Chan and Khandani
Random Variables

Introduction

A lot of the time, we don't actually care too much about the outcome of a single event - whether a specific person does or doesn't catch swine flu is somewhat inconsequential in the grand scheme of things.

We're more interested in statistics that can be calculated on the sample space itself - things like whether being in a certain location is correlated with higher chances of catching swine flu, or how much total mass a volcano has expelled in the last ten years.

These calculations, or functions, that are defined inside a sample space, are called random variables.

Formal Definition

Let S be the sample space of experiment E.
$X(s)$ is a function that assigns a real number to itself based on EVERY elementory event s inside S.

Since the specific outcome s isn't known before the experiment occurs, $X(s)$ is a random variable.

Usually we use capitals for the random variable itself, and lowercase for a particular value

ex. $X(s) = x$

Function

elementary event result

real number that corresponds to s

A random variable is discrete if it doesn't have an infinite number of possible results.

Probability Mass Function

The probability mass function states the chance a discrete random variable X will take on a certain value A.

$$p(a) = P\{X = a\}$$

And as follows, the sum of probability masses of all possible values should be 1 as well.

$$\sum_i p(x_i) = \sum_i P\{x = x_i\} = 1$$

ex. Let's do the simplest case possible; flipping a coin. We can define the random variable X as

$$X = \begin{cases} 0 : \text{heads} \\ 1 : \text{tails} \end{cases}$$

$$\therefore \; X(\text{head}) = 0$$
$$X(\text{tails}) = 1$$

and as such, the probability mass function of p(0), the probability of X(s) resulting in 0, is just

$$p(0) = 1/2$$

as there's a 50% chance of getting a head.

Cumulative Distribution Function

This one's equation is also really self-explanatory.

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

Translation: What's the probability the random variable X will result in a value less than or equal to x?

$$F(a) = \sum_{x \leq a} p(x)$$

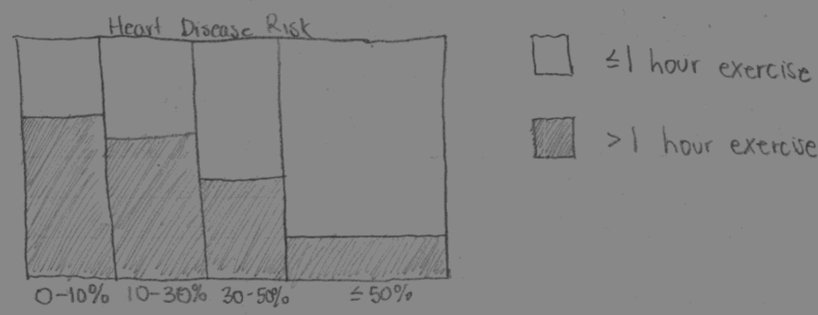
Translation: Again, this is the same. Remember $p(x) = P\{X = a\}$

$$\therefore \quad F(a) = \sum_{x \leq a} P\{X = a\}.$$

and the sum of these probabilities are equivalent just to $P(X \leq a)$.

Why is this useful? Now, we're able to ask questions like: If you exercise for an hour a day, what is the chance that you'll reduce your risk of heart disease to lower than 30%?

ex.



So this sample space indicates heart disease risk according to amount of exercise.

If a is defined as 0.3 (meaning less than or equal to 30% risk), and X is defined as P (heart disease | exercise > 1 hour),

$$F(a) = P(X \leq a)$$
$$F(0.3) = P(X \leq 0.3)$$
$$= P(\text{heart disease} \leq 0.3 \mid \text{exercise} > 1)$$
$$= \frac{\text{area of first two shaded bars}}{\text{total area of shaded bars}}$$

as only the first two correspond to heart disease risk of $\leq 30\%$.

Expected Value

The expected value $E(X)$ is given as

$$\sum_i x_i p(x_i)$$

which sums each possible value of X, multiplied by its probability mass - this is essentially just a weighted average of all possible results to create a value that we would normally expect.

