Consistency and Limiting Distributions (Hogg Chapter Five)

STAT 405-01: Mathematical Statistics I *

Fall Semester 2015

Contents

Tuesday 1 December 2015 – Read Section 5.1 of Hogg

0 Motivation

0.1 The Central Limit Theorem

Our last topic of the semester will be the idea of convergence of a sequence of random variables, with the main motivation being to prove the Central Limit Theorem, which says that if you add many independent, identically distributed random variables, which are not necessarily normal random variables, their sum will be approximately normally distributed. (The same will be true for their average.) We have to be a bit careful here, though. We know that if we have n iid random variables $\{X_i\}$ with mean $\mu = E(X_i)$ and variance $\sigma^2 = \text{Var}(X_i)$, their sum will have mean $E(\sum_{i=1}^n X_i) = n\mu$ and variance $\text{Var}(\sum_{i=1}^n X_i) = n\sigma^2$, so we can't quite talking about convergence to anything as $n \to \infty$. However, if we use the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ to construct

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{0.1}$$

that is a combination whose mean is 0 and whose variance is 1 for all n; the central limit theorem says that this sequence converges to a $N(0,1)$, i.e., standard normal random variable. When we construct a large-sample confidence interval using the percentiles (e.g., z_α) of the standard normal distribution, we are invoking an approximate form of the central limit theorem.

Before we can prove the central limit theorem, though, we need to define what it means to say a sequence $\{X_n\}$ of random variables converges as $n \to \infty$. There are actually two different sorts of convergence we can talk about, *convergence in probability* and *convergence in distribution*. The distinction between them is easiest to understand if we take a step back and consider what it means for two random variables to be equal in the first place.

0.2 Equality of Random Variables

When we talk about two random variables X and Y, and write $X = Y$, what we're saying is that, for every outcome of the experiment, the realizations of X and Y are equal. This implies that[1]

$$P(X = Y) = 1 \qquad \text{if } X = Y \tag{0.2}$$

This kind of equality is the analogue of convergence in probability. Convergence in probability is written $X_n \xrightarrow{P} X$, and we might define "equality in probability" and write something like $X \overset{P}{=} Y$ by analogy but in practice we just call it equality and write it $X = Y$.

We can also say that two random variables are equal in distribution if they have the same probability distribution. For example, any two $N(0,1)$ random variables are equal in distribution, as are any two random variables from a random sample. Formally, we define

$$F_X(x) = P(X \le x) = P(Y \le x) = F_Y(x) \qquad \text{means } X \overset{D}{=} Y \tag{0.3}$$

The analogy of equality in distribution is convergence in distribution, which we write $X_n \xrightarrow{D} X$. This is the sort of convergence we will use when stating the central limit theorem.

Note that it's easy to come up with examples where $X \overset{D}{=} Y$ but $X \neq Y$. For instance, if X is $N(0,\sigma^2)$ so that

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \tag{0.4}$$

and we define $Y = -X$. Obviously, $Y \neq X$, but if we work out the pdf for Y, we find

$$f_Y(y) = \left|\frac{d(-y)}{dy}\right| f_X(-y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(-y)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} \tag{0.5}$$

so Y is also $N(0,\sigma^2)$, and $X \overset{D}{=} Y$.

1 Convergence in Probability

Recalling that equality ($X = Y$) implies $P(X = Y) = 1$, we define convergence in probability as follows:

$$X_n \xrightarrow{P} X \qquad \text{iff} \qquad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1 \text{ for all } \epsilon > 0 \tag{1.1}$$

or equivalently

$$X_n \xrightarrow{P} X \qquad \text{iff} \qquad \lim_{n\to\infty} P(|X_n - X| \ge \epsilon) = 0 \text{ for all } \epsilon > 0 \tag{1.2}$$

[1] Note that this is actually a necessary but not sufficient condition; it's possible for X and Y to disagree for a set of outcomes whose probability is zero, e.g., if X is a $\chi^2(2)$ random variable, and Y is defined to be X if $X \neq 0$, and -1 if $X = 0$.

