1 The Central Limit Theorem

Recall that if X_i are independent $\mathrm{N}(\mu, \sigma^2)$ random variables, then the stability property implies

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ and } T = n\bar{X} = \sum_{i=1}^{n} X_i \sim \mathrm{N}(n\mu, n\sigma^2).$$

The central limit theorem says that this holds in an approximate sense even if X_i are not normally distributed.

Theorem 1 (*Central Limit Theorem*)

Suppose that $X_1, ..., X_n$ are independent random variables, each with a common cumulative distribution function F_X. Suppose further that $\mathbb{E}(X_i) = \mu$, and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then for all $x \in \mathbb{R}$

$$\mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le x\right) \to \Phi(x) \qquad \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{approx}{\sim} \mathrm{N}(0, 1),$$

as $n \to \infty$. Here $\Phi(x)$ is the CDF of a standard normal.

It other words, the CLT implies that if you have

1. have a set of independent and identically distributed random variables,
2. have finite common mean μ and finite common variance,

then the sample mean (or total) can be approximated by normal distribution (by standardization)

$$\bar{X} \overset{approx}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ and } T = n\bar{X} = \sum_{i=1}^{n} X_i \overset{approx}{\sim} \mathrm{N}\left(n\mu, n\sigma^2\right).$$

1.1 Rules of Thumb

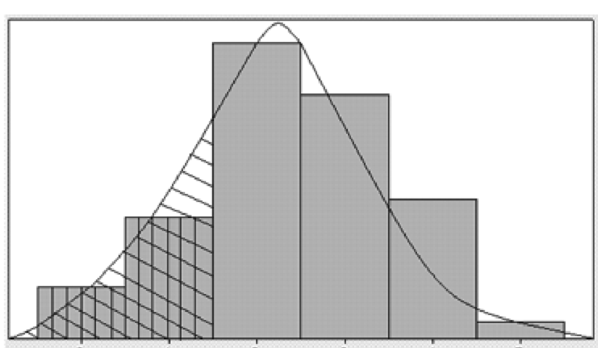
1.1.1 Continuity Correction

When we apply the CLT to *discrete random variables* taking *integer* values (or a subset of consecutive integers), we need to apply a *continuity correction* to account for the fact the random variables can't take non-integer values:

To approximate $\mathbb{P}(a \le T \le b)$ we instead compute $\mathbb{P}(a - 0.5 \le T \le b + 0.5)$.

The latter will give us a better approximation. We subtract or add 0.5 **before standardization to integer valued bounds** a **and** b to our various inequalities when approximating discrete distributions using the CLT, but it should not be applied when approximating continuous distributions.

Example 1. We want to compute $\mathbb{P}(0 \le T \le 1)$, but notice that the area under the continuous approximation of the discrete PMF is closer if we integrate from -0.5 to 1.5.



Remark 1. Whether or not we have strict inequalities matters when applying the continuity correction. To remember which way to apply the continuity correction, we should always write the probabilities using inequalities before applying our rule to widen the interval.

For example, for integer valued x

$$\mathbb{P}(T > x) = \mathbb{P}(T \geq x + 1) \overset{\text{correction}}{\Rightarrow} \mathbb{P}(T \geq x + 1 - 0.5) = \mathbb{P}(T \geq x + 0.5) = \mathbb{P}(T > x + 0.5).$$

Other cases are similar, for example for integer valued x

$$\mathbb{P}(T = x) = \mathbb{P}(x \leq T \leq x) \overset{\text{correction}}{\Rightarrow} \mathbb{P}(x - 0.5 \leq T \leq x + 0.5).$$

1.1.2 Sample Size Requirements

- **General Rule:** In general, the CLT often provides a reasonable approximation when $n > 30$.
- **"Approximately Normal" Distributions:** If the distribution of the observations is "close" to being unimodal, not too skewed, and is "close" to being continuous, then the central limit approximation can be reasonable for $n \in [5, 15]$.
- **"Not Approximately Normal" Distributions:** If the distribution is highly skewed, or very discrete, tcentral limit approximation can be reasonable for $n > 50$.

1.2 Normal Approximations of Important Distributions

We can use the CLT to approximate the distributions of important distributions we have already encountered.

1.2.1 Normal Approximation of the Binomial Distribution

Since $X \sim \text{Bin}(n, p)$ can be written as $X = \sum_{i=1}^{n} X_i$ where the $X_1, \dots, X_n$ are independent $\text{Bin}(1, p)$, we can apply the CLT to $\sum_{i=1}^{n} X_i$.

Theorem 2

If $X \sim \text{Bin}(n, p)$, then for large n

$$\frac{X - np}{\sqrt{np(1-p)}} \overset{approx}{\sim} N(0, 1).$$

1.2.2 Normal Approximation of the Poisson Distribution

If λ is a natural number, then $X \sim \text{Poi}(\lambda)$ can be written as $X = \sum_{i=1}^{\lambda} X_i$ where the $X_1, \dots, X_n$ are independent $\text{Poi}(1)$, we can apply the CLT to $\sum_{i=1}^{\lambda} X_i$.

