The Central Limit Theorem and Normal Approximations: Chapter 10

Chapter Outcomes:

- Understanding the Central Limit Theorem (C.L.T.)
- Introduction to normal approximations
- Moment Generating Functions (mgf’s)

Motivation

Under certain conditions, the normal distribution can be used to approximate probabilities for **linear combinations** of random variables having a *non-normal* probability distribution.

This can be done through the well-known **Central Limit Theorem (C.L.T.)**.

The normal distribution is used as it tends to approximate the distribution of sums of random variables.

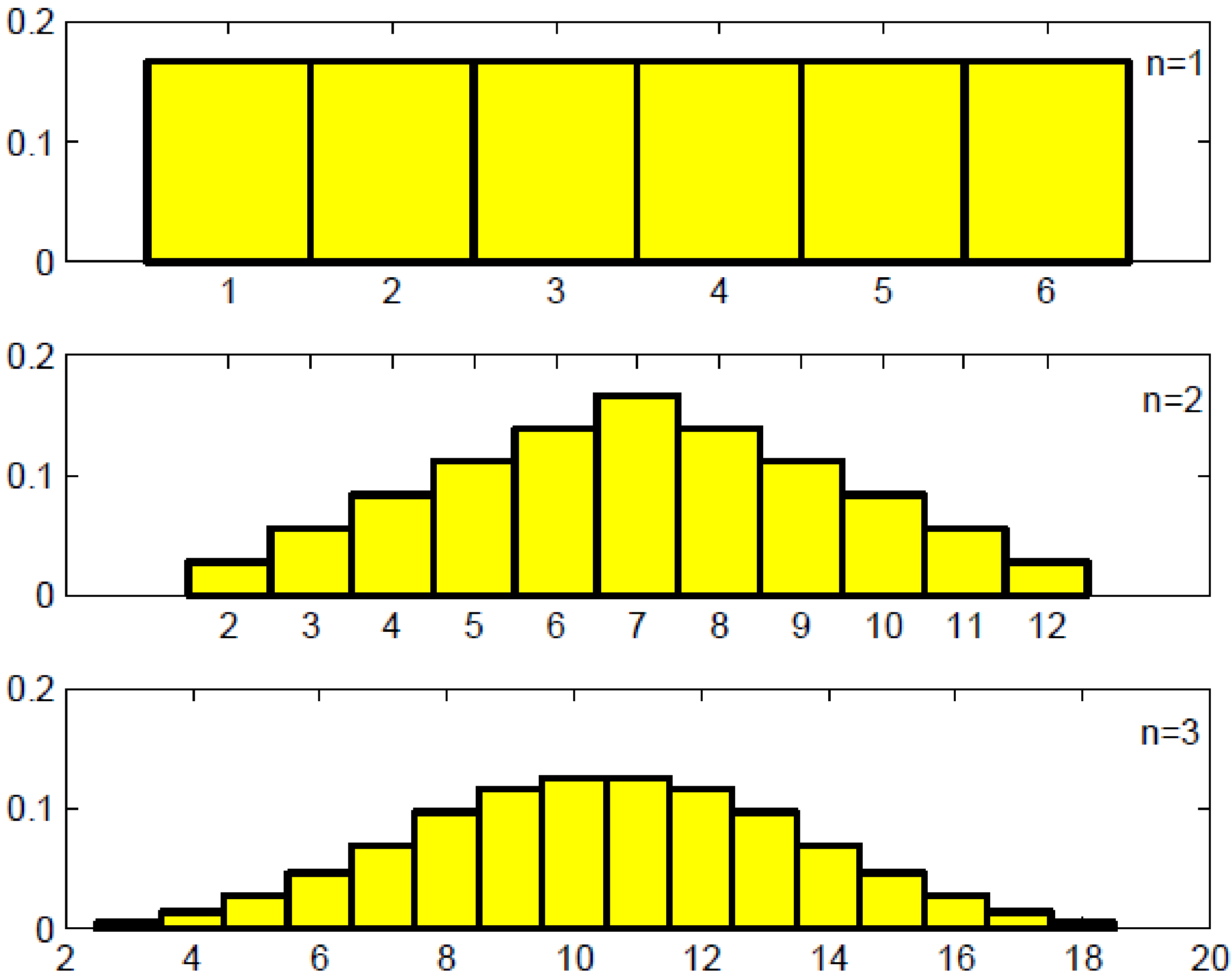


Figure 10.1: Probability histograms for the sum of n rolls of a dice for $n = 1, 2, 3$

