0.0.1 Intuition about convergences

I want to contrast two statements:

1. X and Y are close together.
2. X and Y have similar distributions.

The truth of the first statement depends on the *joint* distribution of X and Y. On the other hand, the truth of the second statement depends only on the *marginal* distributions of X and Y. Both ideas are used in *large sample theory* which is the process of describing mathematically the behaviour of statistical procedures *approximately* in the presence of lots of data.

0.0.2 Relation between convergence and approximation

In this subsection I present some approximations and the limits they come from.

Example: Stirling's approximation for the factorial function is

$$n! \approx \sqrt{2\pi} n^{n+1/2} e^{-n} \equiv s_n$$

Here is a small table which shows that the approximation is actually about ratios; it has small *relative* error and large *absolute* error:

n	$n!$	s_n	$n!/s_n$
5	120	118.019	1.01678
10	3628800	3598695.619	1.008365

Example: The normal approximation to the Binomial distribution. Toss a fair coin 100 times, and let X denote the number of heads. Then

$$P(40 \le X \le 60) \approx \Phi(\frac{60-50}{\sqrt{25}}) - \Phi(\frac{40-50}{\sqrt{25}}) = 0.9544997.$$

In the same context here is a slightly better approximation, usually called a continuity correction:

$$P(40 \le X \le 60) \approx \Phi(\frac{60.5-50}{\sqrt{25}}) - \Phi(\frac{39.5-50}{\sqrt{25}}) = 0.9642712$$

In the same context how should we approximate

$$P(X = 50) \approx ?$$

0.0.3 Associated Limits

Now I want to describe limit theorems which correspond to these approximations. For Stirling's formula we have

Theorem 1

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

The theorem is used when $n = 100$ to give, for instance,

$$100! \approx \sqrt{2\pi 100} 100^{100} e^{-100} = 9.3326215443944152682 \times 10^{157}$$

All those digits are right (but the absolute error which is the approximation minus 100 factorial) is huge.

Theorem 2 *if* $X_n \sim \text{Binomial}(n, 1/2)$ *then*

$$\lim_{n\to\infty} P\left(\frac{X_n - n/2}{\sqrt{n/4}} \leq x\right) = \Phi(x)$$

This theorem is used with $x\sqrt{100/4} + 100/2$ equal to 60 and 40 or 60.5 and 39.5. It is used with 49.5 and 50.5 to get $P(X_{100} = 50)$ approximately.

Summary

- If we want to compute x_{100} we compute $y = \lim_{n\to\infty}$ and approximate $x_{100} \approx y$.
- There are often many different ways to think of x_{100} as an entry in some sequence. These can lead to different approximations. Some of the approximations are lousy while some are great.

Now I want to contrast taking limits of random variables with taking limits of distributions. I want to begin with a motivating example. We do an experiment to measure probability that a dropped tack lands point up. We drop a tack n times and observe $X_n \sim \text{Binomial}(n, p)$, which is the number of times tack lands point up. Here are two common random variables to study:

$$U_n \equiv \frac{X_n - np}{\sqrt{np(1-p)}} \quad \text{and} \quad V_n \equiv \frac{X_n - np}{\sqrt{n\hat{p}(1-\hat{p})}}$$

where $\hat{p} = X/n$.

The first of these is used in hypothesis testing and the second is used to form confidence intervals. One approximation is

$$U_n \approx V_n$$

The idea is that this approximation is correct because the estimated and theoretical standard errors of $\hat{p} = X_n/n$ are very similar. To get a feeling for the meaning of this assertion I did the following experimented. I generated (using the `R` function `rbinom`) 1000 values of X with $n = 100$ and $p = 0.4$. Then for each value of x between the smallest and largest values in the sample I worked out U and V. Figure 1 is a plot of U vs V.

In this case the plot shows

$$P(|U_n - V_n| \text{ is big})$$

is small. (The points are close to the line $y = x$.) I should add that if I had run x from 1 to 99 the graph would look a lot less straight. The two formulas are quite different if $\hat{p}$ is not close to p.

