Notes 25

1

For today's lecture, we are going to apply what we have learned about probability to solve a particular kind of probability problem that arises in all sorts of applications. In the general form of the problem, we have a collection of independent events and we want to get some idea of how many of the events are likely to occur.

Given: mutually independent events $A_1, A_2, \cdots, A_n \subseteq S$.

For example, we can think of A_i as being the event that the ith component of a system fails. Or we can think of A_i as the event that the ith coin is heads when flipping n coins. r anything else–probability student answers the ith question on the final correctly.

What we want to know is, what is the probability that k of the events occur. Let T be the number of events that occur.

The question is easy to answer if each event is equally likely, that is, $\forall i Pr(A_i) = p$. Then $Pr(T = k) = \binom{n}{k} p^k (1-p)^{n-k}$, the binomial distribution. We've already talked about important distribution. Very tightly centered around the mean $= pn$.

Today, we're going to talk about the general case: $Pr(A_i)$ are different.

First thing is to look at the expected number of events to occur. To figure this out, we can use linearity of expectation and indicator variables.

Let T_i be an indicator for the ith event. (1 if A_i occurs, 0 otherwise)

$$\begin{aligned}
T &= T_1 + T_2 + \cdots + T_n \\
E[T] &= E[T_1] + E[T_2] + \cdots + E[T_n] \\
&= Pr(A_1) + Pr(A_2) + \cdots + Pr(A_n) \\
&= \sum_{i=1}^{n} Pr(A_i)
\end{aligned}$$

So to compute the expected number of events, we just add up the probabilities. We don't even need the A_i independent for this, because we just used linearity of expectation.

Example: A_i is the probability the ith component fails. Then the expected number of failures is the sum of the individual component failures, even if there is dependence.

OK, nothing really new here. But say that we want to know the probability that there is one or more failures. That is, say that any fault is critical, so we don't care how many there are, just if there is one. (Once there is one it is all over, so who cares how many more happen?) So we want $Pr(T \geq 1)$.

First let's get a simple upper bound. By Markov,

$$Pr(T \geq 1) \leq \frac{E[T]}{1} = E[T] = \sum_{i=1}^{n} Pr(A_i)$$

So if the sum of the individual event probabilities is small, then the chance that one or more occurs is also small. Note we haven't used independence yet.

We could show this another way, using Boole's inequality. Let A be the event that 1 or more occurs. So $A = A_1 \cup A_2 \cup \cdots \cup A_n$.

$$Pr(T \geq 1) = Pr(A) \leq \sum_{i=1}^{n} Pr(A_i) = E[T]$$

Note we still didn't need independence.

So if the expected number of events to occur is $<< 1$, then the probability at least one occurs is very small. For example, if you buy 1000 lottery tickets, each of which wins with probability 1 in a million, then the probability you have a winner is $\leq 1000/1000000 = 1/1000$.

Now let's look at the case where the expected number of events to occur is bigger than 1. For example if the expected number is 10, it seems likely that at least one will happen. This is the case if the events are independent. The following theorem captures this fact:

Theorem 1.1 *Given independent events $A_1, A_2, \cdots, A_n$, the probability that none occur is $\leq e^{-m}$, where $m = \sum_{i=1}^{n} Pr(A_i) =$ expected number to occur.*

So probability none occur is exponentially small in the expected number of events to occur.

Proof. Probability that no event occurs

$$
\begin{aligned}
&= Pr(\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_n}) \\
&= \prod_{i=1}^{n} Pr(\overline{A_i}) \quad \text{(by independence)} \\
&= \prod_{i=1}^{n} (1 - Pr(A_i)) \\
&\leq \prod_{i=1}^{n} e^{-Pr(A_i)} \\
&= e^{-\sum_{i=1}^{n} Pr(A_i)} \\
&= e^{-m}
\end{aligned}
$$

■

Note independence is critical here. Also, this is another example where Taylor series bounds are important.