To give an example of convergence in probability, and also a way to prove when a sequence converges, we consider a random sample of size n drawn from some distribution with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$, We define the sample mean in the usual way

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.3}$$

where the subscript n is to emphasize that this is the mean of a sample of a particular size. We know that

$$E(\overline{X}_n) = \mu \qquad \text{and} \qquad \text{Var}(\overline{X}_n) = \sigma^2/n \tag{1.4}$$

We will now show that the sequence of sample means actually converges to the distribution mean μ, i.e., $\overline{X}_n \xrightarrow{P} \mu$, as follows. To consider $P(\left|\overline{X}_n - \mu\right| \geq \epsilon)$, we can use Chebyshev's inequality with k chosen so that $k\sigma/\sqrt{n} = \epsilon$, i.e., $k = \epsilon\sqrt{n}/\sigma$, so that

$$P(\left|\overline{X}_n - \mu\right| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \tag{1.5}$$

if we take the limit as $n \to \infty$, we see that indeed

$$\lim_{n\to\infty} P(\left|\overline{X}_n - \mu\right| \geq \epsilon) = 0 \tag{1.6}$$

This is known as the *weak law of large numbers*. It's "weak" because we had to assume $\sigma^2 < \infty$ and therefore it wouldn't work for a sample drawn from e.g., a t distribution with 2 degrees of freedom, which has an infinite second moment. There's actually a strong law of large numbers that shows this works even when the variance doesn't exist, but it's beyond the scope of this course.

Note that most of the useful examples of convergence in probability involve a sequence of random variables converging to a *degenerate random variable*, i.e., one which only takes on one value, and look something like $X_n \xrightarrow{P} a$. We can also see that this Chebyshev method should work to show that $E(X_n)$ converges to its mean (or the limit of its mean) if $\text{Var}(X_n)$ goes to zero as n goes to infinity.

Hogg proves a sequence of theorems about convergence in probability (the weak law of large numbers is actually Theorem 5.1.1), which we'll just state here:

1. Theorem 5.1.2: If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ then $X_n + Y_n \xrightarrow{P} X + Y$.
2. Theorem 5.1.3: If $X_n \xrightarrow{P} X$ then $aX_n \xrightarrow{P} aX$ for any constant a.
3. Theorem 5.1.5: If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ then $X_nY_n \xrightarrow{P} XY$.
4. Theorem 5.1.4: If $X_n \xrightarrow{P} a$ and $g(x)$ is a function which is continuous at $x = a$, then $g(X_n) \xrightarrow{P} g(a)$. So for example, you can show that the sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ converges to σ^2, which implies that $S_n = \sqrt{S_n^2} \xrightarrow{P} \sigma$.

One other definition is the idea of consistency. Recall that we said a statistic T is an unbiased estimator of a parameter θ if $E(T) = \theta$. We say that

$$T_n \text{ is a consistent estimator of } \theta \text{ iff } T_n \xrightarrow{P} \theta \tag{1.7}$$

It's easy to come up with examples where a sequence of estimators is unbiased but not consistent, or vice versa.

- The statistic $T_n = X_n$ is obviously an unbiased estimator of the mean μ, but doesn't converge to anything, so it's not a consistent estimator.
- The maximum likelihood estimator

$$\widehat{\sigma^2}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{n-1}{n}S^2 \tag{1.8}$$

of the variance is a biased estimator, since $E(\widehat{\sigma^2}_n) = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2$, but it is still consistent because $\widehat{\sigma^2}_n \xrightarrow{P} \sigma^2$. (The bias goes to zero in the limit $n \to \infty$.)