Expectations of Functions of Random Variables

Any function that uses other random variables is a random variable itself.

ex. If X is a random variable, and we have some function g, such that $Y = g(X)$, we can then calculate expected values of Y through the same method.

$$E(Y) = E(g(X))$$
$$= \sum_i g(x_i)\, p(x_i)$$

as $g(x_i)$ is every possible result of Y, and since $g(x_i)$ only occurs as frequently as x_i occurs, it is weighted by the probability mass $p(x_i)$.

ex. Given constants a and b, if $Y = (aX + b)$,

$$E(Y) = E(aX + b)$$
$$= a\,E(X) + b$$

Since b is simply added to aX every time, it has no dependence on the result x of X itself. As such, we can just state that the expected value of Y is a times the expected value of X plus b.

Variance

The variance of a random variable X, denoted $Var(X)$ is given as

$$Var(X) = E((X-\mu)^2)$$
$$= E(X^2) - \mu^2$$

where $\mu = E(X)$.

which is a bit strange when you look at it, because this becomes

$$Var(X) = E(X^2) - (E(X))^2$$

Note that we can only ever receive non-negative numbers from this equation, where 0 would mean every result we get from X is exactly the same.

Variance just tells us how far our values deviate from the average - in fact, the square root of variance is a very well known measure - standard deviation.

And again, we can apply any linear operation:

$$\begin{aligned} Var(aX+b) &= E[(aX+b)^2] - (E(aX+b))^2 \\ &= E(a^2X^2 + 2abX + b^2) \quad \text{binomial expansion} \\ &\quad - (aE(X)+b)^2 \\ &\quad \text{by the formula before} \\ &= a^2E(X^2) + 2ab\cancel{E(X)} + \cancel{b^2} \\ &\quad - a^2(E(X))^2 - 2ab\cancel{E(X)} - \cancel{b^2} \\ &= a^2E(X^2) - a^2[E(X)]^2 \\ &= a^2[E(X^2) - (E(X))^2] \\ &= a^2 Var(X) \end{aligned}$$

It's not really strictly necessary to memorize the proof but it's a fairly straightforward expansion/simplification process.

<u>Standard Deviation</u>

Again, this is just

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

There really isn't much difference between the two; standard deviation ends up being in the same units as the original value, which is kind of convenient.

Bernoulli and Binomial Random Variables

The Bernoulli random variable associated with some event E just states that E only has two outcomes:

$$p(0) = \text{failure} = P\{X=0\} = 1-p$$

and

$$p(1) = \text{success} = P\{X=1\} = p$$

↑ this is it right here, X

The Binomial random variable then gives us a way to calculate the probability of having i successes out of n Bernoulli Trials (independent, repeated events):

$$p(i) = \underbrace{\binom{n}{i}}_{\frac{n!}{i!(n-i)!}} p^i (1-p)^{n-i}, \quad i \leq n$$

$\binom{n}{i}$ gives us the number of ways we can have i successes from n trials. $(p^i)(1-p)^{n-i}$ describes the probability of successes and probability of failures. These are disjoint events, so we can simply multiply them.

Let's do a quick example to illustrate.

ex. We want 10 trials: 3 successes and 7 failures. The chance of success is 10%.

$$p(3) = \binom{10}{3}(0.1)^3(0.9)^7$$
$$= \frac{10!}{7!3!}(0.001)(0.478)$$
$$= 5.74\%$$

Properties of Binomial Random Variables

$$E(X) = \overset{\text{trials}}{n}\underset{\text{probability of success}}{p} \qquad Var(X) = np(1-p),$$

both of which can be found by plugging in $p(i)$ into the respective expected value and variance equations.

This is a long and arduous process, so fuck that, let's just believe what the mathematicians tell us. However, it's not actually difficult to understand why these are correct.

Trials	E(X)	Var(X)
1	0.5	0.25
2	1	0.5
3	1.5	0.75
4	2	1
5	2.5	1.25

Let's take flipping a coin, with $X=1$ as heads, and $X=0$ as tails, where $p = 0.5$.

As we flip the coin, the cumulative sum should be around half the trials, meaning $E(X) = 0.5n$.

And if we flip the coin once, we're going to be, on average, $\sqrt{0.25}$ away from the mean, aka ± 0.5 away from 0.5 on the first flip.

$$\underset{\text{Tails}}{0} \xleftrightarrow{SD} \underset{\text{Mean}}{0.5} \xleftrightarrow{SD} \underset{\text{Heads}}{1}$$

Computing the Cumulative Distribution of a Binomial

A few pages back we had some function that calculated the probability of $X \leq a$, the cumulative distribution function.

