5.2 The Law of Large Numbers (Convergence in Probability)

(2 Examples from last lecture)

Example Let $X_{(n)}$ denote the nth order statistic of a random sample $X_1, X_2, ..., X_n$ from a distribution having pdf

$$f(x) = \frac{1}{\theta}$$

for $0 < x < \theta$, $0 < \theta < \infty$.

Show that $X_{(n)}$ converges in probability to θ. Recall from last lect ure, we found that for $0 < x < \theta$,

$$F_n(x) = P[X_{(n)} \leq x] = \prod_{i=1}^{n} P[X_i \leq x]$$

$$= \prod_{i=1}^{n} \int_0^x \frac{1}{\theta} du = \left(\frac{x}{\theta}\right)^n$$

Now, let $0 < \epsilon < \theta$.

$$P[|X_{(n)} - \theta| > \epsilon] = P[X_{(n)} < \theta - \epsilon] = \left(\frac{\theta - \epsilon}{\theta}\right)^n$$

which approaches 0 as n approaches infinity.

Example: Let $X_{(n)}$ denote the nth order statistic of a sample of size n from a uniform distribution on the interval $(0, \theta)$. Show that $Z_n = \sqrt{X_{(n)}}$ converges in probability to $\sqrt{\theta}$.

We know from previous example, that $X_{(n)}$ converges in probability to θ. Also, we know that $g(x) = \sqrt{x}$ is a continuous function on the nonnegative real numbers. So, the fact that Z_n converges in probability to $\sqrt{\theta}$ follows from your Homework Problem.

Suppose we didn't have that general result of applying continuous transformat ion. How would we solve this problem directly?

In that case, we would need to show that for any $\epsilon > 0$,

$$\lim_{n\to\infty} P[|\sqrt{\theta} - \sqrt{X_{(n)}}| > \epsilon] = 0$$

However, notice that

$$P[|\sqrt{\theta} - \sqrt{X_{(n)}}| > \epsilon] = P[\sqrt{\theta} - \sqrt{X_{(n)}} > \epsilon]$$

$$= P[(\sqrt{\theta} - \sqrt{X_{(n)}})(\sqrt{\theta} + \sqrt{X_{(n)}}) > \epsilon(\sqrt{\theta} + \sqrt{X_{(n)}})]$$

$$\leq P[\theta - X_{(n)} > \epsilon\sqrt{\theta}]$$

which we know converges to 0 from previous Example.

5.3 Convergence in Distribution and the Central Limit Theorem

We have seen examples of random variables that are created by applying a function to the observations of a random sample. The distribution of such a **statistic** often depends on n, the size of the sample.

For example, let $\bar{X}$ denote the sample mean of a random sample $X_1, .X_2, ..., X_n$ from the distribution $N(\mu, \sigma^2)$.

We have seen that $\bar{X}$ is distributed $N(\mu, \sigma^2/n)$.

Thus, the distribution of $\bar{X}$ depends on n. In some cases we might wish to denote $\bar{X}$ by $\bar{X}_n$, to emphasize the dependence of the distribution on the size of the sample.

Example 1: Consider a random sample $X_1, X_2, ..., X_n$ where each X_i has distribution $F(x)$.

Let $X_{(n)} = \max\{X_1, X_2, ..., X_n\}$. $X_{(n)}$ is the nth order statistic in a sample of size n. We have seen that the distribution function for $X_{(n)}$ is

$$F_n(x) = [F(x)]^n$$

(In general) Let X_n denote a random variable whose distribution function $F_n(x)$ depends on n for $n = 1, 2, 3, ...$ In many cases, we are interested in knowing if $F_n(x)$ **converges** to some fixed distribution function $F(x)$.

Definition Let $X_1, X_2, \ldots$ be a sequence of random variables with cummulative distributin functions $F_1, F_2, \ldots,$ and let X be a random variable with distribution function F. We say that X_n converges in distribtuion to X if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

at every point x at which $F(x)$ is continuous.

First, let's recall the definition of continuity, and the meaning of a limit.

An infinite sequence of real numbers $a_1, a_2, a_3, \ldots$ has limit a (converges to a), if for any number $\epsilon > 0$ there is an integer n_0 such that for any $n > n_0$

$$|a - a_n| < \epsilon$$

Then we can say

$$\lim_{n \to \infty} a_n = a$$

For example, consider a sequence where $a_n = 1 + \frac{1}{n}$. Show that the limit of the sequence is the number 1.

