STAT-4330 - Class Notes

September 3, 2024

Chapter 1:

September 3, 2024

Tossing a fair coin 100 times, X the number of heads is a Binomial R.V.

$X = 50$

$$P(X = 50) = \binom{100}{50} \left(\frac{1}{2}\right)^{100} = \frac{100!}{50!50!} \frac{1}{2}^{100}$$

Factorials can be estimated with Stirling's Formula

$$n \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

$\sim$ implies

$$\lim_{n \to \infty} \frac{n!}{\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}} = 1$$

i.e. The absolute error might be large (tends to infinity actually) but 'relative error' is small

$45 \leq X \leq 55$

To find $P(45 \leq X \leq 55)$ requires summation over Binomial pdf. Or, X is approximately normal by the central limit theorem.

This can also be done for $P(X = 50)$, by evaluating the density at $x = 50$ in a normal pdf, or integrating from 49.5 to 50.5 maybe.

$X = 70$?

Is the normal approximation still good?

Proof of CLT

'Proofs' for CLT likely proceeds with moment generating functions but glosses over the hard part of the proof.

Moving to a stochastic process

Imagine plotting S_n against n, where S_n is the number of heads - number of tails up to 'time' n.

Now the question of $P(X = 50)$ is asking for the probability that $S_n = 0$ when $n = 100$.

Possible questions

- what is the chance S_n always stays above 0?
- if n goes to infinity, will S_n always hit 0 at some point (yes btw) ? Is there a chance it doesn't hit 0?
- if coin is not fair ($p > \frac{1}{2}$), what will happen to S_n as $n \to \infty$

  - S_n will 'drift' above the axis, n will be above the x axis (proof via **strong law of large numbers**)

- if S_n returns to 0, what do we know about n when it returns (other than n is even)
- is the expected time to return less than 10, between 10 to 20, or 20 to 100?
- let Y be the R.V. denoting the last n at which S_n touched the x axis. What kind of RV is this?
  - it is not obvious, but Y is a symmetrical RV about 50
  - is $Y = 50$ more likely than $Y = 0$ (i.e. $Y = 100$)?
- for 100 tries, is it more likely that the last time $S_n = 0$ occurs at $n = 2$ or $n = 98$ (they are equally likely)
- what fraction of the time is S_n above the axis? What fraction is it below?
- what is $P(S_1 > 0, S_2 > 0, S_3 > 0 \ldots S_2n = r)$ OR, what is the probability that S_n takes some path above the x axis and ends at $(2n, r)$
  - any path of length $2n$ has the same probability $\frac{1}{2}^{2n}$
  - any path that ends at a specified spot as a binomial probability
  - to find the above probability, we count the number of paths that satisfy the condition

Chapter 2:

2.1 Motivating example

Imagine counting votes in n people, p say yes, q say no.

$$n = p + q$$

Define x to be

$$x = p - q$$

Setting $n = 8$, $p = 5$, $q = 3$, $x = 2$.

Imagine taking the votes one by one and plotting x against time t. The ending point is at $(8, 2)$ (i.e. $t = 8, x = 2$).

How many paths are there?

$$\binom{8}{3} \text{ or } \binom{8}{5} \text{ or } \frac{8!}{5!3!}$$

In general, the number of paths from $(0, 0)$ to (n, x) is

$$N_{n,x} = \binom{n}{p} = \binom{n}{q} = \binom{p+q}{q} = \binom{n}{\frac{n+x}{2}} = \binom{n}{\frac{n-x}{2}}$$

How many paths ending in (n, x) are always positive?.

At the first step, the path can only go up, ie $(0, 0) \to (1, 1)$.

Let us denote the number of paths from $(0, 0)$ to (n, x) as $N_{n,x}$.

We try to count the number of paths from $(1, 1)$ to (n, x), i.e. $N_{n-1,x-1}$

$$N_{n-1,x-1} = \binom{p+q-1}{p-1}$$

This is the same as saying we have $n - 1 = p + q - 1$ decisions to make, and $p - 1$ of them are going up.

Of $N_{n-1,x-1}$, we want to know how many paths crosses the horizontal axis.