Central Limit Theorem: If $X_1, X_2, \dots, X_n$ are ***independent*** random variables all having the ***same distribution with mean μ and variance σ^2, then as $n \to \infty$***, the c.d.f. of the r.v.

$$\frac{\left(\sum_{i=1}^{n} X_i\right) - n\mu}{\sigma\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

approaches the **N(0,1)** c.d.f..

Equivalently, the C.L.T. states that the c.d.f. of

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

approaches the **N(0,1)** c.d.f. as $n \to \infty$.

Remarks:

1. Although the C.L.T. deals with a limiting result, we can use this result when n is large (but finite) as an approximation.
2. The C.L.T. works for essentially all distributions except those whose mean and variance *do not exist* (i.e. are not finite). Such distributions exist, but they are *rare*.
3. The *accuracy* of the approximation depends on the value of n (bigger is better!) and also on the actual distribution of $X_1, X_2, \dots, X_n$ (i.e. when the distribution of $X_1, X_2, \dots, X_n$ is close to *symmetric*, the approximation works better for smaller n).
4. $n \geq 30$ is a **general rule of thumb** for the normal approximation to perform reasonably well.
5. The formal proof of the C.L.T. is beyond the scope of this course.

In the previous chapter, we considered the distribution of linear combinations of independent normal random variables.

In particular, if $X_1, X_2, \dots, X_n$ are independent $N(\mu, \sigma^2)$ random variables, then

$$S_n = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

and

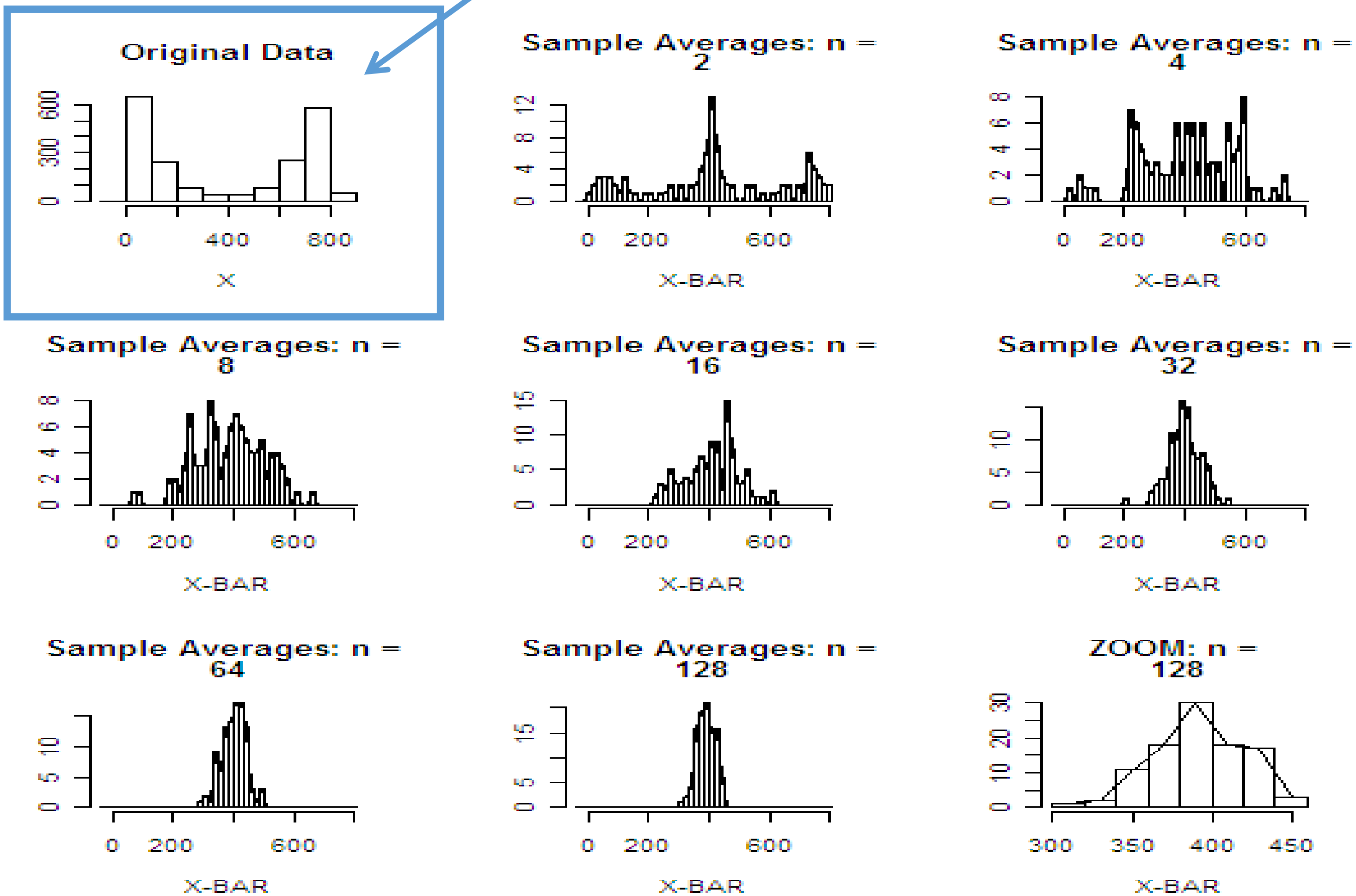
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Therefore, if $X_1, X_2, \dots, X_n$ are normally distributed, then S_n and $\bar{X}$ have **exact** normal distributions **for all** values of n.