Thursday 3 December 2015 – Read Section 5.2 of Hogg

2 Convergence in Distribution

Recall that, in analogy with the two kinds of equality for random variables, namely $X = Y$ which implies $P(X = Y) = 1$, and $X \stackrel{D}{=} Y$ which means $P(X \le x) = P(Y \le x)$, we define two kinds of convergence for a sequence of random variables $\{X_n\}$ as $n \to \infty$. Last time we considered convergence in probability, $X_n \xrightarrow{P} X$, where $\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$ for all $\epsilon > 0$. Now we turn to convergence in distribution, which is defined as

$$X_n \xrightarrow{D} X \qquad \text{iff} \qquad \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \text{wherever } F_X \text{ cts} \tag{2.1}$$

To see why we only consider points where the limiting cdf is continuous, consider the sequence $X_n = a + \frac{1}{n}$. We know that X_n converges in probability to the degenerate random variable $X = a$, and it seems like it should converge in distribution as well. The cdf for each random variable in the sequence is

$$P(X_n \le x) = \begin{cases} 0 & x < a + \frac{1}{n} \\ 1 & a + \frac{1}{n} \le x \end{cases} \tag{2.2}$$

while the cdf for the expected limiting distribution is

$$P(X \le x) = \begin{cases} 0 & x < a \\ 1 & a \le x \end{cases} \tag{2.3}$$

We see that if $x < a$, we do indeed have $P(X_n \le x) = 0$ for all n. If $x > a$, we likewise have, for all $n > \frac{1}{x-a}$, $P(X_n \le x) = 1$. However, if $x = a$, we have $P(X_n \le x) = 0$ for all n, so the sequence of cdfs converges to

$$\lim_{n\to\infty} P(X_n \le x) = \begin{cases} 0 & x \le a \\ 1 & a < x \end{cases} \tag{2.4}$$

which is not quite the cdf of X. (In fact, it isn't a cdf at all, since it's not right continuous.) But they only disagree where $P(X \le x)$ is discontinuous, so the definition of convergence in distribution is satisfied, and $X_n \xrightarrow{D} X = a$.

Note that since convergence in distribution refers only to the cdf of the limiting random variable, we can unambiguously refer to a sequence as converging to a distribution rather than specifying a random variable that obeys that distribution. So e.g., if $X_n \xrightarrow{D} X$ where X is a $\chi^2(5)$ random variable, we can just as well say $X_n \xrightarrow{D} \chi^2(5)$. We refer to this distribution as the *limiting distribution* of the sequence.

Hogg states or proves a number of theorems about convergence in distribution, which we'll state here without proof:

1. Theorem 5.2.1: If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{D} X$. This is to be expected, since convergence in probability is "stronger" than convergence in distribution. (Again, thinking about the analogous point for equality, two random variables which are equal clearly obey the same distribution.)
2. Theorem 5.2.2: If $X_n \xrightarrow{D} a$, then $X_n \xrightarrow{P} a$, i.e.,

$$\lim_{n\to\infty} P(X_n \le x) = \begin{cases} 0 & x < a \\ 1 & a < x \end{cases} \tag{2.5}$$

   Note that we don't need to say what the limit of the cdf is for $x = a$, since the definition of convergence in probability excludes points where the limiting cdf is discontinuous.

We wouldn't expect the converse of Theorem 5.2.1 to hold; convergence in distribution doesn't imply convergence in probability, since different random variables can obey the same distribution. On the other hand, in the special case of convergence to a degenerate random variable which is always equal to the same number, it does work. This also makes sense, since two random variables with the same degenerate distribution are in fact equal (since they each equal the same number with 100% probability).

3. Theorem 5.2.3: If $X_n \stackrel{D}{\rightarrow} X$ and $Y_n \stackrel{P}{\rightarrow} 0$, then $X_n + Y_n \stackrel{D}{\rightarrow} X$.
4. Theorem 5.2.4: If $X_n \stackrel{D}{\rightarrow} X$ and $g(x)$ is a continuous function on the support of X, then $g(X_n) \stackrel{D}{\rightarrow} g(X)$.
5. Theorem 5.2.5: If $X_n \stackrel{D}{\rightarrow} X$, $A_n \stackrel{P}{\rightarrow} a$ and $B_n \stackrel{P}{\rightarrow} b$, then $A_n X_n + B_n \stackrel{D}{\rightarrow} aX + b$. This is known as Slutsky's Theorem.

Convergence in distribution is defined using the cumulative distribution function, which is useful because it works with both continuous and discrete distributions. (This is especially helpful if we want to show that a sequence of discrete random variables converges in distribution to a continuous random variable.) We'll see that the moment generating function can also be used to show convergence. However, there are cases where the mgf doesn't exist, and the cdf is not a convenient function to work with. In these cases, we need to consider the pdf of pmf as appropriate.

