## SESSION 3

# RANDOM DISTRIBUTIONS

# DISCRETE VS. CONTINUOUS DISTRIBUTIONS

- **Random Variable --** a variable which contains the outcomes of a chance experiment

- **Discrete Random Variable --** the set of all possible values is at most a finite or a countably infinite number of possible values

  - Number of new subscribers to a magazine
  - Number of bad checks received by a restaurant
  - Number of absent employees on a given day

- **Continuous Random Variable --** takes on values at every point over a given interval

  - Elapsed time between arrivals of bank customers
  - Percent of the labor force that is unemployed

# CUMULATIVE DISTRIBUTION FUNCTION

Def. The cumulative distribution function (or cdf) $F(.)$ of the random variable $X$ for any real number $b$ in the set of possible values is defined as

$$\boxed{F(b) = P(X \le b)}$$

*Properties of cumulative distribution function*

1. $F(b)$ is a nondecreasing function of $b$.
2. $F(\infty) = 1, F(-\infty) = 0$
3. $\boxed{P(a < X \le b) = F(b) - F(a)}$ for all $a < b$

# DISCRETE RANDOM VARIABLES

**Requirements for a Discrete Probability Function**

- Probabilities are between 0 and 1, inclusively

$$\boxed{0 \le P(X) \le 1} \text{ for all } X$$

- Total of all probabilities equals

$$\boxed{\sum_{\text{over all } X} P(X) = 1}$$

# DISCRETE RANDOM VARIABLES

## *Examples*

- Let $X$ denote the random variable that is defined as the sum of two fair dice; then

  $P(X = 2) = 1/36$; $P(X = 3) = 2/36$; $P(X = 4) = 3/36$; $P(X = 5) = 4/36$;
  $P(X = 6) = 5/36$; $P(X = 7) = 6/36$; $P(X = 8) = 5/36$; $P(X = 9) = 4/36$;
  $P(X = 10) = 3/36$; $P(X = 11) = 2/36$; $P(X = 12) = 1/36$

- Tossing a coin having the probability $p$ of landing head until the first head appears. Let $N$ denote the number of flips required; then

  $P(N = 1) = p$
  $P(N = 2) = (1 - p)p$
  $\dots$
  $P(N = n) = (1 - p)^{n-1} p$

  Note that $\sum_{n=1}^{\infty} P(N = n) = 1$

# DISCRETE RANDOM VARIABLES

## Probability Mass Function

Def. The probability mass function of a discrete random variable $X$, denoted by $P(.)$, is defined as:

$$\boxed{P(a) = P(X = a)}$$

- The cumulative distribution function $F(.)$ can be expressed as:

$$\boxed{F(a) = \sum_{\text{All } x_i \le a} P(x_i)}$$

*Example* ❖ Let $X$ have a probability mass function given by $P(1) = 1/2$; $P(2) = 1/3$; $P(3) = 1/6$ then

$$F(a) = \begin{cases} 0 & a < 1 \\ 1/2 & 1 \le a < 2 \\ 5/6 & 2 \le a < 3 \\ 1 & a \ge 3 \end{cases}$$

# DISCRETE RANDOM VARIABLES

**Mean (Expectation) of a Discrete Distribution**

$$\boxed{\mu = E(X) = \sum_{\text{All } x_i} x_i P(x_i)}$$

*Example*

| $X$ | $P(X)$ | $X\,P(X)$ |
|---|---|---|
| -1 | 0.1 | -0.1 |
| 0 | 0.2 | 0.0 |
| 1 | 0.4 | 0.4 |
| 2 | 0.2 | 0.2 |
| 3 | 0.1 | 0.3 |

**Mean** = 1.0

# DISCRETE RANDOM VARIABLES

## Variance and Standard Deviation of a Discrete Distribution

$$\boxed{\sigma^2 = \sum_{\text{All } x_i} (x_i - \mu)^2 P(x_i)}$$

$$\boxed{\sigma = \sqrt{\sigma^2}}$$

*Example*

| $X$ | $P(X)$ | $X-\mu$ | $(X-\mu)^2$ | $(X-\mu)^2 P(X)$ |
|---|---|---|---|---|
| -1 | 0.1 | -2 | 4 | 0.4 |
| 0 | 0.2 | -1 | 1 | 0.2 |
| 1 | 0.4 | 0 | 0 | 0.0 |
| 2 | 0.2 | 1 | 1 | 0.2 |
| 3 | 0.1 | 2 | 4 | 0.2 |