On the other hand, if $X_1, X_2, \dots, X_n$ **are not** normally distributed, then S_n and $\bar{X}$ have **approximately** normal distributions for **large** values of n.

General Idea of the Central Limit Theorem (C.L.T.)

Original data NOT Normal



Example: Suppose fires reported to a fire station satisfy the conditions for a Poisson process, with a mean of 1 fire every 4 hours.

Find the probability that the 500th fire of the year is reported on the 84th day of the year.

Here is an alternate solution that may make things a little clearer.

The time between a Poisson occurrence has an Exponential distribution. In this case, with 1 fire every 4 hours, we have an average of (1/6) day per fire. To determine the time until the 500th fire we would need to consider the sum of 500 independent Exponential r.v.'s. We haven't learned about this distribution. Instead, we will approximate this probability using the CLT.

To do so, we will consider the average # of days per fire when we set up our calculation.

Again, the average is (1/6) fires each day with a standard deviation of (1/6). Now, we need to focus on time 83 and 84.

Remember that we want the 500th fire to occur some time on the 84th day.

So, to approximate this probability, we want:

$$\mathbf{P}\left(Z < \frac{\frac{84}{500}-\frac{1}{6}}{\left(\frac{1}{6}\right)/\sqrt{500}}\right) - \mathbf{P}\left(Z < \frac{\frac{83}{500}-\frac{1}{6}}{\left(\frac{1}{6}\right)/\sqrt{500}}\right) = P(Z < 0.18) - P(Z < -0.09)$$

$$= 0.57142 - (1 - 0.53586) = 0.10728$$

Note: You can also approximate the probability to this question using the sum of 500 independent Exponential r.v.'s. I have expressed things in terms of averages here. You will still get the same approximate probability.

The previous example used the normal distribution to approximate a sum of **continuous** random variables (i.e. $X_1, X_2, \dots, X_{500}$ were exponentially distributed), each of which is highly skewed and non-symmetric.

When approximating a ***discrete*** r.v., a slight adjustment is required to improve the approximation.