Let $\epsilon > 0$ and select n_0 to be the smallest integer greater than $1/\epsilon$. Then for any $n > n_0$,

$$|1 - a_n| = \frac{1}{n} < \epsilon$$

Using this notion, we say that the function $F(y)$ is **continuous** at y if for any sequence $y_1, y_2, ..$ such that

$$\lim_{n \to \infty} y_n = y$$

we also have

$$\lim_{n \to \infty} F(y_n) = F(y).$$

Example 2 Let $X_{(n)}$ denote the nth order statistic of a random sample $X_1, X_2, ..., X_n$ from a distribution having pdf

$$f(x) = \frac{1}{\theta}$$

for $0 < x < \theta$, $0 < \theta < \infty$.

Clearly $F_n(x) = 0$ for $x < 0$ and $F_n(x) = 1$ for $\theta \leq x < \infty$, and for $0 \leq x < \theta$,

$$F_n(x) = P[X_{(n)} \leq x] = \prod_{i=1}^{n} P[X_i \leq x]$$

$$= \prod_{i=1}^{n} \int_0^x \frac{1}{\theta} du = \left(\frac{x}{n}\right)^n$$

We see that

$\lim_{n \to \infty} F_n(x) = 0$ for $-\infty < x < \theta$ and
$\lim_{n \to \infty} F_n(x) = 1$ for $\theta \leq x < \infty$

However, notice that $F(x) = 0$ for $-\infty < x < \theta$ and $F(x) = 1$ for $\theta \le x < \infty$ is a distribution function, and F_n converges to F.

In particular, F is a distribution function of a discrete type random variable that takes the value θ with probability equal to 1.

A distribution that places all of its mass on a single point is called a **degenerate distribution**.

Let's continue this example, but define a new random variable

$$Z_n = n(\theta - X_{(n)}).$$

Because $X_{(n)} < \theta$ with probability equal to 1, we can say that for $z < 0$, $F_n(z) = 0$ for all n, where F_n is the distribution function of Z_n.

Also, because $X_{(n)} > 0$ with probability 1, we can say that $F_n(z) = 1$ for $z \ge n\theta$.

Now let z take a value in $[0, n\theta)$.

$$F_n(z) = P[Z_n \leq z] = P[n(\theta - X_{(n)}) \leq z]$$

$$= P[X_{(n)} > \theta - \frac{z}{n}] = 1 - P[X_{(n)} \leq \theta - \frac{z}{n}]$$

$$= 1 - \left[\frac{\theta - z/n}{\theta}\right]^n = 1 - (1 - \frac{z}{n\theta})^n$$

To investigate the limit of $F_n(z)$, notice that for any z, $z < n\theta$ when n is taken sufficiently large. Also, we use the fact that for a real number x,

$$\lim_{n\to\infty}(1 + \frac{x}{n})^n = e^x.$$

Thus,

$lim_{n\to\infty}F_n(z) = 0$ for $z < 0$ and
$lim_{n\to\infty}F_n(z) = 1 - e^{-z/\theta}$ for $0 \leq z < \infty$.

Notice that this is the distribution function of a gamma-distribution with $\alpha = 1$ and $\beta = \theta$. More specicially, it is the distribution function of an exponential distribution. In this case F_n converges in distribution to a nondegenerate distribution.

We have focused on distribution functions rather than probability density functions for this notion of convergence in distributions. As it turns out, convergence in distribution may hold when the pdf does not converge to any fixed pdf.

Example 3: Consider a sequence of random variables
$X_1, X_2, X_3, \ldots$, for which the pdf of X_n is given by

$f_n(x) = 1$ for $x = 2 + \frac{1}{n}$ and equals 0 elsewhere.

Then we have for $-\infty < x < \infty$,

$$\lim_{n\to\infty} f_n(x) = 0.$$

We see that f_n converges to the constant function $f(x) = 0$ which is not a pdf.

However,

$F_n(x) = 0$ for $x < 2 + \frac{1}{n}$

and $F_n(x) = 1$ for $x \geq 2 + \frac{1}{n}$.

This converges to the distribution function

$F(x) = 0$ for $x < 2$

$F(x) = 1$ for $x \geq 2$

at all the continuity points of $F(x)$ (all points other than x=2). $F(x)$ is the cdf of a random variable that equals 2 with probability 1.

