IEOR 6711: Stochastic Models I

Fall 2012, Professor Whitt, Tuesday, September 11

Normal Approximations and the Central Limit Theorem

Time on my hands: Coin tosses.

Problem Formulation: Suppose that I have a lot of time on my hands, e.g., because I am on the New York subway travelling throughout the entire system, from one end to another. Fortunately, I have a coin in my pocket. And now I decide that this is an ideal time to see if heads will come up half the time in a large number of coin tosses.

Specifically, I decide to see what happens if I toss a coin many times. Indeed, I toss my coin $1,000,000$ times. Below are various possible outcomes, i.e., various possible numbers of heads that I might report having observed:

1. 500,000
2. 500,312
3. 501,013
4. 511,062
5. 598,372

What do you think of these possible reported outcomes? How believable are these reported outcomes? Which are reasonable answers? How likely are these outcomes?

(a) Answers

We rule out outcome 5; there are clearly too many heads. That first step is easy; it does not require any careful analysis.

The next one is a little tricky: We rule out outcome 1, because it is "too perfect." Even though $500,000$ is the most likely single outcome (the mode of the distribution), it itself is extremely unlikely. It just does not seem credible that it would actually occur in a single experiment. This is inserting judgment, beyond the mathematics. However, the mathematics can quantify how likely is this particular outcome.

Carefully examining the too-perfect case. First, here in these notes we quantify just how unlikely is the "too perfect" outcome of exactly the mean value $500,000$ in 10^6 tosses.

We introduce a probability model. We assume that successive coin tosses are independent and identically distributed (IID) with probability of $1/2$ of coming out heads. (It is important to realize that we actually have performed a modeling step.)

Let S_n denote the number of heads in n coin tosses. The random variable S_n has exactly a binomial distribution. If the probability of heads in one toss were p, then the probability of k heads in n tosses would be

$$P(S_n = k) = b(k; n, p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{(n-k)}$$

Now we are interested in the case and $p = 1/2$ and $k = n/2$ (for n even), i.e.,

$$P(S_n = n/2) = b(n/2; n, 1/2) = \frac{n!}{(n/2)!(n/2)!}(1/2)^n.$$

It is good to be able to roughly estimate these probabilities. To do so, we can use Stirling's formula (see p. 146):

$$n! \approx \sqrt{2\pi n}(n/e)^n.$$

We thus see that

$$P(S_n = n/2) = b(n/2; n, 1/2) \approx \frac{\sqrt{2\pi n}(n/e)^n}{(\pi n)(n/2e)^n}(1/2)^n = \sqrt{2/\pi n} \approx \frac{0.8}{\sqrt{n}}.$$

Hence, the probability of getting the outcome is approximately 0.8/1000, less than 1/1000. Of course, this special outcome is the most likely single outcome, and it could of course occur, but a probability less than 1/1000 is quite unlikely.

But how do we think about the remaining three alternatives?

The other possibilities require more thinking. We can answer the question by applying a **normal approximation**; see p. 41 of the Blue Ross or see Section 2.7, especially pages 79-83, of the Green Ross.

The random variable S_n has exactly a binomial distribution, but it also is approximately normally distributed with mean $np = 500,000$ and variance $np(1-p) = 250,000$. (I stress that the approximation tends to be more useful than the exact distribution.) Thus S_n has standard deviation $SD(S_n) = \sqrt{Var(S_n)} = 500$. Case 2 looks likely because it is less than 1 standard deviation from the mean; case 3 is not too likely, but not extremely unlikely, because it is just over 2 standard deviations from the mean. On the other hand, Case 4 is extremely unlikely, because it is more than 22 standard deviations from the mean. See the Table on page 81 of the Green Ross. From a practical perspective, we see that candidate 2 is reasonable and even likely, candidate 3 is possible but somewhat unlikely, but candidate 4 is extremely unlikely, and thus also not really credible. In summary, we use the analysis to eliminate case 4.

However, even then, we should be careful. Case 4 is conclusive if the model is accurate. In Case 4 we can reject the hypothesis of the coin having probability 1/2 of coming out heads. Indeed, there is recent research showing that the probability a coin flip comes out the same as the initial side up is actually about 0.51. See "Dynamical bias in the coin toss," by Persi Diaconis, Susan Holmes and Richard Montgomery: http://comptop.stanford.edu/preprints/heads.pdf

(b) The Power of the CLT

The normal approximation for the binomial distribution with parameters (n, p) when n is not too small and the normal approximation for the Poisson with mean λ when λ is not too small both follow as special cases of the central limit theorem (CLT). The CLT states that a properly normalized sum of random variables *converges in distribution* to the normal distribution. Let $N(a, b)$ denote a normal distribution with mean a and variance b.