Theorem 3

If $X \sim Poi(\lambda)$, then for large λ

$$\frac{X - \lambda}{\sqrt{\lambda}} \overset{approx}{\sim} N(0, 1).$$

1.3 Example Problems

1.3.1 Applications

Problem 1.1. A carton of wine consists of 20 winebottles. Suppose we can model the volume of wine in each bottle as independent normal random variables $X_1, \dots, X_{20}$ where mean 1.05 litres and and standard deviation $\sqrt{0.0004}$. What is the distribution of the total amount of wine in a carton, say T?

Solution 1.1. Since X_i are normally distributed, the total $T = \sum_{i=1}^{2} 0X_i$ is also normally distributed. We need to find the mean and variance of T,

$$\mathbb{E}[T] = \mathbb{E}\left[\sum_{i=1}^{20} X_i\right] = 20\,\mathbb{E}[X_1] = 20 \cdot 1.05 = 21$$

and by independence,

$$\mathrm{Var}[T] = \mathrm{Var}\left[\sum_{i=1}^{20} X_i\right] = \sum_{i=1}^{20} \mathrm{Var}(X_i) = 20 \cdot 0.0004 = 0.008.$$

Therefore, $T \sim \mathrm{N}(21, 0.008)$.

Problem 1.2. Harold is eating a box of chocolate right now (yes, right now). Each box contains 20 cubes, and it is supposed to have a total of 500 grams of chocolate in it. The weight of each chocolate cube varies a little because they are hand-made from Switzerland. The weight W of each cube is a random variable with mean $\mu = 25$ grams, and the standard deviation $\sigma = 0.1$ grams. Find the probability that a box has at least 500 grams of chocolate in it, assuming that the weight of each cube is independent.

Solution 1.2. Let $W_1, \dots, W_{20}$ be the weights of the cubes and $T = \sum_{i=1}^{20} W_i$ the total weight. The $W_1, \dots, W_{20}$ are independent with mean $\mu = 25$, $\sigma = 0.1$, and by the CLT, T is approximately $N(20\mu, 20\sigma^2)$. Using Z-tables gives us

$$\begin{aligned}\mathbb{P}(T \geq 500) &= \mathbb{P}\left(\frac{T - 20 \cdot \mu}{\sqrt{20\sigma^2}} \geq \frac{500 - 20 \cdot \mu}{\sqrt{20\sigma^2}}\right) \\ &\approx \mathbb{P}(Z \geq 0) = 1 - \Phi(0) = 0.5\end{aligned}$$

Problem 1.3. In February this year, various Youtubers participated in a '100 cup challenge' related to the Roll Up the Rim to Win promotion at Tim Hortons. The advertised cance to win is 1/6. Participants bought 100 promotional cups, and filmed themselves as they found out how many times they'd won. We want to use the central limit theorem to estimate the probability a participant recorded between 15 and 20 wins (inclusive).

1. Compute using the CLT *without* continuity correction.
2. Compute the probability exactly.
3. Compute using the CLT *with* continuity correction.

Note: think about the assumptions we must make when considering real-world examples!

Solution 1.3.

CLT without Correction: Let $X_i \sim \text{Bin}\left(1, \frac{1}{6}\right)$, $i = 1, ..., 100$ and $T = \sum_{i=1}^{100} X_i$ be the total wins. In this case, we have

$$\mathbb{E}[X_i] = p = \frac{1}{6} \text{ and } \text{Var}(X_i) = p(1-p) = \frac{1}{6}\left(1 - \frac{1}{6}\right) = \frac{5}{36}.$$

The CLT says that $T \sim \text{N}(np, np(1-p)) = \text{N}(100 \cdot \frac{1}{6}, 100 \cdot \frac{5}{36})$. Using Z-tables gives us

$$\begin{aligned}
\mathbb{P}(15 \leq T \leq 20) &= \mathbb{P}(T \leq 20) - \mathbb{P}(T < 15) \\
&\approx \mathbb{P}\left(\frac{T - np}{\sqrt{np(1-p)}} \leq \frac{20 - np}{\sqrt{np(1-p)}}\right) - \mathbb{P}\left(\frac{T - np}{\sqrt{np(1-p)}} < \frac{15 - np}{\sqrt{np(1-p)}}\right) \\
&= \mathbb{P}(Z \leq 0.894) - P(Z < -0.447) \\
&= 0.487.
\end{aligned}$$

Exact: We can compute the probability exactly since $T \sim \text{Bin}(100, 1/6)$, and so

$$\begin{aligned}
\mathbb{P}(15 \leq T \leq 20) &= \sum_{x=15}^{20} \binom{100}{x} \left(\frac{1}{6}\right)^x \left(1 - \frac{1}{6}\right)^{100-x} \\
&= 0.561
\end{aligned}$$

which is quite far away from the CLT approximation.

