5.3 Convergence in Distribution and the Central Limit Theorem

Central Limit Theorem Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and positive variance σ^2. Then the random variable

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \sqrt{n}(\bar{X}_n - \mu)/\sigma$$

has a limiting distribution that is normal with mean 0 and variance 1.

Proof: We assume the existence of the mgf

$$M(t) = E[e^{tX}]$$

for $-h < t < h$. Let

$$m(t) = E[e^{t(X-\mu)}] = e^{-\mu t}M(t)$$

be the mgf of the random variable $X - \mu$, which also exists for $-h < t < h$. Since $m(t)$ is the mgf of $X - \mu$ it is clear that

$m(0) = 1$

$m'(0) = E[X - \mu] = 0$, and

$m''(0) = E[(X - \mu)^2] = \sigma^2$.

By Taylor's formula there exists a number λ between 0 and t such that

$$m(t) = m(0) + m'(0)t + \frac{m''(\lambda)t^2}{2}$$

$$= 1 + \frac{m''(\lambda)t^2}{2}$$

If we add and subtract $\frac{\sigma^2 t^2}{2}$,

$$m(t) = 1 + \frac{\sigma^2 t^2}{2} + \frac{[m''(\lambda) - \sigma^2]t^2}{2}$$

Now consider the mgf of Y_n,

$$
\begin{aligned}
M(t;n) &= E\left[exp\left(t\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}\right)\right] \\
&= \prod_{i=1}^{n} E\left[exp\left(t\frac{X_i - \mu}{\sigma\sqrt{n}}\right)\right] \\
&= \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n
\end{aligned}
$$

for $-h < \frac{t}{\sigma\sqrt{n}} < h$.

In the expression

$$m(t) = 1 + \frac{\sigma^2 t^2}{2} + \frac{[m''(\lambda) - \sigma^2]t^2}{2}$$

replace t with $\frac{t}{\sigma\sqrt{n}}$ to obtain

$$m\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + \frac{[m''(\lambda) - \sigma^2]t^2}{2n\sigma^2}$$

for some λ between 0 and $\frac{t}{\sigma\sqrt{n}}$. Then we can write the moment generating function of Y_n by

$$M(t; n) = \left[1 + \frac{t^2}{2n} + \frac{[m''(\lambda) - \sigma^2]t^2}{2n\sigma^2}\right]^n$$

Note that λ converges to 0 as n approaches infinity. By the continuity of m'' we know that this implies

$$\lim_{n\to\infty}[m''(\lambda) - \sigma^2] = 0$$

Thus, we see that

$$\lim_{n\to\infty} M(t;n) = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + \frac{[m''(\lambda) - \sigma^2]t^2}{2n\sigma^2}\right]^n$$
$$= \lim_{n\to\infty}\left[1 + \frac{t^2}{2n}\right]^n = e^{t^2/2}$$

for all real t, which is the moment generating function of a standard normal random variable. Thus, Y_n converges in distribution to a standard normal random variable.

Example 1: Approximate the probability that the mean of a random sample of size 15 from a distribution with pdf $f(x) = 3x^2$, $0 < x < 1$, is between $\frac{3}{5}$ and $\frac{4}{5}$.

$$E[X] = \int_0^1 x(3x^2) = \frac{3}{4}$$

$$E[X^2] = \int_0^1 x^2(3x^2) = \frac{3}{5}$$

$$Var(X) = \frac{3}{5} - \frac{3}{4} \times \frac{3}{4} = \frac{3}{80}$$

From these results, we see that the expected value of $\bar{X}$ is $\frac{3}{4}$ and the standard deviation of $\bar{X}$ is $\sqrt{3/80}/\sqrt{15} = 1/20$.

$$P[\frac{3}{5} < \bar{X} < \frac{4}{5}] =$$

$$P[\frac{3/5 - 3/4}{1/20} < \frac{\bar{X} - 3/4}{1/20} < \frac{4/5 - 3/4}{1/20}]$$

which is approximately,

$$P[-3 < Z < 1] = \Phi(1) - \Phi(-3) = 0.840$$

where Z is a standard normal random variable.

In this example we used the CLT, even though it is not obvious that 15 is sufficiently large for a good approximation. Let's see if that matches a Monte Carlo approximation of the probability of interest.

```
##### Simulate x-bar 10,000 times
##### and save the values

xbar<-c(1:10000)
for( i in 1:10000){
print(i)
sample<-runif(15)
sample<-sample^(1/3)
sampmean<-mean(sample)
xbar[i]<-sampmean
}
```

```
#### Check what proportion of
#### the sample means fall in
### the interval (3/5,4/5)

upper<-xbar<(4/5)
lower<-xbar > (3/5)
proportion<-mean(upper*lower)
[1] 0.8484
```

We can see the the normal approximation and the Monte Carlo approximation nearly agree.

4.6 Approximation Methods

Delta Method: Consider a smooth function $g(x)$, and let $\bar{X}_n$ denote the sample mean of a random sample of size n from a distribution with mean n and variance σ^2.

Since, $\bar{X}_n$ converges in probability to μ, we can use Taylor's formula for the approximation,

$$g(\bar{X}_n) \approx g(\mu) + (\bar{X}_n - \mu)g'(\mu)$$

when $g'(\mu)$ exists and is not 0. From this we can see that

$$E[g(\bar{X}_n)] \approx g(\mu)$$

$$Var[g(\bar{X}_n)] \approx \frac{\sigma^2[g'(\mu)]^2}{n}$$

In fact, the random variable

$$Y_n = \frac{g(\bar{X}_n) - g(\mu)}{\sqrt{[g'(\mu)]^2\sigma^2/n}}$$

converges in distribution to a standard normal random variable.

Example 2: Let $X_1, X_2, ..., X_n$ be a random sample from a Poisson distribution with mean μ. Find the limiting distribution of $\sqrt{\bar{X}}$.

Clearly $\bar{X}$ converges in probability to μ, so that $\sqrt{\bar{X}}$ converges in probability to $\sqrt{\mu}$.

By the delta method, we know that

$$E[\sqrt{\bar{X}}] \approx \sqrt{\mu}$$

and

$$Var(\sqrt{\bar{X}}) \approx [1/(2\sqrt{\mu})]^2 \frac{Var(X)}{n} = \frac{1}{4\mu} \times \frac{\mu}{n}$$

$$= \frac{1}{4n}$$

Furthermore,

$$\frac{\sqrt{\bar{X}} - \sqrt{\mu}}{\sqrt{1/(4n)}}$$

is approximately standard normal.