This is a really powerful result. For example, we immediately get the following corollary:

Corollary 1.2 *Given independence, if we expect 10 or more events to occur, the probability that no event occurs is* $\leq e^{-10} < 1/22000$

Note there is no dependence on n here. This is a really strong result since it says that if we expect some things to happen then they almost certainly will. I.e, if you expect your system to have some faults, then it surely will–this is sort of a proof of Murphy's Law. If there are a bunch of things that can go wrong, even if none of them is very likely, this result says that almost surely something will go wrong.

Example 1 *(Murphy's Law) n components, each fails with probability $p = 10/n$. expected number of failures = 10, probability there is no faults* $\leq e^{-10}$.

So even though the failure probability of any component is small, and only 10 are expected, probability of none is not $1/10$ but $\leq e^{-10}$.

This theorem is in fact so powerful that it will allow me to read your mind.

2 Mind Reading Game

You do the following: Pick a number i from 1 to 9. You then start dealing out a deck of cards one at a time. When you get to card i, you continue counting ahead, this time with the value of that card. 10, J, Q, K, A, and Joker all count as 1. Whatever card you are counting off of when the deck runs out is your card to be mind read.

Example (on a few cards): You pick 3. The sequence of cards in the deck is 4 5 2 3 10 Q 3 7 6 4 2. So the cards you use are 2, 10, Q, 3, 4, and the card to be mind read is the 4.

I get to read your mind because if I do the same thing as you it is likely we will have the same last card. The reason, is that is we ever land on the same card we will have the same last card. Each time one of use hits a card and starts counting again, there is some chance we will land on the others positions. If 1 away, 5/13 chance, if ¿1 away, 1/13 chance, and there are lots of times that we each jump. On average, we will each jump ¿10 times, so 20 about jumps total between the two sides, making it likely that we will collide. In particular, A_i is the event that on the ith jump we collide.

Now the intuition I just went thru was pretty fuzzy. In fact, the theorem doesn't really apply in this example because the events are not independent. If we have not yet collided, it tells us something about the numbers we have had, and since we only have a 52 card deck, something about the numbers that are forthcoming, which conditions the probability that we collide. However, we could analyze it if the cards we selected at random from an infinite number of decks. Then the event of a hit would be independent for every jump.

We don't have time in class to analyze this precisely. It turns out that the probability I read you mind is $> 90\%$ for a 52 card deck.

The point of all this can be summed up as follows:

If we expect several events to occur, then the probability that no events occur is exponentially small if the events are independent. If the events are dependent, then the probability of no events occurring could be anything.

3 Chernoff Bound

Now we are going to prove an even stronger result. If we have a collection of independent events and we expect m of them to occur, then with probability very close to 1, close to m will occur. It's similar to the phenomenon with tossing n coins. We expect $n/2$ heads, and with probability close to 1 we will get close to $n/2$ heads.

For simplicity, we will only look at the case of getting much more than the expected value; a similar method gives similar results for the other side.

Theorem 3.1 *(Chernoff Bound) Let $T = T_1 + T_2 + \cdots + T_n$, where the T_i are independent Bernoulli random variables. Then $\forall c > 1$ $Pr(T \geq cE[T]) \leq e^{-\alpha c E[T]}$, where $\alpha = \ln c + 1/c - 1 \geq 0$.*

In general, this is much stronger than what Markov's theorem gives (1/c); here we get exponential bounds on probability.