**Variance** = 1.2

# DISCRETE RANDOM VARIABLES

**Binomial Distribution**

- Experiment involves *n* identical trials
- Each trial has exactly two possible outcomes: success and failure
- Each trial is independent of the previous trials
  - $p$ is the probability of a success on any one trial
  - $q = (1\text{-}p)$ is the probability of a failure on any one trial
  - p and q are constant throughout the experiment
  - $X$ is the number of successes in the $n$ trials

***X is said to be a binomial random variable***

# DISCRETE RANDOM VARIABLES

## Binomial Distribution

- Applications
  - Sampling with replacement
  - Sampling without replacement -- n < 5% N
- Probability function

$$\boxed{P(X=i) = C_i^n p^i q^{n-i}}$$

where

$$\boxed{C_i^n = \frac{n!}{i!(n-i)!}}$$

# DISCRETE RANDOM VARIABLES

**Binomial Distribution**

- Mean value

$$\boxed{\mu = np}$$

- Variance and standard deviation

$$\boxed{\sigma^2 = npq} \qquad \boxed{\sigma = \sqrt{npq}}$$

# DISCRETE RANDOM VARIABLES

## Binomial Distribution

*Examples*

- Four fair coin are flipped. What is the probability that two heads and two tails are obtained.

Letting $X$ equal the number of heads (successes) that appear, then $X$ is a binomial random variable with parameter (4, 1/2). Therefore,

$$P(X=2)=C_2^4\left(\frac{1}{2}\right)^2\left(1-\frac{1}{2}\right)^{4-2}=\frac{3}{8}$$

# DISCRETE RANDOM VARIABLES

**Binomial Distribution**

*Examples*

- If all items produced by a machine will be defective with probability 0.1, independently of each other. What is the probability that in a sample of three items, at most one will be defective?

If $X$ is the number of defective items in the sample then $X$ is a binomial random variable with parameters (3, 0.1). Hence, the desired probability is given by

$$P(X=0)+P(X=1)=C_0^3(0.1)^0(0.9)^3+C_1^3(0.1)^1(0.9)^2=0.972$$

# DISCRETE RANDOM VARIABLES

## Poisson Distribution

- Describes discrete occurrences over a continuum or interval
- A discrete distribution
- Describes rare events
- Each occurrence is independent any other occurrences
- The number of occurrences in each interval can vary from zero to infinity
- The expected number of occurrences must hold constant throughout the experiment.

# DISCRETE RANDOM VARIABLES

**Poisson Distribution**

Probability Function

$$P(X=i)=e^{-\lambda}\frac{\lambda^i}{i!}$$

$\lambda$: long-run average
$e$: the base of natural logarithms
($e = 2.718282\ldots$)

**Applications**

- **Arrivals at queueing systems**
  - airports -- *people, airplanes, automobiles, baggage*
  - banks -- *people, automobiles, loan applications*
  - computer file servers -- *read and write operations*
- **Defects in manufactured goods**
  - number of defects per 1,000 feet of extruded copper wire
  - number of blemishes per square foot of painted surface
  - number of errors per typed page

# DISCRETE RANDOM VARIABLES

**Poisson Distribution**

- Mean value

$$\boxed{\mu = \lambda}$$

- Variance and standard deviation

$$\boxed{\sigma^2 = \lambda} \qquad \boxed{\sigma = \sqrt{\lambda}}$$

# DISCRETE RANDOM VARIABLES

## Poisson Distribution

*Examples*

- Suppose that the number of typographical errors on a single page of a book has a Poisson distribution with parameter $\lambda = 1$. Calculate the probability that there is at least one error on a certain page.

Let $X$ denote the number of errors on a certain page of the book

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1}\frac{1^0}{0!} = 1 - e^{-1} = 1 - \frac{1}{e} = 0.633$$

- If the number of accidents occurring on a highway each day is a Poisson random variable with parameter $\lambda = 3$, what is the probability that no accidents occur today?

$$P(X = 0) = e^{-3}\frac{3^0}{0!} = e^{-3} = \frac{1}{e^3} = 0.05$$

# DISCRETE RANDOM VARIABLES

## Poisson Distribution

### Poisson Approximation of the Binomial Distribution

- Binomial probabilities are difficult to calculate when $n$ is large.
- Under certain conditions binomial probabilities may be approximated by Poisson probabilities
- Poisson approximation

$$\boxed{\text{If } n > 20 \text{ and } np \leq 7 \quad \Rightarrow \quad \text{Use } \lambda = np}$$

