.01)$.

Then, $P(X = 45) = \binom{49}{4}(0.01)^5(0.99)^{45}$

2.3 Geometric Distribution (5.6)

The Geometric Distribution is just a special case of the Negative Binomial Distribution with $k = 1$. Let $X = \#$ of F's in Bernoulli trials before the first S. $X \sim \text{Geo}(p)$

2.4 Range and Probability Function of the Geometric Distribution

range: $x \in \{0, 1, \ldots\}$

$$\begin{aligned} f(x) &= P(X = x) \\ &= P(\underbrace{\text{F, F, } \ldots}_{\text{all F's}}, \text{S}) \\ &= (1-p)^x p \end{aligned}$$

or sub $k = 1$ into the NB probability function.

Prove $\sum\limits_{\text{all } x} f(x) = 1$

Proof.

$$\begin{aligned} \sum_{x=0}^{\infty}(1-p)^x p = & \underbrace{p + p(1-p) + \ldots}_{\text{(geo. series: } a = p,\, r = 1-p)} \\ &= \frac{p}{1-(1-p)} \\ &= 1 \end{aligned}$$

□

Find the cumulative distribution function.

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= 1 - P(X > x) \\ &= 1 - [f(x+1) + \dots] \\ &= \underbrace{1 - [p(1-p)^{x+1} + p(1-p)^{x+2} + \dots]}_{\text{(geo. series: } a = p(1-p)^{x-1},\ r = 1-p)} \\ &= 1 - \frac{p(1-p)^{x+1}}{1-(1-p)} \\ &= 1 - (1-p)^{x+1} \text{ for } x = 0, 1, \dots \end{aligned}$$

if $x \in \mathbb{R}$, then

$$F(x) = \begin{cases} 1 - (1-p)^{\lfloor x \rfloor + 1}, \text{ if } x \geq 0 \\ 0, \text{ if } x < 0 \end{cases}$$

	Discrete Uniform	Hypergeometric	Binomial	Negative Binomial	Geometric	Poisson
function range parameters	$\text{DU}[a,b]$ $a, a+1, \dots, b$	$\text{Hyp}(N,r,n)$ bad	$\text{Bin}(n,p)$ $0, 1, \dots, n$	$\text{NB}(k,p)$ $0, 1, \dots$	$\text{Geo}(p)$ $0, 1, \dots$	$\text{Poi}(\mu)$ $0, 1, \dots$ $\mu = np$, $\mu = \lambda t$
pf, $f(x)$	$\frac{1}{b-a+1}$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$\binom{n}{x}p^x(1-p)^{n-x}$	$\binom{x+k-1}{k-1}p^k(1-p)^x$	$p(1-p)^x$	$\frac{e^{-\mu}\mu^x}{x!}$
cumulative distribution function, $F(x)$	$\frac{x-a+1}{b-a+1}$				$1-(1-p)^{x+1}$	$e^{\mu}[1 + \frac{\mu^1}{1!} + \cdots \frac{\mu^x}{x!}]$
how to tell	"equally likely" know min. & max.	know total # objects know # S's know # trials without replacement selecting a subset	Bernoulli trials know # trials count # S's	Bernoulli trials "until" "it take... to get" "before" know how many S's we want	"until we get" "before the first"	Bin. with large amount of trials, small prob rate specified (#events/time) no pre-specified max. events occurring at any time (randomly) Poisson process & know time & count events doesn't make sense to ask how often an event did **not** occur

Bernoulli trials:

- independent
- each outcome is a S or F
- P(success)=p=constant

3 Lecture 15

3.1 Summary

Today we started by looking at some examples of identifying which distribution to use.

Then we talked about our last discrete distribution: the Poisson distribution. It arises as a limiting case of the Binomial when n gets large and p gets small,such that the product $np = \mu$. We showed that as n approaches infinity, the probability function $f(x)$ approaches $\frac{e^{-\mu}\mu^x}{x!}$ which is the Poisson pf. In practice, we can use the Poisson to approximate the Binomial when n is reasonably large and p is reasonably close to 0. Like all approximations, the accuracy is better when the conditions are more closely satisfied.

In addition to being a limiting case of the Binomial, it also comes from a Poisson process, which is events occurring throughout time/space with 3 conditions: independence, individuality, and uniformity/homogeneity.

Many different processes can be modelled with a Poisson process, and next time we'll see some examples. See if you can think of any before next class!

3.2 Example

Naomi invites 12 people to her party. If each independently comes with probability p. Let $X = \#$ of guests.

Binomial: $X \sim \text{Bin}(12, p)$

3.3 Example

20 toys in a machine. Each time you grab one with a claw. Let $X = \#$ of tries to get one toy you want.

None.

3.4 Example

Trying to catch a pokemon, each time has a probability p of succeeding. Let $X = \#$ of failed attempts.

Geometric: $X \sim \text{Geo}(p)$

3.5 Example

You have 5 classes randomly scheduled in a row. Let $X = \#$ of classes before your favourite.

range: $0, 1, 2, 3, 4$, and the probability is $1/5$ for each of the range.

Discrete Uniform: $X \sim \text{DU}[1, 4]$

3.6 Poisson Distribution from Binomial (5.7)

Suppose we have a $X \sim \text{Bin}(n, p)$ where n is very large and p is very small. Then, as $n \to \infty$ and $p \to 0$ such that np remains constant, the probability function of X approaches a limit.

Let $np = \mu$, so $p = \frac{\mu}{n}$. Then

$$\begin{aligned}
\lim_{n\to\infty} f(x) &= \lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{x!} \frac{\mu^x}{n^x} \left(1-\frac{\mu}{n}\right)^n \left(1-\frac{\mu}{n}\right)^{-x} \\
&= \frac{\mu^x}{x!} \lim_{n\to\infty} \frac{n}{n}\frac{n-1}{n}\cdots\frac{n-x+1}{n} \left(1-\frac{\mu}{n}\right)^n \left(1-\frac{\mu}{n}\right)^{-x} \\
&= \frac{\mu^x}{x!} \lim_{n\to\infty} \left(1-\frac{\mu}{n}\right)^n \\
&= \frac{e^{-\mu}\mu^x}{x!}
\end{aligned}$$

We write: $X \sim \text{Poi}(\mu)$, range: $0, 1, \ldots$

We can use the Poisson random variable as an approximation to the Binomial when n is large, and p is small. The only thing we need to do is $\mu = np$.

3.7 Example

Tim Hortons roll up the rim says 1 in 6 cups win a prize. Suppose you have 80 cups. Find the probability that you get 10 or fewer winners.

Let $X = \#$ of winning cups. $X \sim \text{Bin}(80, 1/6)$ We want

$$\begin{aligned} F(10) &= P(X \leq 10) \\ &= \sum_{x=0}^{10} f(x) \\ &= \binom{80}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{80} + \cdots + \binom{80}{10}\left(\frac{1}{6}\right)^{10}\left(\frac{5}{6}\right)^{70} \\ &= 0.2002 \text{ (tedious)} \end{aligned}$$

Try a Poisson approximation. $Y \sim \text{Poi}(\mu = np = \frac{80}{6} \approx 13.33)$. Then,

$P(Y \leq 10) = e^{-13.33}\left[1 + \frac{13.33}{1!} + \cdots + \frac{13.33^{10}}{10!}\right] = 0.224$

Not a good approximation since p was too large.

