Notes on the Central Limit Theorem

Cyrus Samii

The Central Limit Theorem is the cornerstone of a lot of the asymptotic theory that we rely on for quantitative inference. These notes sketch out the proof of a simple version of the Central Limit Theorem that relies on the existence of a moment generating function. I follow the presentation in John A. Rice, *Mathematical Statistics and Data Analysis* (Duxbury Press, 1995). I first introduce some important concepts and results, and then proceed with the proof. This is not the most general statement of the Central Limit Theorem—more general versions exist that rely, e.g. only on the assumption of finite first and second moments. Nonetheless, this version applies to many cases of interest.

1 Moment generating functions

First, you need to understand the idea of a moment generating function (mgf). Suppose a random variable $X \sim f$, where f defines a pdf (or pmf for discrete variables). Define the moment generating function of X as $M(t,X) = \mathrm{E}(\exp(tX))$, which exists only for certain f (though for most f that we would be concerned with, e.g. Normal, Poisson, etc.; an exception would be the Cauchy distribution). For $X \sim f$, if $M(t,X)$ exists on an open interval of t containing 0, it uniquely defines f. You can take this on faith—the proof is rather advanced. Trivially, $M(0,X) = 1$. The rth "moment" of X is given by $\mathrm{E}(X^r)$, and the rth "central moment" of X is given by $\mathrm{E}(x - \mathrm{E}(X))^r$. As the name suggests, the mgf yields moments as follows:

$$
\begin{array}{lll}
 & M(t,X) & = \int_{-\infty}^{\infty} \exp(tx) f_X(x) dx \\
\Rightarrow & \frac{d}{dt} M(t,X) & = \int_{-\infty}^{\infty} x \exp(tx) f_X(x) dx \\
\Rightarrow & \frac{d}{dt} M(t,X)|_{t=0} & = \int_{-\infty}^{\infty} x f_X(x) dx = \mathrm{E}(X).
\end{array}
$$

More generally, if the mgf of X exists on an open interval of t containing 0, $\frac{d^r}{dt^r} M(t,X)|_{t=0} = \mathrm{E}(X^r)$, the rth moment of X. Given the mgf, this makes finding the moments of a random variable quite easy. Usually, finding moments requires working with integrals (due to the $\mathrm{E}()$'s), which is hard. Deriving moments from the mgf is done with derivatives, which are easy.

Linear transformations of random variables have straightforward mgf's. Suppose $Y = a + bX$. Then,

$$M(t,Y) = \mathrm{E}(\exp(at + btX)) = \exp(at)\mathrm{E}(\exp(btX)) = \exp(at)M(bt,X) \tag{1}$$

The sum of independent random variables yields a particular mgf. Suppose X and Y are independent random variables and $Z = X + Y$. Then,

$$
\begin{aligned}
M(t,Z) = \mathrm{E}(\exp(tZ)) = \mathrm{E}(\exp(tX)\exp(tY)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(tx)\exp(ty) f_{XY}(x,y) dx dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(tx)\exp(ty) f_X(x) f_Y(y) dx dy \text{ (by independence)} \\
&= \int_{-\infty}^{\infty} \exp(tx) f_X(x) dx \int_{-\infty}^{\infty} \exp(ty) f_Y(y) dy = M(t,X)M(t,Y).
\end{aligned} \tag{2}
$$

By induction, this result can be generalized to the sum of an arbitrary number of independent random variables.

2 Mgf for Standard Normal random variables

Suppose $X \sim \phi$, where ϕ is the standard normal (i.e., N(0,1)) pdf. Then,

$$M(t,X) = \mathrm{E}\left(\exp(tX)\right) = \int_{-\infty}^{\infty} \exp(tx)\phi(x)dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(tx)\exp(-\frac{x^2}{2})dx$$
$$= \text{...solve in Wolfram Alpha ☺...} = \exp(\frac{t^2}{2}). \tag{3}$$

We can check the properties of this mgf:

$$\begin{aligned} \tfrac{d}{dt}M(t,X)|_{t=0} &= t\exp(\tfrac{t^2}{2})|_{t=0} = 0 & \checkmark \\ \tfrac{d^2}{dt^2}M(t,X)|_{t=0} &= t^2\exp(\tfrac{t^2}{2}) + \exp(\tfrac{t^2}{2})|_{t=0} = 1 & \checkmark \end{aligned}$$

3 Central Limit Theorem for a Sum of Scalars under Existence of Mgf

The theorem says that the limiting distribution of a standardized sum of independent random variables is the standard normal distribution.

Theorem 1. *Suppose $\{X_i\}_{i=1}^n$ a sequence of independent random variables with mean 0, variance σ^2, $X_i \sim f$, and $M(t,X_i) = M(t,X)$ defined in the neighborhood of 0 for all i. Let $S_n = \sum_{i=1}^n X_i$. Then,*

$$\lim_{n\to\infty} \Pr\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) = \Phi(x), -\infty < x < \infty,$$

where $\Phi(x)$ is the standard normal CDF.