# CONTINUOUS RANDOM VARIABLES

## Probability Density Function

Def. The probability density function of a discrete random variable $X$, denoted by $f(.)$, is a function such that:

$$\boxed{P(X \in B) = \int_B f(x)dx}$$

From the above definition, it is noted that:

$$\boxed{P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx = 1}$$

# CONTINUOUS RANDOM VARIABLES

**Probability Density Function**

*Notes:*

- $F(a) = P(X \le a) = \int_{-\infty}^{a} f(x)dx$

  $\Rightarrow F(-\infty) = 0, F(\infty) = 1$

- $P(a \le X \le b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$

- $P(X = a) = \int_{a}^{a} f(x)dx = 0$

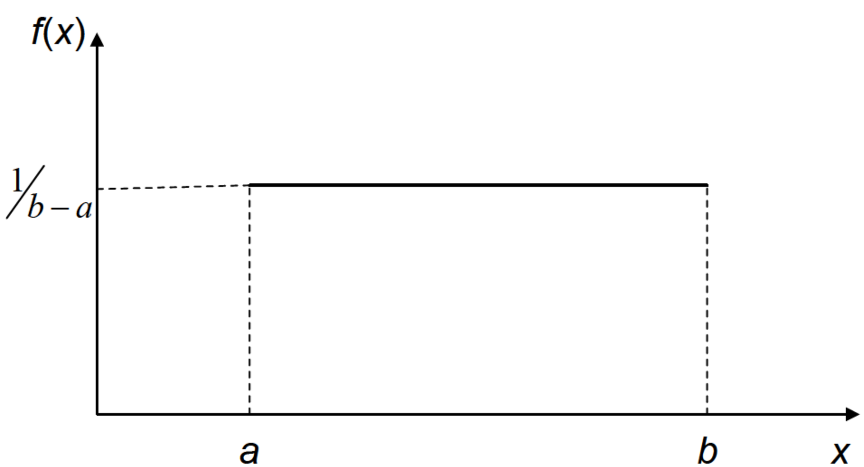
- $P\left(a - \varepsilon/2 \le X \le a + \varepsilon/2\right) = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)dx = \varepsilon f(a)$

# CONTINUOUS RANDOM VARIABLES

## Uniform Distribution

Def. A random variable $X$ is said to be uniformly distributed over the interval $(a, b)$ if its probability density function is given by

$$\boxed{f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}} \qquad (b > a)$$



# CONTINUOUS RANDOM VARIABLES

## Uniform Distribution

Cumulative Distribution Function

$$\boxed{F(x)=\begin{cases} 0 & if & x\le a \\ \dfrac{x-a}{b-a} & if & a<x<b \\ 1 & if & x\ge b \end{cases}}$$

- Mean: $\boxed{\mu=\dfrac{a+b}{2}}$
- Standard Deviation: $\sigma=\dfrac{b-a}{\sqrt{12}}$

# CONTINUOUS RANDOM VARIABLES

## Normal Distribution

<u>Def.</u> A random variable $X$, with parameter $\mu$ and $\sigma^2$, whose probability density function is defined by:

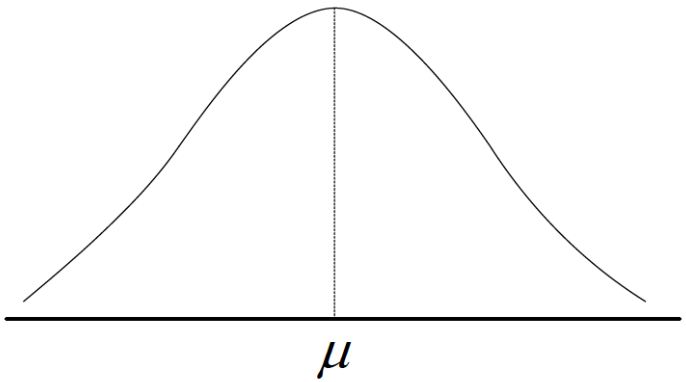
$$\boxed{f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \qquad (-\infty < x < \infty)$$

where: $\mu$ : Mean of $X$.

$\sigma$ : Standard Deviation of $X$.