3.8 Poisson Distribution from Poisson Process (5.8)

The Poisson Process: Suppose events occur randomly in time or space according to three conditions:

(1) Independence: the number of events in one period cannot affect another non-overlapping period

(2) Individuality: events occur one at a time (cannot have two at the exact same time)

(3) Homogeneity or Uniformity: events occur at a constant rate

4 Lecture 16

4.1 Summary

Today we finished up Chapter 5 with more about the Poisson process and the Poisson distribution, and started Chapter 7.

If we define X to be the number of events observed in a time period of length t in a Poisson process with rate λ, then I claimed that X has a Poisson distribution with parameter $\mu = \lambda t$. The proof in the course notes uses all three conditions for the Poisson process, as well as some calculus, and I encourage you to check it out if you enjoy neat proofs.

Many different processes can be modelled with a Poisson process, and all we have to do is identify the values of λ and t (making sure they are measured in the same time units) and define μ to be the product. We looked at some ways to tell when to use the Poisson distribution.

Then we looked at a detailed example of combining other models with the Poisson distribution. These types of problems are often found on tests/exams, and there are lots of other examples in the course notes to practice with. See if you can get the answer to part (d) of the question in class, and I'll post the solution as a follow-up below.

We won't be spending any class time on R (Chapter 6), but I'll post some resources on Learn if you want to find out more about this neat statistical programming language.

Finally we started Chapter 7. We looked at several ways of summarizing data: a frequency table, a frequency histogram, as well as three single numbers: the sample mean (average), the median (middle value), and the mode (most common value). These quantities are not necessarily equal to each other, but by chance they were all 2 when we used some information taken from the class on the number of kids in their family!

I've been running out of time for our SWAGs lately but I'll post 5 and 6 on Learn (about Liar's Dice and Pokémon, respectively.)

Consider a Poisson Process with rate λ, (i.e. λ events occur on average per unit time). Observe the process for t units of time. Let $X = \#$ of events that occur. Then, $X \sim \text{Poi}(\mu)$, where $\mu = \lambda t$. That is,

$$f(x) = \frac{e^{-\mu}\mu^x}{x!}$$

4.2 Example

Request coming in from a web server at a rate of 100 requests per minute. $\lambda = 100, t = \frac{1}{60}$ The $\#$ of requests per second would be

$$\text{Poi}\left(\mu = \frac{100}{60} = \frac{5}{3}\right)$$

4.3 Combining Other Models with the Poisson Process (5.9)

Problems may involve many different random variables!

4.4 Example (Continued)

We say that a second is quiet if it has no requests.

(a) Find probability that a second is quiet

(b) In a minute (60 non-overlapping seconds), find the probability of 10 quiet seconds

(c) Find the probability of having to wait 30 non-overlapping seconds to get 2 quiet seconds

(d) Given (c), find the probability of 1 quiet second in the first 15 seconds

(a) Let $X = \#$ requests in a second. $X \sim \text{Poi}(5/3)$.

We want $P(X = 0) = \frac{e^{-\frac{5}{3}}\left(\frac{5}{3}\right)^0}{0!} = 0.189$

(b) Let $Y = \#$ quiet seconds out of 60. $Y \sim \text{Bin}(60, 0.189)$.

We want $P(Y = 10) = \binom{60}{10}(0.189)^{10}(0.811)^{50} = 0.124$

(c) Let $Z = \#$ non-quiet seconds before getting 2 quiet seconds. $Z \sim \text{NB}(2, 0.189)$.

We want $P(Z = 28) = \binom{29}{1}(0.189)^2(0.811)^{28} = 0.003$

(d) $D_x = \#$ of quiet seconds out of 15. $D_x \sim \text{Bin}(15, 0.189)$.

$$P(D_x = 1) = \binom{15}{1}(0.189)^1(0.811)^{14}$$

We get,

$P(\text{1 quiet second in the first 15 seconds} \mid \text{wait 30 to get 2 quiet}) =$

$$= \frac{P(\text{1 quiet second in the first 15 seconds AND wait 30 to get 2 quiet})}{P(\text{wait 30 to get 2 quiet})} \tag{1}$$

$$= \frac{P(\text{1 quiet second in the first 15 seconds AND wait an additional 15 to get 1 additional quiet})}{P(C)} \tag{2}$$

$$= \frac{P(\text{1 quiet second in the first 15 seconds})P(\text{wait an additional 15 to get 1 additional quiet})}{P(C)} \tag{3}$$

$$= \frac{\binom{15}{1}(0.189)^1(0.811)^{14} \times (0.811)^{14}(0.189)}{\binom{29}{28}(0.189)^2(0.811)^{28}} \tag{4}$$

$$= \frac{\binom{15}{1}}{\binom{29}{28}} \tag{5}$$

$$= \frac{15}{29} \tag{6}$$

In (3) we used the independence of non-overlapping time intervals and constant probability of events.

4.5 Summarizing Data on Random Variables (7.1)

Let $X = \#$ of kids in a family.

Value	Frequency
1	3
2	10
3	1
4	1

4.5.1 Definition (Median)

The *median* of a sample is a value such that half the results are below it and half above it, when the results are arranged in numerical order.

4.5.2 Definition (Mode)

The *mode* of the sample is the value which occurs most often. There is no guarantee there will be only a single mode.

Mean: average $\rightarrow \frac{1\times3+2\times10+3\times1+4\times1}{15}$

Median: 2

Mode: 2

5 Lecture 17*

5.1 Summary

Today we looked at what happens when we replace the relative frequency in the sample mean with a theoretical probability. We get the expected value of X (or theoretical mean) given by: $E[X] = \sum xf(x)$ (where the sum is over all x in the range of X.)

For our MLIW, we noticed that the sample mean in the class was quite different from the theoretical mean, when X was the number of kids in a family. There could be many reasons for this, but likely the most important

is that the sample in the class was not representative of the population of Canada, since that includes many young families with one child that may have more, whereas most of the people in the class will not be gaining any new siblings. Any time you're building a machine learning algorithm, it's only as good as the data you build it on. So if the data is biased and does not reflect reality, the predictions from the model will be biased as well. An important concept in machine learning is data stewardship - making sure the data going in is accurate, representative, and appropriate for the purpose of the model.

In addition to the formal definition of the expected value of X, we may be interested in a function of X, so we also defined the expected value of a function $g(X)$ to be $E[g(X)] = \sum g(x)f(x)$. Expectation is a linear operator so we can split up sums and pull out constants (i.e. $E[aX + b] = aE[X] + b$) but for a general non-linear function, unfortunately $E[g(X)] \neq g(E[X])$.

Then we looked at some applications of expectation, including caching and Roulette. I encourage you to read the other applications in section 7.3 for some more examples.

- If you're interested, see if you can determine how small the probability of a cache hit would have to be (in our example) in order for it not to be worth it to use a cache. Post the answer in the follow-up if you get it.
- In Roulette, a game where there are 38 sections that can be chosen with equal probability, and you can bet on lots of different outcomes. It turns out that no matter what betting strategy you use or how you split up your money, the expected payoff from any $1 bet is always 0.94737, so you essentially lose about 5.3 cents every time you play! (Over Reading Week, I encourage you to imagine a betting strategy and verify this fact - but I do not encourage actually gambling!)
- Of course, different betting strategies will have different amounts of risk, even if the expected value is the same. This is the idea of Variance, which we'll start talking about on Monday after Reading Week. :)

Have a fantastic Reading Week! I recommend setting realistic goals for yourself (including both some dedicated time to relax and dedicated time to catch up / get ahead on school work) and have both a productive and fun week!