Essentially it's just a sum, but

$$P\{X \leq i\} = \sum_{k=0}^{i} \binom{n}{k} p^k (1-p)^{n-k}$$

can be kind of messy, so it's helpful to know that

$$P\{X = k+1\} = \left(\frac{p}{1-p}\right)\left(\frac{n-k}{k+1}\right) P\{X = k\}$$

if you intend on writing anything to compute these values (yay memoization!).

The Poisson Random Variable

So before we actually get into the math, we have to think about and understand what the poisson random variable represents.

The main usage is to count the number of times a certain event occurs in a given period of time. Again, we have to assume the occurrence of an event is independent from every other, and each chunk of time is exactly the same, conditions-wise.

Obviously this is a gross oversimplification of the real world, but we've lived our whole lives doing the exact same, so why not continue?

So maybe we want to count how much mail we get in a day, or the number of earthquakes in a year, or how many times we get woken up by screaming party girls at 4 AM in a week.

Now, back to what the poisson distribution is: it's an APPROXIMATION of the Bernoulli random variable. As such, the expected value $E(X)$ is STILL np. We now call this λ.

$$E(X) = \lambda = \underset{\text{number of trials}}{n}\overset{\text{probability of occurrence}}{p}$$

where in this case, n: number of days/wk, and p: probability screaming party girls wake us up on any given day.

$$\therefore \ \lambda = \left(\frac{\text{days}}{\text{week}}\right)\left(\text{chance of getting woken in a day}^{*}\right)$$

But we run into a problem here with our p: what if we get woken up multiple times in a day? We can't have $p > 1$.

We can change our chance to a smaller and smaller interval, chance of getting woken every hour, every minute, every second. The smaller we make our time interval, the better this approximation gets (as we're increasing the number of trials, too!), and what happens when we increase $n \to$ infinity / decrease our chance's time interval to $\to 0$, we actually come up with the poisson distribution itself.

* notice how this is binary - we're woken up or we're not

So we have the probability of X being some i as

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

where now, since $\lambda = np$, $p = \lambda/n$, giving us:

$$= \frac{n!}{i!\,(n-i)!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

We want to increase the number of trials as much as we possibly can, so we need to put this in a form that'll let us solve for the limit easier

$$= \left[\frac{n!}{(n-i)!\,i!}\right] \left[\frac{\lambda^i}{n^i}\right] \left[\left(1-\frac{\lambda}{n}\right)^n \Big/ \left(1-\frac{\lambda}{n}\right)^i\right]$$

$$= \left[\frac{\overset{\text{reduce}}{\cancel{n!}}}{\cancel{(n-i)!}\,n^i}\right] \left[\frac{\lambda^i}{i!}\right] \left[\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^i}\right]$$

$$= \left[\frac{n(n-1)\dots(n-i+1)}{n^i}\right] \left[\frac{\lambda^i}{i!}\right] \left[\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}\right]$$

Then, we can take the limit as $n \to \infty$:

$$= \left[\overset{1}{\cancel{\frac{n(n-1)\dots(n-i+1)}{n^i}}}\right] \underbrace{\left[\frac{\lambda^i}{i!}\right]}_{\text{can't be reduced}} \left[\frac{\overset{e^{-\lambda}}{\cancel{(1-\lambda/n)^n}}}{\underset{1}{\cancel{(1-\lambda/n)^i}}}\right]$$

$$P\{X=i\} = \frac{\lambda^i}{i!} e^{-\lambda}$$

which is the true form of the Poisson approximation of the binomial random variable.

What we've done is extended the abilities of the binomial random variable from:

What is the chance a given event <u>will occur</u> in some number of trials?

to

What is the chance a given event <u>will occur i times</u> in some number of trials?

Since at the start we've defined $E(X) = \lambda$, this gives us $Var(X) = \lambda$ as well: both the expected value and variance is equal to the parameter λ.

