Lecture 9: The Central limit Theorem and related topics

Reading Assignment for Lectures 7--9: PKT Chapter 5
This lecture addresses some of the material in PKT Chapter 8.2.
Tutorial tomorrow: GAME+Stirling+Central Limit Thm.

Random walk (coin toss) continued: What can you reliably predict about "random" events?

Let's now calculate the full probability distribution of $M = N_+ - N_-$.

I claim: $$P_{N,M} = \left(\frac{1}{2}\right)^N \frac{N!}{N_+!N_-!}$$ "Binomial distribution" for $M = N_+ - N_-$ (note discreteness!)

Proof: Each particular distribution (+-++-+--…), etc., (each "microstate") has a probability

$$p_+^{N_+} p_-^{N_-} = \left(\frac{1}{2}\right)^{N_+ + N_-} = \left(\frac{1}{2}\right)^N.$$ (For homework, you will do the biased walk $p_+=p, p_-=q$.)

But, the number of microstates leading to the same M is the number of distinct ways you can choose N_+ plus-steps from a total of N steps (leaving N_- minus-steps).

The usual argument for this is that you can choose the first + from any of N places, the next from any of the remaining (N-1), etc., giving a total of $N(N-1)\ldots(N-N_+ +1)$ choices. However, this overcounts, since the N_+ pluses will be counted in all possible orders. The number of orders for choosing them is $N_+!$

Thus the number of distinct microstates with N_+, N_- is

$$\frac{N(N-1)(N-2)\ldots(N-N_+ +1)}{N_+!}\cdot\frac{N_-!}{N_-!} = \frac{N!}{N_+!N_-!}.$$ QED

This result is valid for any N; however, it is useful but not particularly illuminating.

What I will show you next is that, when N is large and M is close to <M>=0 (i.e., when we are looking at the "likely" outcomes), then this distribution turns into a Gaussian distribution to a very good approximation. The key ingredient of this proof is a result called Stirling's formula:

$$n! = \sqrt{2\pi n}\, n^n e^{-n}\left(1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right),$$ "Stirling's Formula"

i.e., when n is sufficiently large, you can approximate n! with small factional corrections as shown (note that absolute corre ctions may be large, since prefactor is huge). Alternatively,

$$\ln n! = \left(n + \frac{1}{2}\right)\ln n - n + \frac{1}{2}\ln(2\pi) + \frac{1}{12n} + O\left(\frac{1}{n^2}\right).$$ I will derive this in Tutorial tomorrow.

Now, lets get back to $P_{N,M}$:

Assume that N, N_+, and N_- are all large, i.e., (a) the number of steps N is large and (b) we are looking for parts of the binomial distribution which are NOT way out at the edges. Then, (Stirling)

$$P_{N,M} = \frac{1}{2^N}\frac{N!}{N_+!N_-!} = \frac{\sqrt{2\pi N}}{\sqrt{2\pi N_+}\sqrt{2\pi N_-}}\cdot\frac{e^{-N}}{e^{-N_+}e^{-N_-}}\cdot\frac{1}{2^N}\cdot\frac{N^N}{N_+^{N_+}N_-^{N_-}}\left(1 + O\left(\frac{1}{N},\frac{1}{N_+},\frac{1}{N_-}\right)\right)$$

$$\approx \sqrt{\frac{N}{2\pi N_+ N_-}}\cdot\frac{N^N}{(N+M)^{N_+}(N-M)^{N_-}} = \sqrt{\frac{2N}{\pi(N+M)(N-M)}}\cdot\frac{N^N}{(N+M)^{N_+}(N-M)^{N_-}}$$

$$= \sqrt{\frac{2N}{\pi\left(N^2 - M^2\right)}}\cdot e^{F(N,M)}$$

with $\left(\text{using } N_{\pm} = (N \pm M)/2\right)$

$$F(N,M) \equiv N \ln N - \left(\frac{N+M}{2}\right)\ln(N+M) - \left(\frac{N-M}{2}\right)\ln(N-M)$$

$$= N \ln N - \frac{N}{2}\left(1+\frac{M}{N}\right)\left[\ln N + \ln\left(1+\frac{M}{N}\right)\right] - \frac{N}{2}\left(1-\frac{M}{N}\right)\left[\ln N + \ln\left(1-\frac{M}{N}\right)\right]$$

$$= -\frac{N}{2}\left[\left(1+\frac{M}{N}\right)\ln\left(1+\frac{M}{N}\right) + \left(1-\frac{M}{N}\right)\ln\left(1-\frac{M}{N}\right)\right]$$

Important values of M are going to be of order $\sqrt{N}$, so the ratio $\frac{M}{N}$ is small, and we expand:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots, \text{ so } (1+x)\ln(1+x) = x + \frac{x^2}{2} - \frac{x^3}{6} + \frac{x^4}{12} - \ldots$$

In computing F(N,M) the odd terms all cancel and the even term double, so

$$F(N,M) = -\frac{N}{2}\left[\left(\frac{M}{N}\right)^2 + \frac{1}{6}\left(\frac{M}{N}\right)^4 + O\left(\frac{M}{N}\right)^6\right].$$

Now, the first term (in the exponent) sets $M \sim \sqrt{N}$ (as we were expecting!). In this regime, the higher order terms in the bracket go as 1/N, $1/N^2$, etc., and can be neglected. Thus, finally:

$$\boxed{P_{N,M} \approx \sqrt{\frac{2}{N\pi}}\, e^{-\frac{M^2}{2N}}}.$$ (Discrete probability distribution)

Note that N in the exponent is the variance, $\sigma^2 = N$, calculated at p. 8.4 above.

