(Discrete) Probability Distributions Exercises

Pages	Suggested Reading
159 – 164	Sections 4.1, 4.2, and 4.3

Pages	Problems
184 – 193	(***Section 4.15***) Note: "Expected Value" is just another term for "average". It's a fun and appropriate term, since it's the value you would "expect" when you run an experiment. 1(a,b), 2, 3, 4, 5, 6, 38, 40, 42

E1. Games are called "statistically fair" if their expectations are equal to zero (that way, neither the house nor the player loses, nor gains, any money). Refer back to the probability distribution from Chuck – A – Luck that we used in class, and adjust the payout for a roll of three 5's so that the game is statistically fair (leave the other 3 payouts at -\$1, \$0, and \$1).

E2. Refer back to the Chuck – A – Luck discussion from class. This time, I want you to keep the same payouts (nothing for no 5's, \$1 for one 5, \$2 for 2 fives, and \$10 for 3 fives), but adjust the ***ante*** to make the game statistically fair.

A Symantec study in 2010 found that about 90% of all email sent is spam. But why? Well, I decided to do a little digging, and here's what I found:

- In America, \$8 billion "earned" annually by spammers lowball from many estimates…no one seems to know)
- America has about 150 million email users, of whom 12% (or, 18.5 million) knowingly "click through" spam links (Radicati Group, Messaging Anti-Abuse Working Group, Internet World Stats)

So, the average cost of loss per American, per spam link clicked, is around \$430

- 90 ***trillion*** emails are sent every year (Radicati group, 2010).
- 90% of all email is spam (Symantec et.al.).
- America accounts for about 12% of internet usage worldwide (Internet World Stats)

So, around 10 trillion spams are sent to America every year.

Here's a T – table probability distribution of the above facts (assuming random distribution of spam email):

X = amount "earned", per spam email, in America	P(x)
\$430	$\frac{18.5 \text{ million}}{10 \text{ trillion}} \approx 0.00000185$
\$0	$1 - \frac{18.5 \text{ million}}{10 \text{ trillion}} \approx 0.99999815$

E3. Find the average amount a spammer earns per spam, based on this distribution.

E4. "Spammer X" released a book where he documented his 5 years of spamalicious behavior. In his book, he claims that he sent 40 million spam emails a week, 50 weeks a year. How much revenue did he earn in 5 years, assuming he made the average amount per spam that you found in E3?

E5. Now, he also had some bills to pay (hosting, bots, etc.). These amounted to about \$11,000 per week. Assuming his career of 250 weeks, how much profit did he clear after expenses?

Answers.

E1. and **E2**. Can't help you, as we randomly generated the probabilities in class. Feel free to ask, though!

E3. ≈\$0.0008 per spam email sent.

E4. About \$8 million.

E5. Just over \$5.2 million. No ***wonder*** he retired.

(Discrete) Probability Distributions Quizzes

Quiz 1.

Let's say you and I play the following game: You flip a coin four times. If you get 4 heads or 4 tails, I'll give you \$1. If you flip exactly three heads or exactly 3 tails, you give me \$1. If you flip 2 of each, neither of us pays the other anything.

(**2 points**) Compete the following distribution (grayed boxes). To get the probabilities, I'll give you the entire sample space for the experiment!

So, you can see that the probability of, say, two heads and two tails is 6 out of 16, or 3/8. OK! Have at it!

Event	**Payout**	**Probability**
4 heads		
3 heads, 1 tail		
2 heads, 2 tails		
1 head, 3 tails		
4 tails		

1. (**2 points**) Find the mathematical expectation (AKA, average) of this distribution.

2. (**2 points**) What does it mean?

3. (**4 points**) (**w**) If a game's distribution has an expectation of 0, it is said to be "fair". What should your friend pay you when you roll 4 heads or 4 tails, to make this game fair? Assume you still have to pay him a dollar if you roll 3 and 1, and neither of you pays the other for 2 and 2. Hint: Create an equation, using the idea of a weighted mean. Or, if you can't stand algebra, solve it a different way…but describe ***exactly*** what you did.

Quiz 2.

Refer to problem E2 above. Let's answer it!

I've started you off below – I've added the probabilities of matching X "4's" in three rolls of a die (they're a little different than what we got in class, most likely – we used a Monte Carlo simulation, and, here, I used the binomial distribution. Soon, we'll learn that together!). I'm letting the ante (that we're solving for) be "A"…therefore, if you bet on 4, and no 4's come up, your "winnings" would be "**–A**" (as indicated in the distribution). Likewise, if you get one 4, you'll earn back $1, so your profit for that round would be "**1 – A**".