Dynamic Programming Optimizations

Again, if we'd like to write a program to calculate these, we could use the relationship

$$P\{X = i+1\} = \frac{\lambda}{i+1} P\{X = i\}$$

Geometric Probability Distribution

Geometric probability describes the probability that an event E will occur for the first time on the n^{th} try.

$$P\{X = n\} = \underbrace{(1-p)^{n-1}}_{n-1 \text{ failures}} \underbrace{p}_{\text{single success}}$$

~~*You might notice that the equation itself doesn't actually care about the order of tries - it actually states the probability of having one success in n tries.~~

~~n-1 failures followed by 1 success is mathematically equivalent to 1 success then n-1 failures, etc.~~

* Ignore this, it's incorrect

$$E(X) = 1/p \qquad Var(X) = \frac{1-p}{p^2}$$

<u>Negative Binomial Random Variable</u>

So instead of 1 success, what if we wanted r successes, where n is the number of trials to reach r?

Remember the regular binomial distribution:

$$P\{X=i\} = \binom{n}{i} p^i (1-p)^{n-i}$$

which described the probability of i successes in n trials.

Now, our new equation:

$$P\{X=n\} = \binom{n-1}{r-1} \underset{\substack{\uparrow \\ \text{successes}}}{p^r} \underset{\substack{\uparrow \\ \text{failures}}}{(1-p)^{n-r}}$$

The only thing that has changed is the binomial coefficient - why is that?

It's got everything to do with how we've defined these probabilities. Let's compare

Binomial	Negative Binomial
Want probability of i successes in n trials	Want probability that the r^{th} success lands exactly on the n^{th} trial
↓	↓
H H H T T	H H T T $\boxed{H}$
3 successes (heads) in 5 trials	3^{rd} success (head) is on th 5^{th} trial

As we can see, in the negative binomial, we can GUARANTEE the last trial is a success.

That means we have $(n-1)$ spaces to distribute $(r-1)$ remaining successes, giving us the final

$$P\{X=n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

$$E(X) = r/p. \qquad Var(X) = \frac{r(1-p)}{p^2}$$

Unsurprisingly the expected values and variance are just the geometric distribution's, except multiplied by the r successes → $r=1$ yields the same results

Hypergeometric Distribution

Aside from having an almost inappropriately futuristic name, the hypergeometric distribution also has a somewhat daunting equation, so let's construct it ourselves instead.

So let's say we're at a fashion event. There's a giveaway, where people's names are randomly drawn, and separate prizes given to men and women

There are 455 people in attendance - 215 men, and 240 women. They are giving away 12 prizes. What is the probability that we get 9 men and 3 women chosen?

$$P = \frac{\text{samples with 9 men and 3 women}}{\text{all samples with 12 people}}$$

Hopefully so far this makes sense. The number of ways to pick any 12 people from 455 is $\binom{455}{12}$.

$$P = \frac{\text{samples with 9 men, 3 women}}{\binom{455}{12}}$$

Now, if we want any 9 men from our 215, this is $\binom{215}{9}$. Obviously, 3 women from 240 is the same, $\binom{240}{3}$. As such, our equation looks like this:

$$P = \frac{\binom{215}{9}\binom{240}{3}}{\binom{455}{12}} = 3.66\%$$

Now, we can look at the actual hypergeometric distribution equation:

$$P\{X=i\} = \frac{\overset{\text{choose } i \text{ things of type 1}}{\binom{m}{i}}\overset{\text{choose } n-i \text{ things of type 2}}{\binom{N-m}{n-i}}}{\underset{\text{choose } n \text{ things out of } N}{\binom{N}{n}}}$$

which, now, isn't very scary at all.

$$E(X) = \frac{nm}{N} \qquad Var(X) = \left(\frac{N-n}{N-1}\right) np(1-p)$$

where $p = m/N$

Of course, we can extend this by adding more terms in the numerator, creating the multivariate hypergeometric distribution, given by

$$P = \frac{\overbrace{\prod}^{\Sigma \text{ for multiplication}}{}_{i=1}^{\overset{\text{number of types}}{c}} \binom{K_i}{k_i}}{\binom{N}{n}}$$

K_i ← number of things of type i

k_i ← things chosen of type i

N ← number of things

n ← number of things chosen

Properties of the Cumulative Distribution Function

1) If some constant a is less than b, $CDF(a) \leq CDF(b)$.