Theorem 0.1 (central limit theorem) *Let $\{X_n : n \geq 1\}$ be a sequence of independent and identically distributed (IID) random variables with finite mean and variance: $EX_n = m < \infty$ and $Var(X_n) = \sigma^2 < \infty$. Let*

$$S_n \equiv X_1 + \cdots + X_n, \quad n \geq 1 \ .$$

Then

$$\frac{S_n - nm}{\sqrt{n\sigma^2}} \Rightarrow N(0,1) \ .$$

Where does the **sum** appear in our application? A random variable that has a binomial distribution with parameters (n,p) can be regarded as the sum of n IID random variables with a Bernoulli distribution having parameter p; i.e., we can assume $P(X_n = 1) = 1 - P(X_n = 0) = p$. (In our case, $p = 1/2$.) Similarly, a random variable having a Poisson distribution with mean λ can be regarded as the sum of n IID random variables, each with a Poisson distribution with mean λ/n (for any n). (In an exercise you will be showing that the sum of independent Poisson random variables is again Poisson with a mean equal to the sum of the means.)

And what about the **normalization**? We simply subtract the mean of S_n and divide by the standard deviation of S_n to make the normalized sum have mean 0 and variance 1. Note that

$$\frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} = \frac{S_n - nm}{\sqrt{n\sigma^2}} \tag{1}$$

has mean 0 and variance 1 whenever

$$S_n \equiv X_1 + \cdots + X_n \ ,$$

where $\{X_n : n \geq 1\}$ is a sequence of IID random variables with mean m and variance σ^2. (It is crucial that the mean and variance be finite.) But we can make any random variable with finite mean and variance have mean 0 and variance 1 by first subtracting the mean and then dividing by the standard deviation. For a random variable with a normal distribution, such scaling produces a new random variable, again with a normal distribution. This normalization is useful to produce a well-defined limit; the distribution of S_n has mean and variance growing proportional to n. The normalization is important to formulate a proper limit, but it does not explain the normality. The normalization produces a scaled random variable with mean 0 and variance 1 independent of n. It is thus not too surprising that the distribution might actually converge as well. But the CLT shows that the limit is always $N(0,1)$.

The CLT applies much more generally; it has remarkably force. **The random variables being added do not have to be Bernoulli or Poisson; they can have any distribution.** We only require that the distribution have finite mean m and variance σ^2. The statement of a basic CLT is given on p. 41 of the Blue Ross and in Theorem 2.2 on p. 79 of the Green Ross. The conclusion actually holds under even weaker conditions. The random variables being added do not actually have to be independent; it suffices for them to be "weakly dependent;" and the random variables do not have to be identically distributed; it suffices for no single random variable to be large compared to the sum. But the statement then need adjusting: the first expression in (1) remains valid, but the second does not.

What does the CLT say? The precise mathematical statement is a **limit** as $n \to \infty$. It says that, as $n \to \infty$, the normalized sum in (1) **converges in distribution** to $N(0,1)$, a random variable that has a normal distribution with mean 0 and variance 1, whose distribution is given in the table on page 81 of the Green Ross.

What does convergence in distribution mean? It means that the cumulative distribution functions (cdf's) converge to the cdf of the normal limit, i.e.,

$$P\left(\frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} \leq x\right) \to P\left(N(0,1) \leq x\right) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \, dy$$

for all x. Note that convergence in distribution means convergence of cdf's, which means convergence of functions. Here the functions converge pointwise.

However, in general, convergence in distribution of random variables is not simply pointwise convergence of cdf's. There is a slight twist. For a general sequence of random variables $\{Z_n : n \geq 1\}$, we say that Z_n converges in distribution to another random variable Z, and write

$$Z_n \Rightarrow Z \quad \text{as} \quad n \to \infty$$

if

$$F_n(x) \to F(x) \quad \text{as} \quad n \to \infty$$

for all x that are continuity points of the limiting cdf F, where F_n is the cdf of Z_n, $n \geq 1$, and F is the cdf of Z, i.e.,

$$F_n(x) \equiv P(Z_n \leq x) \quad \text{and} \quad F(x) \equiv P(Z \leq x) \quad \text{for all} \quad x \; .$$

But we only require that convergence for those points (arguments) x that are continuity points of the limiting cdf F. (A point x is a continuity point of F if F is continuous at x.) The first homework assignment discusses the need for the extra qualification.

