ECON30580 Economics of Betting Markets
The Law of Large Numbers and the Central Limit Theorem

Karl Whelan

School of Economics, UCD

Spring 2025

Part I

The Law of Large Numbers

Averages from Large Samples

- We have seen that just because you have a probability p of winning, that doesn't mean there is a golden rule that p will be average fraction of times you win when playing for a given length. But what if you play for a really long time? Sure on average if $p = 0.2$, then I will on average win close to 20% of the time?
- Yes: This result was also first formally proved by Jacob Bernoulli, again published posthumously in *Ars Conjectandi* (1713). He called his result "the golden theorem" but today it is known as the (Weak) Law of Large Numbers.

Weak Law of Large Numbers (WLLN): A Formal Statement

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with expected value μ and finite variance σ^2. Define the sample mean as $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for any $\varepsilon > 0$, the Weak Law of Large Numbers states that

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| \geq \varepsilon\right) = 0.$$

In other words, as the sample size n goes to infinity, the sample mean $\bar{X}_n$ converges in probability to the true mean μ.

How Long Does Convergence Take?

- Ok, but we don't have infinite amounts of time.
- How long will it take for us to be pretty sure our sample mean will be close to the true probability?
- And what determines this length of time?
- We can see from the last equation of the previous slide that the variance of X_i series plays a role in the result.
- The higher is σ^2, the lower is the probability that $\left|\bar{X}_n - \mu\right| < \epsilon$ for any given ϵ.
- This means that we have two different series $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ with the X_i having lower variance than the Y_i, then $\bar{X}_n$ will converge faster to the true mean than $\bar{Y}_n$.
- It turns out this has some important implications for gambling.

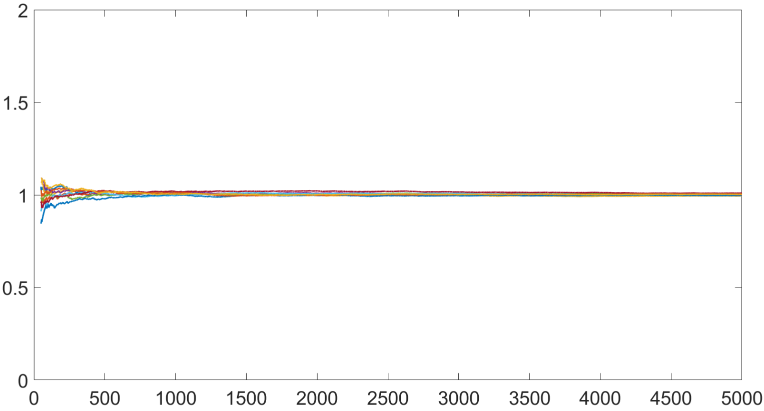
Implications for Gambling and Longshot Bets

- We have shown before that when bets have an expected payoff of k and probability of winning p, the variance of the payoff is

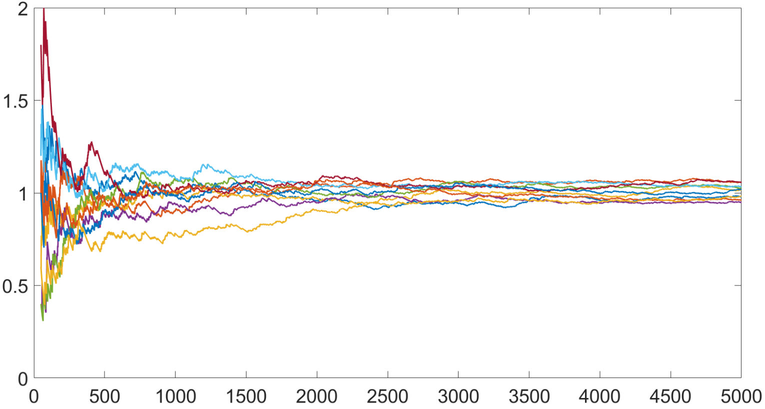
$$\text{Var}\left(X\right) = k^2\left(\frac{1}{p} - 1\right)$$