By the Reflection Principle, there is a one-to-one mapping of paths from $(1, 1)$ to (n, x) that cross the axis to paths from $(1, -1)$ to (n, x). This number is

$$N_{n-1,x+1} = \binom{p+q-1}{p}$$

Note: we use $p + q - 1$ choose p because to get to $(1, -1)$ from $(0, 0)$, we used a 'no' vote, but we still have p 'yes' votes.
The final answer is

$$N_{n-1,x-1} - N_{n-1,x+1}$$

2.2 Extending from the previous example

Let $X_1, \ldots, X_{2n}$ be iid RV where

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$$

Let $S_0 = 0$, and

$$S_k = \sum_{i=1}^{k} X_i \text{ for } k = 1, \ldots, 2n$$

How would we find

$$P(S_1 > 0, S_2 > 0, \ldots, S_{2n} > 0)$$

Consider S_{2n}, the possible values for S_{2n} are $\{-2n, \ldots, -4, -2, 0, 2, 4, \ldots 2n\}$

We can first find

$$P(S_1 > 0, S_2 > 0, \ldots, S_{2n} = 2r) \text{ where } r = 1, 2, 3, \ldots, n$$

This is equivalent to looking for a path from $(0, 0)$ to $(2n, 2r)$ that stays above the axis. The number of paths is

$$N_{2n-1,2r-1} - N_{2n-1,2r+1}$$

The probability of getting any single path is

$$\left(\frac{1}{2}\right)^{2n}$$

The required probability is

$$P(S_1 > 0, S_2 > 0, \ldots, S_{2n} = 2r) = (N_{2n-1,2r-1} - N_{2n-1,2r+1}) \left(\frac{1}{2}\right)^{2n}$$

To answer the initial question

$$\begin{aligned}
&P(S_1 > 0, S_2 > 0, \ldots, S_{2n} > 0) \\
=&\sum_{r=1}^{n} P(S_1 > 0, S_2 > 0, \ldots, S_{2n} = 2r) \\
=&\sum_{r=1}^{n} (N_{2n-1,2r-1} - N_{2n-1,2r+1}) \left(\frac{1}{2}\right)^{2n} \\
=&\left(\frac{1}{2}\right)^{2n} \sum_{r=1}^{n} (N_{2n-1,2r-1} - N_{2n-1,2r+1}) \\
=&\left(\frac{1}{2}\right)^{2n} (N_{2n-1,1} - N_{2n-1,3} + N_{2n-1,3} + N_{2n-1,5} + \ldots + N_{2n-1,2n-1} - N_{2n-1,2n+1}) \\
=&\left(\frac{1}{2}\right)^{2n} (N_{2n-1,1} - N_{2n-1,2n+1}) \\
=&\left(\frac{1}{2}\right)^{2n} (N_{2n-1,1} - 0) \text{ since there are 0 ways to get to } 2n+1 \text{ with } 2n-1 \text{ steps} \\
=&N_{2n-1,1} \left(\frac{1}{2}\right)^{2n}
\end{aligned}$$

Chapter 3: Class 3

September 3, 2024

3.1 Review of 'no return' problems

Recall the example from last week. Let $X_1 \dots X_{2n}$ be iid RVs where $P(X_i = 1) = \frac{1}{2}, P(X_i = -1) = \frac{1}{2}$.

Let $S_0 = 0$, $S_j = \sum_{i=0}^{j} X_i$ for $j = 1, \dots, 2n$

Problem: Find $P(S_1 > 0, S_2 > 0, \dots, S_{2n} > 0)$

Step 1: Breaking up the event into a set of disjoint events and summing up their probability

The required probability is

$$\begin{aligned} &P(S_1 > 0, S_2 > 0, \dots, S_{2n} > 0) \\ &= \sum_{r=1}^{\infty} P\left(S_1 > 0, S_2 > 0, \dots, S_{2n} = 2r\right) \end{aligned}$$

Step 2: Count the number of paths from $(0, 0)$ to $(2n, 2r)$ that do not touch the x-axis.

$$\begin{aligned} &P(S_1 > 0, S_2 > 0, \dots, S_{2n} > 0) \\ &= \sum_{r=1}^{\infty} P\left(S_1 > 0, S_2 > 0, \dots, S_{2n} = 2r\right) \\ &= \sum_{r=1}^{\infty} (\text{Number of paths from origin to } (2n, 2r) \text{ that do not touch the axis}) \left(\frac{1}{2}\right)^{2n} \end{aligned}$$

Step 3: Count the total number of paths from $(0, 0)$ to $(2n, 2r)$, and the number of paths that touch the x-axis.

$$\begin{aligned} &P(S_1 > 0, S_2 > 0, \dots, S_{2n} > 0) \\ &= \sum_{r=1}^{\infty} P\left(S_1 > 0, S_2 > 0, \dots, S_{2n} = 2r\right) \\ &= \sum_{r=1}^{\infty} (\text{Number of paths from origin to } (2n, 2r) \text{ that do not touch the axis}) \left(\frac{1}{2}\right)^{2n} \\ &= \sum \left(N_{2n-1,2r-1} - N_{2n-1,2r+1}\right) \left(\frac{1}{2}\right)^{2n} \end{aligned}$$

Step 4: Telescopic cancellation.

