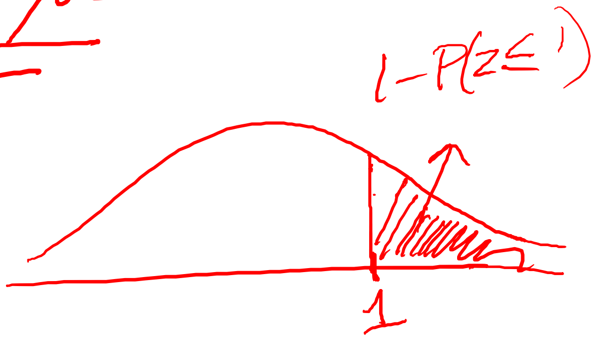
QUANTITATIVE TECHNIQUES FOR MANAGERIAL DECISION - 1 (QTMD1G21-1)

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



PROBLEM 5.21, TEXTBOOK PG. 213

Suppose in an apartment complex consisting of 45 families, the total drinking water usage for a day was 1350 lit. If the water usage per family is distributed according to normal distribution with a standard deviation 5 lit,

a) what percentage of families used more than 35 lit of water?

$\mu = E(X) = \frac{1350}{45} = 30.$

X = Water usage $\quad X \sim N(\mu, \sigma^2) \quad \sigma = 5,$

$$P(X>35) = P\left(\frac{X-30}{5} > \frac{35-30}{5}\right) = P(Z>1)$$

$$= 1 - P(Z \le 1)$$

$$= 1 - \text{pnorm}(1)$$

$$= 0.1586553 \rightarrow 15.86\%$$

$45 \to > 35$

PROBLEM 5.21, TEXTBOOK PG. 213

Suppose in an apartment complex consisting of 45 families, the total drinking water usage for a day was 1350 lit. If the water usage per family is distributed according to normal distribution with a standard deviation 5 lit,

b) what is the probability that exactly 5 families used more than 35 lit of water?

Y = # of families using more than 35 lit

$Y \sim Bin(n, p)$, $n = 45$, $p = 0.1587$

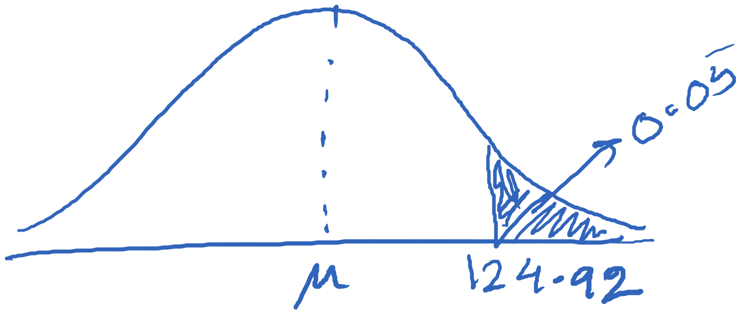
$$P(Y=5) = \binom{45}{5}(0.1587)^5(1-0.1587)^{40}$$

$$= \text{dbinom}(5, 45, 0.1587)$$

$$= 0.122435$$

PROBLEM

$$Z = \frac{X-\mu}{3}$$



A food processor packages instant coffee in small jars. The weights of the jars are normally distributed with a standard deviation of 3 grams.

If 5% of the jars weigh more than 124.92 grams, then what is the mean weight of the jars?

Note: $\Phi(1.645) = 0.95$.

X = Weights of jars

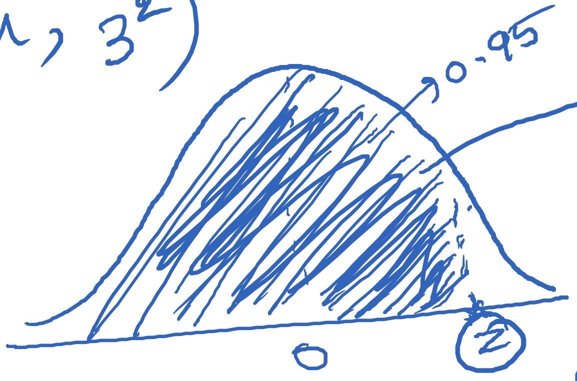
$X \sim N(\mu, 3^2)$

$$P(X > 124.92) = 0.05$$

$$\Rightarrow P(X \le 124.92) = 0.95$$

$$\Rightarrow P\left(Z \le \frac{124.92-\mu}{3}\right) = 0.95$$

$$\Rightarrow \Phi\left(\frac{124.92-\mu}{3}\right) = \Phi(\quad)$$



$= \text{qnorm}(0.95)$

$$z = \text{qnorm}(0.95) = 1.645$$

$$\Phi\left(\frac{124.92-\mu}{3}\right) = \Phi(1.645)$$

$$\frac{124.92-\mu}{3} = 1.645$$

$$\Rightarrow \mu = 119.985$$

STOCHASTIC PROCESSES

STOCHASTIC PROCESSES

A stochastic process is a sequence of random variables, finite or infinite, usually denoted X_1, X_2, … (Or sometimes X_0, X_1, X_2, …)