We can define convergence in distribution in terms of probability measures instead of cdf's. That extension is important for more general random objects, like random functions (often called stochastic processes), because then the probability distribution is not characterized by a cdf. We might then want to say that

$$Z_n \Rightarrow Z \quad \text{as} \quad n \to \infty$$

if

$$P(Z_n \in A) \to P(Z \in A) \quad \text{as} \quad n \to \infty$$

for all subsets A in the sample space. First, we must require that the subset be measurable, but actually that is not enough. We need to require that A be what is called a "continuity set," where its boundary has probability 0 with respect to the limiting probability measure; see Sections 3.2 and 11.3 of my book Stochastic-Process Limits (available online: http://www.columbia.edu/ ww2040/jumps.html. The book is mostly beyond the scope of this course, but a few sections may be helpful. The relevant sections are Sections 3.2, 3.4 (in Chapter 3) and 11.3, 11.4 (in Chapter 11). Proofs omitted there mostly appear in P. Billingsley, Convergence of Probability Measures (two editions: 1968 and 1999). There is a long Internet Supplement to my book as well; Chapter 1 might be of interest (http://www.columbia.edu/ ww2040/supplement.html) It has proofs of some of the key fundamental results. (Again mostly beyond the scope of this course.)

Our specific example of coin tossing, where S_n has exactly a binomial distribution, illustrates why we do not want to require convergence for all measurable subsets A. Because we can have

$$P\left(\frac{S_n - nm}{\sqrt{n\sigma^2}} \in A\right) = 1, \quad \text{while} \quad P(N(0,1) \in A) = 0 \; .$$

It suffices to let A be the set of all possible values of the normalized sum on the left, which is a countably infinite set. In particular, it suffices to let

$$A \equiv \{(k - nm)/\sqrt{n\sigma^2} : 0 \leq k \leq n, \quad n \geq 1\} \; .$$

How do we apply the CLT? We *approximate* the distribution of the normalized sum in (1) by the distribution of $N(0,1)$. The standard normal (with mean 0 and variance 1) has no parameters at all; its distribution is given in the Table on page 81 of the Green Ross. By scaling, we can reduce other normal distributions to this one.

(c) An Application of the CLT: Modelling Stock Prices

Given the generality of the CLT, it is nice to consider an application where the random variables being added in the CLT are not Bernoulli or Poisson, as in many applications. Hence we consider such an application now.

(i) An Additive Model for Stock Prices

We start by introducing a random-walk (RW) model for a stock price. Let S_n denote the price of some stock at the end of day n. We then can write

$$S_n = S_0 + X_1 + \cdots + X_n \ , \tag{2}$$

where X_i is the *change* in stock price between day $i-1$ and day i (over day i) and S_0 is the initial stock price, presumably known (if we start at current time and contemplate the evolution of the stock price into the uncertain future.

We now make a probability model. We do so by assuming that the successive changes come from a sequence $\{X_n : n \geq 1\}$ of IID random variables, each with mean m and variance σ^2. This is roughly reasonable. Moreover, we do not expect the distribution to be Bernoulli or Poisson. The stochastic process $\{S_n : n \geq 0\}$ is a **random walk** with steps X_n, but a general random walk. If the steps are Bernoulli random variables, then we have a simple random walk. But here the steps can have an arbitrary distribution.

We now can apply the CLT to deduce that the model implies that we can approximate the stock price on day n by a normal distribution. In particular,

$$P(S_n \leq x) \approx P(N(S_0 + nm, n\sigma^2) \leq x) = P(N(0,1) \leq (x - S_0 - nm)/\sigma x) \ .$$

How do we do that last step? Just re-scale: subtract the mean from both sides and then divide by the standard deviation for both sides, inside the probabilities. The normal variable is then transformed into $N(0,1)$. We can clearly estimate the distribution of X_n by looking at data. We can investigate if the stock prices are indeed normally distributed.

(ii) A Multiplicative Model for Stock Prices

Actually, many people do not like the previous model, because they believe that the change in a stock price should be somehow proportional to the price. (There is much much more hard-nosed empirical evidence, not just idle speculation.) That leads to introducing an alternative multiplicative model of stock prices. Instead of (2) above, we assume that

$$S_n = S_0 \times X_1 \times \cdots \times X_n \ , \tag{3}$$

where the random variables are again IID, but now they are random daily multipliers. Clearly, the random variable X_n will have a different distribution if it is regarded as a multiplier instead of an additive increment.

But, even with this modification, we can apply the CLT. We obtain an additive model again if we simply take logarithms (using any base, but think of standard base $e = 2.71828....$). Note that

$$\log{(S_n)} = \log{(S_0)} + \log{(X_1)} + \cdots + \log{(X_n)} \ , \tag{4}$$

so that

$$\log(S_n) \approx N(\log(S_0) + nm, n\sigma^2) \ , \tag{5}$$

where now (with this new interpretation of X_n)

$$m \equiv E[\log(X_1)] \quad \text{and} \quad \sigma^2 \equiv Var(\log(X_1) \ . \tag{6}$$

As a consequence, we can now take exponentials of both sides of (5) to deduce that

$$S_n \approx e^{N(\log S_0 + nm, n\sigma^2)} \ . \tag{7}$$

That says that S_n has a *lognormal* distribution. Some discussion of this model appears on page 608 of the Green Ross. It underlies *geometric Brownian motion*, one of the fundamental stochastic models in finance; see p. 368-9 of the Blue Ross.