5.2 Expectation of a Random Variable (7.2)

Imagine we know the theoretical probability of each # of kids in a family.

x	1	2	3	4	5
$f(x)$	0.43	0.4	0.12	0.04	0.01

Now we replace the observed proportion in the sample mean with $f(x)$

$$\sum_{\text{all } x} xf(x) = (1)(0.43) + (2)(0.4) + (3)(0.12) + (4)(0.04) + (5)(0.01) = 1.8$$

which is the theoretical mean.

Why do we have sample mean > theoretical mean?

- urban vs rural population
- income level
- sampled max family size but theoretical includes growing families
- selection bias (if you randomly select people rather than families, people with lots of siblings will be over-represented

5.2.1 Definition (Expected Value)

Let X be a discrete random variable and probability function $f(x)$. The *expected value* (also called the mean or the expectation) of X is given by

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

Remark 1. μ will be within the range but not necessarily equal to a possible value of x.

We might be interested in the expected value of some function of X, $g(X)$.

5.3 Example

Tax credit of $\$1000$ plus $\$250$ per kid. Find the average cost.

x	1	2	3	4	5
$g(x)$	1250	1500	1750	2000	2250

Average cost = weighted average of $g(x)$ values $= (1250)(0.43) + \cdots + (2250)(0.01) = 1450$

5.3.1 Theorem

Let X be a discrete random variable and probability function $f(x)$. The expected value of a some function $g(X)$ of X is given by

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

Note that $g(x) = 1000 + 250x$ from last example.

$$E[g(X)] = 1000 + 250E[X] = 1450$$

What if tax credit = $\frac{2000}{x}$

$$E[g(X)] = (2000)(0.43) + (1000)(0.40) + \cdots + (400)(0.01) = 1364$$

But $\frac{2000}{E[X]} = \frac{2000}{1.8} = 1111.11$. Therefore

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

unless g is a linear function. That is, if $g(X) = aX + b$, then $E[g(X)] = aE[X] + b$

5.4 Example

A web server has a cache. Takes 10ms to check, 20% of the requests are found (cache hit) and immediately shown. If it's not found (cache miss), it takes $\underbrace{50}_{\text{to server}} + \underbrace{70}_{\text{lookup}} + \underbrace{50}_{\text{to client}}$ additional milliseconds to get info and display. Is it worth it? Let $X = \#$ of milliseconds to display the information.

x	10	10+50+70+50=180
$f(x)$	0.2	0.8

$$E[X] = (10)(0.2) + (180)(0.8) = 146\text{ms}$$

Time no cache $= 50 + 70 + 50 = 170$ms.

Since $146\text{ms} < 170\text{ms}$, it's worth it!

5.5 Example

Roulette: each of 38 numbers is equally likely

(1) If you bet 1 dollar on number $7 \rightarrow$ pays $35:1$

OR

(2) If you bet 50 cents on red $\rightarrow$ pays $1:1$ and 50 cents on first 12 numbers $\rightarrow$ pays $2:1$

(1)

x	0	36
$f(x)$	$37/38$	$1/38$

(2)

y	0	1	1.50	2.50
$f(y)$	$\underbrace{14/38}_{\text{neither}}$	$\underbrace{12/38}_{\text{red}}$	$\underbrace{6/38}_{\text{black}}$	$\underbrace{6/38}_{\text{both red}}$

$$E[X] = 0(\frac{37}{38}) + 36(\frac{1}{38}) = 0.94737$$

$$E[Y] = 0(\frac{14}{38}) + 1(\frac{12}{38}) + 1.5(\frac{6}{38}) + 2.5(\frac{6}{38}) = 0.94737$$

6 Lecture 18

6.1 Means and Variances of Distributions (7.4)

The mean $E[X]$ tells us where the distribution is on average. We also need a way to describe how spread out a distribution is. Variance could be $E[X - \mu] = 0$.

What about $E[|X - \mu|]$

- need cases to evaluate
- non-differentiable at point $X - \mu$
- linear penalty for being away from the mean

Instead we use $E[(X - \mu)^2]$

6.1.1 Definition (Variance)

The *variance* of a random variable X, denoted by $Var(X)$ or by σ^2, is

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

6.2 Example

$X = \#$ on fair 6-sided die

$E[X] = 3.5$

$E[(x - 3.5)^2]$

$E[X]^2 - 3.5^2$

x	1	2	3	4	5	6
x^2	1	4	9	16	25	36

Alternate form (calculation form)

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - 2(E[X])^2 + E[X]^2 \\ &= E[X^2] - E[X]^2 \\ &= \sum_{\text{all } x} x^2 f(x) - \left(\sum_{\text{all} x} x f(x) \right)^2 \end{aligned}$$

6.3 Example (Roulette)

$X = 0$ or 36 (dollars)

x	0	36
$f(x)$	$37/38$	$1/38$

$$E[X] = 0.94737$$

$$Var(X) = E[X^2] - 0.94737^2 = 36^2(\frac{1}{36}) - 0.94737^2 = 33.207 \text{ dollars}^2$$

To interpret the variance better, we often take the square root to get the same units of the original variable.

6.3.1 Definition (Standard Deviation)

The *standard deviation* of a random variable X is

$$\sigma = SD(X) = \sqrt{Var(X)}$$

$$SD(X) = \sqrt{33.207} = 5.76$$

What if we bet \$1 on red. Y=winnings

y	0	2
$f(y)$	$20/38$	$18/38$

$$E[Y] = 0.94737$$
$$Var(Y) = E[Y^2] - 0.94737^2 = 0.97723$$
$$SD(Y) = 0.9986$$

6.4 Linear Transformations

If $Y = aX + b$, and we know $E[X]$ and $Var(X)$, what can we say about $E[Y]$ and $Var(Y)$.

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

$$\begin{aligned} Var(Y) &= E[(Y - E[Y])^2] \\ &= E[(aX + b) - (aE[X] + b)^2] \\ &= E[a^2X^2 - 2XE[X] + E[X]^2] \\ &= a^2E[(X - E[X])^2] \end{aligned}$$

$$Var(Y) = a^2 Var(X)$$
$$SD(Y) = |a| SD(X)$$

7 Lecture 19*

7.1 Summary

Today we started with some examples of calculating variance and verified the results for linear transformation of variables using a specific example.

Then we derived the mean of the Binomial and Poisson distributions. Both of them used similar tricks ($x! = x(x-1)!$, and manipulating our sum until it became a pf so the sum was 1) and gave us results that make logical sense: the mean of a $\text{Bin}(n,p)$ is np and the mean of a $\text{Poi}(\mu) = \mu$. There are similar results for the means of the other named distributions, but there are easier ways to get them as we'll see later.

Then we derived the variance of the Poisson distribution, which uses similar tricks to the proofs for the mean. Interestingly, the Poisson variance is equal to the mean μ. If you want to give it a try, you can find that the Binomial variance is $np(1-p)$ which is symmetric around $p = 0.5$, and smallest when p is near 0 or 1.

I ran out of time for our SWAG about League of Legends, but the sheet is posted on Learn with lots of details about the calculations if you are interested.

On Friday we'll review Chapters 5 and 7 before turning our attention to continuous random variables in Chapter 8.

7.2 Example

Suppose X has probability function:

x	0	1	2	3	4
y	1	3	5	7	9
$f(x)$	0.1	0.1	0.1	0.5	0.2

Let $Y = 2X + 1$.