$$
\begin{aligned}
&P(S_1 > 0, S_2 > 0, \ldots, S_{2n} > 0) \\
&= \sum_{r=1}^{\infty} P\left(S_1 > 0, S_2 > 0, \ldots, S_{2n} = 2r\right) \\
&= \sum_{r=1}^{\infty} (\text{Number of paths from origin to } (2n, 2r) \text{ that do not touch the axis}) \left(\frac{1}{2}\right)^{2n} \\
&= \sum \left(N_{2n-1,2r-1} - N_{2n-1,2r+1}\right) \left(\frac{1}{2}\right)^{2n} \\
&= \sum_{r=1}^{\infty} \left(N_{2n-1,2r-1} - N_{2n-1,2r+1}\right) \left(\frac{1}{2}\right)^{2n} \\
&= N_{2n-1,1} \left(\frac{1}{2}\right)^{2n}
\end{aligned}
$$

Note: The upper bound for summation, whether $r = n$ or ∞, does not matter, because

- there are exactly 0 ways to get to $2n - 1$ with $2n + 1$ steps 'up'

Step 5: Evaluate $N_{2n-1,1}$

Recall that $N_{2n-1,1}$ represents $2n-1$ steps, and ending up at 1. That means there were $(2n-1+1) \div 2 = n$ steps up, $n - 1$ steps down

$$
\begin{aligned}
N_{2n-1,1} &= \binom{2n-1}{n} \\
&= \frac{(2n-1)!}{n!(n-1)!} \\
&= \frac{(2n-1)!}{n!(n-1)!} \times \frac{2n}{n} \times \frac{1}{2} \\
&= \frac{(2n)!}{n!n!} \times \frac{1}{2} \\
&= \frac{1}{2} \times \binom{2n}{n}
\end{aligned}
$$

The required probability is

$$
\begin{aligned}
&P(S_1 > 0, S_2 > 0, \ldots, S_{2n} > 0) \\
&= \sum_{r=1}^{\infty} P\left(S_1 > 0, S_2 > 0, \ldots, S_{2n} = 2r\right) \\
&= \sum_{r=1}^{\infty} (\text{Number of paths from origin to } (2n, 2r) \text{ that do not touch the axis}) \left(\frac{1}{2}\right)^{2n} \\
&= \sum \left(N_{2n-1,2r-1} - N_{2n-1,2r+1}\right) \left(\frac{1}{2}\right)^{2n} \\
&= \sum_{r=1}^{\infty} \left(N_{2n-1,2r-1} - N_{2n-1,2r+1}\right) \left(\frac{1}{2}\right)^{2n} \\
&= N_{2n-1,1} \left(\frac{1}{2}\right)^{2n} \\
&= \binom{2n-1}{n} \left(\frac{1}{2}\right)^{2n} \\
&= \frac{1}{2} \times \binom{2n}{n} \times \left(\frac{1}{2}\right)^{2n}
\end{aligned}
$$

Step 6: Recognize that $\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}$ is $P(S_{2n}=0)$

$$\begin{aligned}
&P(S_1>0,S_2>0,\ldots,S_{2n}>0)\\
&=\sum_{r=1}^{\infty}P\left(S_1>0,S_2>0,\ldots,S_{2n}=2r\right)\\
&=\sum_{r=1}^{\infty}(\text{Number of paths from origin to }(2n,2r)\text{ that do not touch the axis})\left(\frac{1}{2}\right)^{2n}\\
&=\sum\left(N_{2n-1,2r-1}-N_{2n-1,2r+1}\right)\left(\frac{1}{2}\right)^{2n}\\
&=\sum_{r=1}^{\infty}\left(N_{2n-1,2r-1}-N_{2n-1,2r+1}\right)\left(\frac{1}{2}\right)^{2n}\\
&=N_{2n-1,1}\left(\frac{1}{2}\right)^{2n}\\
&=\binom{2n-1}{n}\left(\frac{1}{2}\right)^{2n}\\
&=\frac{1}{2}\times\binom{2n}{n}\times\left(\frac{1}{2}\right)^{2n}\\
&=\frac{1}{2}P\left(S_{2n}=0\right)
\end{aligned}$$

Extension idea: How would we think about

$$P\left(S_1\neq 0,S_2\neq 0,\ldots,S_{2n}\neq 0\right)$$

We can break it up into two disjoint pieces

$$\begin{aligned}
&P\left(S_1\neq 0,S_2\neq 0,\ldots,S_{2n}\neq 0\right)\\
=&\left(S_1>0,S_2>0,\ldots,S_{2n}>0\right)+\\
&\left(S_1<0,S_2<0,\ldots,S_{2n}<0\right)+\\
=&P\left(S_{2n}=0\right)
\end{aligned}$$

I.E. tossing the coin 100 times, the probability of never having the same number of coins is

$$\binom{100}{50}\times\left(\frac{1}{2}\right)^{100}$$

3.2 First return problems

Problems about the 1st return to 0. This can only happen on an even toss.