You finish it up!

(2 points for each of the remaining 2 grayed boxes; 6 points for solving for A, in any way you can!) (w) Complete the distribution below (last two grayed boxes) and then answer E2! I'd (personally) recommend creating an equation using the property that $\mu = \Sigma x \cdot p(x)$.

(# of 4's observed)	X = winnings ($)	P(X)
0	– A	0.578704
1	1 – A	0.347222
2		0.069444
3		0.00463

Quiz 3.

I've mentioned (a couple of times, I think) something called the "Gambler's Ruin" in class.

a. (**2 points**) **Google that phrase, and give me a one – sentence synopsis (no math yet – just the big idea).**

Now, I'd like to illustrate the idea of the Gambler's Ruin. Here's a situation that models it very well – you walk into a casino with a wad of money, and you change it all for $1 chips. You then proceed to a craps table with a minimum bet of $1, lay a $1 chip down, and make what's called a "pass line" bet…such a bet has a 244/495 chance of winning, and a 1:1 payout (you might want to Google this if you're interested in why – or ask me. I'd be happy to show you! It's the fairest bet on a game of chance in Vegas[a]).

Suppose you win your first bet (I mean, heck – there's almost a 50% chance you will, right?). Congrats! The dealer gives you $2 back, and you've profited $1 (said another way, it cost you $1 to make $1). Not bad! But, suppose you lose – what do you do? Well, to model the Gambler's Ruin, let's suppose that whoever's playing continues to play, but he also really wants to win $1. So, after losing the first round (and, thus, being in the hole $1), he bets $2 (so he's now in the hole $3 total). However, if he wins on the next round, he'll get $4 back from the dealer. $3 of that will pay off his accumulated debt, and he'll have $1 left in profit. So, this time, it cost $3 to earn $1.

Assuming he won on the second roll. What if he didn't? He'll double his bet, again, to $4. He's now in the hole $7, but, if he wins on the next round, he'll get $8 from the dealer, and profit $1.

I think you see the pattern now. It's a surefire way to win $1!

Number of Games It Takes To Win	Accumulated Costs	Winnings on Round n	Profit
1	$1	$2	$1
2	$3	$4	$1
3	$7	$8	$1
4	$15	$16	$1
5	$31	$32	$1
6	$63	$64	$1
7	$127	$128	$1
8	$255	$256	$1
9	$511	$512	$1
10	$1023	$1024	$1
…	…	…	$1
n	$\$1 + \$2 + \$4 + \ldots + \$2^{n-1}$	2^n	$1

This is an example of a distribution called the **geometric** – it's like the binomial without a fixed number of trials. That's why, sometimes, statisticians deal with PDFs (**probability density functions**) when representing distributions – with infinite ones, like this, it's easier to create an equation to represent a distribution than to infinitely write tables.

So, what we'll now do is calculate, on average, how much it costs to win $1 on a pass line bet in craps using the above betting strategy. Here's the distribution we'll use:

[a] **FYI**: "fairest" doesn't mean "fair". ☺

Number of games?	Amount it Costs to Win $1	P(X)
1	**$1**	**0.492929**
2	**$3**	**0.24995**
3	**$7**	**0.126742**
4	**$15**	**0.064267**
5	**$31**	**0.032588**
6	**$63**	**0.016524**
7	**$127**	**0.008379**
8	**$255**	**0.004249**
9	**$511**	**0.002154**
10	**$1023**	**0.001092**

b. (**3 points**) Find the average of this distribution (that is, the average amount it will cost you to win back your dollar within 10 games).

Now, that's not exactly mathematically correct, is it? Technically, the number of possible games doesn't stop at 10…let's let it go a bit further (it might take you more games to win!):

Number of games?	Amount it Costs to Win $1	P(X)
1	**$1**	**0.492929**
2	**$3**	**0.24995**
3	**$7**	**0.126742**
4	**$15**	**0.064267**
5	**$31**	**0.032588**
6	**$63**	**0.016524**
7	**$127**	**0.008379**
8	**$255**	**0.004249**
9	**$511**	**0.002154**
10	**$1023**	**0.001092**
11	**$2047**	**0.000554**
12	**$4095**	**0.000281**
13	**$8191**	**0.000142**
14	**$16383**	**7.22E-05**
15	**$32767**	**3.66E-05**
16	**$65535**	**1.86E-05**
17	**$131071**	**9.42E-06**
18	**$262143**	**4.77E-06**
19	**$524287**	**2.42E-06**
20	**$1048575**	**1.23E-06**

c. (**3 points**) Find the average of ***that*** distribution (that is, the average amount it will cost you to win back your dollar within 20 games).

