Lecture 26:
Weak law of large numbers & Central limit theorem
Part II

Satyajit Thakor
IIT Mandi

Example

- Example (application of Chebyshev inequality): The time taken to serve a customer at a fast-food restaurant has a mean of 75.0 seconds and a standard deviation of 7.3 seconds. Use Chebyshev inequality to calculate the time interval that has 75% probability of containing a particular service time.

μ σ at least

- Chebyshev inequality: $P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}$.
- Let X be the time taken to serve a customer.
- Find a time interval that has 75% prob. of containing a particular service time.
- Note that, $P(|X-\mu| \geq a) = 1 - P(|X-\mu| < a)$

$$\Rightarrow \quad 1 - P(|X-\mu| < a) \leq \frac{\sigma^2}{a^2}$$

$$\Rightarrow \quad P(|X-\mu| < a) \geq 1 - \frac{\sigma^2}{a^2}$$

Example

- To find the desired interval, let

$$1 - \frac{\sigma^2}{a^2} = 0.75 \implies a^2 = \frac{(7.3)^2}{.25} = 213.16$$

$$\implies a = 14.6$$

- Hence,

$$P(|X - \mu| < 14.6) \geq 0.75$$

$$\implies P(-14.6 < X - 75 < 14.6) \geq 0.75$$

$$\implies P(60.4 < X < 89.6) \geq 0.75.$$

- Thus, for X in the interval (60.4, 89.6) the probability is at least 75%. (by chebyshev inequality)

Central limit theorem

- Consider a sequence $X_1, \ldots, X_n$ of independent identically distributed random variables. Suppose that $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$ for $i = 1, \ldots, n$.
- If we define

  average r.v.: $$\bar{X} = \frac{X_1 + \ldots + X_n}{n}$$

  then $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

  Proved in Lecture 25.
- The central limit theorem states that regardless of the actual distribution of the individual random variables X_i, the distribution of their average $\bar{X}$ is closely approximated by a $\mathcal{N}(\mu, \sigma^2/n)$ distribution.
- That is, average of a set of independent identically distributed random variables is always approximately normally distributed.
- The accuracy of the approximation improves as n increases.

Central limit theorem (CLT)

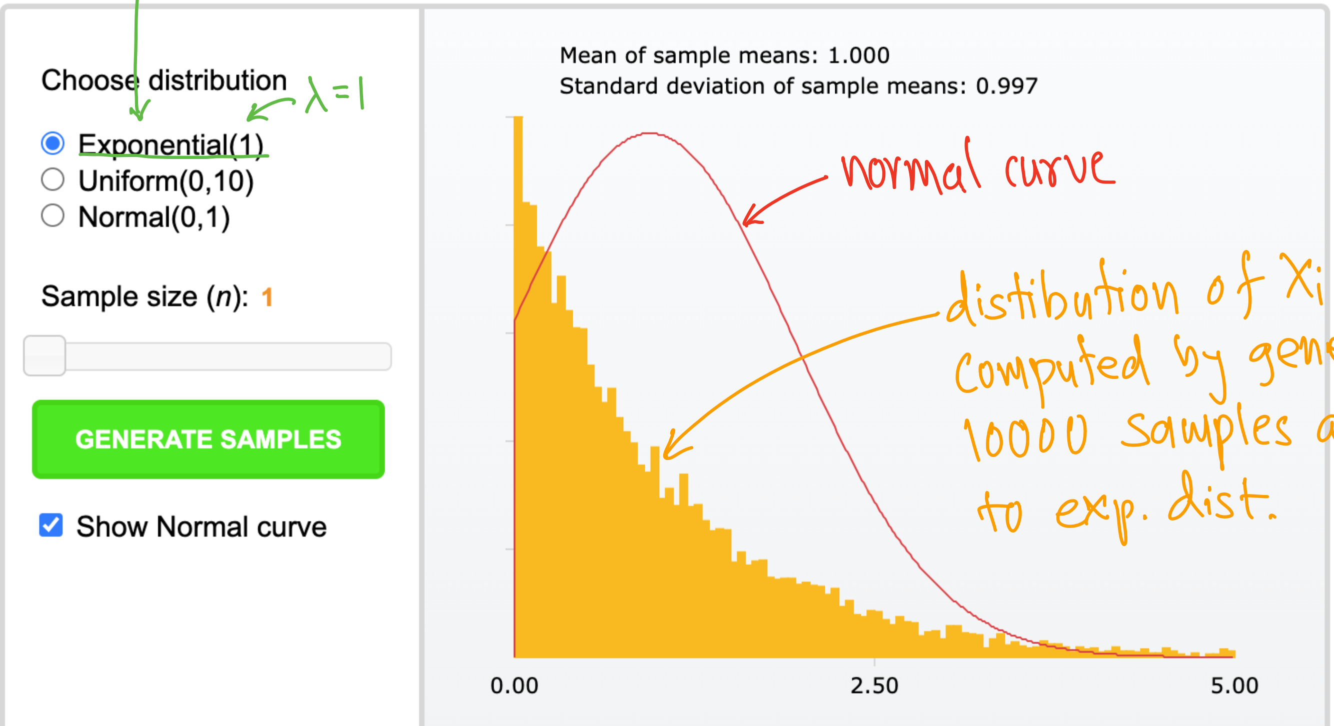
- If $X_1, \ldots, X_n$ is a sequence of independent identically distributed random variables with a mean μ and a variance σ^2, then the distribution of their average $\bar{X}$ can be approximated by
$$\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$
- This is equivalent to: If $X_1, \ldots, X_n$ is a sequence of independent identically distributed random variables with a mean μ and a variance σ^2, then the distribution of the sum $X_1 + \ldots + X_n$ can be approximated by
$$\mathcal{N}(n\mu, n\sigma^2).$$
- The proof is beyond the scope of this course. We will study the statement with examples.