Corollary 3.2 *With T as above, $Pr(T \geq eE[T]) \leq e^{-E[T]}$.*

Proof. $c = e \Rightarrow \alpha = \ln e + 1/e - 1 = 1/e \Rightarrow \alpha c = 1$ ∎

Example 2 *If $E[T] = 100$, $Pr(T \geq 272) \leq e^{-100}$.*

Example 3 *1 million people play "Pick 4". Each player picks a random 4 digit decimal number. So $Pr(winning) = 1/10000$, $E[\#winners] = 1M/1000 = 100$. We assume random picks for the sake of argument. If picks are independent, then $Pr(\geq 272winners) \leq e^{-100}$, which is really, really, tiny. Note answer is not dependent on the number of people playing or the individual probability of winning.*

Of course, the result is not true if the picks are dependent. If everyone picks the same number, then the $Pr(\geq 272winners) = 1/10000 =$ chance that number comes up. (In effect, we've proved that it is incredibly unlikely that 272 pick the same number if they choose independently.)

Now in truth with lottery, people tend to pick in a structured fashion. (Digression) They pick the date, or something they saw on TV, or yesterdays number, etc. In fact, Chernoff himself (who was once on the faculty here) developed a strategy to make money on the state lottery. He did a statistical study of unpopular numbers and showed that by betting on them you would expect to earn money! This only works in lotteries where payoff is fraction of total pot. If zillions of people all bet on few numbers and you bet on all the rest, you are likely to win their zillions of dollars. Way it turned out made 7 percent return. Now state sets it up to give you random number.

Back to the proof of Chernoff's theorem:

Proof. This uses an exponential transform of Markov's theorem.

$$Pr(T \geq cE[T]) = Pr(c^T \geq c^{cE[T]})$$

This is because it is the same event, just stuffed in an exponent on one side. Then by Markov,

$$\leq \frac{E[c^T]}{c^{cE[T]}}$$

Need to evaluate $E[c^T]$:

$$
\begin{aligned}
T &= T_1 + T_2 + \cdots + T_n \\
\Rightarrow c^T &= c^{T_1} c^{T_2} \cdots c^{T_n} \\
\Rightarrow E[c^T] &= E[c^{T_1}]E[c^{T_2}] \cdots E[c^{T_n}] \quad \text{(by product rule } (T_i\ ind. \Rightarrow c^{T_i}\ ind.)) \\
&= \prod_{i=1}^{n} E[c^{T_i}]
\end{aligned}
$$

So just need to evaluate $E[c^{T_i}]$:

$$
\begin{aligned}
E[c^{T_i}] &= c^1 Pr(T_i = 1) + c^0 Pr(T_i = 0) \\
&= cPr(T_i = 1) + (1 - Pr(T_i = 1)) \\
&= 1 + (c-1)Pr(T_i = 1) \\
&\leq e^{(c-1)Pr(T_i=1)} \quad \text{(since } 1 + x \leq e^x) \\
&= e^{(c-1)E[T_i]} \quad \text{(since indicator variables)}
\end{aligned}
$$

Now we can plug in:

$$
\begin{aligned}
E[c^T] &\leq \prod_{i=1}^{n} e^{(c-1)E[T_i]} \\
&= e^{\sum_{i=1}^{n} (c-1)E[T_i]} \\
&= e^{(c-1)\sum_{i=1}^{n} E[T_i]} \\
&= e^{(c-1)E[T]}
\end{aligned}
$$

Now we can go back and plug into the bounds from Markov:

$$
\begin{aligned}
Pr(T \geq cE[T]) &\leq \frac{e^{(c-1)E[T]}}{c^{cE[T]}} \\
&= e^{(c-1)E[T] - c\ln cE[T]} \\
&= e^{-c(\ln c + 1/c - 1)E[T]} \\
&= e^{-\alpha c E[T]} \quad where\ \alpha = \ln c + 1/c - 1
\end{aligned}
$$

■

This is the deepest result in the class. Don't feel like you'd have to come up with this on your own–it's pretty tricky. But the basic method is not so bad:

$$
Pr(A > B) = Pr(f(A) > f(B)) \leq \frac{E[F(A)]}{f(B)}
$$

and then just plug and chug.

You should try to remember the implications of the result. If you have a sum of 0-1 variables or a collection of independent events, then almost surely the number of events is close to the expected number. "Tails" are exponentially small.