This looks a lot like a Gaussian distribution; however, if you are alert, you will see that the normalisation looks wrong (more below).

This is still a discrete distribution for the variable M, which (depending on N) consists of odd or even integers.

Summary of validity:

(a) It is only valid when N, N_+, and N_- are all large (Stirling). Even for large N, it is not a good approximation out in the region were M~+N or M~-N, where $P_{M=\pm N} = \frac{1}{2^N} = e^{-N\ln 2}$ (and the approximation gives—incorrectly-- $\sqrt{\frac{2N}{\pi}}\, e^{-N/2}$).

(b) As long as $M \sim \sqrt{N}$, the corrections are of relative order 1/N.

(c) Will now show that, for N large, this region includes all outcomes that have any appreciable probability.

Question: Does this satisfy the (discrete) normalization condition $1 = \sum_{M=-N}^{N} P_{N,M}$ (M's all even/odd)?

It's OK to replace $P_{N,M}$ by $P_N(M)$, since (for large N) the variation with $M \to M \pm 2$ is smooth. However in going to a continuous variable we want to be sure that $P_{N,M} = P_N(M)dM$ for each interval $dx = 2$, so that each discrete state is counted once. It follows that:

$$P_N(M) = \frac{1}{2} P_{N,M} = \frac{1}{\sqrt{2\pi N}} e^{-\frac{M^2}{2N}},$$ (Equivalent continuous probability distribution)

which is properly normalized (in the continuum), showing that at large N, the states like M=+/- N (where Stirling fails) carry no significant weight.

Important Comments:

1. Can you predict exactly the result M of the N-coin toss experiment for large N?
Answer: No. You can only expect a range of results with a width of order $\sqrt{N}$. But, the probability, e.g., that the result is M=0 (for N even) is indistinguishable from the probability that it is M=2 or M=-2.

2. Not withstanding Point 1, if what you want to predict is the average coin-toss value $m \equiv M/N$ (your mean winning/toss), then the story is different:

Note that in the *discrete* representation $P_{N,m} = P_{N,M} = \sqrt{\frac{2}{N\pi}} e^{-\frac{Nm^2}{2}}$.

However, now increasing M by 2 means increasing m by 2/N, so in going to the continuum representation $P_{N,m} = \frac{2}{N} P_N(m)$, and

$$P_N(m) = \sqrt{\frac{N}{2\pi}} e^{-\frac{Nm^2}{2}}.$$ (note continuum normalization)

This formula shows that the "prediction" m=0 for a single run of N coin tosses is EXACT, in the sense that any other predicted outcome (m=δ?) will have zero probability of proving correct in the limit $N \to \infty$. This result shows that, for large N, you can predict with virtual certainty—not just the average of many N-coin-toss trials but—the outcome of a single (probabilistic) trial. The reason is not that anomalous outcomes (all heads) do not exist but that the probability that a particular trial will be anomalous becomes vanishingly small as N becomes large. This is important in discussing the foundations of statistical mechanics.

3. The Central Limit Theorem:
Is there something special about coin tosses that makes this so?
A: No. This result is very general for average observations of uncorrelated, statistically independent single events.
I will ask you on homework to show that it is true for the biased coin toss, $p_+ = p, p_- = q\ (= 1 - p)$.

Consider N continuous random variables $\{x_n\}_{n=1}^{N}$ each one of which is governed independently by ANY single-variable probability distribution $p(x_k)$ with

$$\langle x \rangle_p = \int_{-\infty}^{\infty} dx\, x p(x)$$

$$\langle x^2 \rangle_p = \int_{-\infty}^{\infty} dx\, x^2 p(x)$$ Note: $p(x)$ is arbitrary; however, it is crucial that $\langle x \rangle, \langle x^2 \rangle$ exist.

$$\sigma_p^2 = \langle x^2 \rangle_p - \langle x \rangle_p^2.$$

Question: What is the distribution of the (random) variable $X \equiv \sum_{n=1}^{N} x_n$? **9.4**

Claim: $P_N(X) = \frac{1}{\sqrt{2\pi N\sigma_p^2}} e^{-\frac{\left(X-N\langle x\rangle_p\right)^2}{2N\sigma_p^2}}$ (normalized) or, equivalently, for the average $x = \frac{X}{N}$,

$$P_N(x) = \sqrt{\frac{N}{2\pi\sigma_p^2}}\, e^{-\frac{N\left(x-\langle x\rangle_p\right)^2}{2\sigma_p^2}}.$$ (normalized)

This result is remarkably general. If we have time in tutorial Tuesday, I'll derive it for you. Derivation depends on Fourier transforms, Dirac's delta function, etc., and is not part of this course.