theorem statement

Central limit theorem

$f_X(x) = \lambda e^{-\lambda x}, x \geq 0.$ $\mu_X = \frac{1}{\lambda}$, $\sigma^2 = \frac{1}{\lambda^2}$.

- Numeral exponential distribution: X_i



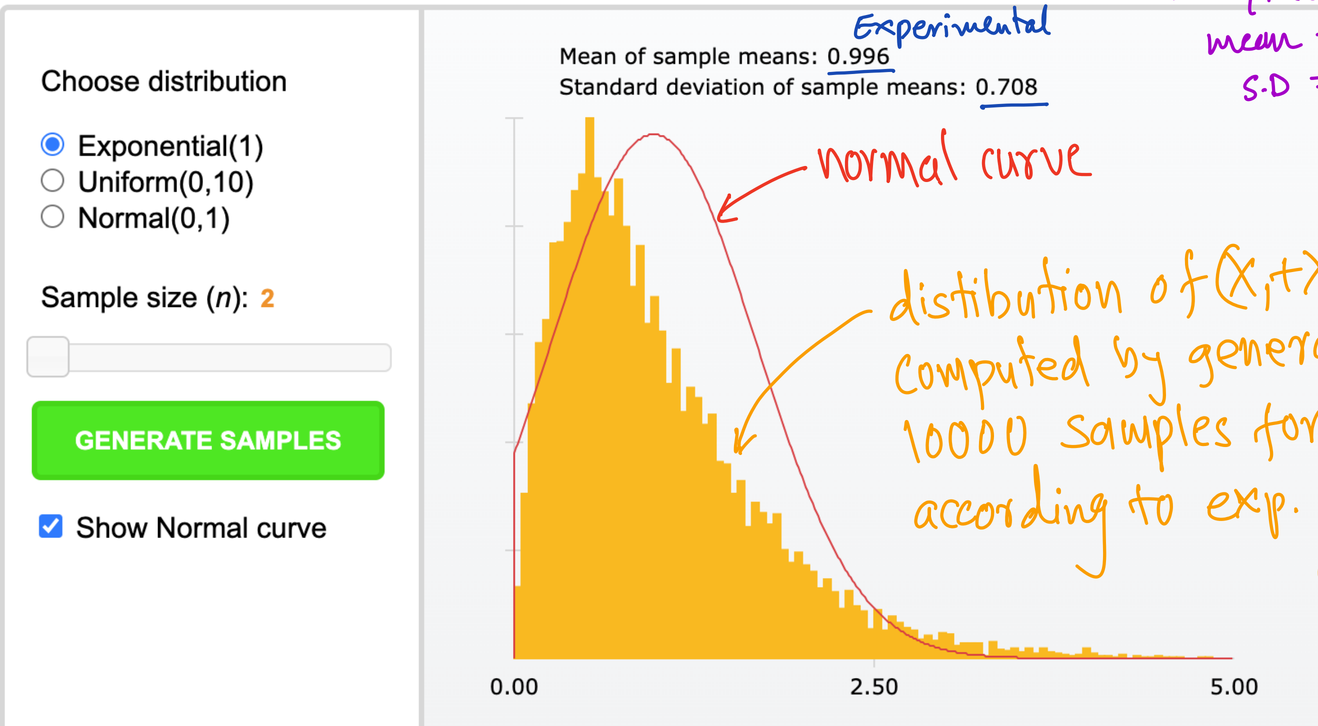
Source: `https://digitalfirst.bfwpub.com/stats_applet/stats_applet_3_cltmean.html`