This correction is called the **"continuity correction".**

It is **ONLY** applied when using a **continuous** distribution to approximate a **discrete** one.

Instead of guessing or remembering whether to add or subtract 0.5, it is often easier to draw a quick sketch.

When approximating a probability such as $P(X = 50)$ where $X \sim Bin(100, 0.5)$, it is essential to first apply the continuity correction (since without it, we simply get an approximation of $P(X = 50)$ = 0, which is meaningless).

Normal Approximation to the Poisson Distribution

If $X \sim Poisson(\mu)$, then

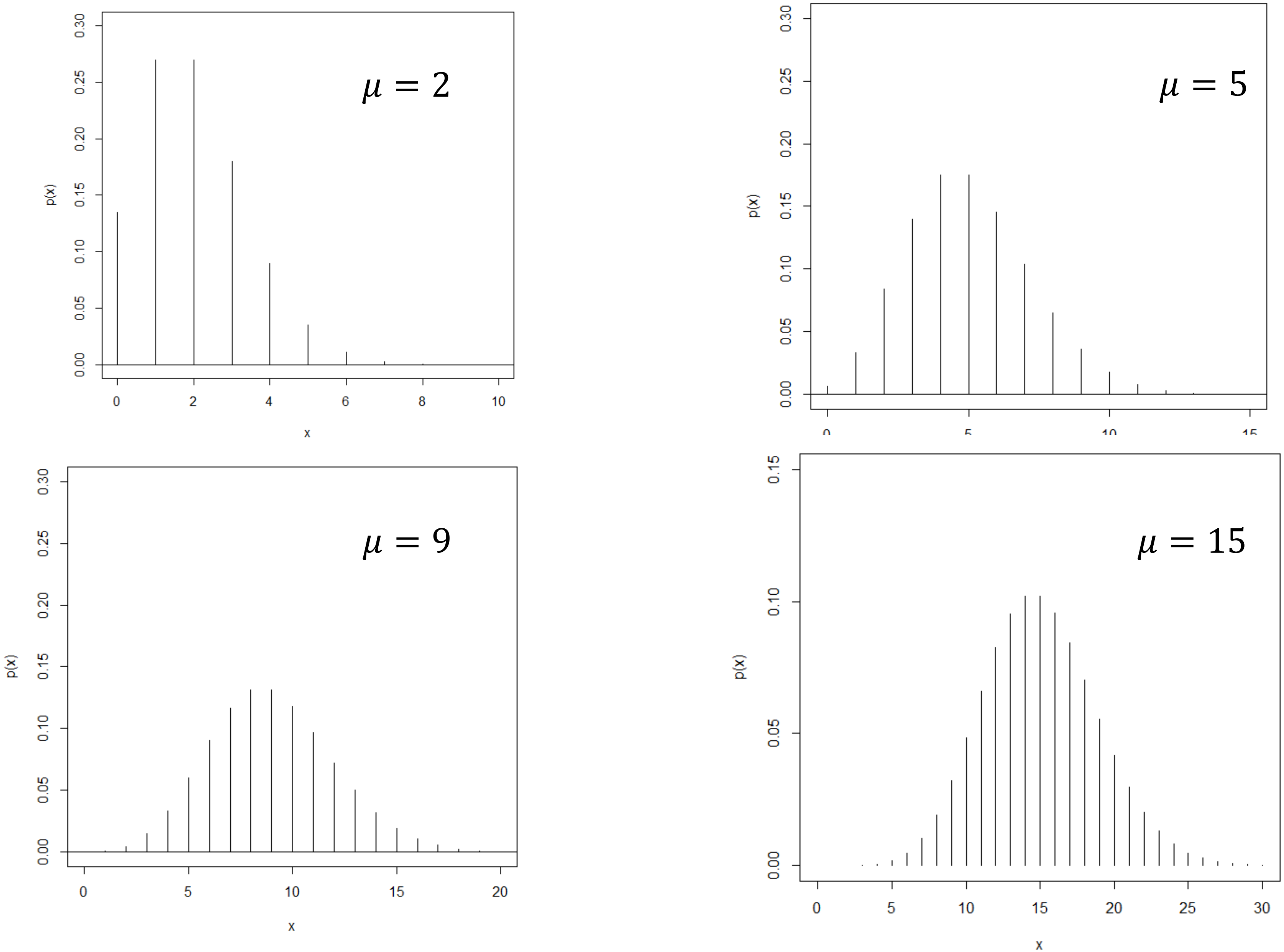
$$Z = \frac{X - \mu}{\sqrt{\mu}}$$

is approximately **N(0,1)** for large values of μ.