- The lower the value of p, the higher the variance will be.
- So different bets may have the same expected payoff k but the convergence of the average payoff to k will be slower for bets with low values of p.
- We use simulations to illustrate this. We let the computer do virtual "coin flip" exercises to figure out how often people will win for various values of p and then to calculate average payoffs for various game lengths.
- The next 3 pages show multiple simulations of average payoffs for various game lengths when fair value bets ($k = 1$), first for $p = 0.1$, then for $p = 0.9$ and then for $p = 0.01$. (The first 50 periods are omitted).
- The typical convergence of the average payoffs to 1 gets slower as p declines (note that the range of the graph is different for $p = 0.01$).
- In the long-run the average payoff is always close to 1 but for "longshot" bets, the long run is very very long.

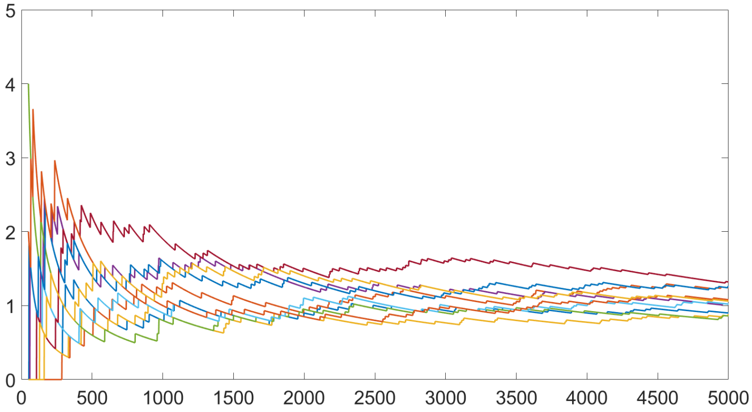
Average Payoffs From Fair Value Bets When $p = 0.9$



Average Payoffs From Fair Value Bets When $p = 0.1$



Average Payoffs From Fair Value Bets When $p = 0.01$



Part II

The Central Limit Theorem

The Central Limit Theorem

- The Law of Large Numbers tells us that the average of your series will gradually converge on the true sample mean.
- But you may want answers to more precise questions: For example, as the sample gets large, what is the probability that the sample mean is above or below a certain number?
- These questions can be answered by the **Central Limit Theorem**, one of the most incredible results in all of science. First proved by Pierre-Simon Laplace in 1810, it explains why we use Normal distributions so much.

The Central Limit Theorem: A Formal Statement

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with mean μ and variance $\sigma^2 < \infty$. Define the sample mean as $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, as n approaches infinity, the distribution of the standardized sum

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

converges in distribution to the standard normal distribution, $N(0, 1)$.

Example of How to Use the CLT

- Consider a series of observations $X_1, X_2, ..., X_n$ where n is large.
- Assume the X_i each have a mean of 5 and a standard deviation of 2.
- Because n is large, we can assume that

$$Z_n = \frac{\bar{X}_n - 5}{2\sqrt{n}} \sim N(0,1)$$

where $\sim$ means "distributed as" and $N(0,1)$ is the standard Normal distribution, meaning it has mean zero and standard deviation 1.

How Likely the Sample Mean to be Less Than 4.5?

- What is the chance that $\bar{X}_n$ will be less than 4.5 when $n = 64$?
- We calculate

$$Z_n = \frac{4.5 - 5}{2/\sqrt{64}} = \frac{0.5}{0.25} = -2$$

- Checking standard normal distribution tables shows this probability is 2.2%

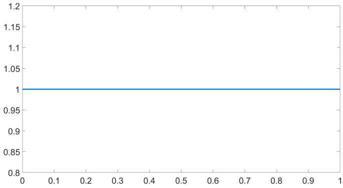
The Wonder That Is The Central Limit Theorem

- This is one of the most amazing results in science.
- Pause for a moment and think about what it means.
- Take samples of a random variable drawn from a distribution. No matter what the distribution is, if the samples are large enough, then the averages from those samples are distributed according to a Normal distribution.
- I don't have any simple intuition for the CLT. I recommend treating it like I do as some kind of miracle result.