But technically, ***that's*** not right, either. The distribution, technically, would have to go on forever, and ever…because, technically, even though it's unlikely to need to take 20 games (and, therefore, spend over a million dollars) to win back $1 – it's still *possible*.

Let's look at it a different way – here's an old school look at the calculation of the average of the previous distribution (like we've done a few times before):

X	P(X)	X*P(X)
$1	0.492929	**$0.492929**
$3	0.24995	**$0.74985**
$7	0.126742	**$0.887196**
$15	0.064267	**$0.96401**
$31	0.032588	**$1.01023**
$63	0.016524	**$1.041041**
$127	0.008379	**$1.064142**
$255	0.004249	**$1.083439**
$511	0.002154	**$1.100915**
$1023	0.001092	**$1.117576**
$2047	0.000554	**$1.133934**
$4095	0.000281	**$1.15025**
$8191	0.000142	**$1.166659**
$16383	7.22E-05	**$1.183229**
$32767	3.66E-05	**$1.199998**
$65535	1.86E-05	**$1.216987**
$131071	9.42E-06	**$1.234206**
$262143	4.77E-06	**$1.251664**
$524287	2.42E-06	**$1.269367**
$1048575	1.23E-06	**$1.287319**
…	…	**…**

Now remember that, to find the average like this, we use the weighted mean formula $\mu = \Sigma X \cdot P(X)$. That means that we have to add up each of those elements, one by one, in the "X*P(X)" column at left. See how, the farther down we go, the larger they get?

So, in other words, it appears that if we continue to add more and more numbers of games (as we have to do to be mathematically correct, we will continually add monetary amounts, each getting slightly bigger, forever.

But then, we'll never be done adding these amounts.

d. (**2 points**) So then…what's the average amount it will take us to win back our $1, if we use the complete, infinite distribution[b]? Hint: it's not a number – more of an "*abstract idea describing something without any limit*".

(Wikipedia's words, not mine – but they're pretty good words).

And ***that***, my friends, is the Gambler's Ruin.

---

[b] For two extra points, come up with the probability density function (PDF) for this distribution!

Quiz 4.

Ever flipped a coin to decide something? Great! Has it ever gone something like this?

<u>You</u>: "Let's flip to decide who takes the trash out!"
<u>Friend</u>: "OK!"
<u>You</u>: "Call it in the air!"
(*you flip*)
Friend: "Heads!"
(*it lands "tails"*)
<u>Friend</u>: "Best of 3!"

Happens with my son quite a bit. ☺ I've started to wonder, though – if you give in to your friend, what's the chance they'll win in a best of 3, knowing that they're already one flip down?

a. (**6 points**) Well, what do you say? Answer fully, explaining each step of the way. If you had me for MTH 105, this would be a great time to utilize tree diagrams!

b. (**2 points**) Suppose you play 3 games, and your friend loses 2 – 1 (suppose the order went, "You won, they won, you won"). (S)he class out, "Best of 5!" What's the chance (s)he wins, if you agree?

c. (**2 points**) Suppose, in the original "best of 3", you won the first two games (thereby winning "best of 3". If you friend calls out "Best of 5!" and you agree, what's the chance (s)he wins now?

Quiz 5.

I like lottery games. I think it's because my Grandpa always played those little scratch off games, and used to let me do the scratching. As I've gotten older, I've realized they're rigged…but there's always something kinda cool about what's going on behind the scenes. In this quiz, we'll analyze one of the biggies: PowerBall!

In case you've never heard of Powerball, here's the gist: you go to a store that sells PowerBall tickets and give the clerk either $1 (ante for a traditional PowerBall game) or $2 (ante for the "Power Play" PowerBall game). Then, you get your ticket (either by telling the clerk 5 numbers and a 6th PowerBall number, or, as many do, by having the clerk randomly generate 6 numbers of you). There's an example of one at right – this person bet on the 5 numbers 7, 22, 28, 37, and 45, and a "PowerBall" number 14.

Then you wait for the next drawing and see how your ticket did! Depending on how many (and of which type of) numbers you match, you win varying prize amounts, as follows:

Match	Prize $	Prize $ w/ Power Play	Odds 1 in:
○○○○○●	Jackpot	Jackpot	175,223,510
○○○○○	$1,000,000	$2,000,000	5,153,633
○○○○●	$10,000	$40,000	648,976
○○○○	$100	$200	19,088
○○○●	$100	$200	12,245
○○○	$7	$14	360
○○●	$7	$14	706
○●	$4	$12	110
●	$4	$12	55

This looks like a probability distribution to me! Hold on – lemme make it look better in Excel!