Proof. Let $Z_n = \frac{S_n}{\sigma\sqrt{n}}$. By results (1) and (2) above, $M(t,Z_n) = \left[M(\frac{t}{\sigma\sqrt{n}},X)\right]^n$. Note that (i) a Taylor series expansion of $M(s,X)$ around zero yields,

$$M(s,X) = M(0,X) + s\frac{d}{ds}M(s,X)|_{s=0} + \frac{1}{2}s^2\frac{d^2}{ds^2}M(s,X)|_{s=0} + \varepsilon_s,$$

where the approximation error vanishes relative to s^2 as s is drawn closer to zero(i.e. $\frac{\varepsilon_s}{s^2} \to 0$ as $s \to 0$). Also, note that (ii) $M(0,X) = 1$ trivially, $\frac{d}{ds}M(s,X)|_{s=0} = \mathrm{E}\left(X_i\right) = 0$, and $\frac{d^2}{ds^2}M(s,X)|_{s=0} = \mathrm{Var}\left(X_i\right) = \sigma^2$. Putting (ii) and (iii) together, we have,

$$M\left(\frac{t}{\sigma\sqrt{n}},X\right) = 1 + \frac{1}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2\sigma^2 + \varepsilon_n = 1 + \frac{t^2}{2n} + \varepsilon_n \Rightarrow M(t,Z_n) = \left(1 + \frac{t^2}{2n} + \varepsilon_n\right)^n,$$

where the approximation error vanishes relative to $\frac{t^2}{n\sigma^2}$ as $\frac{t}{\sigma\sqrt{n}}$ is drawn closer to zero, which implies that the approximation error vanishes as n becomes large (i.e. $\frac{\varepsilon_n}{t^2/n\sigma^2} \to 0$ as $n \to \infty$). By the properties of the exponential function, we know that[1]

$$\lim_{n\to\infty}\left(1 + \frac{k}{n}\right)^n = \exp(k).$$

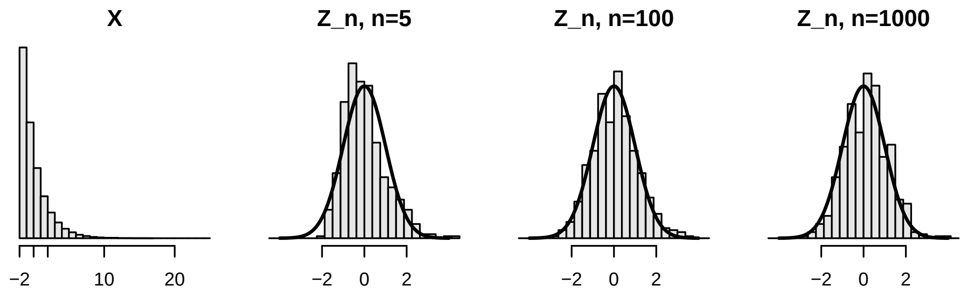
Putting this all together, we get

$$\lim_{n\to\infty} M(t,Z_n) = \exp(\frac{t^2}{2}),$$

which by (3) is the mgf of the standard normal distribution. □

[1]See, e.g. http://mathworld.wolfram.com/ExponentialFunction.html

Figure 1: **Illustration of CLT on Gamma Distributed X**



The leftmost histogram shows the distribution of the underlying X. The next three histograms show the distribution of 500 standardized sums for sample sizes (n) of 5, 100, and 1000. A standard normal distribution is overlaid. It is evident that the distribution of the standardized sums approaches the standard normal distribution as n increases.

4 Illustration

As an ilustration, I sample from a population of X's where $X = \tilde{X} - \mu_{\tilde{X}}$, $\tilde{X} \sim$ Gamma(1,2), and by the properties of the Gamma distribution, $\mu_{\tilde{X}} = 2$. This distribution of the X is highly skewed and truncated. Nonetheless, the distribution of standardized sums approaches the standard normal distribution. This is shown in Figure 1.

```
> rm(list = ls())
> set.seed(1234)
> n.pop <- 1e+05
> index <- seq(1:n.pop)
> k <- 1
> theta <- 2
> mu.x.sc <- k * theta
> x.sc <- rgamma(n.pop, shape = k, scale = theta)
> x <- x.sc - mu.x.sc
> sigma.x <- sqrt(k * theta^2)
> par(mfrow = c(1, 4))
> par(pty = "s")
> par(mar = c(1, 0, 1, 0) + 1)
> hist(x, freq = F, main = "X", axes = F, xlab = "", ylab = "",
+     breaks = 25, col = gray(0.9))
> axis(1, at = c(-2, 0, 2, 10, 20), labels = c(-2, 0, 2, 10, 20))
> n.samp <- c(5, 100, 1000)
> z <- as.list(rep(NA, length(n.samp)))
> n.sim <- 500
> for (j in 1:length(n.samp)) {
+     for (i in 1:n.sim) {
```

```
+         x.samp <- x[sample(index, n.samp[j])]
+         z[[j]][i] <- sum(x.samp)/(sigma.x * sqrt(n.samp[j]))
+     }
+ }
> for (j in 1:length(n.samp)) {
+     xlim <- max(max(unlist(z)), abs(min(unlist(z))))
+     hist(z[[j]], freq = F, main = paste("Z_n, n=", n.samp[j],
+         sep = ""), xlim = c(-xlim, xlim), ylim = c(0, 0.5), axes = F,
+         xlab = "", ylab = "", col = gray(0.9), breaks = seq(from = -xlim,
+             to = xlim, length = 25))
+     axis(1, at = c(-2, 0, 2), labels = c(-2, 0, 2))
+     curve(dnorm(x), -4, 4, add = T, lwd = 2)
+ }
```

5 Application to Sample Mean

The CLT above is applied to a sample sum. A corollary is that all sample statistics that are linear transformations of the sample sum will exhibit a limiting normal distribution. Consider the sample mean. In that case, we substitute $\frac{X_i}{n}$ for the X_i in the theorem, in which case S_n is the sample mean, $\bar{X}$; we substitute $\frac{\sigma^2}{n^2}$ for the variance. Then, the appropriate Z_n becomes $\sqrt{n}\frac{\bar{X}}{\sigma}$.