$\pi = 3.14159$

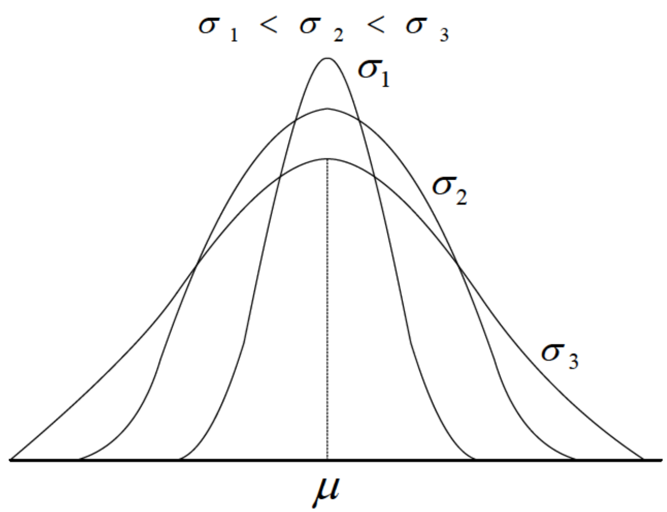
$e = 2.718282$



# CONTINUOUS RANDOM VARIABLES

## Normal Distribution

**Normal Curves for Different Standard Deviations**




- The cumulative distribution function of a normal distribution does not have an explicit expression.

- The cumulative distribution function of a normal distribution is given in tabular form for the case $\mu = 0, \sigma = 1$ (which is called the *standardized normal distribution*)

# CONTINUOUS RANDOM VARIABLES

## Normal Distribution

**Standardized Normal Distribution**

- A normal distribution with *a mean of zero, and a standard deviation of one*
- $Z$ Formula *standardizes any normal distribution*
- $Z$ Score
  - computed by the $Z$ Formula
  - the number of standard deviations which a value is away from the mean

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

**Use of Normal Table**

*Follow instructions in class*

# CONTINUOUS RANDOM VARIABLES

## Normal Distribution

### Normal Approximation of the Binomial Distribution

- The normal distribution can be used to approximate binomial probabilities

- Procedure

    1. Convert binomial parameters to normal parameters
    2. Does the interval $\mu \pm 3\sigma$ lie between 0 and n? If so, continue; otherwise, do not use the normal approximation.
    3. Correct for continuity
    4. Solve the normal distribution problem

# CONTINUOUS RANDOM VARIABLES

## Normal Distribution

### Normal Approximation of the Binomial Distribution

*Conversion equations*

$$\boxed{\mu = np} \qquad \boxed{\sigma = \sqrt{npq}}$$

*Correcting for Continuity*

| Value being determined | Correction |
|---|---|
| $X >$ | +.50 |
| $X \geq$ | -.50 |
| $X <$ | -.50 |
| $X \leq$ | +.05 |
| $\leq X \leq$ | -.50 and +.50 |
| $< X <$ | +.50 and -.50 |

# CONTINUOUS RANDOM VARIABLES

## Normal Distribution

### Normal Approximation of the Binomial Distribution

*Example*

The binomial probability $P\left(X \geq 25 \middle| n = 60 \text{ and } p = .30\right)$

- Convert parameters: $\mu = np = 18$, $\sigma = \sqrt{npq} = 3.55$
- Check: $\mu \pm 3\sigma = (7.35, 28.65) \in (0, 60)$
- Approximation: $P\left(X \geq 24.5 \middle| \mu = 18 \text{ and } \sigma = 3.55\right)$

# CONTINUOUS RANDOM VARIABLES

## Exponential Distribution

- Continuous
- Family of distributions
- Skewed to the right
- $X$ varies from 0 to infinity
- Steadily decreases as $X$ gets larger

Probability function

$$\boxed{f(x) = \begin{cases} \lambda e^{-\lambda x} & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases}}$$

# CONTINUOUS RANDOM VARIABLES

## Exponential Distribution

Cumulative Distribution Function

$$\boxed{F(x) = P(X \le x) = \int_0^x f(x)\,dx = 1 - e^{-\lambda x}} \quad (x > 0)$$

*Example*

- $P(X \ge 2 \mid \lambda = 1.2) = 1 - P(X < 2 \mid \lambda = 1.2) = e^{-\lambda x} = e^{-1.2*2} = .0907$

# CONTINUOUS RANDOM VARIABLES

## Gamma Distribution

Def. A random variable $X$ whose probability density function is defined by:

$$\boxed{f(x)=\begin{cases}\dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & \text{if } x\geq 0\\ 0 & \text{if } x<0\end{cases}}$$

$$(\lambda,\alpha>0)$$

is said to be a Gamma distribution with parameters $\alpha,\lambda$.