Examples:

1. Starting at some point of time, X_i is the amount withdrawn at an ATM in the campus by the i^{th} visitor.

2. Starting at some day, X_i is the daily closing price of a stock at the end of the i^{th} day

3. Starting at some point of time, X_i is the number of days (hours?) between the $(i-1)^{th}$ and i^{th} fatal accident in Jamshedpur

In short, any list of variables in a sequence, usually recorded in order of time or space.

THE SIMPLEST STOCHASTIC PROCESS

$Y_1 \sim$

$Y_2 \sim$

$Y_3 \sim$

$\vdots$

$Y_n \sim$

Simplest possible structure: All X_is are independent of each other, and have the same structure/ distribution.

→Independent and Identically Distributed variables (IID)

Examples:

1. The ATM withdrawals. X_i = amount withdrawn by the i^{th} individual
2. Number of tails between consecutive heads in a long sequence of coin tosses, where X_i = number of tails between $(i-1)^{th}$ and i^{th} head
3. No. of errors in each page of a book, where X_i = number of errors in Page i

SAMPLE MEAN

SAMPLE MEAN

$\bar{X} = f(x_1, \ldots, x_n)$

r.v. r.v.

For a simple random sample $X_1, X_2, \ldots, X_n$ from some population, the sample mean is given by

$$\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

→ average of random variables (r.v.)

Note: $X_1, X_2, \ldots, X_n$ are independent and identically distributed (IID) when chosen with replacement. This is the case we'll restrict to from now on.

EXPECTATION AND VARIANCE OF SAMPLE MEAN

$= E(\bar{X})$ $= V(\bar{X})$

$E(aX) = aE(X)$

$E(X+Y) = E(X) + E(Y)$

$V(aX) = a^2 V(X)$

$V(\underbrace{X+Y}_{\text{ind}}) = V(X) + V(Y)$

$X_1, X_2, \ldots, X_n$: IID

$E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$ (identical)

$V(X_1) = V(X_2) = \cdots = V(X_n) = \sigma^2$ (")

$$E(\bar{X}) = E\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right)$$
$$= \frac{1}{n} E(X_1 + X_2 + \cdots + X_n)$$
$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right]$$
$$= \frac{1}{n}(n\mu) = \mu$$

** $\boxed{E(\bar{X}) = \mu}$

$$\mathrm{Var}(\bar{X}) = V\left(\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right)$$
$$= \frac{1}{n^2} V(X_1 + \cdots + X_n)$$
$$= \frac{1}{n^2}\left[V(X_1) + V(X_2) + \cdots + V(X_n)\right]$$
$$\therefore = \frac{1}{n^2}(n\sigma^2)$$
$$= \frac{\sigma^2}{n}$$

$\boxed{\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}}$

$X_1, X_2, \ldots, X_{20} \rightarrow$ IID

$N(296, 8^2) \rightarrow$ Ind

DISTRIBUTION OF SAMPLE MEAN

The distribution of $\overline{X}$ is given by the probability distribution of the values $\overline{X}$ can take.

Example: The quantity of soft drinks put in a soft drink bottle is supposed to be 300ml, but suppose actually it follows a normal distribution with mean 296 ml and sd 8 ml. What is the probability that randomly chosen 20 bottles contain 300ml or more on average?

X_i = Content of ith bottle

$X_1 \sim N(296, 8^2)$

$X_2 \sim N(296, 8^2)$

$\vdots$

$X_{20} \sim N(296, 8^2)$

For this, we need to answer:

What is the distribution of the average quantity in 20 randomly chosen bottles?

$P(\overline{X} > 300) =$

$$\boxed{E(\overline{X}) = \mu = 296, \quad V(\overline{X}) = \frac{8^2}{20} = \frac{\sigma^2}{n}}$$

True for any IID r.v.