CLT with Correction: If we use correction, then we need to compute

$$\begin{aligned}
\mathbb{P}(14.5 \leq T \leq 20.5) &= \mathbb{P}(T \leq 20.5) - \mathbb{P}(T < 14.5) \\
&\approx \mathbb{P}\left(\frac{T - np}{\sqrt{np(1-p)}} \leq \frac{20.5 - np}{\sqrt{np(1-p)}}\right) - \mathbb{P}\left(\frac{T - np}{\sqrt{np(1-p)}} < \frac{14.5 - np}{\sqrt{np(1-p)}}\right) \\
&= \mathbb{P}(Z \leq 1.029) - \mathbb{P}(Z < -0.581) \\
&= 0.568.
\end{aligned}$$

which is very close to the probability of 0.561.

Problem 1.4. Suppose $X \sim \text{Poi}(\mu)$. Use the normal approximation to approximate

$$\mathbb{P}(X > \mu)$$

and compare this approximation with the true value when $\mu = 9$.

Solution 1.4. Since X is discrete, we need to use the normal approximation with continuity correction. We have

$$\mathbb{P}(X > 9) = \mathbb{P}(X \geq 10) \Rightarrow \mathbb{P}(X \geq 10 - 0.5) = \mathbb{P}(X \geq 9.5)$$

If we apply the normal approximation:

$$\frac{X - \mu}{\sqrt{\mu}} \overset{approx}{\sim} Z \sim \text{N}(0, 1)$$

we can see that, with $\mu = 9$:

$$\begin{aligned}\mathbb{P}(X > 9.5) &= \mathbb{P}\left(\frac{X-\mu}{\sqrt{\mu}} > \frac{9.5-\mu}{\sqrt{\mu}}\right) \\ &\approx \mathbb{P}(Z > 0.17) = 0.432.\end{aligned}$$

Remark 2. We can compute the probability exactly

$$\mathbb{P}(X > 9) = 1 - \mathbb{P}(X \le \mu) = 1 - \left(e^{-9} + 9e^{-9} + ... + \frac{9^9}{9!}e^{-9}\right) = 0.4126,$$

which is quite close to the normal approximation with correction. If we didn't apply the continuity correction, then

$$\mathbb{P}(X > \mu) = \mathbb{P}\left(\frac{X-\mu}{\sqrt{\mu}} > \frac{\mu-\mu}{\sqrt{\mu}}\right) \approx \mathbb{P}(Z > 0) = 0.5,$$

which is quite far off from the true value.

Problem 1.5. Let p be the proportion of Canadians who think Canada should adopt the US dollar.

(a) Suppose 400 Canadians are randomly chosen and asked their opinion. Let X be the number who say yes. Find the probability that the proportion, $\frac{X}{400}$, of people who say yes is within 0.02 of p, if $p = 0.20$.

(b) Suppose for a future opinion poll we want to determine the number, n, to survey to ensure that there is a 95% $\frac{T}{n}$ lies within 0.02 of p. Suppose $p = 0.20$ is known.

(c) Repeat (b) when the value of p is unknown. (Note that this would be the more realistic situation in the case of conducting an opinion poll.)

Solution 1.5. We want to apply the CLT.

Part (a): We have $X \sim \text{Bin}(n = 400, p = 0.2)$. Notice that

$$\mathbb{E}[X] = np = 80 \text{ and } \text{Var}(X) = np(1-p) = 64.$$

We want to compute

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \le 0.02\right) = \mathbb{P}\Big(n(p-0.02) \le X \le n(p+0.02)\Big) = \mathbb{P}\Big(72 \le X \le 88\Big).$$

Since X is takes integer values and the bounds are also integer valued we need to apply the continuity correction, so we instead compute

$$\mathbb{P}\Big(72 - 0.5 \le X \le 88 + 0.5\Big) = \mathbb{P}\Big(71.5 \le X \le 96.5\Big).$$

Splitting the probabilities, and standardizing gives us

$$\begin{aligned}\mathbb{P}\Big(71.5 \le X \le 88.5\Big) &= \mathbb{P}\Big(X \le 96.5\Big) - \mathbb{P}\Big(X < 71.5\Big) \\ &= \mathbb{P}\left(\frac{X-80}{\sqrt{64}} \le \frac{88.5-80}{\sqrt{64}}\right) - \mathbb{P}\left(\frac{X-80}{\sqrt{64}} < \frac{71.5-80}{\sqrt{64}}\right)\end{aligned}$$

Using the CLT, this is approximately equal to

$$\begin{aligned}&\mathbb{P}\left(Z \le \frac{88.5-80}{\sqrt{64}}\right) - \mathbb{P}\left(Z < \frac{71.5-80}{\sqrt{64}}\right)\\ &= \mathbb{P}(Z \le 1.0625) - \mathbb{P}(Z \le -1.0625) = 2\,\mathbb{P}(Z \le 1.0625) - 1 = 2(0.85543) - 1 = 0.71086.\end{aligned}$$

Remark 3. Notice that we applied the continuity correction to X and not $\frac{X}{n}$, since X is the random variable that takes integer values.