K! Go ahead and open the sheet "PowerBall" that accompanies this quiz. Make sure you're on the "normal" tab (this is the $1 game). In this sheet, I've replicated the above par sheet into a fairer form – because I also included the cells where you lose!

1. (**1 point**) What would the "X" and "P(X)" values be for the losing cells? That is, what values are in cells C12 and D12?

OK! Now, what we're really interested in is how much money we lose, on average, on a PowerBall ticket. But, in order to do this, we need to know what the jackpot value is. Today (11.11.15) it's **50 million dollars**. Go ahead and place that value in cell C3.

Our distribution is complete! Now, we can average it – in class, we did this by evaluating $\mu = \sum X * P(X)$, and that's totally cool – but Excel has a great command called "**sumproduct**" that'll do it for us! Check it out! Type the following formula in cell H3:

=SUMPRODUCT(C3:C12,D3:D12)

Don't forget the equals sign!

2. (**1 point**) What's the average?
3. (**2 points**) What does it mean?

OK! Now, let's analyze the "PowerPlay" option! Click over to the "PowerPlay" tab, fill in the "losing cell'" values (remember the Power Play game costs $2), and calculate the average of THAT distribution!

4. (**1 point**) What's the PowerPlay average?
5. (**3 points**) How can it be that the average is lower in the PowerPlay when they're giving away larger prizes?!?!

One last interesting point – I've noticed that, occasionally, when the jackpot gets ***really*** big, we start hearing about it on the news. For example, the largest PowerBall jackpot in history (as of this writing) was a whopping $590.5 million!

Go back into your "Normal" spreadsheet and replace the jackpot with this $590.5 million!

6. (**1 points**) What's the "normal" average now?
7. (**1 points**) What's the "PowerPLay" average now?

How neat is that! The expectation went ***positive***!

I have a theory – that we, as humans, start "feeling it" right around that time, and that's when we start hearing about PowerBall on the news – right about when the expected value of the tickets switches from negative to positive (right around $112 million).

Of course, it's not all puppies and rainbows, even for the jackpot winners:

- The jackpot rarely gets close to these amounts that make expectations positive; people usually hit them when they're small. The median jackpot is around $36 million.
- Remember that the government takes about 40% of your prize money (anything over $600) due to "taxes"[c].
- Don't forget state "income" tax! Oregon's is 8%.
- Remember also that, historically in PowerBall, when the jackpots get big, more people play, and the jackpot then is split. That means you have to reevaluate the average again for multiple winners.
- Speaking of the "jackpot"...it's horribly inflated. For example, for the current jackpot, the reported value is "$50 million"...but in little type under that, it says "$30.5 million cash value". The reported jackpot is misleading, due to financial reasons I can't understand[d].

So, just like all the other gambling games, they get you whether you win or not. But hey – you can't win if you don't play, right? ☺

(*this quiz was written before October, 2015. That'll be pertinent on the followup quiz*!)

---

[c] I've often considered this to be one of the most bogus acts in government. The $1 you played with was already taxed. I can't believe they can hit you again, because they're considering the lottery win an "investment". And you can't claim the ticket price as an investment ***loss***. Geez...

[d] The same reason I can't do my own taxes, I'm sure.

Quiz 6.

Thursday, January 14, 2016.
Powerball mania reached a fever pitch Wednesday, with at least three winning tickets finally being sold in the largest jackpot in U.S. history. The lucky overnight mutli-millionaires will be splitting a whopping $1,586,400,000, blowing the previous record of **$656 million** out of the water. Jan 14, 2016

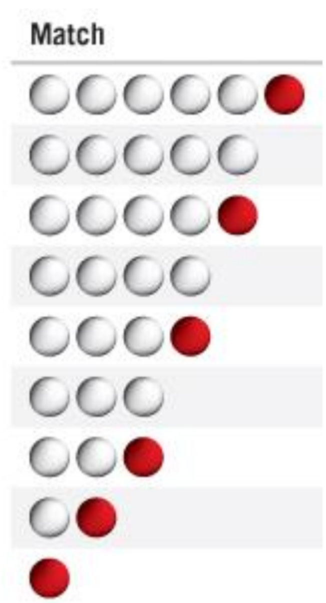
Meet the lucky winners of the largest lottery jackpots in U.S. ...
abc7.com/society/meet...**winners**...**largest-lottery-jackpots**...**history**/513434/