$X_1, X_2, \ldots, X_n \sim IID$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

Case I

$X_1, X_2, \ldots, X_n \sim IID\ N(\mu, \sigma^2)$

Problem

$P(\bar{X} > 300) = ?$

$X_1, X_2, \ldots, X_{20} \sim IID\ N\left(296, \frac{8^2}{20}\right) \to \sigma^2$

pnorm(x, μ, σ)

$\sigma^2 = \frac{8^2}{20} \Rightarrow \sigma = \sqrt{\frac{64}{20}}$

$P(\bar{X} > 300) = 1 - P(\bar{X} \leq 300)$

$= 1 -$ pnorm (300, 296, sqrt (64/20))

$= 0.126$

A RESULT

If $X_1, X_2, \ldots, X_n$ is a sequence of independent $N(\mu,\sigma^2)$ (i.e., IID $N(\mu,\sigma^2)$) random variables, then

$E(X_1+X_2+\cdots+X_n) = n\mu$

$$X_1 + \ldots + X_n = \sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$$

$V(X_1+X_2+\cdots+X_n) = n\sigma^2$

$V(\bar{X})$

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

$E(\bar{X})$

$E(X_1+X_2+\cdots+X_n) = n\mu$

$\Rightarrow E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \frac{n\mu}{n}$

$\Rightarrow E(\bar{X}) = \mu$

$V(X_1+\cdots+X_n) = n\sigma^2$

$\therefore V\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{n\sigma^2}{n^2} \Rightarrow V(\bar{X}) = \frac{\sigma^2}{n}$

WHAT IF THE RANDOM VARIABLES ARE NOT NORMAL?

Sample Mean in General Case:

Consider a sequence of n independent and identically distributed (IID) random variables: $X_1, X_2, \ldots, X_n$

Let all of them be IID with mean μ and variance σ^2. (But not necessarily normal.)

Then $\boxed{E(\overline{X}) = \mu,\ V(\overline{X}) = \frac{\sigma^2}{n}}$

But how is $\overline{X}$ "distributed"?

DISTRIBUTION OF SAMPLE MEAN IN GENERAL CASE

There are three lunch specials in a restaurant: A, B and C, which cost Rs.100, Rs. 140 and Rs. 150 respectively. A student, who lunches in that restaurant every day, chooses these three specials with probabilities 60%, 20% and 20% respectively. He chooses one lunch special every day independently of his previous decisions.

a) Obtain the distribution of the student's daily expenditure on lunch.

X = Daily expenditure

1 day!

x	100	140	150
$P(X=x)$	0·6	0·2	0·2

DISTRIBUTION OF SAMPLE MEAN IN GENERAL CASE

There are three lunch specials in a restaurant: A, B and C, which cost Rs.100, Rs. 140 and Rs. 150 respectively. A student, who lunches in that restaurant every day, chooses these three specials with probabilities 60%, 20% and 20% respectively. He chooses one lunch special every day independently of his previous decisions.

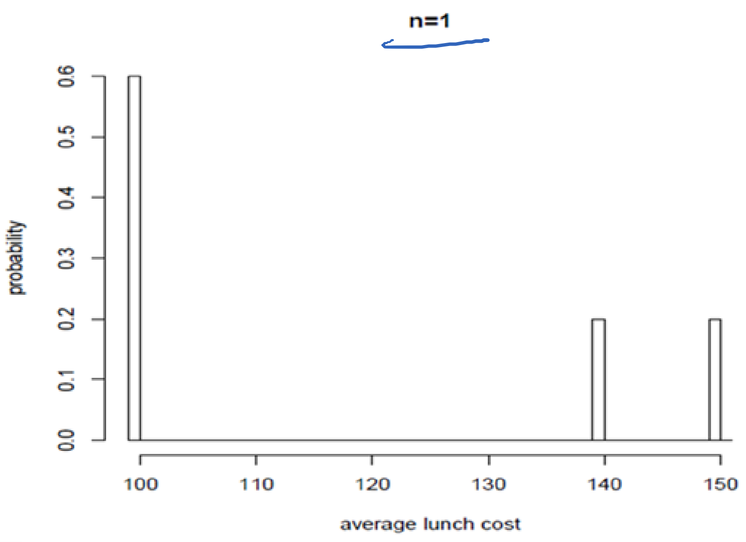
b) Obtain the distribution of the student's average expenditure on lunch over two days.