For example, let $\{T_n\}$ be a sequence where T_n obeys a t-distribution with n degrees of freedom. We know that the pdf is

$$f_T(t;n) = \frac{\Gamma([n+1]/2)}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-[n+1]/2} \tag{2.6}$$

We can't just consider the limit of the pdf, though; we're supposed to work with the cdf, which is

$$P(T_n \le t) = \int_{-\infty}^{t} f_T(y;n)\,dy \tag{2.7}$$

Now, the limit is

$$\lim_{n\to\infty} P(T_n \le t) = \lim_{n\to\infty} \int_{-\infty}^{t} f_T(y;n)\,dy = \int_{-\infty}^{t} \lim_{n\to\infty} f_T(y;n)\,dy \tag{2.8}$$

It is a non-trivial thing to interchange the limit and the integral like that, since the integral is itself a limit as the lower limit goes to $-\infty$. It turns out that in this case the form of $f_T(y;n)$ is nice enough that it works; see Hogg for more details. Now we can talk about

$$\lim_{n\to\infty} f_T(y;n) = \left(\lim_{n\to\infty} \frac{\Gamma([n+1]/2)}{\sqrt{n\pi}\,\Gamma(n/2)}\right) \left(\lim_{n\to\infty} \left[1 + \frac{y^2}{n}\right]^{-\frac{n+1}{2}}\right) \tag{2.9}$$

Now, the first factor is just a number, so in principle we could work it out after the fact by requiring the limiting pdf to be normalized. You will show on the homework that it is $\sqrt{\frac{1}{2\pi}}$, using *Stirling's formula* which says

$$n! = \Gamma(n+1) \approx n^n e^{-n} \sqrt{2\pi n} \tag{2.10}$$

See for example section 4.2 of `http://ccrg.rit.edu/~whelan/courses/2004_1sp_A410/notes02.pdf` for a motivation of this.

The non-constant part is

$$\lim_{n\to\infty} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}} = \lim_{n\to\infty} \left(1 + \frac{y^2}{n}\right)^{-n/2} = \left(e^{y^2}\right)^{-1/2} = e^{-y^2/2} \tag{2.11}$$

so

$$\lim_{n\to\infty} f_T(y;n) = \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2} \tag{2.12}$$

and

$$\lim_{n\to\infty} P(T_n \le t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy = \Phi(t) \tag{2.13}$$

which is the cdf of a standard normal distribution, so

$$T_n \xrightarrow{D} N(0,1) \tag{2.14}$$

I.e, the sequence of t-distributed random variables has a limiting distribution which is standard normal.

2.1 Moment Generating Function Technique

We know that the moment generating function

$$M(t) = E(e^{tX}) \tag{2.15}$$

if it exists in a neighborhood $-h < t < h$ is useful not only for computing the moments of a distribution but also for identifying random variables that follow a certain distribution. The same is true for limits of sequences; if $M(t;n) = E(e^{tX_n})$ exists and $\lim_{n\to\infty} M(t;n) = M(t) = E(e^{tX})$, then $X_n \xrightarrow{D} X$. We'll use this to prove the central limit theorem, but first let's use it on a limit which is not a case of the CLT.

For example, suppose that Y_n is $b(n,\mu/n)$ for all integer $n > \mu$ where μ is some positive constant, i.e., consider a sequence of binomial random variables with the same mean. The mgf for $b(n,p)$ is $M(t) = (pe^t + (1-p))^n$ for all real t. If we take $p = \mu/n$ we get

$$M(t;n) = \left(\frac{\mu}{n}e^t + 1 - \frac{\mu}{n}\right)^n = \left(1 + \frac{\mu}{n}(e^t - 1)\right)^n \tag{2.16}$$

so the limit is

$$\lim_{n\to\infty} M(t;n) = \exp(\mu(e^t - 1)) = M(t) \tag{2.17}$$

which we see is the mgf of a Poisson distribution with mean μ, so a sequence of $b(n,\mu/n)$ random variables has a Poisson with mean μ as its limiting distribution.