Since CDF calculates the probability some random variable X will be $\leq a$, if we increase the scope of what we're looking for, CDF will stay the same or increase as well.

2) $\lim_{b \to \infty} CDF(b) = 1$

This is a similar argument as above. If we encompass every possible value, we'll eventually reach a probability of 1 because X must be $\leq \infty$.

So whenever B becomes the largest possible value of X, $CDF(B) = 1$.

3) $\lim_{b \to -\infty} CDF(b) = 0$

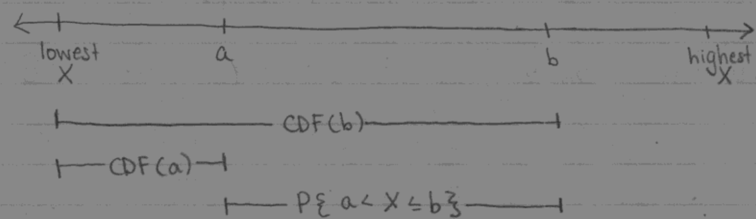
Again, if we make $b <$ the lowest value of X, the probability of $X \leq b$ is 0

4) For any b, and DECREASING sequence a_n, $n \geq 1$, that converges to (eventually gets to or approaches) b,

$$\lim_{n \to \infty} CDF(a_n) = CDF(b)$$

This one is kind of a no brainer. If we decrease a_n until it reaches b, obviously $CDF(a_n)$ will be $CDF(b)$.

5) $P\{a < X \leq b\} = CDF(b) - CDF(a)$



I thought this kind of thing is best thought about visually. The top line represents all possible values of X.

CDF(b) would be from b down to the lowest point.
CDF(a) " " a " "

As such, if you cut off all the parts that CDF(a) and CDF(b) intersect, you end up with exactly $P\{a < X \leq b\}$.

Damn, that's it. That was a ton of fucking information. It's example time - time to see whether we actually learned anything.

ex. A lot of items has some fraction θ of defective items. Let X be the random variable for the probability of needing n inspections to find 2 defective items. Find the probability distribution and the mean

Since it doesn't look like we remove the item when inspecting, we can use the negative binomial distribution here.

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

We don't know what n is, but $r = 2$, and $p = \theta$

$$= \binom{n-1}{1} \theta^2 (1-\theta)^{n-2}$$
$$= (n-1)\, \theta^2 (1-\theta)^{n-2}$$

And as such,

$$E(X) = \sum_i x_i\, p(x_i)$$
$$= \sum_{n=2}^{\infty} n \left[(n-1)\, \theta^2 (1-\theta)^{n-2} \right]$$

we can find 2 defective items in 2 up to ∞ inspections

So how do we actually turn this into a real number? At this point we've solved the logical portion, and now it's just a math problem, in which I don't really understand how to solve, but Wolfram and the lecture notes assure me the answer is $2/\theta$.

Let's take a look at how they actually did it

They started off with this:

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$$

which hopefully something we've got memorized already. Taking the derivative of both sides, we get:

$$\sum_{i=1}^{\infty} i\, x^{(i-1)} = \frac{1}{(1-x)^2}$$

And one more time:

$$\sum_{i=2}^{\infty} i(i-1)\, x^{(i-2)} = 2\left(\frac{1}{(1-x)^3}\right)$$

Now, we substitute $x = 1-\theta$:

$$\sum_{i=2}^{\infty} i(i-1)(1-\theta)^{i-2} = \frac{2}{(1-(1-\theta))^3}$$

$$= \frac{2}{\theta^3}$$

From here, we move θ^2 from the denominator of the RHS to the numerator of the LHS.

$$\sum_{i=2}^{\infty} i(i-1)\,\theta^2 (1-\theta)^{i-2} = \frac{2}{\theta}$$

Voila! How I'd think of that myself, I've got no clue.