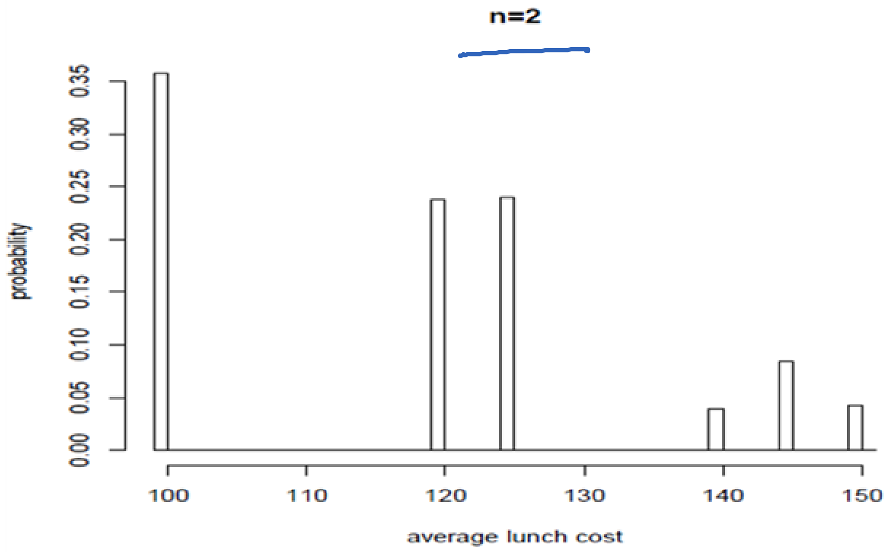
	(100,100)	(100,140) (140,100)	(100,150) (150,100)	(140,140)	(140,150),(150,140)	(150,150)
$\bar{X}$	100	120	125	140	145	150
$P(\bar{X}=x)$	0.6×0.6 $=0.36$	$0.6\times0.2\times2$ $=0.24$	$0.6\times0.2\times2$ $=0.24$	0.04	0.08	0.04

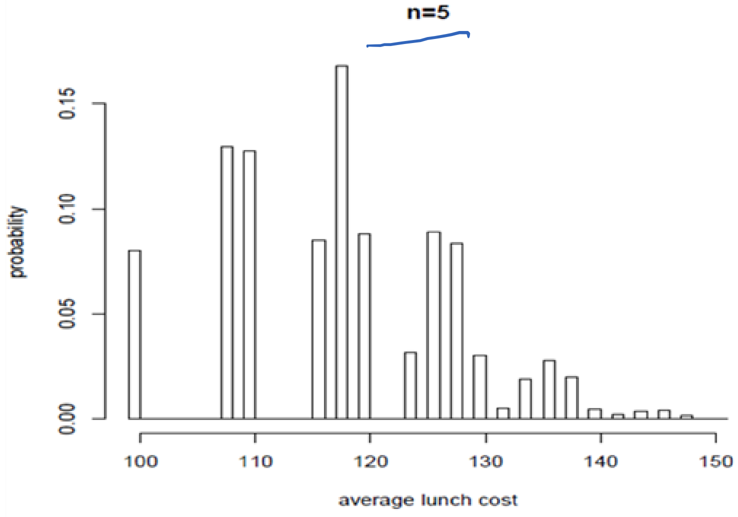
DISTRIBUTION OF SAMPLE MEAN IN GENERAL CASE

There are three lunch specials in a restaurant: A, B and C, which cost Rs.100, Rs. 140 and Rs. 150 respectively. A student, who lunches in that restaurant every day, chooses these three specials with probabilities 60%, 20% and 20% respectively. He chooses one lunch special every day independently of his previous decisions.

c) Obtain the distribution of the student's average expenditure on lunch over 30 days.