Example 4: Let $X_1, X_2, ..., X_n$ be a random sample from the distribution $N(\mu, \sigma^2)$. Show that

$$Z_n = \sum_{i=1}^{n} X_i$$

does not have a limiting distribution.

First, note that Z_n has distribution $N(n\mu, n\sigma^2)$.

$$F_n(z) = P[Z_n \leq z]$$

$$= P[\frac{Z_n - n\mu}{\sigma\sqrt{n}} \leq \frac{z - n\mu}{\sigma\sqrt{n}}] = \Phi(\frac{z - n\mu}{\sigma\sqrt{n}})$$

where Φ is the distribution function of a $N(0, 1)$ distributed random variable. Suppose the $\mu < 0$. Then for $-\infty < z < \infty$

$$\lim_{n\to\infty} \Phi(\frac{z - n\mu}{\sigma\sqrt{n}}) = 1$$

If $\mu = 0$,

$$\lim_{n\to\infty} \Phi(\frac{z - n\mu}{\sigma\sqrt{n}}) = 1/2.$$

If $\mu > 0$,

$$\lim_{n\to\infty} \Phi(\frac{z - n\mu}{\sigma\sqrt{n}}) = 0.$$

In all three cases, F_n converges to a constant value, and does not converge to a distribution function.

Theorem: Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables such that X_n converges in distribution to a random variable X, and Y_n converges in probability to a random variable c. Then, it follows that

a. $X_n + Y_n$ converges in distribution to $X + c$

b. $Y_n X_n$ converges in distribution to cX

c. X_n/Y_n converges in distribution to X/c

Theorem A Continuity Theorem: Let F_n be a sequence of cdf's with corresponding mgf's M_n. Let F be a cdf with the mgf M. If $M_n(t) \to M(t)$ for all t in an open interval containing zero, then $F_n(x) \to F(x)$ for all continuity points of F.

We may use the notation: $lim_{n\to\infty} M(t;n) = M(t)$ for $M_n(t) \to M(t)$.

In applying this theorem, we make much use of the form

$$\lim_{n\to\infty}\left[1+\frac{b}{n}+\frac{\psi(n)}{n}\right]^{cn}$$

where b and c do not depend on n and

$$\lim_{n\to\infty}\psi(n) = 0.$$

An important result in analysis is that under these assumptions,

$$\lim_{n\to\infty}\left[1+\frac{b}{n}+\frac{\psi(n)}{n}\right]^{cn} = \lim_{n\to\infty}\left(1+\frac{b}{n}\right)^{cn} = e^{bc}$$

Example 5: Let Y_n have a distribution that is $b(n, p_n)$ where $p_n = \mu/n$ for some $\mu > 0$. The mgf of Y_n is

$$M(t; n) = E(e^{tY_n}) = [(1 - p_n) + p_n e^t]^n$$

$$= [(1 - \mu/n) + \mu e^t/n]^n = [1 + \frac{\mu(e^t - 1)}{n}]^n$$

Using the preceding result we have

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

for all t. However, this is the moment generating function of a Poisson distribution with mean μ.

This example indicates that we may use a Poisson distribution to approximate probabilities of events concerning binomial random variables when p is small and n is very large.

For example suppose that Y has distribution $b(50, .04)$.

$$P[Y \leq 1] = \left(\frac{24}{25}\right)^{50} + 50\left(\frac{1}{25}\right)\left(\frac{24}{25}\right)^{49} = 0.400.$$

However, we can say that Y has approximately the same distribution of a Poisson random variable W that has mean $\mu = np = 50(.04) = 2$. For example,

$$P[W \leq 1] = e^{-2} + 2e^{-2} = 0.406.$$

We can use the computer to compare the cumulative distribution functions of Y and W.

```
y<-c(0:7)
Fy<-pbinom(y,50,.04)
Fw<-ppois(y,2)

cbind(y,Fy,Fw)
```

```
     y        Fy        Fw
[1,] 0 0.1298858 0.1353353
[2,] 1 0.4004812 0.4060058
[3,] 2 0.6767140 0.6766764
[4,] 3 0.8608692 0.8571235
[5,] 4 0.9510285 0.9473470
[6,] 5 0.9855896 0.9834364
[7,] 6 0.9963899 0.9954662
[8,] 7 0.9992186 0.9989033
```

Example 6: Let Z_n be $\chi^2(n)$. Then the mgf of Z_n is $(1-2t)^{-n/2}$ for $t < 1/2$. The mean and variance of Z_n are n and $2n$, respectively.