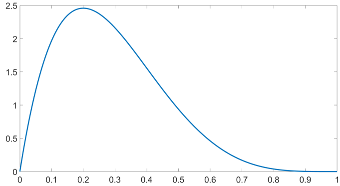
The CLT at Work for 4 Different Distributions

- On the next page, we describe 4 different probability density functions: A uniform distribution (each point being equally likely) and 3 different versions of what is known as the beta distribution. You can see that these distributions have very different shapes.
- Now consider drawing a random sample of size $n = 50$ from each of these distributions and calculating the average.
- The page after shows histograms for the dataset obtained by doing that process 100,000 times. They all look like Normal distributions.

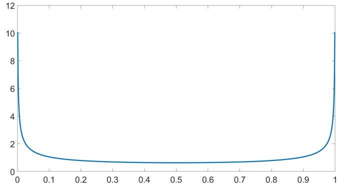
Four Different Distributions



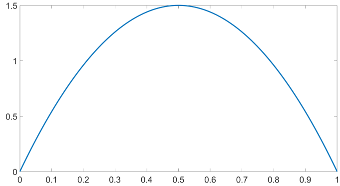
(a) Uniform Distribution



(b) Beta Distribution with $\alpha = 2$, $\beta = 0.5$

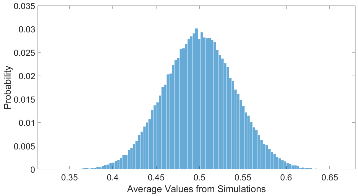


(c) Beta Distribution with $\alpha = 0.5$, $\beta = 0.5$

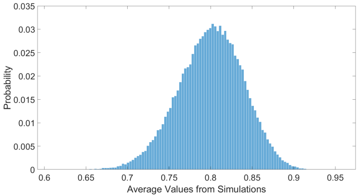


(d) Beta Distribution with $\alpha = 2$, $\beta = 2$

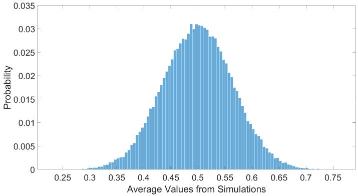
Histograms for Means Drawn From the 4 Distributions with Sample Size $N = 50$ (100,000 simulations)



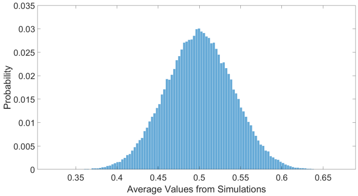
(a) Uniform Distribution



(b) Beta Distribution with $\alpha = 2$, $\beta = 0.5$



(c) Beta Distribution with $\alpha = 0.5$,



(d) Beta Distribution with $\alpha = 2$,

Asymptotic Distribution of the Sample Average

- Technically, the Central Limit Theorem says that the thing that is Normally distributed is the "z score".

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma\sqrt{n}} \sim N(0,1)$$

- But of course, we will often be interested, not in the z score, but in making probabilistic statements about the sample average itself.
- Two basic results in statistics are that when a is a fixed number and X is a random variable then

$$\begin{aligned} E(X-a) &= E(X)-a \\ \mathrm{Var}(aX) &= a^2\mathrm{Var}(X) \end{aligned}$$

- This means the following will apply to the sample mean.

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

(For those who know their asymptotics, yes there are various hand waving qualifiers needed here ...)