Tuesday 8 December 2015 – Read Section 5.3 of Hogg

3 Central Limit Theorem

The central limit theorem states that if $\{X_i\}$ is a sample of size n drawn from a distribution with mean $E(X_i) = \mu$ and variance $\operatorname{Var}(X_i) = \sigma^2$, then

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1) \tag{3.1}$$

We can prove this using the moment generating function. I'll show this using the cumulant generating function

$$\psi(t) = \ln E(\exp[tX_i]) \tag{3.2}$$

for the distribution from which the sample is drawn. We know that $\psi(0) = 0$, $\psi'(0) = E(X_i) = \mu$, and $\psi''(0) = \operatorname{Var}(X_i) = \sigma^2$, so we can write the truncated McLaurin series for $\psi(t)$ as

$$\psi(t) = \psi(0) + t\psi'(0) + \frac{t^2}{2}\psi''(0) + o(t^2) = \mu t + \frac{\sigma^2 t^2}{2} + o(t^2) \tag{3.3}$$

where $o(t^2)$ is some expression such that $\lim_{t\to 0} \frac{o(t^2)}{t^2} \to 0$; we also sometimes write this as $\mathcal{O}(t^3)$. Hogg writes that more precisely as[2]

$$\psi(t) = \psi(0) + t\psi'(0) + \frac{t^2}{2}\psi''(\xi(t)) = \mu t + \frac{\sigma^2 t^2}{2} + \frac{\psi''(\xi(t)) - \sigma^2}{2} t^2 \tag{3.4}$$

[2] Actually, Hogg just calls this ξ, which leads to confusion when the argument of the mgf changes from t to $\frac{t}{\sigma\sqrt{n}}$.

where $\xi(t)$ is some number between 0 and t. Since

$$\lim_{t \to 0} \psi''(\xi(t)) = \psi''(0) = \sigma^2 \tag{3.5}$$

the last term is indeed $o(t^2)$. Now we consider the cumulant generating function

$$\begin{aligned}
\Psi(t;n) &= \ln E(\exp[tY_n]) = \ln E\left(\exp\left[\frac{t}{\sigma\sqrt{n}}\left\{\sum_{i=1}^{n}(X_i - \mu)\right\}\right]\right) \\
&= \ln \prod_{i=1}^{n} E\left(\exp\left[\frac{t}{\sigma\sqrt{n}}\{X_i - \mu\}\right]\right) \\
&= \sum_{i=1}^{n} \ln E\left(\exp\left[\frac{t}{\sigma\sqrt{n}}\{X_i - \mu\}\right]\right) \\
&= n\left[\psi\left(\frac{t}{\sigma\sqrt{n}}\right) - \frac{t}{\sigma\sqrt{n}}\mu\right]
\end{aligned} \tag{3.6}$$

If we use the Taylor expansion (3.4) we have

$$\Psi(t;n) = n\left[\frac{1}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 \psi''\left(\xi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)\right] = \frac{t^2}{2\sigma^2}\psi''\left(\xi\left(\frac{t}{\sigma\sqrt{n}}\right)\right) \tag{3.7}$$

When we take the limit $n \to \infty$, the argument of $\psi''()$ goes to zero, and thus

$$\lim_{n \to \infty} \Psi(t;n) = \frac{t^2}{2\sigma^2}\psi''(0) = \frac{t^2}{2} \tag{3.8}$$

which is the natural log of the mgf of a $N(0,1)$ random variable, so by the mgf method we've shown $Y_n \xrightarrow{D} N(0,1)$.