Central limit theorem

- Numeral exponential distribution: $(X_1 + X_2)/2$

Analytical:
mean $= \mu_X$
S.D $= \sigma_X/\sqrt{n}$

Experimental

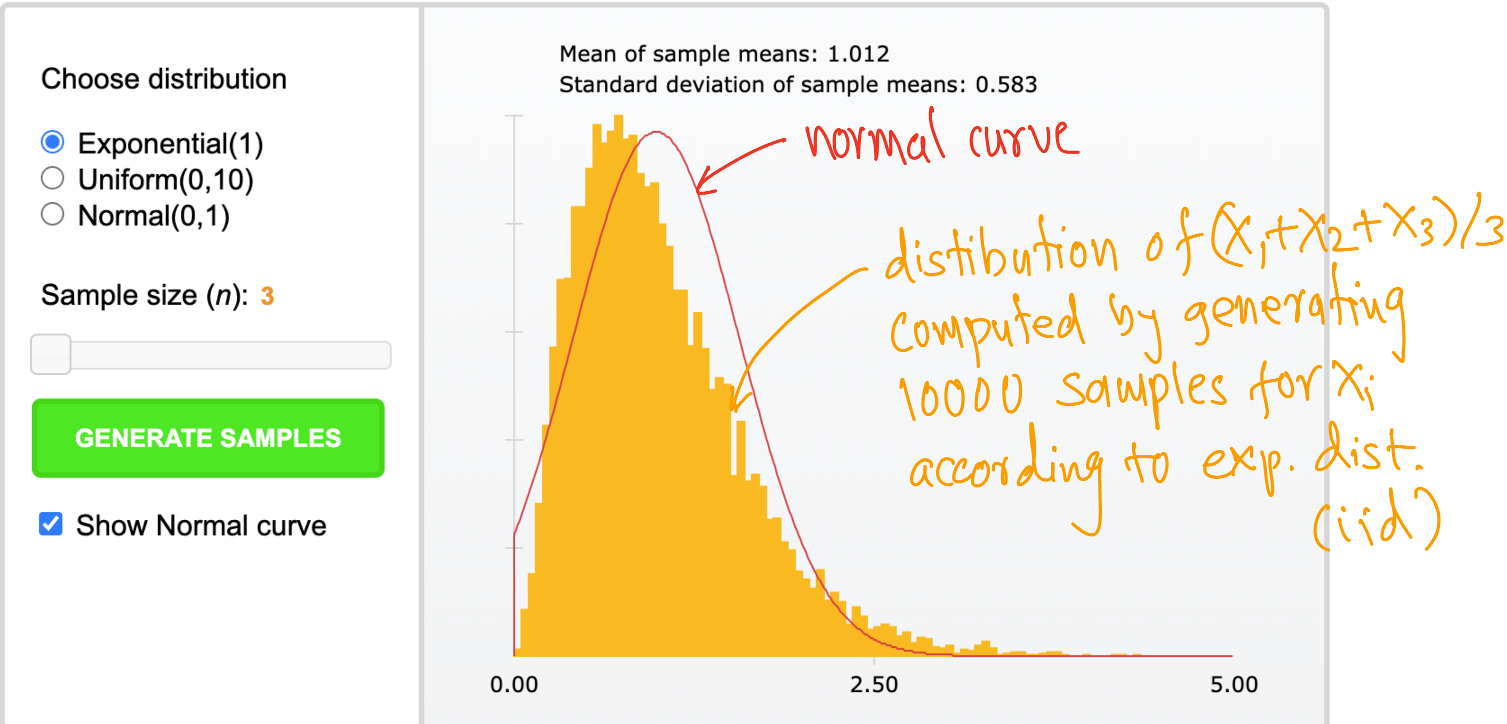


normal curve

distibution of $(X_1+X_2)/2$ computed by generating 10000 samples for X_i according to exp. dist. (iid)