Application to Gambling Problems

- Consider a series of bets with profits $X_1, X_2, ..., X_n$ where n is large.
- Whatever the process is determining these profits (and losses), if you know the mean and the variance of the X_i, then you know how the sample average behaves when n is large.
- Previously, we derived a formula for the variance of payouts from \$1 bets when the probability of success was p and the expected payout was k.
- Because the profit is the payout minus the constant \$1 stake, this formula for the variance also applies to the profits from these bets.
- This gives us a mean and variance for profits

$$E(X_i) = k - 1$$
$$\text{Var}(X_i) = k^2 \left(\frac{1}{p} - 1 \right)$$

- This means that for large n

$$\bar{X}_n \sim N\left(k - 1, k\sqrt{\frac{1}{n}\left(\frac{1}{p} - 1 \right)} \right)$$

The Central Limit Theorem and Accumulated Profits

- You may not always be interested in your average profit. You may care about your total accumulated profit or loss.
- You may want to know the probability of your accumulated profit or loss being below some negative number i.e. how likely are you to lose a particular amount.
- Again assume a is a fixed number, we can use the following two basic results in statistics

$$\begin{aligned} E(aX) &= aE(X) \\ \text{Var}(aX) &= a^2\text{Var}(X) \end{aligned}$$

- Multiplying the mean by n to get the sum, we can say that when n is large enough

$$\sum_{i=1}^{n} X_i = n\bar{X}_n \sim N\left((k-1)\,n, k\sqrt{n\left(\frac{1}{p}-1\right)}\right)$$

- This allows us to make probabilistic statements about total cumulated profits or losses.

An Example of the Gambler Having an Edge

- Suppose the gambler is offered a set of bets with the following property. They can win an amount K with probability p and lose 1 with probability $1 - p$ so that their expected profit is $\mu > 0$. This means

$$pK + (1 - p)(-1) = \mu$$

- This means

$$p = \frac{1 + \mu}{K + 1}$$

- For example if $\mu = 0.01$ and $K = 1$, then $p = 0.505$.
- In this case, for each gamble we have $E(X_i) = \mu$ and using the same technique as we did in the previous notes, you can show that

$$\text{Var}(X_i) = (1 + \mu)(K + 1) - 1 - \mu^2$$

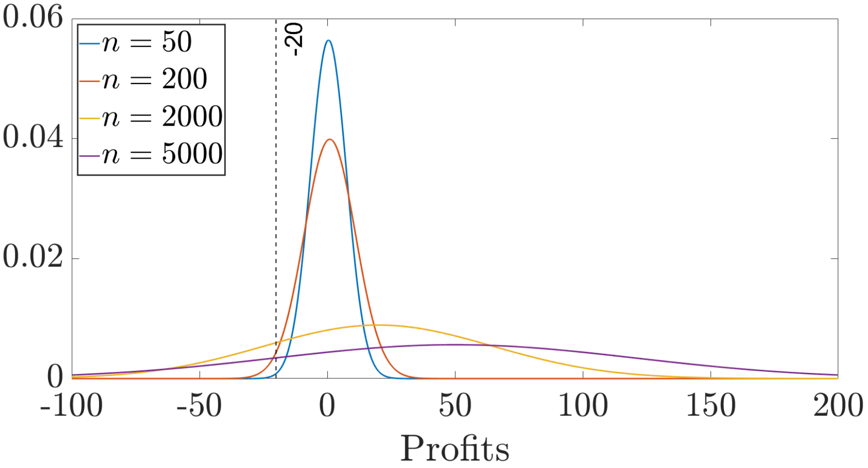
- This means the sum of accumulated profits is distributed as

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, \sqrt{[(1 + \mu)(K + 1) - 1 - \mu^2]\, n}\right)$$