Note: When using the Normal approximation to the Poisson, you should apply a **continuity correction**.

Why?

We are using a continuous distribution to approximate a discrete distribution's probability.



$\mu = 2$
$\mu = 5$
$\mu = 9$
$\mu = 15$
p(x)
x
0.00 0.05 0.10 0.15 0.20 0.25 0.30
0 2 4 6 8 10
0 5 10 15
0 5 10 15 20
0 5 10 15 20 25 30

Example: Suppose $\mathrm{X} \sim \mathrm{Poisson}(9)$. Use the normal approximation to approximate $P(X > 9)$ and compare this approximation with its true value.

Normal Approximation to the Binomial Distribution

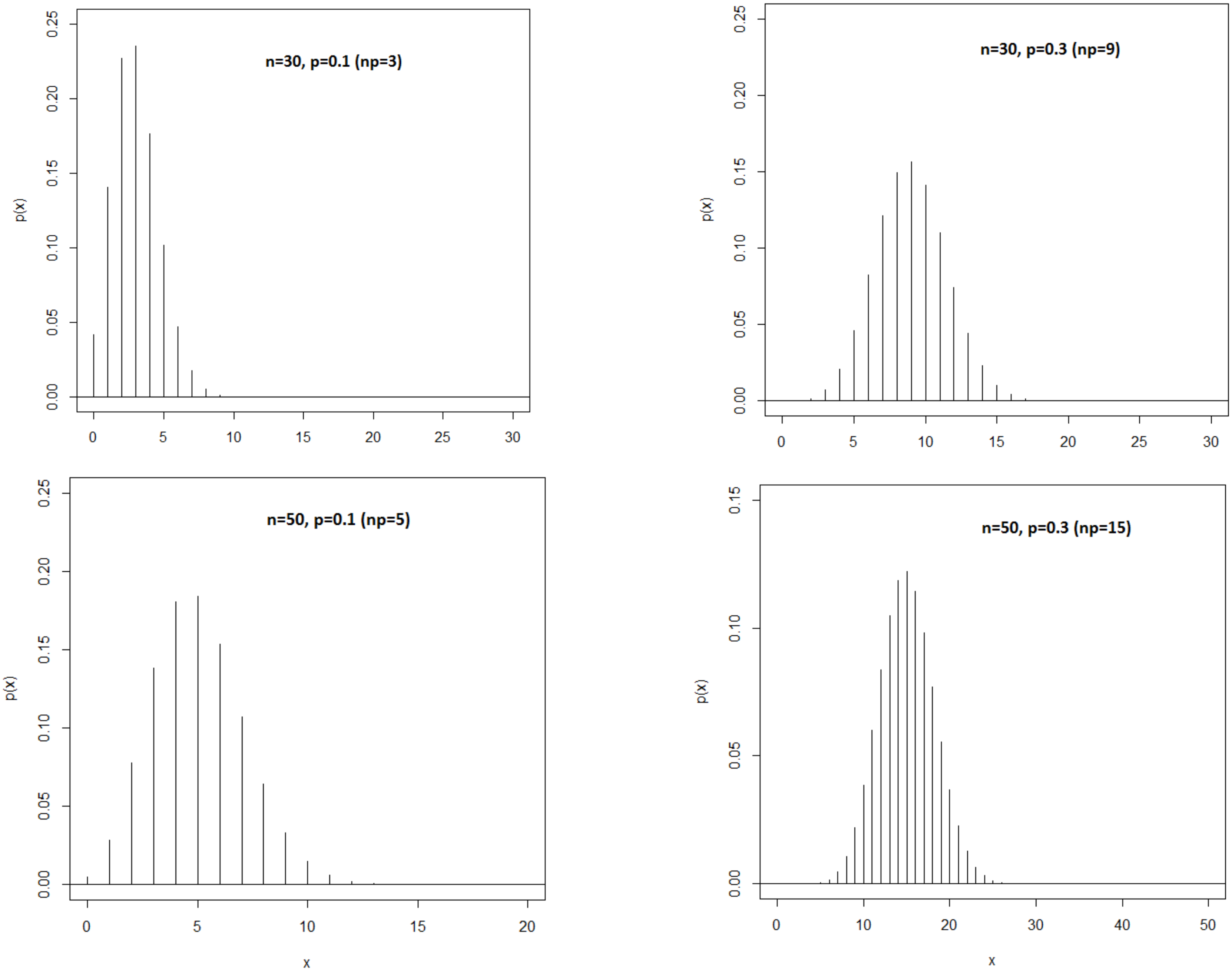
If $X \sim Bin(n, p)$, then

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately **N(0,1)** for large values of n.

Should we apply a continuity correction for this approximation?

Yes! (See Normal approximation to the Poisson Distribution if you are unsure).


n=30, p=0.1 (np=3)
n=30, p=0.3 (np=9)
n=50, p=0.1 (np=5)
n=50, p=0.3 (np=15)
p(x)
x
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
40
50

Example: Suppose $X \sim Bin(20, 0.4)$. Use the normal approximation to approximate $P(4 \leq X \leq 12)$ and compare this approximation with its true value.

Example: 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 in the sample, $X/400$, of people who say yes is within 0.02 of p, if p is 0.20.

b) Find the minimum number, n, who must be surveyed so that there is a 95% chance that the sample proportion, X/n, lies within 0.02 of p, when p is unknown.

****Remember that when p is UNKNOWN, assuming p = 0.5 will give you the largest value of n. It's the safest guess with no information.**

10.2 Moment Generating Functions:

So far we have looked at two functions which characterize a distribution:

1. Probability function: $f(x) = P(X = x)$
2. Cumulative distribution function:

$$F(x) = P(X \leq x)$$

Now we will talk about a third function:

Moment Generating Function, which **uniquely** determines a distribution.

- **Definition**: Consider a discrete random variable X with probability function $f(x)$. The moment generating function (m.g.f) of X is defined as:

$$M(t) = E[e^{tX}] = \sum_x e^{tx} f(x)$$

The m.g.f is assumed to be defined and finite for values of $t \in [-a, a]$ for a>0

- The m.g.f can be used to evaluate the moments of the random variable X, where the **moments** of X are defined as the **expectations** of the **functions** X^r for r=1, 2, …
- $E[X^r]$ is the r^th^ moment of X

- Example:
- The mean: $\mu = E[X]$ is the first moment of X.
- $E[X^2]$ is the second moment of X and so on.

- **Theorem:** Let the random variable X have mg.f. M(t). Then:

$$\boldsymbol{E[X^r] = M^{(r)}(0) \quad r = 1, 2, \dots}$$

Where $M^{(r)}(0)$ stands for $\frac{d^r M(t)}{dt^r}$ evaluated at t=0.

Proof:

Finding $E[X^r]$ using the m.g.f.

Theorem: Let the random variable X have m.g.f, M(t).

Then: $E[X^r] = M^{(r)}(0)$, r = 1, 2, ……

Where $M^{(r)} = \left(\frac{d}{dt}\right)^r M(t)$

(the r^{th} derivative of the mgf, with respect to t).

Note: Start by finding the first, second, and maybe the third derivative with respect to r, and see if you can pick up the pattern above.