Part (b): It is a bit tricky to apply the continuity correction in this problem because we are solving for n and we don't necessarily know if $n \cdot 0.02$ will be an integer. However, we will see that n is quite large so the effect of the continuity correction will be small.

Let $\bar{X} = \frac{X}{n}$ be the sample mean. We want to find a n such that

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| \le 0.02\right) = \mathbb{P}\left(|\bar{X} - p| \le 0.02\right) = 0.95$$

Since

$$\mathbb{E}[\bar{X}] = p \text{ and } \mathrm{Var}(\bar{X}) = \frac{p(1-p)}{n}$$

the CLT implies that

$$\mathbb{P}\left(|\bar{X} - p| \le 0.02\right) = \mathbb{P}\left(\frac{|\bar{X}-p|}{\sqrt{\frac{p(1-p)}{n}}} \le \frac{0.02}{\sqrt{\frac{p(1-p)}{n}}}\right) \approx \mathbb{P}\left(|Z| \le \frac{0.02}{\sqrt{\frac{p(1-p)}{n}}}\right).$$

We want the right hand side to be 0.95, so we use quantiles to find the critical value. Since

$$\begin{aligned}\mathbb{P}(|Z| \le x) = \mathbb{P}(-x \le Z \le x) &= \mathbb{P}(Z \le x) - \mathbb{P}(Z \le -x)\\ &= \mathbb{P}(Z \le x) - (1 - \mathbb{P}(Z \le x)) = 2\,\mathbb{P}(Z \le x) - 1\end{aligned}$$

using the quantile table, we have that

$$\mathbb{P}(|Z| \le x) = 0.95 \iff 2\,\mathbb{P}(Z \le x) - 1 = 0.95 \iff \mathbb{P}(Z \le x) = 0.975 \iff x = F_Z^{-1}(0.975) = 1.96.$$

Therefore, we require that (with $p = 0.2$)

$$\frac{0.02}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.02}{\sqrt{\frac{0.2(1-0.2)}{n}}} = 1.96 \implies n = 1536.64$$

so we should take at least $n = 1537$ samples.

Remark 4. By the 95% rule we know that approximately 95% of the probability lies within two standard deviations. In the computations above, we carefully showed this result by finding the n such that the $\frac{0.02}{\sqrt{\mathrm{Var}(\frac{X}{n})}} = 1.96$, which is roughly two standard deviations.

Part (c): If p is unknown, then we want to make sure that our sample size is large enough so that we can guarantee a 0.95 accuracy for any p. That is, we want

$$\mathbb{P}\left(|\bar{X} - p| \le 0.02\right) \ge 0.95$$

for all p. If $p = 0$ or $p = 1$, then the problem is trivial. If $p \in (0,1)$, then the computations from above implies that

$$\mathbb{P}\left(|\bar{X} - p| \le 0.02\right) \approx \mathbb{P}\left(|Z| \le \frac{0.02}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

This is larger than 0.95 whenever $\frac{0.02}{\sqrt{\frac{p(1-p)}{n}}} \ge 1.96$. Thus, for this to be larger than 0.95 to hold for all p, we want to find n such that

$$\frac{0.02}{\sqrt{\frac{p(1-p)}{n}}} \ge 1.96$$

for any all p. Notice that $p(1-p)$ is maximized when $p = \frac{1}{2}$, so we have

$$\frac{0.02}{\sqrt{\frac{p(1-p)}{n}}} \ge \frac{0.02}{\sqrt{\frac{0.5(1-0.5)}{n}}}$$

with equality when $p = 0.5$. Since

$$\frac{0.02}{\sqrt{\frac{0.5(1-0.5)}{n}}} = 1.96 \implies n = 2401$$

we need $n = 2401$ to achieve the desired accuracy for any $p = 0.5$. Therefore, by monotonicity

$$\mathbb{P}\left(|Z| \le \frac{0.02}{\sqrt{\frac{p(1-p)}{n}}}\right) \ge \mathbb{P}\left(|Z| \le \frac{0.02}{\sqrt{\frac{0.5(1-0.5)}{n}}}\right) = 0.95$$

which gives us the required accuracy for all p.

Remark 5. By the 95% rule we know that approximately 95% of the probability lies within two standard deviations. In the computations above, we found that in the worst case scenario when $p = \frac{1}{2}$ taking n such that the $\frac{0.02}{\sqrt{\mathrm{Var}(\frac{X}{n})}} = 1.96$ achieved the required accuracy. Since n is taken large enough to be close in the worst case scenario, we get the required accuracy in all scenarios.

