Chapter 5. Multiple Random Variables

5.11: Proof of the CLT

Slides (Google Drive) Alex Tsun Video (YouTube)

In this optional section, we'll prove the Central Limit Theorem, one of the most fundamental and amazing results in all of statistics, using MGFs!

5.11.1 Properties of Moment Generating Functions

Let's first recall the properties of MGFs (this is just copied from 5.6):

Theorem 5.11.1: Properties and Uniqueness of Moment Generating Functions

For a function $f : \mathbb{R} \to \mathbb{R}$, we will denote $f^{(n)}(x)$ to be the n^{th} derivative of $f(x)$. Let X, Y be independent random variables, and $a, b \in \mathbb{R}$ be scalars. Then MGFs satisfy the following properties:

1. $M_X'(0) = \mathbb{E}[X]$, $M_X''(0) = \mathbb{E}\left[X^2\right]$, and in general $M_X^{(n)} = \mathbb{E}[X^n]$. This is why we call M_X a *moment generating* function, as we can use it to generate the moments of X.
2. $M_{aX+b}(t) = e^{tb} M_X(at)$.
3. If $X \perp Y$, then $M_{X+Y}(t) = M_X(t) M_Y(t)$.
4. (**Uniqueness**) The following are equivalent:
   (a) X and Y have the same distribution.
   (b) $f_X(z) = f_Y(z)$ for all $z \in \mathbb{R}$.
   (c) $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$.
   (d) There is an $\varepsilon > 0$ such that $M_X(t) = M_Y(t)$ for all $t \in (-\varepsilon, \varepsilon)$ (they match on a small interval around $t = 0$).

   That is M_X uniquely identifies a distribution, just like PDFs or CDFs do.

5.11.2 Proof of the Central Limit Theorem (CLT)

Here is a restatement of the CLT from 5.7 that we will prove:

Theorem 5.11.2: The Central Limit Theorem (CLT)

Let $X_1, \dots X_n$ be a sequence of independent and identically distributed random variables with mean μ and (finite) variance σ^2. Then, the standardized sample mean approaches the standard Normal distribution:

$$\text{As } n \to \infty, \quad \overline{Z}_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$$

Proof of The Central Limit Theorem. Our strategy will be to compute the MGF of $\overline{Z}_n$ and exploit properties of the MGF (especially uniqueness) to show that it must have a standard Normal distribution!

Suppose $\mu = 0$ (without loss of generality), so:

$$\mathbb{E}\left[X_i^2\right] = \mathsf{Var}\left(X_i\right) + \mathbb{E}\left[X_i\right]^2 = \sigma^2$$

Now, let:

$$\overline{Z}_n = \frac{\overline{X}_n}{\sigma/\sqrt{n}} = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} X_i$$

Note there is no typo above: the $\frac{1}{n}$ from $\overline{X}_n$ changes the division by $\sqrt{n}$ to a multiplication.

We will show $M_{\overline{Z}_n}(t) \to e^{t^2/2}$ (the standard normal MGF) and hence, $\overline{Z}_n \to \mathcal{N}(0,1)$ by uniqueness of the MGF.

1. First, for an ***arbitrary*** random variable Y, since the MGF exists in $(-\varepsilon, \varepsilon)$ under "most" conditions, we can use the 2^{nd} order Taylor series expansion around 0 (quadratic approximation to a function):

$$\begin{aligned}
M_Y(s) &\approx M_Y(0)\cdot\frac{s^0}{0!} + M_Y'(0)\cdot\frac{s^1}{1!} + M_Y''(0)\cdot\frac{s^2}{2!} \\
&= \mathbb{E}\left[Y^0\right] + \mathbb{E}\left[Y\right]s + \mathbb{E}\left[Y^2\right]\frac{s^2}{2} && \left[\text{Since } M_Y^{(n)}(0) = \mathbb{E}\left[Y^n\right]\right] \\
&= 1 + \mathbb{E}\left[Y\right]s + \mathbb{E}\left[Y^2\right]\frac{s^2}{2} && \left[\text{Since } Y^0 = 1\right]
\end{aligned}$$

2. Now, let M_X denote the common MGF of all the X_i's (since they are iid).

$$\begin{aligned}
M_{\overline{Z}_n}(t) &= M_{\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i}(t) && \left[\text{Definition of } \overline{Z}_n\right] \\
&= M_{\sum_{i=1}^n X_i}\left(\frac{t}{\sigma\sqrt{n}}\right) && \left[\text{By Property 2 of MGFs above, where } a = \frac{1}{\sigma\sqrt{n}}, b = 0\right] \\
&= \left[M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n && \left[\text{By Property 3 of MGFs above}\right]
\end{aligned}$$

3. Recall Step 1, and now let $Y = X$ and $s = \frac{t}{\sigma\sqrt{n}}$ so we get a Taylor approximation of M_X. Then:

$$\begin{aligned} M_X\left(\frac{t}{\sigma\sqrt{n}}\right) &\approx 1 + \mathbb{E}[X]\frac{t}{\sigma\sqrt{n}} + \mathbb{E}\left[X^2\right]\frac{\left(\frac{t}{\sigma\sqrt{n}}\right)^2}{2} && \text{[Step 1]} \\ &= 1 + 0 + \sigma^2\frac{t^2}{2\sigma^2 n} && \text{[Since } \mathbb{E}[X] = 0 \text{ and } \mathbb{E}\left[X^2\right] = \sigma^2\text{]} \\ &= 1 + \frac{t^2/2}{n} \end{aligned}$$

4. Now we combine Steps 2 and 3:

$$\begin{aligned} M_{\overline{Z}_n}(t) &= \left[M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n && \text{[step 2]} \\ &\approx \left(1 + \frac{t^2/2}{n}\right)^n && \text{[step 3]} \\ &\to e^{t^2/2} && \left[\text{Since } \left(1 + \frac{x}{n}\right)^n \to e^x\right] \end{aligned}$$

Hence, $\overline{Z}_n$ has the same MGF as that of a standard normal, so must follow that distribution! □