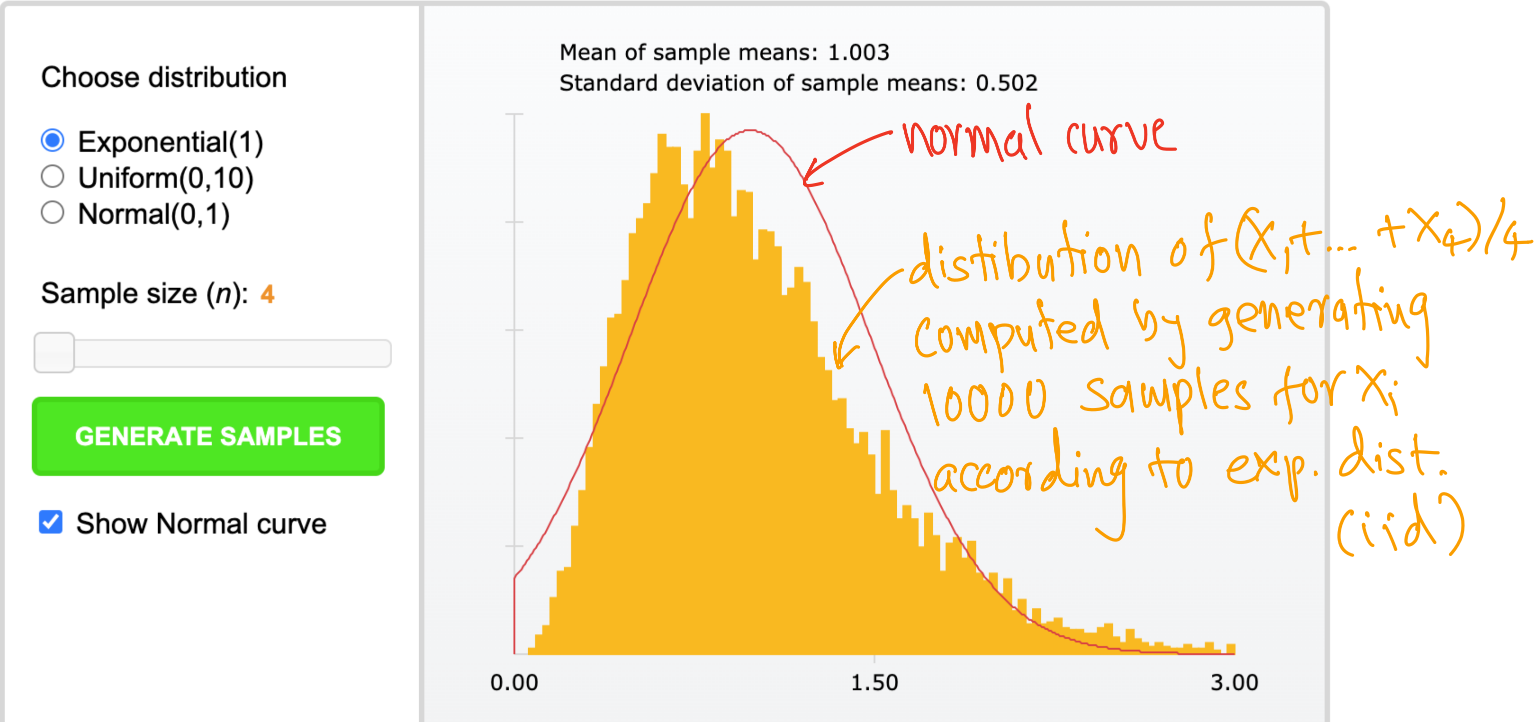
Central limit theorem

- Numeral exponential distribution: $(X_1 + X_2 + X_3)/3$



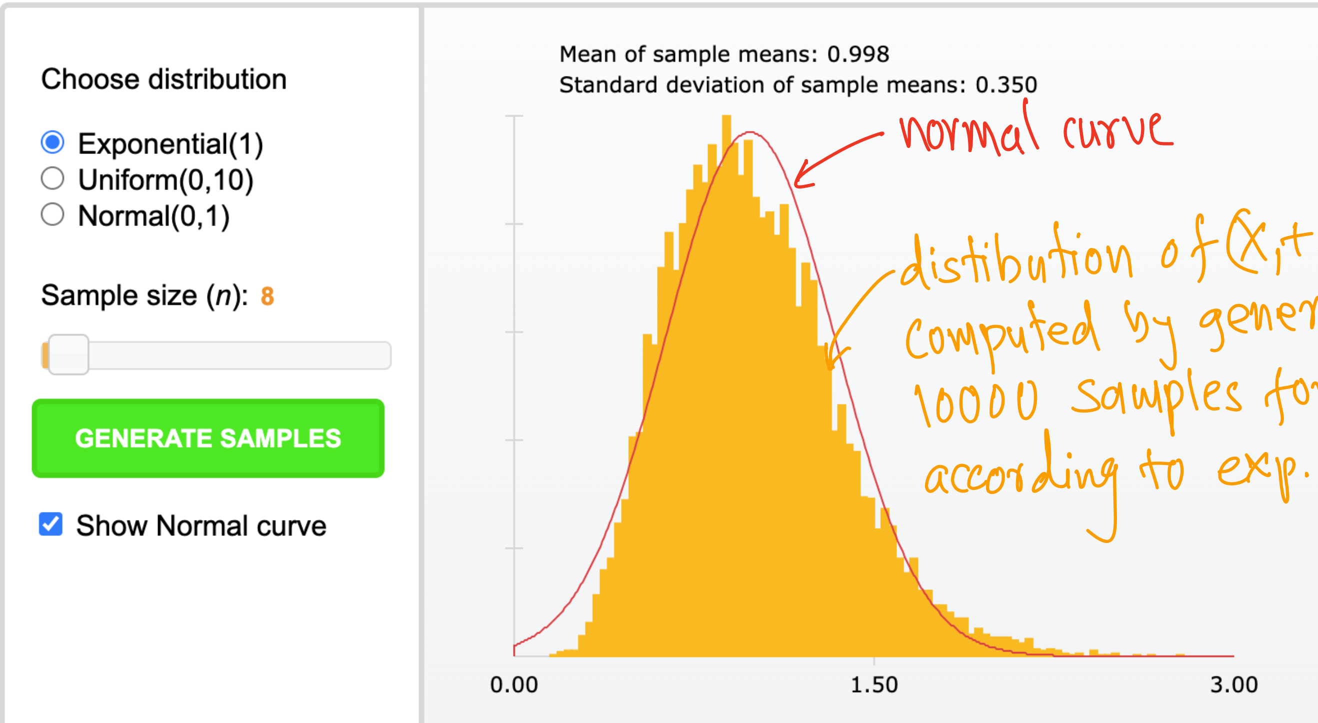
Central limit theorem

- Numeral exponential distribution: $(X_1 + \ldots + X_4)/4$



Central limit theorem

- Numeral exponential distribution: $(X_1 + \ldots + X_8)/8$

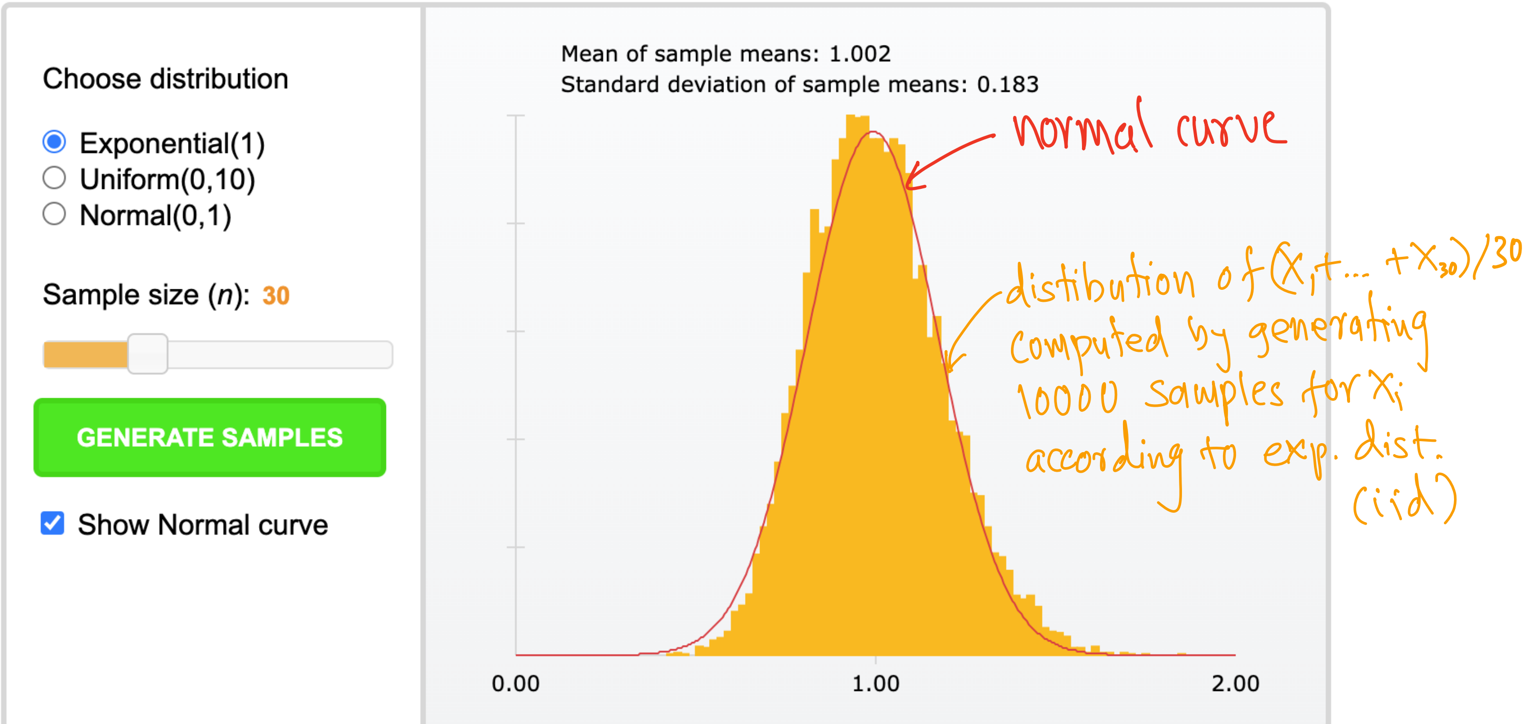


normal curve

distibution of $(X_1 + \ldots + X_8)/8$ computed by generating 10000 samples for X_i according to exp. dist. (iid)

Central limit theorem

- Numeral exponential distribution: $(X_1 + \ldots + X_{30})/30$



Example: CLT

- Recall Problem 2, Assignment 6: Milk containers have label printed "2 liters". But, the PDF of the amount of milk deposited in a milk container by a dairy factory is

$$f_X(x) = \begin{cases} 40.976 - 16x - 30e^{-x}, & 1.95 \le x \le 2.20; \\ 0, & \text{elsewhere.} \end{cases}$$