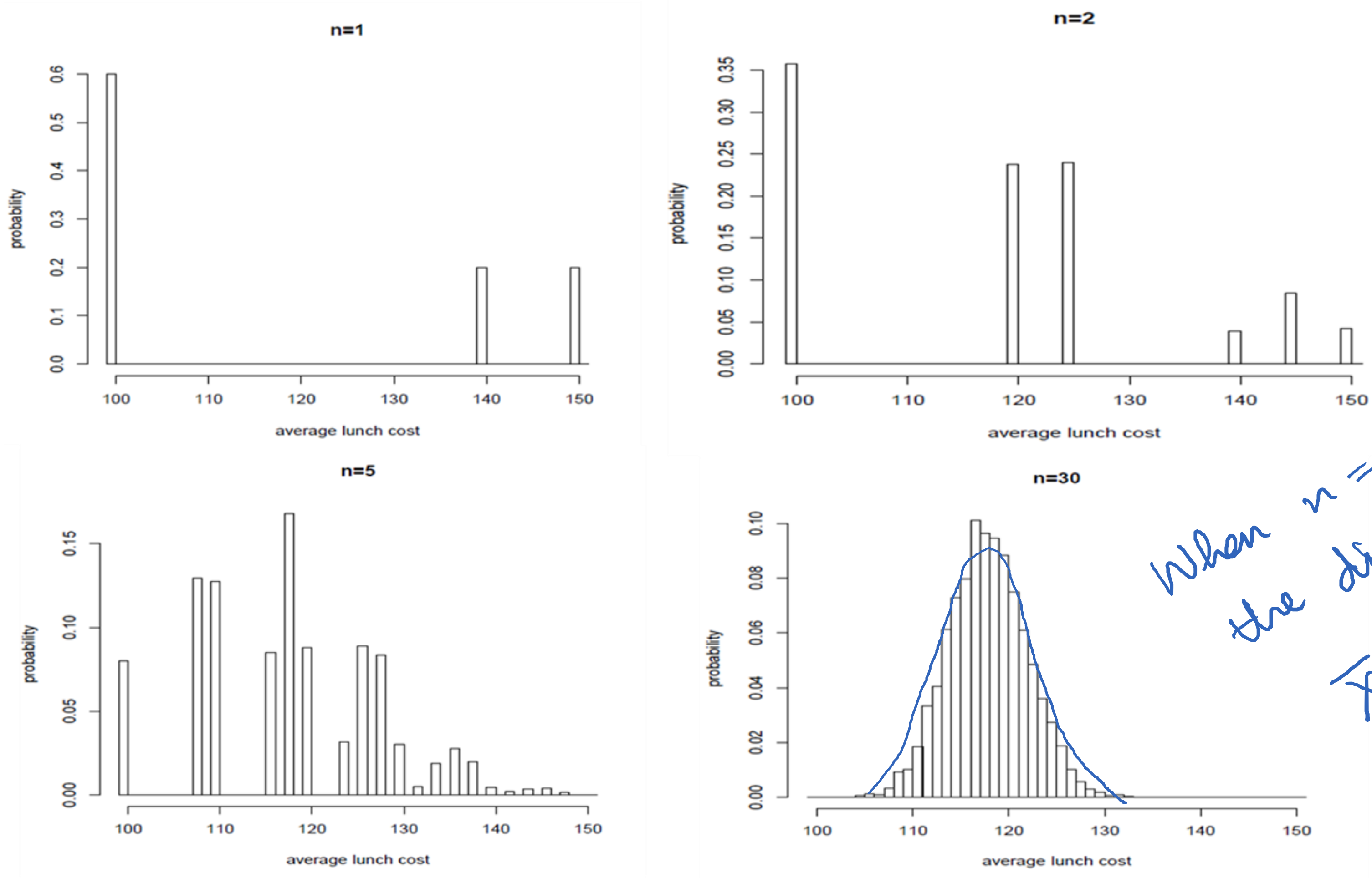






n=30?

DISTRIBUTION OF $\overline{X}$

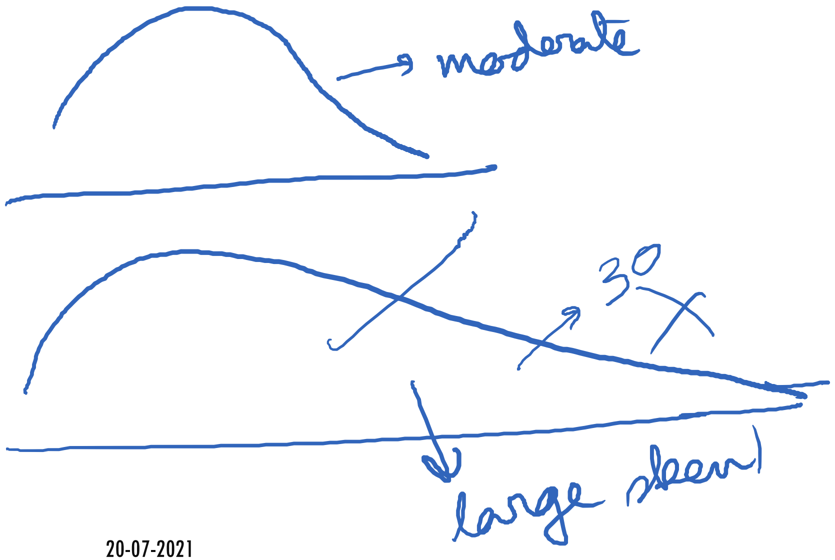


When $n = 30$, the distribution of $\bar{X}$ is approximately normal!

DISTRIBUTION OF $\bar{X}$

* Works if the original distribution has little, moderate or no skew!

THE CENTRAL LIMIT THEOREM



$X_1, X_2, \ldots X_n$: IID sample with mean μ and variance σ^2

For large n, $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$ approximately.

OR

For large* sample (n) size, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ approximately

Typical benchmark for "large": $n \geq 30$
(works in most cases, but not all)

BACK TO OUR EXAMPLE

X	100	140	150
$P(X=x)$	0.6	0.2	0.2

$E(X) = 118$

μ and σ:

$\mu = 118$, $\sigma^2 = 496$ (check), $\sigma = 22.27$.