For a Binomial distribution it can be shown that if X~Bin(n, p) then

$$M(t) = (pe^t + 1 - p)^n$$

This can then be used to prove:

$E(X) = np$ (shown in class) and

$Var(X) = np(1-p)$

Var(X) = E(X²) – [E(X)]²

E(X²) = $\boldsymbol{M''(0)}$

We found $\boldsymbol{M'(t) = [pe^t + 1 - p]^{n-1} * npe^t}$

$$\boldsymbol{M''(t) = [(n-1)(pe^t + 1 - p]^{n-2} * pe^t] * npe^t + [pe^t + 1 - p]^{n-1} * npe^t}$$

$$\boldsymbol{M''(0) = n(n-1)p^2 + np = (np)^2 - np^2 + np}$$

Var(X) = $\boldsymbol{(np)^2 - np^2 + np}$ **-** $\boldsymbol{(np)^2 = np - np^2 = np(1-p)}$

(as required)

You try Solution:

For a Poisson distribution it can be shown that if X~Pois(μ) then:

$$M(t) = M_X(t) = e^{-\mu} * e^{\mu e^t} = e^{\mu(e^t-1)}$$

$$M(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} f(x) = \sum_{x=0}^{\infty} e^{tx} \left(\frac{e^{-\mu}\mu^x}{x!}\right)$$

$$M(t) = e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu e^t)^x}{x!} = e^{-\mu} * e^{\mu e^t} = e^{\mu(e^t-1)}$$

$$M'(t) = \mu e^t e^{\mu(e^t-1)}$$

This can then be used to prove:

$$E(X) = \mu$$

$$E(X) = M'(0) = \mu e^0 e^{\mu(e^0-1)} = \mu$$

You try Solution (cont'd):

For a Poisson distribution it can be shown that if X~Pois(μ) then:

$M(t)$ = $M_X(t) = e^{-\mu} * e^{\mu e^t} = e^{\mu(e^t-1)}$

M(t) = $E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} f(x) = \sum_{x=0}^{\infty} e^{tx} \left(\frac{e^{-\mu}\mu^x}{x!}\right)$

M(t) = $e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu e^t)^x}{x!} = e^{-\mu} * e^{\mu e^t} = e^{\mu(e^t-1)}$

$M'(t) = \mu e^t e^{\mu(e^t-1)}$

$M''(t) = \mu e^t e^{\mu(e^t-1)} + \mu e^t [\mu e^t * e^{\mu(e^t-1)}]$

This can then be used to prove:

$Var(X) = \mu$

E(X²) = $M''(0) = \mu e^0 e^{\mu(e^0-1)} + \mu e^0 \left[\mu e^0 * e^{\mu(e^0-1)}\right] = \mu + \mu^2$

So, Var(X) = $\mu + \mu^2$ **-** μ^2 **=** μ

Example: Suppose we are given the following pf in table form:

x	0	1	2	3	4
f(x) = P(X = x)	0.1	0.2	0.2	0.3	0.2

Determine the mgf of X.

The mgf of X is given by $E[e^{tX}]$.

In this case, $E[e^{tX}] = \sum_{x=0}^{4} e^{tX} * f(x)$

$E[e^{tX}] = (0.1)e^{t(0)} + (0.2)e^{t(1)} + (0.2)e^{t(2)} + (0.3)e^{t(3)} + (0.2)e^{t(4)}$

So, mgf is given by:

$M_X(t) = E[e^{tX}] = 0.1 + 0.2e^{t} + 0.2e^{2t} + 0.3e^{3t} + 0.2e^{4t}$

$M'(t) = 0.2e^{t} + 0.4e^{2t} + 0.9e^{3t} + 0.8e^{4t}$

$M'(0) = 0.2 + 0.4 + 0.9 + 0.8$ = 2.3 (as required)

- Note: The m.g.f uniquely identifies a distribution in the sense that two different distributions CANNOT have the same m.g.f.
- For example: If you can show that a random variable X has m.g.f :

$$e^{2(e^t-1)}$$

- Then you can immediately tell that X~Poi(2).
- Hence m.g.fs are often used to identify a given distribution.

It follows that if two random variables have the same m.g.f. then they have the same distribution.

Where the m.g.f.'s must match for all values of t, not just at a few points.