$E[X] = 2.6$

$E[X^2] = 6.2$

$E[Y] = 6.2$

$E[Y^2] = 94.2$

$Var(X) = 8.2 - 2.6^2 = 1.44$

$SD(X) = 1.2$

$Var(Y) = 44.2 - 6.2^{=}5.76$

$SD(Y) = 2.4$

Now we can verify,

$$\begin{aligned} E[Y] &= E[2X+1] \\ &= 2E[X] + 1 \\ &= 2(2.6) + 1 \\ &= 6.2 \end{aligned}$$

$$Var(Y) = 2^2 Var(X) = 4(1.44) = 5.76$$

$$SD(Y) = |2| SD(X) = 2(1.2) = 2.4$$

Let $X \sim \text{Bin}(n, p)$. Find $E[X]$.

$$E[X] = \sum_{\text{all } x} xf(x) \tag{1}$$
$$= \sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x} \tag{2}$$
$$= \sum_{x=1}^{n} x\binom{n}{x}p^x(1-p)^{n-x} \tag{3}$$
$$= \sum_{x=1}^{n} x\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \tag{4}$$
$$= \sum_{x=1}^{n} x\frac{n!}{x(x-1)!(n-x)!}p^x(1-p)^{n-x} \tag{5}$$
$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!}p^x(1-p)^{n-x} \tag{6}$$
$$= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)![(n-1)-(x-1)]!}pp^{x-1}(1-p)^{(n-1)-(x-1)} \tag{7}$$
$$= np(1-p)^{n-1}\sum_{x=1}^{n}\binom{n-1}{x-1}p^{x-1}(1-p)^{-(x-1)} \tag{8}$$
$$= np(1-p)^{n-1}\sum_{x=1}^{n}\binom{n-1}{x-1}\left(\frac{p}{1-p}\right)^{x-1} \tag{9}$$

From (2) to (3) we used to fact that when $x = 0$ the value of the expression is 0. Provided that $x \neq 0$, we can expand $x!$ as $x(x-1)!$ as seen from (4) to (5). Let $y = x - 1$, we get

$$E[X] = np(1-p)^{n-1}\sum_{x=1}^{n}\binom{n-1}{y}\left(\frac{p}{1-p}\right)^{y} \tag{10}$$
$$= np(1-p)^{n-1}\left(1+\frac{p}{1-p}\right)^{n-1} \tag{11}$$
$$= np(1-p)^{n-1}\frac{(1-p+p)^{n-1}}{(1-p)^{n-1}} \tag{12}$$
$$= np \tag{13}$$

From (10) to (11) we used the Binomial Theorem.

Let $X \sim \text{Poi}(\mu)$. Find $E[X]$.

$$E[X] = \sum_{\text{all } x} xf(x) \tag{1}$$
$$= \sum_{x=0}^{\infty} x\frac{e^{-\mu}\mu^x}{x!} \tag{2}$$
$$= \sum_{x=1}^{\infty} x\frac{e^{-\mu}\mu^x}{x(x-1)!} \tag{3}$$
$$= \sum_{x=1}^{\infty} \mu\frac{e^{-\mu}\mu^{x-1}}{(x-1)!} \tag{4}$$
$$\tag{5}$$

Let $y = x - 1$, we get

$$E[X] = \mu e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^y}{y!} \tag{6}$$
$$= \mu e^{-\mu} e^{\mu} \tag{7}$$
$$= \mu \tag{8}$$

From (6) to (7) we used the fact that $e^x = \sum_{y=0}^{\infty} \frac{x^y}{y!}$.

Similarly,

$$X \sim \text{DU}[a, b],\ E[X] = \frac{a+b}{2}$$
$$X \sim \text{Hyp}(N, r, n),\ E[X] = \frac{nr}{N}$$
$$X \sim \text{NB}(k, p),\ E[X] = \frac{k(1-p)}{p}$$
$$X \sim \text{Geo}(p),\ E[X] = \frac{1-p}{p}$$

Let $X \sim \text{Poi}(\mu)$. Find $Var(X)$.

Since there's $x!$ in the denominator of $f(x)$, let's find $E[X(X-1)]$.

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)\frac{\mu^x e^{-\mu}}{x!}$$
$$= \sum_{x=2}^{\infty} x(x-1)\frac{\mu^x e^{-\mu}}{x(x-1)(x-2)!}$$
$$= \mu^2 e^{-\mu} \sum_{x=2}^{\infty} \frac{\mu^{x-2}}{(x-2)!}$$

Let $y = x - 2$, we get

$$E[X(X-1)] = \mu^2 e^{-\mu} \sum_{x=2}^{\infty} \frac{\mu^y}{y!}$$
$$= \mu^2$$

$$Var(X) = E[X(X-1)] + E[X] - E[X]^2$$
$$= \mu^2 + \mu - \mu^2$$
$$= \mu$$

8 Lecture 20*

8.1 Summary

We started with one last Chapter 7 example (using the mean and variance of the Poisson, and the properties of linear transformations of random variables) and our SWAG.

Then we summarized everything we know about discrete random variables, which turned out to be quite a lot! I talked about which aspects will remain the same and which will change when we move into the continuous case.

Surprisingly, a lot stays the same. The main differences come from the fact that we have an uncountable number of values the variable can take on, so $P(X = x) = 0$ for all x, and all our sums become integrals. We found that we need the derivative of the cumulative distribution function (which we call the probability density function or pdf $f(x)$) to tell us information about the random variables local behaviour, and so the relationship between $F(x)$ and $f(x)$ is now a derivative-integral relationship rather than a sum-difference relationship like in the discrete case. Most of the properties of $F(x)$ and $f(x)$ remain the same, but not all. Many of the discrete distributions have a similar distribution in the continuous world: in this course we'll only talk about 3: Uniform, Exponential, and Normal. Everything about expected value and variance remains the same, including how linear transformations work, we just have to evaluate them with integrals instead of sums.

Next time a guest lecturer will formally define all these quantities and properties, and look at some examples of continuous random variables.

8.2 Example

Suppose the amount of data you use on your phone (in units of 100MB) has a Poisson distribution with mean 7 per month. You pay 15 per month plus 3 per 100MB. Find the standard deviation of random month's phone bill.

Let $X = \#$ of units of data used. $X \sim \text{Poi}(7)$. Let $Y = 15 + 3X \rightarrow E[Y] = 15 + 3(7) = 36$.

$SD(Y) = 3SD(X) = 3\sqrt{7} = 7.94$.

9 Lecture 21

9.1 Summary

Today you talked more about the properties of $F(x)$ and $f(x)$ in the continuous case. For continuous random variables, the cdf $F(x)$ has the same definition and properties as in the discrete case. In addition, since any specific point x has $P(X = x) = 0$, $P(a < X < b) = P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b)$ and $F(x)$ is a continuous function of x.

You also defined the probability density function $f(x) = F'(x)$. It represents the relative likelihood of X being "near" the value x in the following way: $f(x) * \epsilon \approx P(x \leq X \leq x + \epsilon)$ for a small ϵ. The pdf has similar properties to the pf of a discrete random variable, but not quite the same as it is not a probability. It must be non-negative, and integrates to 1 over the range of X. We can also get $F(x)$ from $F(x)$ using an integral from the bottom of the range to x.

Finally, you talked about how to calculate expectation and variance in the continuous case, and how to find percentiles of a continuous distribution.

A *continuous random variable* X maps points in a continuous sample space to real numbers such that the range is uncountably infinite.

Examples of continuous random variables

Let X be the number the point spots at.

(1) temperature of a day

(2) length of time until a bus arrives

(3) height of a random person

(4) average height of 10 people

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

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

9.2 Example

For $x < 0$, no chance of the point stopping at a number < 0.

For $x > 4$, $F(x) = 1$ since the point is certain to stop at a number below 4.