1.3.2 Derivations and Proofs

Problem 1.6. If

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{approx}{\sim} \mathrm{N}(0,1)$$

show that

$$\bar{X} \overset{approx}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ and } T = n\bar{X} = \sum_{i=1}^{n} X_i \overset{approx}{\sim} \mathrm{N}\left(n\mu, n\sigma^2\right).$$

Solution 1.6. Let $Z \sim \mathrm{N}(0,1)$. By standardization, we have

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{approx}{\sim} Z \implies \bar{X} = \frac{\sigma}{\sqrt{n}} Z + \mu$$

so $\bar{X} \overset{approx}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Likewise, using the fact that $T = n\bar{X}$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{approx}{\sim} Z \implies T = n\bar{X} = \sqrt{n}\sigma Z + n\mu$$

so $T \overset{approx}{\sim} \mathrm{N}\left(n\mu, n\sigma^2\right)$.

Problem 1.7. Prove the following Normal approximation results

1. If $X \sim \text{Bin}(n,p)$, then for large n

$$\frac{X - np}{\sqrt{np(1-p)}} \overset{approx}{\sim} N(0,1).$$

2. If $X \sim Poi(\lambda)$, then for large λ

$$\frac{X - \lambda}{\sqrt{\lambda}} \overset{approx}{\sim} N(0,1).$$

Solution 1.7.

1. Since $X \sim \text{Bin}(n,p)$ can be written as $X = \sum_{i=1}^n X_i$ where the $X_1, \dots, X_n$ are independent $\text{Bin}(1,p)$ (which have mean p and variance $p(1-p)$), we can apply the CLT to $\sum_{i=1}^n X_i$,

$$\sum_{i=1}^n X_i \sim \text{N}(np, np(1-p))$$

so the result now follows from standardization.

2. If $\lambda = n$ is a natural number, then $X \sim \text{Poi}(n)$ can be written as $X = \sum_{i=1}^\lambda X_i$ where the $X_1, \dots, X_n$ are independent $\text{Poi}(1)$ (which have mean 1 and variance 1), we can apply the CLT to $\sum_{i=1}^\lambda X_i$,

$$\sum_{i=1}^n X_i \sim \text{N}(n, n)$$

so the result now follows from standardization.

2 Moment Generating Functions

So far we have seen that the distribution of a random variable can be characterized by the PMF/PDF and the CDF. We now show that it is also possible to encode the distribution through another object called the moment generating function.

Definition 1. The *moment generating function* (MGF) of a random variable X is given by the function

$$M_X(t) = \mathbb{E}[e^{tX}],$$

provided the expression exists in a neighbourhood of zero, say for $t \in (-a, a)$.

If X is discrete with PMF $f_X(x)$, then

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{x \in X(S)} e^{tx} f_X(x)$$

and if X is continuous with PDF $f_X(x)$, then

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx.$$

Remark 6. The probability generating function is another generating function we have already seen. It is defined by $G_X(s) = \mathbb{E}[s^X]$ and it is just a change of variables of the MGF since

$$G_X(s) = \mathbb{E}[s^X] = \mathbb{E}[e^{\ln(s^X)}] = \mathbb{E}[e^{\ln(s)X}] = M_X(\ln(s)).$$

2.1 Properties

If X has finite support on $\{x_1, \ldots, x_n\}$ with PMF f_X, then

$$M_X(t) = f_X(x_1)e^{tx_1} + \cdots + f_X(x_n)e^{tx_n}$$

so it is easy to read off the PMF given the MGF. We see that even for more complicated distributions the MGF still encodes the distribution of the random variables.

1. **Moments:** The MGF encodes all the moments of X. Assuming that $M_X(t)$ is defined in a neighbourhood of $t = 0$,

$$\frac{d^k}{dt^k} M_X(0) = \mathbb{E}[X^k] \text{ for all } k \geq 0.$$

   This follows from Taylor's theorem and linearity holds even with infinite sums provided that $M_X(t)$ is defined in a neighbourhood of $t = 0$

$$M_X(t) = \mathbb{E}\left[\sum_{j=0}^{\infty} \frac{t^j X^j}{j!}\right] = \sum_{j=0}^{\infty} \frac{t^j\, \mathbb{E}[X^j]}{j!}.$$

2. **Uniqueness Theorem:** If X and Y have MGFs $M_X(t)$ and $M_Y(t)$ defined in neighbourhoods of the origin, and $M_X(t) = M_Y(t)$ for all t where they are defined, then

$$X \sim Y.$$

3. **Independent Sums:** Suppose that X and Y are independent and each have moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of $X + Y$ is

$$M_{X+Y}(t) = \mathbb{E}\left(e^{t(X+Y)}\right) = \mathbb{E}\left(e^{tX}\right)\mathbb{E}\left(e^{tY}\right) = M_X(t)M_Y(t).$$

   This is a useful property because the distribution of the sum of random variables can now be easily computed (which required computing iterated sums).