Definition: A sequence of random variables X_n *converges in probability* to a random variable X if for every $\epsilon > 0$ we have

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

Definition: A sequence of random variables X_n *converges almost surely* to a random variable X if

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

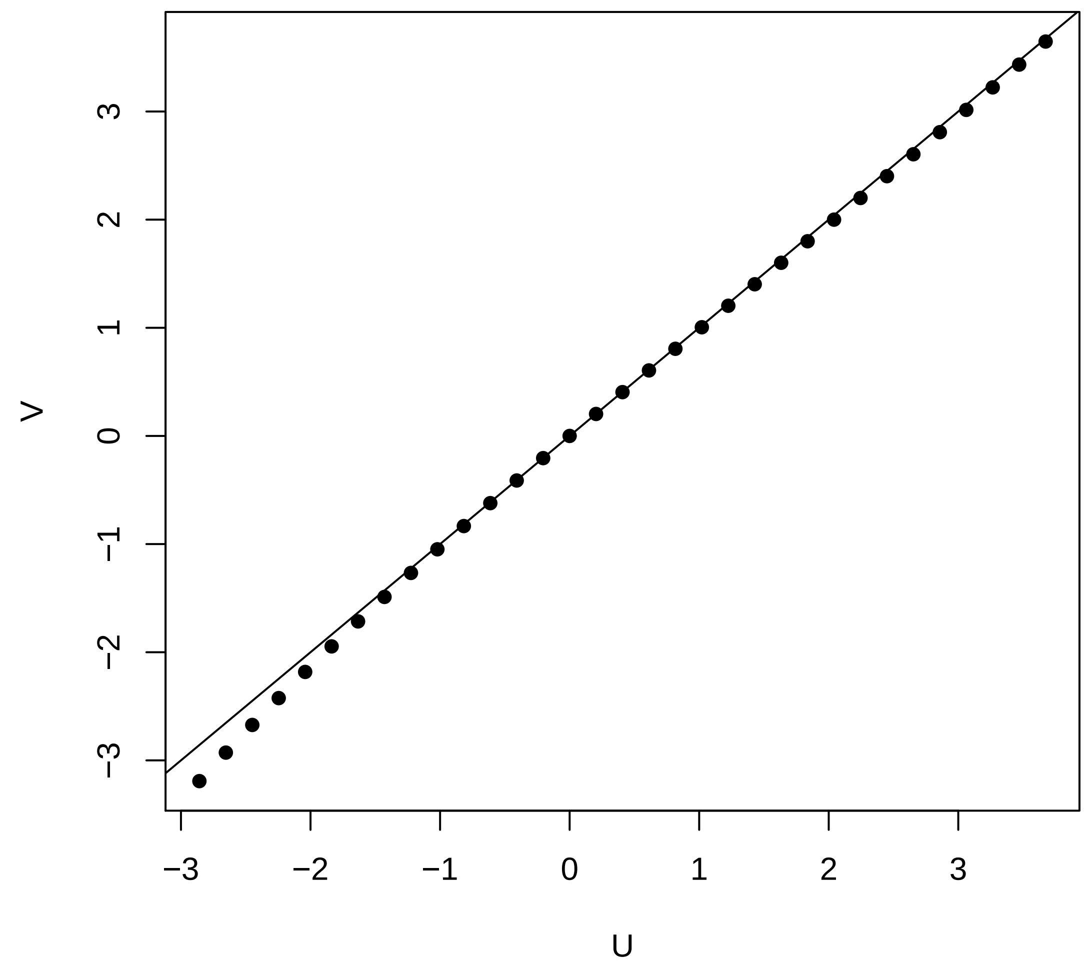
Definition: A sequence of random variables X_n *converges in mean* or *converges in L_1* to a random variable X if

$$\lim_{n\to\infty} \mathrm{E}(|X_n - X|) = 0.$$

Definition: A sequence of random variables X_n *converges in quadratic mean* or *converges in L_2* to a random variable X if

$$\lim_{n\to\infty} \mathrm{E}(|X_n - X|^2) = 0.$$

Figure 1: Plot comparing two approximate pivots for the binomial distribution. On the x-axis we have the pivot using true standard error of $\hat{p}$ while on the y-axis we use the estimated standard error.



For pth mean we use

$$\lim_{n\to\infty} \mathrm{E}(|X_n - X|^p) = 0.$$

Now I return to our example. In fact $U_n - V_n$ converges to 0 in probability and $U_n - V_n$ converges to 0 almost surely. But they do not converge in pth mean because V_n does not have a finite mean ($P(\hat{p} = 0) > 0$).