Feedback

So, if you did the PowerBall quiz last week, you might have chuckled at my use of the word "whopping" when I referred to the then-unheard-of jackpot prize amount of right around half a billion dollars. Unbeknownst to me, at almost the exact same time that I was creating that quiz (October 2015), the folks that run the PowerBall game were scheming up a new way to change the game. Historically, the game has changed over the years – the basic premise is the same (5 white balls and one red), but the number of balls from which these are drawn has gone through some changes:

Year	Total White Balls	Total Red Balls	Total Number of Tickets[e]
1992	45	45	54,979,154
1997	49	42	80,089,127
2002	53	42	120,526,769
2005	55	42	146,107,961
2009	59	39	195,249,054
2012	59	35	175,223,510
2015	69	26	292,201,338

Notice the whopping number of tickets that are now possible (about 67% more!) That explains the huge buildup in time (and money) for the most recent whopper jackpot. I'd expect more. ☺



From Powerball's "How To Play" website:

"This change (<u>note</u>: *the 2015 change*) means bigger jackpots. It also means that your odds of winning some prize are improved."

1. (**1 point**) Do a little Googling and figure out ***which*** of these prizes' odds improve. I've given an image of the prizes at right; if you like, you could copy and paste this image into an editing program (like Paint) and circle them electronically.

Like last time, I've created a spreadsheet based on these new probabilities. Go ahead and open it – if you did the quiz last week, it'll look familiar. If not, you'll still be able to get the gist. Make sure you're in the "normal" tab for starters.

---

[e] If you're wondering where these come from, apply the Fundamental Counting Principle from Project #3. Or, you know – come ask me. ☺

Once you're there, do these things!

- Today (2.20.16) the jackpot value[f] is $**212 million**. Go ahead and place that value in cell C3.
- Fill in the values for cells C12 and D12. Note: the "normal" game costs **$2**

2. (**3 points**) **Based on the probability that you just placed in cell D12, do you agree with this PowerBall claim: "The overall odds of winning a prize are 1 in 24.87"? Why or why not, in one or two sentences.**

Just like last time, type the following into cell H3: **=SUMPRODUCT(C3:C12,D3:D12)**

Don't forget the equals sign! And, just like last time, let's also analyze the "PowerPlay" option (once again, I'm using the 2x multiplier). Click over to the "PowerPlay" tab and calculate the average of THAT distribution!

3. (**2 points**) **Use the two averages you just found to fill in the following chart** (**grayed cells**):

Average	**Pre – October 2015**	**Post October 2015**
Normal	**– $0.43**	
PowerPLay	**– $0.93**	

Amazing. Just amazing. Look at how much money they make on these tickets!

One last thing – in the "normal" sheet, plug in the value $1,586,400,000 in of the jackpot.

4. (**2 points**) **What does the average change to now?**

5. (**2 points**) **What makes this average different from the ones above? And what would this big change mean? Make sure you use the word "average" in your explanation!**

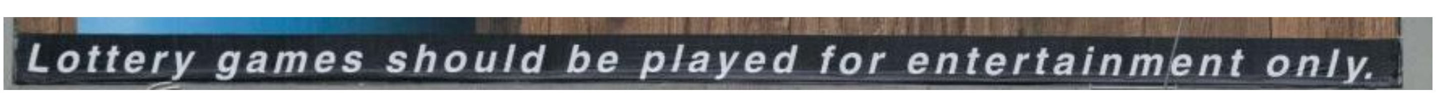
Hey! That's cool, right? We talked about this in class – how those kids from New York beat that Keno – like game! Let's just get a whole bunch of people together and buy a whole bunch of tickets!

6. (**3 extra points**) **Suppose we were able to get folks to go into this scheme – that is, with the PowerBall jackpot at the $1,586,400,000, we convinced people to buy, say, a million tickets (so we're out $2 million at the start). How much would we walk away with, on average, after our investment scheme?**

What the?!? Sounds amazing!

But remember – one big, big reason that the NY kids' plan worked so well was that they were winning small prizes that didn't need to be taxed. In Oregon, any prize over $5000 gets a 25% tax hit from the feds and a state hit of 8%. I went ahead and reran the updated prize numbers for ya in the tab labeled "Don't Forget Taxes!". I also divided that huge jackpot by 3, since 3 people hit it.

Yep – it's like they (should) always say:



[f] As I'm typing this (finishing it up on 2.21.16), the jackpot increased to more than 20 million dollars. Get ready, kiddos.