Investigate the limiting distribution of the random variable

$$Y_n = (Z_n - n)/\sqrt{2n}$$

The mgf of Y_n is

$$M(t;n) = E\left(\exp\left[t\left(\frac{Z_n - n}{\sqrt{2n}}\right)\right]\right)$$

$$= e^{-tn/\sqrt{2n}} E(e^{tZ_n/\sqrt{2n}})$$

$$= e^{-tn/\sqrt{2n}} (1 - 2\frac{t}{\sqrt{2n}})^{-n/2}$$

This holds for for $t < \sqrt{n}2$. If we write $e^{-tn/\sqrt{2n}}$ as

$$e^{-tn/\sqrt{2n}} = (e^{t\sqrt{2/n}})^{-n/2}$$

we see that

$$M(t;n) = \left(e^{t\sqrt{2/n}} - t\frac{\sqrt{2}}{\sqrt{n}}e^{t\sqrt{2/n}}\right)^{-n/2}$$

To simplify this, we apply Taylor's formula and note that

$$e^{t\sqrt{2/n}} = 1 + t\sqrt{2/n} + \frac{1}{2}(t\sqrt{2/n})^2 + \frac{e^{\lambda(n)}}{6}(t\sqrt{2/n})^3$$

where $\lambda(n)$ is some number between 0 and $t\sqrt{2/n}$.

If we substitute this expression for $e^{t\sqrt{2/n}}$ in the equation

$$M(t;n) = \left(e^{t\sqrt{2/n}} - t\frac{\sqrt{2}}{\sqrt{n}}e^{t\sqrt{2/n}}\right)^{-n/2}$$

we see that

$$M(t;n) = \left(1 - \frac{t^2}{n} + \frac{\psi(n)}{n}\right)^{-n/2}$$

where

$$\psi(n) = \frac{\sqrt{2}t^3e^{\lambda(n)}}{3\sqrt{n}} - \frac{\sqrt{2}t^3}{\sqrt{n}} - \frac{2t^4e^{\lambda(n)}}{3n}$$

Since $\lambda(n) \rightarrow 0$ and $e^{\lambda(n)} \rightarrow 1$ as $n \rightarrow \infty$, we see that $\lim_{n\rightarrow\infty} \psi(n) = 0$. This implies that

$$\lim_{n\rightarrow\infty} M(t;n) = e^{t^2/2}$$

Recall that the mgf for a $N(\mu, \sigma^2)$ distribution is

$$M(t) = \exp[\mu t + \frac{\sigma^2 t^2}{2}]$$

If we choose $\mu = 0$ and $\sigma^2 = 1$, we see that $M(t; n)$ converges to the mgf of a standard normal distribution for all real numbers t.

Thus, Y_n converges in distribution to a standard normal random variable.

The preceding example is really a special case of the **central limit theorem**. In the next lecture, we will state and prove the central limit theorem. However, we'll see an empirical illustration of it here.

Example 7: Consider a random sample $X_1, X_2, ..., X_n$ from a distribution with pdf

$$f(x) = \frac{3}{2}x^2$$

for $-1 < x < 1$.

From this we find

$$F(x) = \int_{-1}^{x} \frac{3}{2} t^2 dt = \frac{1}{2} + \frac{x^3}{2}$$

for $-1 < x < 1$.

$$E[X] = \int_{-1}^{1} \frac{3x}{2} x^2 dx = 0$$

$$Var(X) = \int_{-1}^{1} \frac{3}{2} x^4 dx = \frac{3}{5}$$

Now consider the distribution of

$$Z_n = \frac{\sqrt{n}\bar{X}_n}{\sqrt{3/5}}$$

where

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

for $n = 1, 5, 25$.

In this case, we investigate these distributions by simulating enormous samples for each studied value of n. By inspecting the resulting histograms, we can approximate the distributions.

To simulate observations with distribution function $F(x)$, we find F^{-1} and apply this to a uniformly distributed random variable.

$$u = F(x) = \frac{1}{2} + \frac{x^3}{2}$$

for $-1 < x < 1$. Thus,

$$x = F^{-1}(u) = (2u - 1)^{1/3}$$

for $0 < u < 1$.