  (a) Is f_X a valid PDF?
  (b) What is the probability that a container produced by the dairy factory is underweight?
- Recall that the probability of "a container produced by the dairy factory is underweight" is .261 (solution of 2(b)).

Example: CLT

- Example: What is the distribution of the number of underweight containers X in a box of 20 containers? Find the
  (a) exact and
  (b) approximate (using CLT) value
  of the probability "a box contains no more than three underweight containers".

$\underbrace{\text{A container is underweight}}$ with prob. .261.

denote as r.v. W : this is a Bernoulli r.v. with $p = 0.261$.

$W = 1$ (underweight) with prob. p

$W = 0$ (not underweight) with prob. $1-p$.

$\underbrace{\text{No. of underweight containers in a box of 20.}}$

denote as r.v. X : this is a $\underbrace{\text{Binomial r.v.}}$ $B(\underset{n}{20}, \underset{p}{.261})$

Recall: sum of iid Bernoulli r.v.s is Binomial.

Example: CLT

(a) $P(X \le 3) = \sum_{K=0}^{3} \binom{20}{K} (.261)^K (1-.261)^{20-K} = \underbrace{0.1934}_{\text{exact}}$

(b) Note that a Binomial r.v. with the distribution $B(n,p)$ is a sum of n i.i.d Bernoulli r.v.s. Hence, by the CLT, we can approximate:

$$X \sim B(n,p) \approx \mathcal{N}(n\mu_W, n\sigma_W^2)$$

$$\Rightarrow X \sim B(20, .261) \approx \mathcal{N}(20 \times .261, 20 \times .261 \times (1-.261))$$

$$\approx \mathcal{N}(5.22, 3.86) \sim Y = \mu + \sigma Z$$

$$\Rightarrow P(X \le 3) = P(X \le 3.5) \approx P(Y \le 3.5)$$

$$= P(\mu + \sigma Z \le 3.5)$$

$$= P(5.22 + \sqrt{3.86}\, Z \le 3.5)$$

$$= P(Z \le -.8754) = \underbrace{.1922}_{\text{Very close to the exact value.}}$$