3.1 Applications

The Central Limit Theorem is a statement about the limit of a sequence of distributions, i.e.,

$$\lim_{n \to \infty} P(Y_n \le z) = P(Z \le z) = \Phi(z) \tag{3.9}$$

where $Y_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ and Z is a standard normal ($N(0,1)$) random variable. For practical applications, of course, we don't live in the limit $n \to \infty$, and it's usually used to say that when n is "large",

$$P(Y_n \le z) \approx \Phi(z) \qquad \text{for large } n \tag{3.10}$$

or equivalently

$$P(\overline{X}_n \le x) \approx \Phi\left(\frac{x - \mu}{\sigma/\sqrt{n}}\right) \qquad \text{for large } n \tag{3.11}$$

or

$$P\left(\sum_{i=1}^{n} X_i \le y\right) \approx \Phi\left(\frac{y - n\mu}{\sigma\sqrt{n}}\right) \qquad \text{for large } n \tag{3.12}$$

One important caveat is that we haven't said anything about how good the approximation is for large finite n, or even what "large" means. Such questions are beyond the scope of this course, but one piece of conventional wisdom to keep in mind is that convergence is slowest out on the tails, i.e., when the argument of the standard normal cdf $\Phi(z)$ is greater than 3 or so, or less than -3 or so.

We can use the central limit theorem to show that many distributions behave approximately like the normal distribution for certain ranges of their parameters. For instance, the following random variables can be written as sums of random samples:

- A $b(n,p)$ (binomial) random variable is the sum of n iid $b(1,p)$ (Bernoulli) random variables

- A Poisson random variable with mean $n\mu$ is the sum of n iid Poisson random variables each with a mean μ
- A $\Gamma(n\alpha, \beta)$ (Gamma) random variable is the sum of n iid $\Gamma(\alpha, \beta)$ random variables
- A $\chi^2(n)$ (chi square) random variable is the sum of n iid $\chi^2(1)$ random variables.

so in each case, for large enough n, we can approximate the random variable in question as normally-distributed with the same mean and variance as the original random variable.

We should be careful not to get too carried away, though, since such an argument could be misinterpreted to tell us that *any* Poisson random variable is apparently normally distributed. For example, we can write a Poisson random variable with mean 1.5 as the sum of a million iid Poisson random variables with mean 1.5×10^{-6}, which is certainly a large sample. The reason the central limit theorem doesn't let us approximate this sum as normally-distributed is that, in order to get a large sample, we had to make the Poisson parameter for each member of the sample very small, making that random variable nearly degenerate (each of these random variables has a probability of about 0.9999985 of being zero). In effect, we've tried to take the limit of the sum of n Poisson random variables each with mean μ/n, as $n \to \infty$. The central limit theorem doesn't apply because we're not holding the distributions of the random variables in the sample constant. The same thing happens if we take $b(n, p)$ with n large and p small; that is not approximately normal (we showed last time it was approximately Poisson).

We can, however, approximate the binomial distribution with the normal distribution as long as np and $n(1-p)$ are large enough. The usual guideline is for each of them to be greater than about 5. If these are large but not incredibly large, we have to worry a bit about approximating a discrete distribution ($b(n, p)$) with a continuous one ($N(np, np[1-p])$). For example, suppose we flip a fair coin 100 times and let X be the number of heads. We know that X is a $b(100, 0.5)$ random variable, with $E(X) = 100(0.5) = 50$ and $\operatorname{Var}(X) = 100(0.5)(1-0.5) = 25$, so its distribution should be approximately the same as that of Y, a $N(50, 25)$ random variable. But since X is discrete and can only take on integer values, we know

$$P(X \le 40) = P(X < 41) \tag{3.13}$$

while this is not true for the corresponding continuous random variable:

$$\begin{aligned} P(Y \le 40) &= \Phi\left(\frac{40-50}{5}\right) = \Phi(-2) \\ &\neq P(Y < 41) = \Phi\left(\frac{41-50}{5}\right) = \Phi(-1.8) \end{aligned} \tag{3.14}$$

The convention, to get a good approximation, is to take the value in the middle of the possible interval from 40 to 41, i.e.,

$$\begin{aligned} P(X \le 40) &= P(X < 41) = P(X < 40.5) \\ &\approx P(Y < 40.5) = \Phi\left(\frac{40.5-50}{5}\right) = \Phi(-1.9) \end{aligned} \tag{3.15}$$

This is known as the *continuity correction.*

Thursday 10 December 2015 – Review for Final Exam

The exam is comprehensive, but with relatively more emphasis on chapters four and five. Please come with questions and topics you'd like to go over.