$P(0 < x \leq 1) = \frac{1}{4} = F(1)$

$$F(x) = \begin{cases} 0, \, x < 0 \\ \frac{x}{4}, \, 0 \leq x \leq 4 \\ 1, \, x > 4 \end{cases}$$

PROPERTIES OF $F(x)$

(1) For all x, $P(X = x) = 0$. So,

$$\begin{aligned} P(a < x \leq b) &= P(a \leq x \leq b) \\ &= P(a < x < b) \\ &= P(a \leq x < b) \end{aligned}$$

Remark 2. End points don't matter.

(2)

$$\begin{aligned} \lim_{\epsilon \to 0} F(x) - F(x - \epsilon) &= \lim_{\epsilon \to 0} P(x - \epsilon < X \leq x) \\ &= P(X = x) \\ &= 0 \end{aligned}$$

Thus $\lim_{\epsilon \to 0} F(x - \epsilon) = F(x)$, so $F(x)$ is continuous.

(3) $F(x)$ is non-decreasing.

(4) $\lim_{x \to +\infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$

(5) $0 \leq F(x) \leq 1$

9.2.1 Definition (Probability Density Function)

The *probability density function* (p.d.f) $f(x)$ for a continuous random variable X is the derivative

$$f(x) = \frac{d}{dx} F(x)$$

where $F(x)$ is the cumulative distribution function for X.

Remark 3. $f(x)$ is not a probability. It can be > 1 relative likelihood that X takes a value near X.

PROPERTIES OF $f(x)$

(1)

$$\begin{aligned} P(a \leq X \leq b) &= F(b) - F(a) \\ &= \int_a^b f(x)dx \end{aligned}$$

(2)

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)dx &= F(+\infty) - F(-\infty) \\ &= 1 - 0 \\ &= 1 \end{aligned}$$

(3) $f(x) \geq 0$ (since $F(x)$ is non-decreasing, it's derivative is non-negative)

(4)

$$F(x) = \int_{-\infty}^{x} f(u)du$$

9.3 Example

Suppose a continuous random variable X is on the range $[0, 1]$ has the cumulative distribution function $F(x) = x^2$.

What is the probability density function?

$f(x) = \frac{d}{dx}F(x) = 2x$.

What is $P(X = 0.25)$?

$P(X = 0.25) = 0$

What is $P(X \leq 0.25)$?

(1) $P(X \leq 0.25) = F(0.25) = (0.25)^2 = 0.0625$

(2)

$$P(X \leq 0.25) = \int_0^{0.25} f(x)dx = \int_0^{0.25} 2xdx = 0.625$$

Expectation:

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_{x \in \text{range}} xf(x)dx$$

Variance:

$$Var(X) = E[X^2] - E[X]^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx - \int_{-\infty}^{+\infty} xf(x)dx$$

9.3.1 Definition (Percentiles)

The p^{th} *percentile* of a distribution x_p such that $F(x_p) = p$.

10 Lecture 22

10.1 Example

$F(x) = x^2$ for $0 < x < 1$.

FIND THE MEAN, MEDIAN, AND MODE

Mean:

$$E[X] = \int_0^1 x2xdx = \int_0^1 2x^2dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}$$

Median: $x_{0.5}$ satisfies $F(x_{0.5}) = 0.5 \implies (x_{0.5})^2 = 0.5 \implies x_{0.5} = \sqrt{0.5} = 0.707$

Mode: 1 (x value that maximizes $f(x)$)

10.2 Continuous Uniform Distribution (8.2)

A *continuous* random variable takes real values between a and b with $a < b$ such that any interval of fixed size is equally likely.

NOTATION

$X \sim U(a, b)$

Remark 4. Can include or exclude endpoints, doesn't matter.

FIND $f(x)$

$f(x) = c$, (since it can't depend on x). We need

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_a^b c\,dx = 1$$

$[cx]_a^b = 1 \implies c(b-a) = 1 \implies c = \frac{1}{b-a}$

So,

$$f(x) = \begin{cases} \frac{1}{b-a},\ a \le x \le b \\ 0,\ \text{otherwise} \end{cases}$$

FIND $F(x)$

$$F(x) = \int_{-\infty}^{x} f(u)du = \int_a^x \frac{1}{b-a}du = \left[\frac{u}{b-a}\right]_a^x = \frac{x-a}{b-a}$$

$$F(x) = \begin{cases} \frac{x-a}{b-a},\ a < x < b \\ 0,\ x < a \\ 1,\ x > b \end{cases}$$

FIND THE MEAN, MEDIAN AND MODE

Mean:

$$E[X] = \int_a^b x \frac{1}{b-a} dx = \left[\left(\frac{x^2}{2} \right) \left(\frac{1}{b-a} \right) \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

Median: is also $\frac{a+b}{2}$

Mode: no unique mode

Similarly,

$$Var(x) = \frac{(b-a)^2}{12}$$

SPECIAL CASE

$U \sim U(0,1)$ (i.e. $a = 0$, $b = 0$)

$$f(u) = \begin{cases} 1, \ 0 < u < 1 \\ 0, \ \text{otherwise} \end{cases}$$

$$F(u) = \begin{cases} u, \ 0 < u < 1 \\ 0, \ u < 0 \\ 1, \ u > 1 \end{cases}$$

$U(0,1)$ random variables are easy to generate.

10.3 Change of Variables

Suppose you know the distribution of X and you want the distribution of $Y = g(X)$.

1. Write the cumulative distribution function of Y in terms of the cumulative distribution function of X
2. Sub in what we know about X, then differentiate to get the pdf
3. Determine the range of Y

10.4 Example (Change of Variable)

Let $X \sim U(0,4)$, $F_X(x) = \frac{x}{4}$, $f_X(x) = \frac{1}{4}$ $x \in (0,4)$

Let $Y = \frac{1}{X}$

1.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P\left(\frac{1}{X} \leq y \right) \\ &= P\left(X > \frac{1}{y} \right) \\ &= 1 - F_X\left(\frac{1}{y} \right) \end{aligned}$$

2.

$$\begin{aligned} F_Y(y) &= 1 - F_X\left(\frac{1}{y}\right) \\ &= 1 - \frac{\frac{1}{y}}{4} \\ &= 1 - \frac{1}{4y} \end{aligned}$$

$f_Y(y) = \frac{d}{dx}F_Y(y) = \frac{1}{4y^2}$

OR differentiate $F_Y(y)$ before substituting in the information about X. You need the chain rule!

$$\frac{d}{dy}\left[1 - F_X\left(\frac{1}{y}\right)\right] = -f_X\left(\frac{1}{y}\right)\left(-\frac{1}{y^2}\right) = \frac{1}{4y^2}$$

3. $y \in (\frac{1}{4}, \infty)$

11 Lecture 23

11.1 Example

Let $Y \sim U(0,1) \implies f_Y(y) = \frac{1}{1-0} = 1$, $F_Y(y) = \frac{y-0}{1-0} = y$

$X = 2\sqrt[3]{Y}$.

FIND $f_X(x)$

1.

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(2\sqrt[3]{Y} \leq x\right) \\ &= P\left(Y \leq \frac{x^3}{8}\right) \\ &= F_Y\left(\frac{x^3}{8}\right) \end{aligned}$$

2.

$$\begin{aligned} F_X(x) &= F_Y\left(\frac{x^3}{8}\right) \\ &= \frac{x^3}{8} \end{aligned}$$

$f_X(x) = \frac{d}{dx}F_X(x) = \frac{3}{8}x^2$

3. $x \in (0, 2)$

11.2 Exponential Distribution (8.3)

Suppose we have a Poisson Process with rate λ. Let $X =$ time until the next event occurs. X has an exponential distribution.