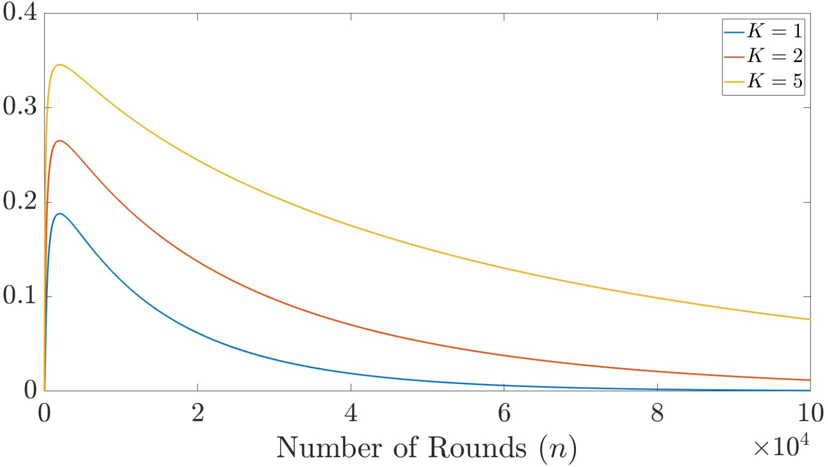
How Likely Are You to Lose a Specific Amount?

- Let's assume you choose repeated €1 bets gambles with $p = 0.5$ and $\mu = 0.01$.
- What is the probability after n bets that you will have lost €20?
- You have an "edge" here—on average you expect to make a profit, so you might expect that the longer you play, the less likely you are to lose €20.
- And indeed, the mean of your accumulated profit gradually increases—that makes it less likely that you can lose €20.
- But so does the variance—and that makes it more likely you can lose €20.
- On the next page, you can see the CLT-based probability distributions for cumulated profits for $n = 50$ and bigger samples.
- As your sample size increases from small values, your chance of losing €20 increases and it only starts to fall when $n = 2000$.
- The page after shows the probability of losing €20 after n rounds with $\mu = 0.01$ for 3 different values of K. The higher K gets (so the lower the probability of winning) the more likely you are to lose a fixed amount at each point in time. This is because the variance of profits increases as K increases.

Distribution of cumulated profits after playing N Rounds with an expected profit per round of $\mu = 0.01$



Probability of losing 20 after varying numbers of rounds with an expected profit per round of $\mu = 0.01$



Comment on The Last Chart (Not on the exam)

- Note that the peak value for the probability of losing 20 occurred for playing precisely 2000 rounds for each value of K.
- Why was this?
- This is because we are charting the cumulative probability distribution of accumulated profits at -20 which depends negatively on the z-score

$$z_n = \frac{-20 - \mu n}{\sqrt{\left[(1+\mu)(K+1) - 1 - \mu^2\right] n}}$$

- The term $\frac{1}{\sqrt{(1+\mu)(K+1)-1-\mu^2}}$ multiplies this value but does not have an impact on which value of n attains the minimum z score. So the minimum value doesn't depend on K (hence the peak is the same for each value of K in the chart).
- Differentiating z_n with respect to n and setting it equal to zero, one can show this minimum value occurs at $n = \frac{20}{\mu}$, which is 2,000 in the case
- This is why the probability of losing 20 reaches a peak after $n = 2000$ rounds of play.

Implications for Repeated Gambles

- If you make a series of gambles of the same size and you know the expected return on your gambles and the variance of your profits, then you can make probability statements about the average profit per bet and also about your total accumulated profit.
- As the length of your gambling streak increases, the distribution of the average profit centres more and more closely around the true mean.
- But the distribution of your accumulated total profits has higher and higher variance as the length of your streak increases. And this is particularly true if you are making long odds bets.
- If you have an edge, so the expected profit is positive, you might imagine that all you have to do to avoid losses is to play for long enough and let this long-run advantage kick in.
- This is true in the very very long run but if your edge is modest—like the 1% edge modelled here—you can be playing a long time and still have lost money.
- And this assumes that you can somehow keep absorbing losses long enough for your long-run advantage to kick in.
- What happens if you lose all your money and can't gamble anymore?