Let's do one more very simple one.

ex. Find the mean and variance of the Poisson random variable Y, if it is 3 times as likely for $Y=4$ than $Y=2$.

$$P\{Y=i\} = \frac{\lambda^i}{i!} e^{-\lambda}$$

$$P\{Y=4\} = \frac{\lambda^4}{4!} e^{-\lambda}$$

Since we're given that $P\{Y=4\} = 3P\{Y=2\}$, so

$$\frac{\lambda^4}{4!} e^{-\lambda} = \frac{3\lambda^2}{2!} e^{-\lambda}$$

$$\frac{e^{-\lambda}\lambda^4}{24} = \frac{3e^{-\lambda}\lambda^2}{2}$$

$$\cancel{e^{-\lambda}}\lambda^4 = 36\cancel{e^{-\lambda}}\lambda^2$$

$$\lambda^2 = 36$$

$$\lambda = 6$$

And since the poisson random variable's λ is defined as both the mean and the variance,

$$E(Y) = Var(Y) = \lambda = 6$$

ex. Let's play Lotto 649! You choose 6 numbers from [1, 49], no repetitions. If you choose the ones we chose, you get [LARGE SUM OF MONEY]!

a) What is the probability that k numbers, $k=[3,6]$ match k?
Give eight significant digits.

We'll be using the hypergeometric dist., because we're choosing numbers that match, or don't match.

$$P\{X=k\} = \frac{\overset{\text{numbers you chose}}{\binom{6}{k}} \overset{\text{number of non-matching numbers}}{\binom{43}{6-k}}}{\binom{49}{6}}$$

numbers that match

numbers you chose that don't match

6 numbers from 49

At this point it's kind of worthless to plug in the numbers, you don't need explanations for that

b) What is the probability all of my numbers are [1, 9]?

$$\frac{\binom{9}{6}}{\binom{49}{6}}$$

numbers from [1, 9]

ones I've chosen

6 from 49

c) How about that 2 of my numbers are [1, 9], and I have one number for each of [10, 19], [20, 29], [30, 39], [40, 49]?

Once more, we can use the hypergeometric (multivariate):

$$\frac{\underset{\text{2 from [1,9]}}{\binom{9}{2}} \overset{[10,19]}{\binom{10}{1}} \overset{[20,29]}{\binom{10}{1}} \overset{[30,39]}{\binom{10}{1}} \overset{[40,49]}{\binom{10}{1}}}{\binom{49}{6}} = \frac{\binom{9}{2}\binom{10}{1}^4}{\binom{49}{6}}$$

6 from 49

ex. Let's play some more Russian Roulette. Refresher on the rules: one bullet, six chambers. Spin, fire, pass to the next guy. Continue till someone dies.

What's the probability someone will be killed on the i^{th} trial?

If we think about this - this is $i-1$ no one gets killed, and 1 person gets killed at the end. Remind you of something? It should - the geometric distribution.

$$P\{X = n\} = (1-p)^{n-1}\, p$$

where p is the probability of a kill: $(1/6)$. Therefore,

$$P\{X = i\} = (5/6)^{i-1}\,(1/6)$$

ex. A Waterloo student got a shitty co-op, and just handles day-to-day activities like sorting mail. At the end of his rope, he decides to randomly distribute n pieces of mail in n mailboxes.

How many people, on average, will receive the correct pieces of mail?

Each piece of mail has a $1/n$ chance of being placed in the right box, and a $1-1/n$, or $n-1/n$ chance of being put in the wrong box.

Note that this is a binomial distribution - we're either right or wrong.

The expected value for the binomial distribution is simply

$$E(X) = \overset{\text{trials}}{n}\, \underset{\text{\% success}}{p}$$

$$= n\left(\frac{1}{n}\right)$$

$$= 1$$

so on average we'd get 1 piece of mail in the right place.

Alternatively, you could manually find $E(X)$ using

$$E(X) = \sum_i x_i\, p(x_i)$$

from

$$p(x_i) = P\{X = x_i\}$$
$$= \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}$$

but that's way harder. Let's do something more difficult.

ex. Little Jimmy loves Yu-Gi-Oh! The creators have found out their playerbase will buy packs even if the return is abysmal, so they've reduced the size of packs to 1 card. In the new set, there are r total cards. If Lil' Jimmy buys n packs, what is the probability he'll complete the entire set $(n \geq r)$?

This question doesn't even really belong in this chapter, it's more of a chapter 2 question. But let's do it anyway.