FIND RANGE, $F(x)$, AND $f(x)$

$x \in (0, \infty)$

$$\begin{aligned}
F(x) &= P(X \leq x) \\
&= P(\text{time to next event} \leq x) \\
&= P(\text{number of events in } (0,x) \geq 1 \\
&= P(Y \geq 1) \qquad Y \sim \text{Poi}(\lambda x) \\
&= 1 - P(Y = 0) \\
&= 1 - \frac{(e^{-\lambda x})(\lambda x)^0}{0!} \\
&= \begin{cases} 1 - e^{-\lambda x}, \ x > 0 \\ 0, \ x \leq 0 \end{cases}
\end{aligned}$$

Alternate forms: $\theta = \frac{1}{\lambda}$, so

$$F(x) = 1 - e^{-\frac{x}{\theta}}$$

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$$

We say $X \sim Exp(\theta)$.

12 Lecture 24

FIND THE MEAN AND VARIANCE

$$E[X] = \int_0^\infty x \frac{1}{\theta} e^{-\frac{x}{\theta}} dx \text{ IBP}$$

Trick: **Gamma Function**

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

where $\alpha > 0$

PROPERTIES OF THE GAMMA FUNCTION

(1) if $\alpha > 1$, then $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$

(2) if α is an integer ≥ 1,

$$\Gamma(1) = 1$$

$$\Gamma(2) = 1\Gamma(1) = 1$$

$$\Gamma(3) = 2\Gamma(2) = 2$$

$$\Gamma(4) = 3\Gamma(3) = 6$$

In general,

$$\Gamma(\alpha) = (\alpha - 1)!$$

So, back to our example:

$$\begin{aligned} E[X] &= \int_0^\infty x\frac{1}{\theta}e^{-\frac{x}{\theta}}dx \qquad y = \frac{x}{\theta} \implies x = (\theta y) \wedge \theta dy = dx \\ &= \int_0^\infty (\theta y)\frac{1}{\theta}e^{-y}\theta dy \\ &= \theta\int_0^\infty y^{2-1}e^{-y}dy \qquad \Gamma(2) = (2-1)! = 1 \\ &= \theta \end{aligned}$$

$$E[X] = \theta = \frac{1}{\lambda}$$

Why? If λ is higher, events happen more often, which means shorter wait time.

To find $Var(X)$,

$$\begin{aligned} E[X]^2 &= \int_0^\infty x^2\frac{1}{\theta}e^{-\frac{x}{\theta}}dx \qquad y = \frac{x}{\theta} \implies x^2 = (\theta y)^2 \wedge \theta dy = dx \\ &= \int_0^\infty (\theta y)^2\frac{1}{\theta}e^{-y}\theta dy \\ &= \theta^2\int_0^\infty y^{3-1}e^{-y}dy \qquad \Gamma(3) = (3-1)! = 2 \\ &= 2\theta^2 \end{aligned}$$

So $Var(X) = 2\theta^2 - \theta^2 = \theta^2$, $SD(X) = \theta = E[X]$

12.1 Memoryless Property

12.2 Example

Suppose busses follow a Poisson process with average 5 per hour.

(a) Find the probability that you wait > 15 mins.

Let $X =$ time until next bus. $X \sim Exp(12)$

$P(X > 15) = 1 - F(X \leq 15) = 1 - \left(1 - e^{-\frac{15}{12}}\right) = e^{-\frac{15}{12}} \approx 0.2865$

(b) If you have been waiting 6 minutes already, what is the probability that you wait another > 15 more minutes.

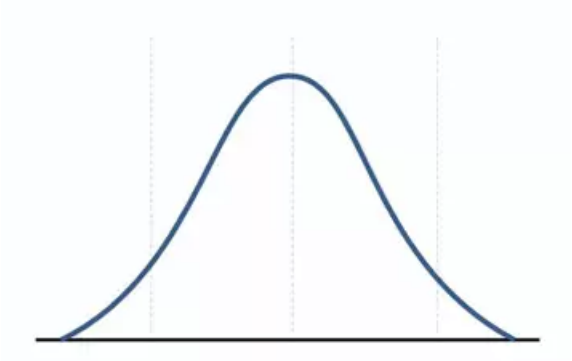
$$\begin{aligned} P(X > 21 \mid X > 6) &= \frac{P(X > 21 \text{ AND } X > 6)}{P(X > 6)} \\ &= \frac{P(X > 21)}{P(X > 6)} \\ &= \frac{1 - F(21)}{1 - F(6)} \\ &= \frac{1 - (1 - e^{-\frac{21}{12}})}{1 - (1 - e^{-\frac{6}{12}})} \\ &= e^{-\frac{15}{12}} \approx 0.2865 \end{aligned}$$

The memoryless property says the past is irrelevant in the future distribution. In general, if $s, t > 0$:

$$P(X > t + s \mid X > s) = P(X > t)$$

12.3 Normal Distribution (8.5)

Many natural phenomena tend to follow a shape like this:



- amount of precipitation
- heights/weights of large populations
- measurement errors
- grades in courses

A normal rv X with parameters μ and σ^2 has pdf

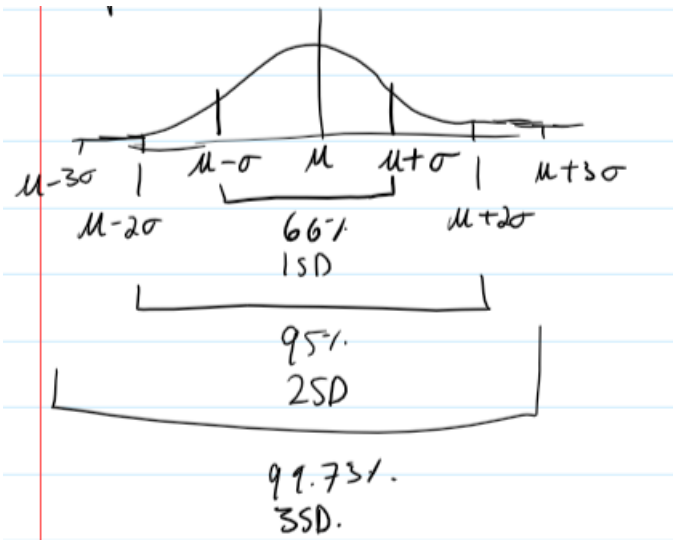
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$

for $x \in \mathbb{R}$

- symmetric around μ
- both tails go to zero quickly
- $\frac{1}{\sqrt{2\pi}\sigma}$ makes it integrate to 1.

We can show that $E[X] = \mu$ and $Var(X) = \sigma^2$

12.4 Empirical rule



FIND $f(x)$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(u-\mu)^2/(2\sigma^2)} du$$

- not analytically integrable
- look it up or numerically evaluate

Standard Normal rv (special case with $\mu = 0$, $\sigma^2 = 1$)

$Z \sim N(0, 1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$F(z)$ still has no closed form.

13 Lecture 25

Recall $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$

Standard normal: $Z \sim N(0, 1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

No integrable cumulative distribution function, use table.

FIND $P(Z \leq 2.31)$

$$P(Z \leq \underbrace{2}_{\text{row}} . \underbrace{31}_{\text{col}}) = 0.98956$$

FIND $P(Z \leq -0.63)$

$$P(Z \leq -0.63) = P(Z > 0.63) = 1 - P(Z \leq 0.63) = 1 - 0.73565 = 0.26435$$

13.1 Example

A voltage of $+2$ means 1 and -2 means 0. The connection adds a $N(0,1)$ distribution amount of noise. If they receive a voltage of > 0.5, they interpret as 1, < 0.5 as 0.