2.2 Moment Generating Functions of Common Distributions

1. **Discrete Uniform:** If $X \sim \mathrm{U}[a,b]$ (discrete) then
$$M_X(t) = \frac{1}{b-a+1}\sum_{x=a}^{b} e^{tx} \text{ for } t \in \mathbb{R}$$
2. **Binomial:** If $X \sim \mathrm{Bin}(n,p)$ then
$$M_X(t) = (pe^t + (1-p))^n \text{ for } t \in \mathbb{R}$$
3. **Negative Binomial:** If $X \sim \mathrm{NegBin}(k,p)$ then
$$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^k \text{ for } t \leq -\ln(1-p)$$
4. **Poisson:** If $X \sim \mathrm{Poi}(\lambda)$ then
$$M_X(t) = e^{\lambda(e^t-1)} \text{ for } t \in \mathbb{R}$$
5. **Continuous Uniform:** If $X \sim \mathrm{U}[a,b]$ (continuous) then
$$M_X(t) = \begin{cases} \frac{e^{bt}-e^{at}}{(b-a)t} & t \neq 0 \\ 1 & t = 0 \end{cases}$$
6. **Exponential:** If $X \sim \mathrm{Exp}(\theta)$ (waiting time parametrization) then
$$M_X(t) = \frac{1}{1-\theta t} \text{ for } t < \frac{1}{\theta}$$
7. **Normally Distributed:** If $X \sim \mathrm{N}(\mu,\sigma^2)$, then
$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \text{ for } t \in \mathbb{R}$$

2.3 Example Problems

2.3.1 Applications

Problem 2.1. Suppose that X has MGF
$$M_X(t) = 0.3 + 0.2e^t + 0.5e^{2t}$$
What is the PMF of X?

Solution 2.1. We have
$$M_X(t) = 0.3 + 0.2e^t + 0.5e^{2t} = 0.3e^{0\cdot t} + 0.2e^{1\cdot t} + 0.5e^{2\cdot t}.$$
Then reverse engineering the PMF from the MGF implies that
$$f_X(0) = \mathbb{P}(X=0) = 0.3, \qquad f_X(1) = \mathbb{P}(X=1) = 0.2 \qquad F_X(2) = \mathbb{P}(X=2) = 0.5.$$
It is clear that at least for finitely supported PMF, the MGF is simply another way of writing encoding the distribution.

Problem 2.2. Use the MGFs to show that if $X \sim \text{Poi}(\lambda)$, then $\mathbb{E}[X] = \lambda$ and $\text{Var}(X) = \lambda$

Solution 2.2. The MGF is

$$M_X(t) = e^{\lambda(e^t-1)}.$$

Taking derivatives, we have

$$M_X'(t) = e^{\lambda(e^t-1)}\lambda e^t \Rightarrow \mathbb{E}(X) = M_X'(0) = \lambda.$$

To compute the variance, we take the second derivative

$$M_X''(t) = e^{\lambda(e^t-1)}\lambda e^t + e^{\lambda(e^t-1)}(\lambda e^t)^2 \Rightarrow \mathbb{E}(X^2) = M_X''(0) = \lambda^2 + \lambda$$

and conclude

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Problem 2.3. Use MGFs to show that if $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$ are independent, then $X+Y \sim \text{Poi}(\lambda+\mu)$.

Solution 2.3. We know that $M_X(t) = e^{\lambda(e^t-1)}$ and $M_Y(t) = e^{\mu(e^t-1)}$. Since X and Y are independent, the MGF of $X+Y$ is

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\lambda(e^t-1}e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$$

which we recognize as the MGF of a $\text{Poi}(\lambda+\mu)$ random variables. By uniqueness of the MGF, we conclude $X+Y \sim \text{Poi}(\lambda+\mu)$.

Problem 2.4. Use MGFs to show that if $X \sim \text{N}(\mu_1, \sigma_1^2)$ and $Y \sim \text{N}(\mu_2, \sigma_2^2)$ are independent, then $X+Y \sim \text{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Solution 2.4. We know that $M_X(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}}$ and $M_Y(t) = e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}}$. Since X and Y are independent, the MGF of $X+Y$ is

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}}e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}} = e^{(\mu_1+\mu_2)t + \frac{(\sigma_1^2\sigma_2^2)t^2}{2}}$$

which we recognize as the MGF of a $\text{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$ random variables. By uniqueness of the MGF, we conclude $X+Y \sim \text{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Problem 2.5. Use MGFs to show that for large n and small p such that $\mu = np$, we can approximate the $\text{Bin}(n,p)$ distribution with a $\text{Poi}(\mu)$ distribution.