Of course, Lil' Jimmy starts off with no cards from the set.

$$P(\text{getting a specific card from one pack}) = 1/r$$

Conversely,

$$P(\text{getting anything BUT a specific card}) = 1 - 1/r$$

So let's define an event here:

E_i = Lil' Jimmy does not own card i.

It would then follow that

$$E = \bigcup_{i=1}^{r} E_i$$

= Lil' Jimmy doesn't own any cards from the new set

What we want to find is

$P(E)$ = the probability that Jimmy doesn't own any card after n packs

and then

$1 - P(E)$ = the probability that Jimmy owns all cards after n packs

Since $P(E)$ is the probability of the union of some events, we can follow the procedure I outlined back in chapter 2 (as these aren't independent)*:

① Add the probability of all events

$$P(E_i) = (1 - 1/r)^{n \leftarrow \text{trials}} \; ; \quad \therefore$$

$$\sum_{i=1}^{r} (1 - 1/r)^n$$

② Subtract every combination of 2 events' intersections

$$P(E_i \cap E_j) = (1 - 2/r)^n$$

, because we want the chances of getting every card EXCEPT these two, $\therefore$

$$-\sum_{i<j} (1 - 2/r)^n$$

③ And we continue this flip flop until we reach $i = r$, where $(1 - i/r)^n$ simply turns into 0

$$\therefore \; P(E) = \sum_{i=1}^{r} (1 - 1/r)^n - \sum_{i<j} (1 - 2/r)^n + \sum_{i<j<k} (1 - 3/r)^n - \ldots + 0$$

$$= \underbrace{\binom{r}{1}(1 - 1/r)^n}_{\text{individual events}} - \underbrace{\binom{r}{2}(1 - 2/r)^n}_{\text{intersection of 2 events}} + \underbrace{\binom{r}{3}(1 - 3/r)^n}_{\text{intersection of 3 events}} - \ldots + 0$$

And as such,

$$1 - P(E) = 1 - \sum_{k=1}^{r} (-1)^{k+1} \binom{r}{k} (1 - k/r)^n$$

* getting a new card affects the chances of getting a new card next pack

ex. A single unbiased die is rolled n times. Let r_1 = the number of 1s, and r_2 = the number of 2s. What is $E(r_1 \cap r_2)$?

This is an odd question that requires a lesson in the examples:

<u>Multinomial Distribution</u>

As the name implies, this is a multivariate version of the binomial distribution. It's given by

$$\left(\frac{n!}{x_1! \dots x_k!}\right)(p_1)^{x_1} \dots (p_k)^{x_k}$$

where n = trials
x_i = number of successes of type i
p_i = probability of success of type i

Since it's just another expression of the binomial distribution,

$$E(x_i) = np_i \qquad Var(x_i) = np_i(1-p_i)$$

which is the same as the binomial distribution. Now that's out of the way, we can just plug and play.

$$P(r_1 \cap r_2) = \frac{n!}{r_1!\, r_2!}(1/6)^{r_1}(1/6)^{r_2}$$

$$E(r_1 \cap r_2) = n(1/6)(1/6)$$

That's it! The notes' solution is actually kind of odd, and is almost a lesson, itself.

ex. You're trapped! There's 3 doors:

A: to safety in 3 hours
B: back to the start in 5 hours
C: back to the start in 7 hours.

If you're an idiot, and don't realize the doors are leading you back to the start, how long will it take to get out, on average?

Let's let X be the number of hours, ∴ we want $E(X)$.

Obviously
$$E(X \mid A) = 3 \text{ hours.}$$
$$E(X \mid B) = 5 + E(X)$$
$$E(X \mid C) = 7 + E(X)$$
go back to the start

$$\therefore \quad E(X) = E(X \mid A)(1/3) + E(X \mid B)(1/3) + E(X \mid C)(1/3)$$
$$E(X) = 1 + (5 + E(X))(1/3) + (7 + E(X))(1/3)$$
$$E(X) = 1 + 5/3 + 7/3 + (2/3)E(X)$$

At this point it's a simple algebraic solve.

$$1/3 \, E(X) = 5$$
$$E(X) = 15$$

That's it for chapter 4!