FIND P(ERROR) IF A 1 WAS SENT Let $R =$ received $= 2 + Z$. $Z \sim N(0,1)$

$$\begin{aligned} P(\text{error}) &= P(R < 0.5) \\ &= P(2 + Z < 0.5) \\ &= P(Z < -1.5) \\ &= P(Z > 1.5) \\ &= 1 - P(Z \leq 1.5) \\ &= 1 - 0.93319 \\ &= 0.06681 \end{aligned}$$

Similarly P(error) if 0 was sent: $R = -2 + Z$,

$$\begin{aligned} P(\text{error}) &= P(R > 0.5) \\ &= P(-2 + Z > 0.5) \\ &= P(Z > 2.5) \\ &= 1 - P(Z \leq 2.5) \\ &= 1 - 0.99379 \\ &= 0.06621 \end{aligned}$$

FIND PERCENTILES OF $N(0,1)$

Suppose we want c such that $P(Z < c) = 0.85$

- look in body of table for ≈ 0.85 and read off row and column: c is between 1.03 and 1.04
- use reverse table, look up row and column: 1.0364

TRANSFORMING A NORMAL RV

Suppose $X \sim (\mu, \sigma^2)$, $\mu, \sigma^2 < \infty$.

Claim: if

$$Z = \frac{X - \mu}{\sigma}$$

then $Z \sim N(0,1)$

Proof.

1.

$$\begin{aligned} F_Z(z) &= P(Z \le z) \\ &= P\left(\frac{X-\mu}{\sigma} \le z\right) \\ &= P\left(X \le z\sigma + \mu\right) \\ &= F_X(\sigma z + \mu) \end{aligned}$$

2. Differentiate

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} F_Z(z) \\ &= \frac{d}{dz} F_X(\sigma z + \mu) \\ &= f_X(\sigma z + \mu)\sigma \qquad \text{Chain Rule} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-((\sigma z+\mu)-\mu)^2/(2\sigma^2)}\right)\sigma \\ &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \end{aligned}$$

3. range of Z is $\mathbb{R}$, so $Z \sim N(0,1)$

13.2 Example

MCAT scores are normal with mean 25.3 and standard deviation 6.5.

A score of 41 is how good?

Find $P(X > 41)$ where $X \sim N(25.3, 6.5^2)$

$$P\left(\frac{X-25.3}{6.5} > \frac{41-25.3}{6.5}\right) = P(Z > 2.42) = 1 - 0.99202 = 0.00798$$

13.3 Example

You want 98% of the population to use a ride by height. $X =$ height $\sim N(69, 2.4^2)$. That is, find h such that $P(X < h) = 0.98$, so

$$P\left(\frac{X-69}{2.4} < \frac{h-69}{2.4}\right) = 0.98 \implies P\left(Z < \frac{h-69}{2.4}\right) = 0.98$$

Set $F(\frac{h-69}{2.4}) = 0.98$, and solve for h. You can also take F^{-1} on each side.

$$2.0537 = \frac{h-69}{2.4} \implies h = (2.0537)(2.4) + 69 = 73.93 \text{ inches}$$

In general,

$$x_p = \sigma z_p + \mu$$

14 Lecture 26

14.1 Example

If $Z \sim N(0,1)$, find d such that $P(|Z| < d) = 0.9$.

$$\begin{aligned} P(-d < Z < d) &= 1 - P(-d \geq Z \geq d) \\ &= 1 - P(Z \leq -d) \\ &= 1 - P(Z > d) \\ &= 1 - [1 - P(Z \leq d)] \\ &= P(Z \leq d) \end{aligned}$$

$$P(Z \leq d) = 0.9$$

So $d = 1.286$.

14.2 Basic Terminology and Techniques (9.1)

We have models for a single RV (both discrete or cts) but we often care about two or more RV's at the same time (and their relationship) Examples:

- two stock returns
- heights and weights
- number of cards of a rank vs number of a suit
- treatment vs recovery time
- all machine learning classification and regression

In this course, we focus on all discrete random variables

> The joint probability function of two random variables X and Y is
>
> $$f(x,y) = P(X = x, Y = y)$$
>
> for all (x,y) in the joint range.

14.3 Example

Suppose we flip a coin 3 times. Let $X = \#$ heads. Let

$$Y = \begin{cases} 1, & \text{if first flip is a H} \\ 0, & \text{otherwise} \end{cases}$$

Find $f(x,y)$.

$y \backslash x$	0	1	2	3
0	$1/8$	$2/8$	$1/8$	$0/8$
1	$0/8$	$1/8$	$2/8$	$1/8$

$f(x,y)$ can be represented in a table or as a function of x and y (not usually a histogram).

In general, the joint pf of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

Properties:

- $\sum_x \sum_y f(x,y) = 1$
- $f(x,y) \geq 0$ for all (x,y)

Now suppose we are only interested in one of the random variables. e.g. suppose we are only want to find out about X.

$$P(X = x) = f(0,0) + f(0,1) = \frac{1}{8} + 0 = \frac{1}{8}$$

In general, we define the marginal pf of X as

$$f_X(x) = \sum_y f(x,y) = P(X = x)$$

and the marginal pf of Y is

$$f_Y(y) = \sum_x f(x,y) = P(Y = y)$$

Two discrete random variables X and Y are independent iff

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

$$f(x,y) = f_X(x)f_Y(y)$$

for all (x,y)

From example: Are X and Y independent? No. $f(0,0) = \frac{1}{8} \neq f_X(0)f_Y(0) = \frac{1}{8} \cdot \frac{1}{2}$ Shortcut: any 0 in your table $\rightarrow$ dependent.

15 Lecture 27

Recall (on overhead):

- joint and marginal pfs, independence
- conditional pfs
- functions of two (or more) random variables

Questions from ch 5,7,9 (1 question)

Recall (on board):

joint pf $f(x,y) = P(X = x, Y = y)$ (non negative, sum up all pairs $x, y = 1$)

marginal pf:

$f_X(x) = \sum_y f(x,y)$

$f_Y(y) = \sum_x f(x,y)$

independence: two variables are independent iff

$$f(x,y) = f_X(x)f_Y(y) \qquad \forall (x,y)$$

15.1 Thought Question

For a full-time UW Math Faculty student, let $X =$ number of courses taking and $Y = 1$ if in co-op, or 0 if in regular. The joint pf is given by (this is real data)

$y\backslash x$	3	4	5	6	$f_Y(y)$
0	0.09	0.17	0.22	0.01	
1	0.05	0.10	0.32	0.04	0.51
$f_X(x)$			0.54		1

Are X and Y independent?

a) Yes, b) No, c) Not enough information

Correct answer is b): No. $f(5,1) = 0.32 \neq f_X(5)f_Y(1) = (0.54)(0.51) = 0.2754$

15.2 Example

Imagine you have a card game with a total of 12 cards. Classified in three different categories: 5 cards (money), 4 cards (action), 3 cards (useless). Draw a hand of them, in this case 3 without replacement, and let $X = \#$ of useless, $Y = \#$ action.