The m.g.f. can also be used to determine that a sequence of distributions gets closer and closer to some limiting distribution.

We can use an mgf to determine the distribution of a sum of independent random variables.

Suppose Z = X + Y, where X and Y are independent random variables.

$$M_Z(t) = E[e^{tZ}] = E\left[e^{t(X+Y)}\right] = E[e^{t(X+Y)}]$$
$$M_Z(t) = E[e^{tX}]E[e^{tY}] = M_X(t) * M_Y(t)$$

You can see that we can easily determine the mgf of Z here, and could be able to identify the distribution of Z.

Example:

Let X ~ Bin(n, p) and let Y ~ Bin(m, p).

Let Z = X + Y.

Recall: From chapter 9, we saw that Z = X + Y in this case Z ~ Bin(n + m, p).
Let's check this using the mgf method.

$M_Z(t) = E[e^{tZ}] = E\left[e^{t(X+Y)}\right] = E[e^{t(X+Y)}]$
$M_Z(t) = E[e^{tX}]E[e^{tY}] = M_X(t) * M_Y(t)$
$M_X(t) = (pe^t + 1 - p)^n$ and $M_Y(t)$ = $(pe^t + 1 - p)^m$
$M_Z(t) = M_X(t) * M_Y(t) = \ [(pe^t + 1 - p)^n][(pe^t + 1 - p)^m]$
$M_Z(t) = (pe^t + 1 - p)^{n+m}$

We recognize that this is the mgf of a Binomial r.v. with n + m trials and probability of success equal to p.

Moment Generating Function of a Continuous Random Variable

Definition:

Consider a continuous random variable X with probability density function $f(x)$. The M.G.F. of X is defined as

$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

Again we assume that the m.g.f is defined and finite for values t in an interval around 0.

You try Solution:

Suppose $X \sim Exponential(\theta)$ show that the m.g.f is given by:

$$M(t) = \frac{1}{1 - \theta t}$$

For $t < \frac{1}{\theta}$

Remember to use the appropriate pdf here.

In this case, with X ~ Exponential (θ):

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, x > 0.$$

To start, we know that $\boldsymbol{E[e^{tX}] = \int_0^\infty e^{tx} \left(\frac{1}{\theta} e^{-x/\theta}\right) dx}$

You try Solution (cont'd):

Suppose $X \sim Exponential(\theta)$ show that the m.g.f is given by:

$$M(t) = \frac{1}{1 - \theta t}$$

For $t < \frac{1}{\theta}$

To start, we know that $E[e^{tX}] = \int_0^\infty e^{tx} \left(\frac{1}{\theta} e^{-x/\theta}\right) dx$

$M(t) = E[e^{tX}] = \frac{1}{\theta} \int_0^\infty e^{x(t - \frac{1}{\theta})} dx$

(for t < $\frac{1}{\theta}$**, this integral is convergent)**

$$M(t) = \frac{1}{\theta} \left[\frac{1}{t - \frac{1}{\theta}} e^{x(t - \frac{1}{\theta})} \right] \Big|_{x=0}^{x=\infty} = \left(\frac{1}{\theta}\right) \left(\frac{\theta}{\theta t - 1}\right) [0 - 1]$$

M(t) = $\frac{1}{1 - \theta t}$, $t < \frac{1}{\theta}$**, as required.**

You try Solution (cont'd):

Suppose $X \sim Exponential(\theta)$ show that the m.g.f is given by:

$$M(t) = \frac{1}{1 - \theta t}$$

For $t < \frac{1}{\theta}$

Let's use the mgf to verify that E(X) = θ and Var(X) = θ^2.

E(X) = $\boldsymbol{M'(0) = \frac{\theta}{(1-\theta(0))^2} = \theta}$

$\boldsymbol{E[X^2] = M''(0) = \frac{2\theta^2}{(1-\theta(0))^3}}$ **=** $\boldsymbol{2\theta^2}$

Var(X) = $\boldsymbol{E[X^2] - (E(X))^2 = 2\theta^2 - \theta^2 = \theta^2}$

(as required)