Solution 2.5. By the uniqueness theorem, we can prove this by showing that the MGF of $X \sim \text{Bin}(n,p)$ converges to the MGF of $Y \sim \text{Poi}(\mu)$ as $n \to \infty$ where $\mu = np \Leftrightarrow p = \frac{\mu}{n}$. Indeed,

$$M_X(t) = (pe^t + 1 - p)^n = (1 + p(e^t-1))^n = \Big(1 + \frac{\mu}{n}(e^t-1)\Big)^n.$$

Since for every x

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

we have

$$\lim_{n\to\infty} M_X(t) = \lim_{n\to\infty}\Big(1+\frac{\mu}{n}(e^t-1)\Big)^n = e^{\mu(e^t-1)} = M_Y(t),$$

2.3.2 Proofs and Derivations

Problem 2.6. Prove the following formulas for the MGFs of common distributions

1. **Poisson:** If $X \sim \text{Poi}(\lambda)$ then
$$M_X(t) = e^{\lambda(e^t-1)} \text{ for } t \in \mathbb{R}$$
2. **Normal Distribution:** If $X \sim \text{N}(\mu, \sigma^2)$, then
$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \text{ for } t \in \mathbb{R}$$
3. **Negative Binomial:** If $X \sim \text{NegBin}(k, p)$ then
$$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^k \text{ for } t \le -\ln(1-p)$$

Solution 2.6. The computation of all the MGFs use similar tricks we have seen before (summation formulas, reducing to sums of PMF/integral PDF, etc)

Poisson: The MGF is computed using the exponential sum

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{-\lambda} e^{e^t\lambda} = e^{\lambda(e^t-1)}.$$

Normal Distribution: The MGF is computed by completing the square. We have

$$M_X(t) = \mathbb{E}(e^{tX}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2(\mu+t\sigma^2)x+\mu^2}{2\sigma^2}}\, dx.$$

We want to rewrite the integral in terms of the integral of a PDF, so we complete the square

$$\begin{aligned} x^2 - 2(\mu+t\sigma^2)x + \mu^2 &= x^2 - 2(\mu+t\sigma^2)x + (\mu+t\sigma^2)^2 - (\mu+t\sigma^2)^2 + \mu^2 \\ &= (x-(\mu+t\sigma^2))^2 - 2t\sigma^2\mu - t^2\sigma^4 \end{aligned}$$

so

$$\begin{aligned} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2(\mu+t\sigma^2)x+\mu^2}{2\sigma^2}}\, dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+t\sigma^2))^2-2t\sigma^2\mu-t^2\sigma^4}{2\sigma^2}}\, dx \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+t\sigma^2))^2}{2\sigma^2}}\, dx}_{=1 \text{ sum of PDF of } \text{N}(\mu+t\sigma^2, \sigma^2)}. \end{aligned}$$

Negative Binomial: We first consider the case when $Y \sim \text{NegBin}(1, p) = \text{Geo}(p)$. In this case, using the geometric series (which exists for $|(1-p)e^t| < 1 \implies t < -\ln(1-p)$)

$$M_Y(t) = \sum_{x=0}^{\infty} e^{tx} p(1-p)^x = \sum_{x=0}^{\infty} p((1-p)e^t)^x = \frac{p}{1-(1-p)e^t}.$$

We recall that $X \sim \text{NegBin}(k, p)$ is the sum of k independent $\text{Geo}(p)$ random variables. Therefore, $X = Y_1 + \cdots + Y_k$ and the sum of independent random variables formula for MGFs give

$$M_X(t) = M_{Y_1+\cdots+Y_k}(t) = M_{Y_1}(t) \cdots M_{Y_k}(t) = M_Y^k(t) = \left(\frac{p}{1-(1-p)e^t}\right)^k,$$

which exists under the same condition as $M_Y(t)$.

Problem 2.7. ($\star$) Suppose that $M_X(t)$ is defined for all $t \in [-a, a]$ for some $a > 0$. Prove that

$$\mathbb{E}[X^k] = \frac{d^k}{dt^k} M_X(0) = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0} \qquad \forall k \geq 0.$$

Solution 2.7. Assuming that we can interchange the differentiation and integration, we have

$$\frac{d^k}{dt^k} M_X(t) = \frac{d^k}{dt^k}\,\mathbb{E}[e^{tX}] = \mathbb{E}\left[\frac{d^k}{dt^k} e^{tX}\right] = \mathbb{E}[X^k e^{tX}]$$

Therefore,

$$\frac{d^k}{dt^k} M_X(0) = \mathbb{E}[X^k e^{0\cdot X}] = \mathbb{E}[X^k].$$

The interchange between the differentiation and expectation is justified by a result called the dominated convergence theorem. The technical assumption that $M_X(t)$ is defined for all $t \in [-a, a]$ for some $a > 0$ is required to find a dominating function. In the case when X has finite support, then the interchange of differentiation and expectation simply follows from the linearity of the differentiation operator.