FIND THE JOINT PF AND RANGE

$y\backslash x$	0	1	2	3	$f_Y(y)$
0	$10/220$	$30/220$	$15/220$	$1/220$	$56/220$
1	$40/220$	$60/220$	$12/220$	0	$112/220$
2	$30/220$	$18/220$	0	0	$48/220$
3	$4/220$	0	0	0	$4/220$
$f_X(x)$	$84/220$	$108/220$	$27/220$	$1/220$	1

Range: $x \in \{0,1,2,3\}$, $y \in \{0,1,2,3\}$ such that $x + y \leq 3$

$f(0,0)$ (no useless, no action)$=P$(all money)

$$\frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}$$

$f(1,1)$ (1 useless, 1 action)$=P$(one of each type)

$$\frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220}$$

$$f(x,y) = \frac{\binom{3}{x}\binom{4}{y}\binom{5}{3-x-y}}{\binom{12}{3}}$$

Find marginal pfs (sum), $X \sim \text{Hyp}(12,3,3)$. $Y \sim \text{Hyp}(12,4,3)$. Check that the marginal pfs match.

Are they independent? No (don't have a cartesian product)

Recall: conditional probability:

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

15.3 Definition (Conditional pf)

The conditional pf of X given $Y = y$ is

$$f(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x, y)}{f_Y(y)}$$

for all y.
Similarly, the conditional pf of Y given $X = x$ is

$$f(y \mid x) = \frac{f(x, y)}{f_X(x)}$$

for all x.

15.4 Example

What is the probability that someone taking 4 courses is a co-op student?

In other words, $P(Y = 1 \mid X = 4) = \frac{0.1}{0.27} = 0.37$

For 6 courses, $\frac{0.04}{0.05} = 0.80$.

15.5 Example

If you have 1 action card, find the pf of the number of useless cards.

i.e. the pf of $X \mid Y = 1$

x	0	1	2
$f(x \mid 1)$	$40/112$	$60/112$	$12/112$

15.6 Functions of 2 or more random variables

Suppose U is some function of X and Y, e.g. $U = X - Y$. To find the pf of U.

$y \backslash x$	3	4	5	6
0	3	4	5	6
1	2	3	4	5

1. determine the possible values of U for each pair (x, y)

so the range is $u \in \{2, 3, 4, 5, 6\}$

2. $f(u)$ is the sum of $f(x, y)$ for all combos that map to u.

$$f(u) = \sum_{(x,y) \text{ s.t. } x-y=u} f(x, y)$$

u	2	3	4	5	6
$f(u)$	0.05	0.19	0.49	0.24	0.01

$u = 2 \to (x = 3, y = 1) = 0.05$

$u = 3 \to (x = 4, y = 1) + (x = 3, y = 0) = 0.1 + 0.09 = 0.19$

Using the earlier table:

$y \backslash x$	3	4	5	6
0	0.09	0.17	0.22	0.07
1	0.05	0.1	0.32	0.04

16 Lecture 28

Recall

- joint, marginal, conditional pfs, functions of rvs
- multinomial distribution (9.2)
- joint pf
- marginal pfs
- conditional pfs
- pf of sum
- MLIW document content analysis

16.1 Thought Question

Suppose $X = \#$ apple products and $Y = \#$ Microsoft products (given at least one of each) have a joint pf:

$y \backslash x$	1	2	3
1	0.30	0.17	0.20
2	0.17	0.10	0.06

Find $P(X + Y = 4)$

a) 0.10, b) 0.20, c) 0.30, d) 0.40, e) none

Correct answer is c): $P(X + Y = 4) = (3, 1) + (2, 2) = 0.20 + 0.10 = 0.30$

16.2 Sums of rvs

Suppose $T = X + Y$, and X, Y are non-negative.

The range of T is $0, 1, \ldots, \max(X) + \max(Y)$ pf of T is

$$\begin{aligned} f_T(t) &= \sum \sum_{x+y=t} f(x, y) \\ &= f(0, t) + f(1, t-1) + f(2, t-2) + \cdots + f(t, 0) \\ &= \sum_{x=0}^{t} f(x, t-x) \end{aligned}$$

If X and Y are independent, then

$$f_T(t) = \sum_{x=0}^{t} f_X(x) f_Y(y)(t-x)$$

This can be used to prove:

- sum of two independent Poission is a Poission rv
- sum of k independent $\mathrm{Geo}(P)$ is $\mathrm{NB}(k, p)$

16.3 Multinomial Distribution (9.2)

An extension of Binomial, where each independent trial can have k possible outcomes.

The probability of type i is p_i which is constant.

$$p_1 + p_2 + \cdots + p_k = 1$$

We do n trials and let $X_i = \#$ of outcome i's that occur.

$$X_1 + X_2 + \cdots + X_k = n$$

where n is the total number of trials.

Then we say $X_1, \ldots, X_k \sim \text{Mult}(n, p_1, p_2, \ldots p_k)$.

Remark 5. X_k can be written as $n - \sum_{i=1}^{k-1} x_i$ and p_k can be written as $1 - \sum_{i=1}^{k-1} p_i$

16.4 Example

Roll a fair 6-sided die 10 times. $X_1 = \#$ 1's $X_2 = \#$ composites (4,6) $X_3 = \#$ primes (2,3,5)

Find range: $X_i \in \{0, \ldots, n\}$ $n = 10$ in this case. So,

$$X_1 + X_2 + X_3 = 10$$

Find joint pf: $f(x_1, x_2, x_3) = P(X_1 1's, X_2 C's, X_3 P's)$. So,

$$\underbrace{\frac{10!}{x_1!x_2!x_3!}}_{\text{arrangements}} \underbrace{\left(\frac{1}{6}\right)^{x_1}\left(\frac{2}{6}\right)^{x_2}\left(\frac{3}{6}\right)^{x_3}}_{\text{outcomes}}$$

In general,

$$f(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

for $x_1 + \cdots + x_k = n$ OR

$$f(x_1, \ldots, x_{k-1}) = \frac{n!}{x_1! \cdots x_{k-1}!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}}$$

for $x_1 + \cdots + x_{k-1} \leq n$

Find marginal pf of x_1.

$$\begin{aligned}
f_1(x_1) &= \sum_{x_2}^{x_3} f(x_1, x_2, x_3) \\
&= \sum_{x_2=0}^{10-x_1} f(x_1, x_2) \\
&= \sum_{x_2=0}^{10-x_1} \frac{10!}{x_1!x_2!(10-x_1-x_2)!}\left(\frac{1}{6}\right)^{x_1}\left(\frac{1}{3}\right)^{x_2}\left(\frac{1}{2}\right)^{(10-x_1-x_2)} \\
&= \frac{10!}{x_1!(10-x_1)!}\left(\frac{1}{6}\right)^{x_1} \sum_{x_2=0}^{10-x_1} \frac{(10-x_1)!}{x_2!(10-x_1-x_2)!}\left(\frac{1}{3}\right)^{x_2}\left(\frac{1}{2}\right)^{(10-x_1-x_2)} \\
&= \binom{10}{x_1}\left(\frac{1}{6}\right)^{x_1} \sum_{x_2=0}^{10-x_1} \binom{10-x_1}{x_2}\left(\frac{1}{3}\right)^{x_2}\left(\frac{1}{2}\right)^{(10-x_1-x_2)} \\
&= \binom{10}{x_1}\left(\frac{1}{6}\right)^{x_1}\left(\frac{1}{3}+\frac{1}{2}\right)^{10-x_1}
\end{aligned}$$

$$f(x_1) = \binom{10}{x_1} \left(\frac{1}{6}\right)^{x_1} \left(\frac{5}{6}\right)^{10-x_1} \sim \text{Bin}(10, 1/6)$$

In general:

$$X_i \sim \text{Bin}(n, p_i)$$