Now work out the approximate distribution of the sample mean $\bar{X}$ for $n = 30$ using CLT:

$$\bar{X} \sim N(\overset{\mu}{118}, \overset{=\frac{\sigma^2}{n}}{496/30}) = N(118, 16.53) = N(118, 4.07^2)$$

$$\bar{X} \sim N\left(118, \frac{496}{30}\right)$$

BACK TO OUR EXAMPLE

$$\bar{X} \sim N\left(118, \frac{496}{30}\right) \qquad \sqrt{} = 4.07$$

What is the probability that over 30 days, *the average spend is at least Rs.122?* $= \bar{x}$

$$P(\bar{X} \geq 122) = 1 - P(\bar{X} < 122) = 1 - \text{pnorm}(122, 118, 4.07)$$
$$= 0.1628$$

What is the probability that over 30 days, *total spend is at least Rs. 4000?* $\rightarrow \sum_{i=1}^{30} X_i$

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

$$P\left(\sum_{i=1}^{30} X_i > 4000\right) = 1 - \text{pnorm}(4000, 30 \times 118, \text{sqrt}(30 \times 496))$$

For values x ≥ 3.5, can assume Φ(x) ≈ 1.

$$= 0.00008$$

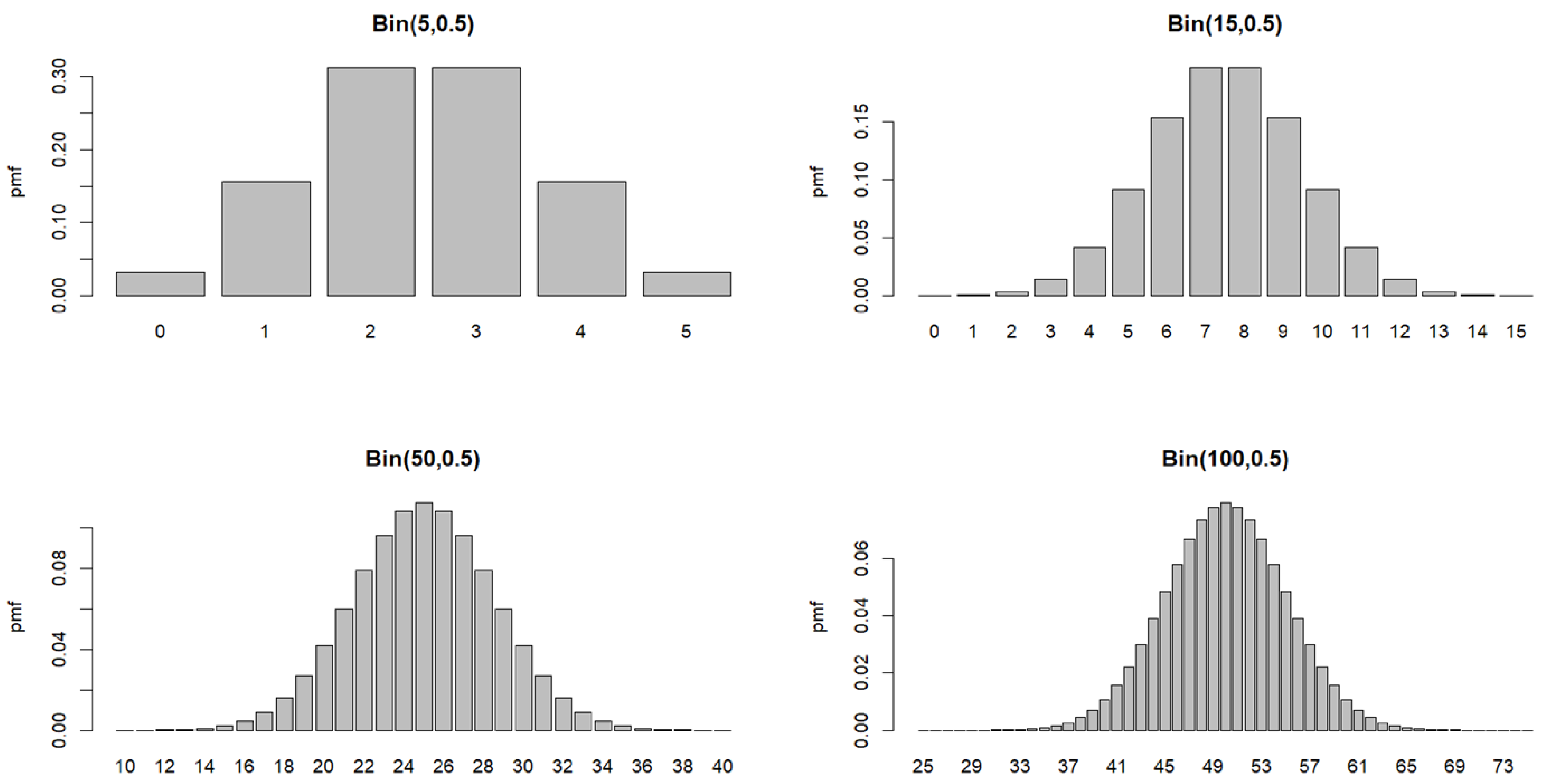
ISSUES WITH THE CLT

When can we use it:

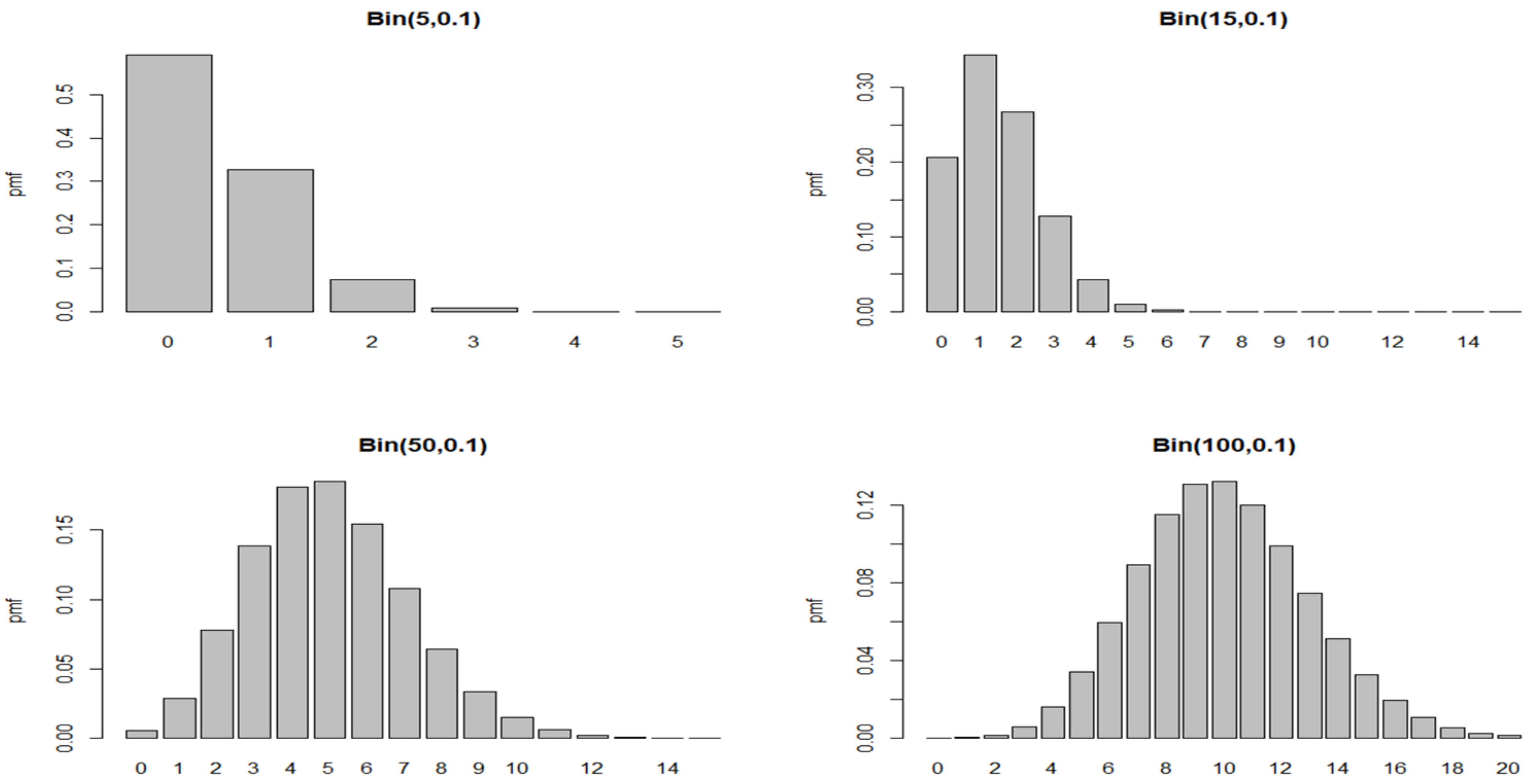
- If samples are from IID distributions

When we may not use it:

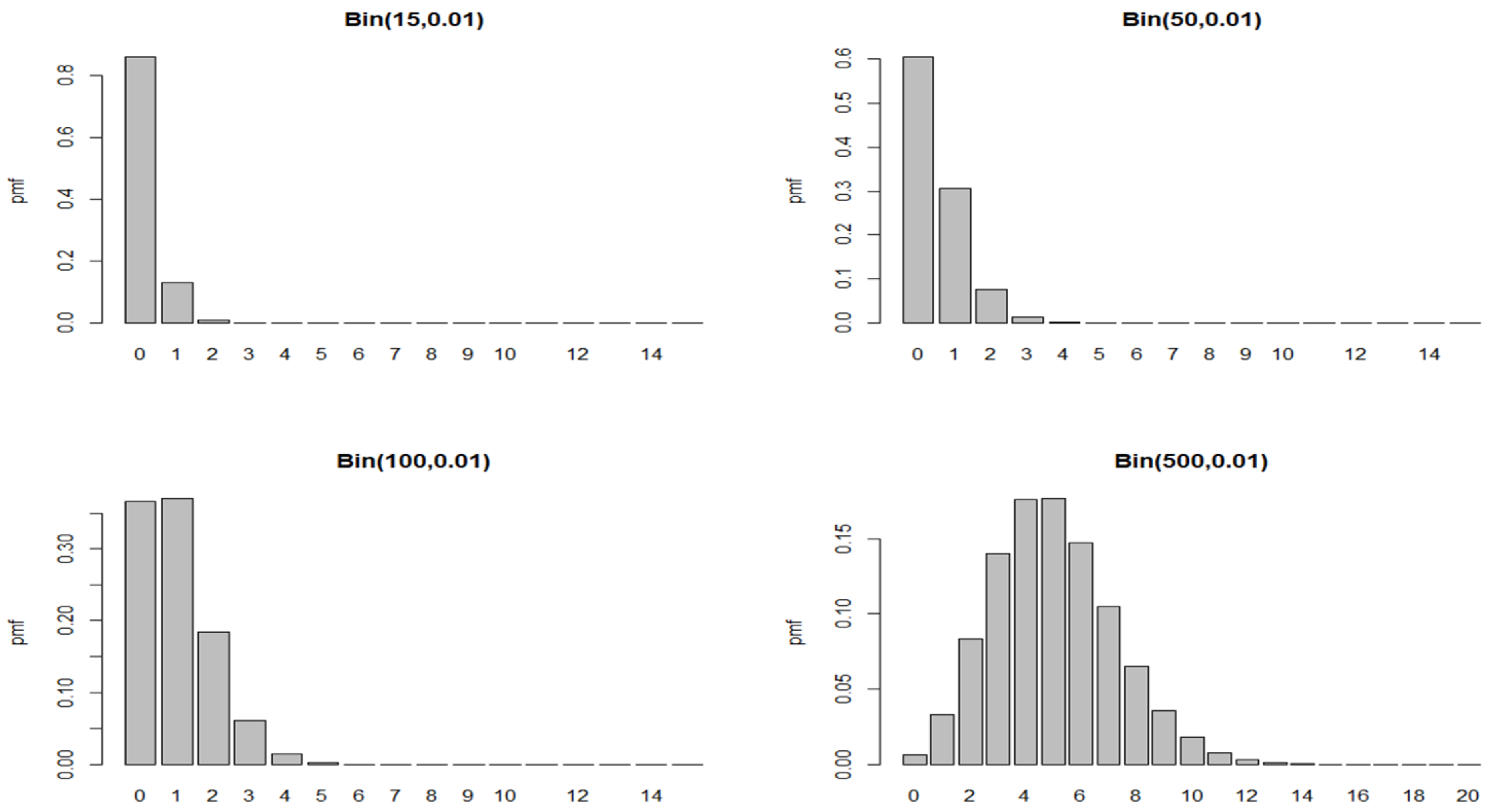
- Don’t use it if the distributions are not IID.
- Errors may be large for small samples from skewed distributions



Example: Binomial(n,p) with **p = 0.5**, and various n.



Example: Binomial(n,p) with **p = 0.1**, and various n



Example: Binomial(n,p) with **p = 0.01**, and various n

HOW GOOD ARE THE NORMAL APPROXIMATIONS?

Depends!

- Larger skew: need larger sample size

→No skew: even 15 is a reasonable sample size to use the normal approximation

→Moderate skew: need 30 or more

→High skew: need 50 or more

→Severe skew (Example: binomial with large n and small p so that Poisson approximation holds): might need very high sample size, in the range of several hundred or even higher