Now use a computer to simulate 10,000 observations of Z_n for each value of n, and plot the histograms.

```
## Case n=1
m<-10000
n<-1
u<-runif(m*n)
## Since S-plus seems to dislike cube
## roots of negative numbers, I need
## to trick it
x<-abs(2*u-1)^(1/3)
x<-x*((2*u-1)>0)-x*((2*u-1)<0)
z<-x/sqrt(3/5)
postscript("n1hist.ps")
hist(z,nclass=50)
dev.off()
```

```
## Case n=5
m<-10000
n<-5
### construct m datasets
u<-runif(m*n)
x<-abs(2*u-1)^(1/3)
x<-x*((2*u-1)>0)-x*((2*u-1)<0)
x<-matrix(x,m,n)

## find sample means and z
z<-apply(x,1,mean)
z<-z/sqrt((3/5)/n)
postscript("n5hist.ps")
hist(z,nclass=50)
dev.off()
```

```
## Case n=25
m<-10000
n<-25
### construct m datasets
u<-runif(m*n)
x<-abs(2*u-1)^(1/3)
x<-x*((2*u-1)>0)-x*((2*u-1)<0)
x<-matrix(x,m,n)

## find sample means and z
z<-apply(x,1,mean)
z<-z/sqrt((3/5)/n)
postscript("n25hist.ps")
hist(z,nclass=50)
dev.off()
```

Figure 1: Histogram of Z_n for $n = 1$

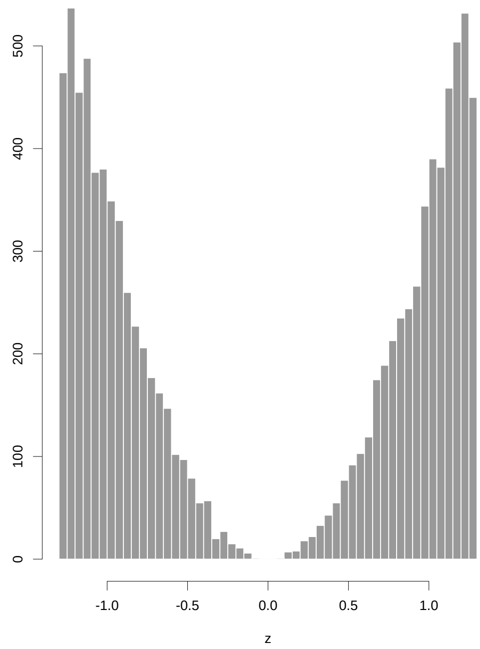

Figure 2: Histogram of Z_n for $n = 5$

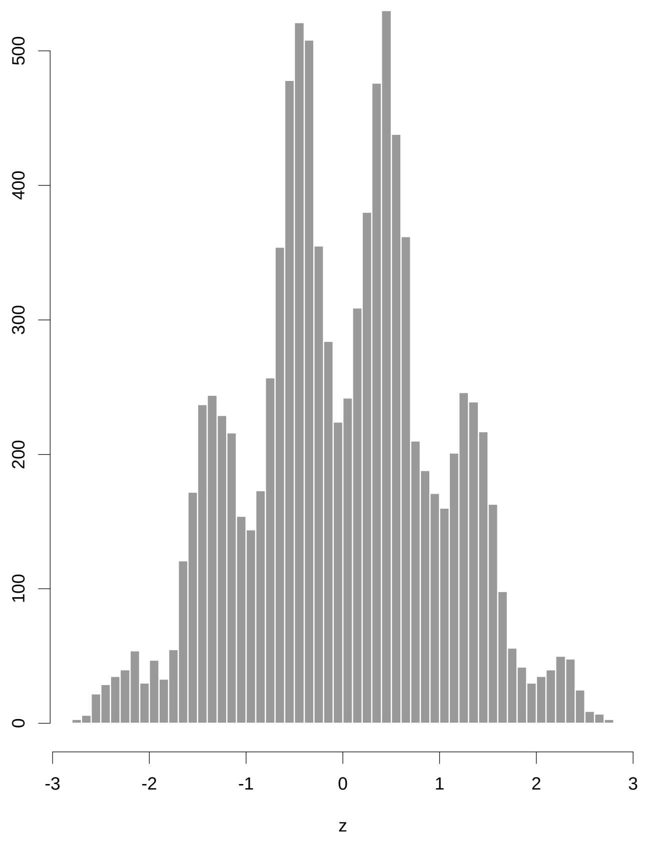

Figure 3: Histogram of Z_n for $n = 25$