Problem 2.8. ($\star$) Prove the uniqueness theorem: If X and Y have MGFs $M_X(t)$ and $M_Y(t)$ defined in neighbourhoods of the origin, and $M_X(t) = M_Y(t)$ for all t where they are defined, then

$$X \sim Y.$$

Solution 2.8. The full proof of this statement is too advanced for this course. However, we can do a simpler proof of this result for discrete random variables. We have already seen that if X has finite support, then we can simply read off the PMF from the MGF. In particular, if X and Y are supported on finitely many points and they have the same MGFs, then they define the same PMFs so they have the same distribution.

We will prove that this logic extends to X that are supported on the natural numbers. Consider the probability generating function $G_X(s) = M_X(\ln(s)) = \mathbb{E}[s^X]$ and suppose that $G_X(s)$ is finite for some s_0. By definition, we have

$$G_X(s) = \sum_{x=0}^{\infty} s^x f_X(x) = f_X(0) + s^1 f_X(1) + s^2 f_X(2) + \dots$$

Notice that

$$G_X(0) = f_X(0) = \mathbb{P}(X = 0), \quad G'_X(0) = f_X(1) = \mathbb{P}(X = 1), \quad G''_X(s) = 2f_X(2) = 2\,\mathbb{P}(X = 2).$$

Continuing inductively, we see that $\frac{d^k}{ds^k} G_X(0) = k!\,\mathbb{P}(X = k)$. In particular, the moment generating function, and the corresponding probability generating function encodes the PMF. So if X and Y are supported on finitely many points and they have the same MGFs, then they encode the same PMFs so they have the same distribution.

Problem 2.9. ($\star$) Prove the central limit theorem under the additional assumption that the MGF of X is finite for all $t \in [-a, a]$ for some $a > 0$. Suppose that $X_1, ..., X_n$ are independent random variables, each with a common cumulative distribution function F_X. Suppose further that $\mathbb{E}(X_i) = \mu$, and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then for all $x \in \mathbb{R}$

$$\mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) \to \Phi(x)$$

as $n \to \infty$.

Solution 2.9. We will show that as $n \to \infty$, the MGF of $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ converges to the MGF of the standard normal distribution $Z \sim \mathrm{N}(0,1)$,

$$M_Z(t) = e^{\frac{t^2}{2}}.$$

Consider the normalized random variables $Y_i = \frac{X_i - \mu}{\sigma}$. Notice that Y_i satisfies

$$\mathbb{E}[Y_i] = \mathbb{E}\left[\frac{X_i - \mu}{\sigma}\right] = \frac{\mathbb{E}[X_i] - \mu}{\sigma} = 0$$

and

$$\mathrm{Var}(Y_i) = \mathrm{Var}\left[\frac{X_i - \mu}{\sigma}\right] = \frac{1}{\sigma^2}\mathrm{Var}(X_i) = 1.$$

We have

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

By the independence property of the MGFs,

$$M_{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}(t) = \mathbb{E}[e^{\frac{t}{\sqrt{n}}\sum_{i=1}^n Y_i}] = \prod_{i=1}^n \mathbb{E}[e^{\frac{t}{\sqrt{n}}Y_i}] = \prod_{i=1}^n M_{Y_i}\Big(\frac{t}{\sqrt{n}}\Big) = M_Y^n\Big(\frac{t}{\sqrt{n}}\Big). \tag{1}$$

By Taylor's theorem, we have that $M_Y(t)$ satisfies,

$$M_Y(t) = M_Y(0) + tM_Y'(0) + \frac{t^2}{2}M_Y''(0) + o(t^2)$$

where the error term $o(t^2)$ satisfies $\lim_{t\to 0}\frac{o(t^2)}{t^2} = 0$. Since the derivatives of the MGFs encodes the moments, we have that

$$M_Y(0) = 1,\ M_Y'(0) = \mathbb{E}[Y] = 0,\ M_Y''(0) = \mathbb{E}[Y^2] = \mathrm{Var}(Y) + (\mathbb{E}[Y])^2 = 1.$$

Therefore,

$$M_Y(t) = 1 + \frac{t^2}{2} + o(t^2).$$

Combining this with (1) results in

$$M_{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}(t) = \Big(1 + \frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\Big)^n.$$

Since $\lim_{n\to\infty}(1 + \frac{x}{n})^n = e^x$, we have that (the error term $o(\frac{t^2}{n})$ will not affect the limit)

$$\lim_{n\to\infty} M_{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}(t) = \lim_{n\to\infty}\Big(1 + \frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\Big)^n = e^{\frac{t^2}{2}} = M_Z(t),$$

so the uniqueness property implies that $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ has the same asymptotic distribution as a standard normal.

Remark 7. We used the assumption that the MGF exists and is finite on an interval around 0 to use the uniqueness and differentiation properties of the MGF. We were a bit sloppy in the usage of the error terms $o(t^2)$ and the application of Taylor's theorem for a random remainder term, but this can be justified using an interchange of limit and expected value by the dominated convergence theorem.