Notation:

$$f_{2k}=\text{ chance of 1st return to 0 is at time }2k$$

More concretely

$$f_2=P(\mathrm{H},\ \mathrm{T})+P(\mathrm{T},\ \mathrm{H})=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$$

In general,

$$f_{2k}=P\left(S_1\neq 0,S_2\neq 0,\ldots,S_{2k-1}\neq 0,S_{2k}=0\right)$$

Idea 1: Breaking required event into two pieces

Consider the events $(S_1 \neq 0, \ldots S_{2k-1} \neq 0)$

$$A = (A \cap B) \cup (A \cap B^C)$$
$$(S_1 \neq 0, \ldots S_{2k-1} \neq 0) = (S_1 \neq 0, \ldots S_{2k} \neq 0) \cup (S_1 \neq 0, \ldots, S_{2k} = 0)$$

Converting events to probability

$$P(S_1 \neq 0, \ldots S_{2k-1} \neq 0) = P(S_1 \neq 0, \ldots S_{2k} \neq 0) + P(S_1 \neq 0, \ldots, S_{2k} = 0)$$

Idea 2: Draw connection to 'no return' problems discussed above

Note the LHS, since S_{2k-1} cannot be equal to 0, the required probability is

$$P(S_1 \neq 0, \ldots S_{2k-1} \neq 0) = P(S_1 \neq 0, \ldots, S_{2k-2} \neq 0) = u_{2k-2}$$

Note on the RHS,

$$P(S_1 \neq 0, \ldots, S_{2k} \neq 0) = u_{2k}$$

And the last term

$$P(S_1 \neq 0, \ldots, S_{2k-1} \neq 0, S_{2k} = 0) = f_{2k}$$

Hence

$$P(S_1 \neq 0, \ldots S_{2k-1} \neq 0) = P(S_1 \neq 0, \ldots S_{2k} \neq 0) + P(S_1 \neq 0, \ldots, S_{2k} = 0)$$
$$u_{2k-2} = u_{2k} + f_{2k}$$

Hence

$$f_{2k} = u_{2k-2} - u_{2k} \text{ for } k = 1, 2, \ldots$$

Idea 3: Taking the infinite union of disjoint events

The statement "You eventually return" is the union of "Return for the first time at time 2", "Return for the first time at time 4", ...

The probability of eventually returning is

$$\begin{aligned} & f_2 + f_4 + f_6 + \ldots \\ =& (u_0 - u_2) + (u_2 - u_4) + \ldots \\ =& u_0 \\ =& 1 \end{aligned}$$

Note: $u_0 = P(S_0 = 0) = 1$

Therefore we have used **elementary, combinatorial** arguments to show that **we will always return to 0**.

3.3 Expected time to first return

Note that there is a caveat in calling T, the time to first return, a r.v. because r.v.s map from sample space to the real numbers, but the sample space contains a perhaps infinite string of tails (i.e. T will be infinite), which is 'extended real numbers'

Let T be the time to first return

$$E[T] = \sum_{k=1}^{\infty} 2k f_{2k}$$

where

$$\begin{aligned} f_{2k} &= u_{2k-2} - u_{2k} \\ &= \binom{2k-2}{k-1}\left(\frac{1}{2}\right)^{2k-2} - \binom{2k}{k}\left(\frac{1}{2}\right)^{2k} \end{aligned}$$

As with before, we want to express combinatorials in the same 'basis'. We rewrite

$$\begin{aligned} \binom{2k-2}{k-1} &= \frac{(2k-2)!}{(k-1)!(k-1)!} = \frac{(2k-2)!}{(k-1)!(k-1)!} \times \frac{2k(2k-1)}{k \times k} \times \frac{k^2}{(2k)(2k-1)} \\ &= \frac{(2k)!}{k!k!} \times \frac{k}{4k-2} \end{aligned}$$

FILL IN THE BLANKS FOR ALGEBRA LATER

$$\begin{aligned} f_{2k} &= u_{2k-2} - u_{2k} \\ &= \binom{2k-2}{k-1}\left(\frac{1}{2}\right)^{2k-2} - \binom{2k}{k}\left(\frac{1}{2}\right)^{2k} \\ &= \frac{1}{2k-1}\binom{2k}{k}\left(\frac{1}{2}\right)^{2k} \end{aligned}$$

The expected value for T is

$$\begin{aligned} E\left[T\right] &= \sum 2k f_2 k \\ &= \sum \frac{2k}{2k-1}\binom{2k}{k}\left(\frac{1}{2}\right)^{2k} \end{aligned}$$

To understand the behavior of this, we need to understand $\binom{2k}{k}$.

By **Stirling's formula**

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$