The Gamma function: $\Gamma(\alpha)=\int_0^\infty e^{-x}x^{\alpha-1}dx$

When $\alpha$ is an integer, say $\alpha=n$, the Gamma distribution becomes the Erlang distribution and it is easy to show, by induction, that

$$\Gamma(n)=(n-1)!$$

# CONTINUOUS RANDOM VARIABLES

## Chi-square ($\chi^2$) Distribution

Def. If $z_1, z_2, \ldots, z_k$ are normally and independently $N(0,1)$ and

$$x = z_1^2 + z_2^2 + \ldots + z_k^2$$

then $x$ follows chi-square distribution with $k$ degrees of freedom.

$$\boxed{f(x) = \frac{1}{2^{k/2}\,\Gamma\left(\frac{k}{2}\right)} x^{k/2-1} e^{-x/2}} \qquad (x > 0)$$

Mean: $\mu = k$

Variance: $\sigma^2 = 2k$

# CONTINUOUS RANDOM VARIABLES

**Chi-square ($\chi^2$) Distribution**

*<u>Example</u>:* If $y_1, y_2, \ldots, y_n$ is a random sample from an $N\left(\mu, \sigma^2\right)$ then

$$\frac{SS}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}{\sigma^2} \sim \chi^2_{n-1}$$

# CONTINUOUS RANDOM VARIABLES

## Student ($t$) Distribution

Def. If $z$ and $\chi_k^2$ are independent standard normal and chi-square random variables then the random variable

$$t_k = \frac{z}{\sqrt{\chi_k^2 / k}}$$

follows the $t$ distribution with $k$ degrees of freedom

$$\boxed{f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \frac{1}{\left[ t^2/k + 1 \right]^{(k+1)/2}}} \qquad (-\infty < t < +\infty)$$

Mean: $\mu = 0$

Variance: $\sigma^2 = k/_{k-2}$ for $k > 2$

- If $k \to \infty$: the $t$ distribution becomes the standard normal distribution

# CONTINUOUS RANDOM VARIABLES

## Fisher ($F$) Distribution

Def. If $\chi_u^2$ and $\chi_v^2$ are independent chi-square random variables then

$$F_{u,v} = \frac{\chi_u^2 / u}{\chi_v^2 / v}$$

follows the $F$ distribution with $u$ numerator degrees of freedom and $v$ denominator degrees of freedom

$$\boxed{f(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{u/2} x^{u/2-1}}{\Gamma\left(\frac{u}{2}\right)\Gamma\left(\frac{v}{2}\right)\left[\left(\frac{u}{v}\right)x+1\right]^{\frac{u+v}{2}}}} \qquad (0 < x < \infty)$$

# CONTINUOUS RANDOM VARIABLES

## Fisher ($F$) Distribution

*Example*: If $y_{11}, y_{12}, \ldots, y_{1n_1}$ and $y_{21}, y_{22}, \ldots, y_{2n_2}$ are random samples from two independent normal populations with common variance $\sigma^2$ then

$$\frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}$$

where $S_1^2, S_2^2$ are the two sample variances.

# FUNCTION OF RANDOM VARIABLES

*Def.*: A function defined on values of a random variable

**Expectation of a Function of a Random Variable**

In general,

$$\boxed{E\left[g(X)\right]=\sum_{x} g(x)\,p(x)}$$

$$\boxed{E\left[g(X)\right]=\int_{x} g(x)\,f(x)\,dx}$$

# FUNCTION OF RANDOM VARIABLES

*Example*

1. Suppose $X$ has the following probability mass function

$$p(0)=0.2 \qquad p(1)=0.5 \qquad p(2)=0.3$$

Then: $$E\left[X^2\right]=\sum_x x^2 p(x)=0*0.2+1*0.5+4*0.3=1.7$$

2. Let $X$ be uniformly distributed over $(0,1)$

Then $$E\left[X^3\right]=\int_0^1 x^3 f(x)dx=\int_0^1 x^3 dx=\frac{1}{4}$$

# PROPERTIES OF EXPECTATION & VARIANCE

1. $E(c) = c$
2. $E(cy) = cE(y) = c\mu$
3. $V(c) = 0$
4. $V(cy) = c^2V(y) = c^2\sigma^2$
5. $E(y_1 + y_2) = E(y_1) + E(y_2)$
6. $Cov(y_1, y_2) = E\left[(y_1 - \mu_1)(y_2 - \mu_2)\right]$
7. $V(y_1 \pm y_2) = V(y_1) + V(y_2) \pm 2Cov(y_1, y_2)$
8. If $y_1, y_2$ are independent:

$$Cov(y_1, y_2) = 0$$
$$V(y_1 \pm y_2) = V(y_1) + V(y_2)$$
$$E(y_1.y_2) = E(y_1)E(y_2)$$

# MARKOV'S INEQUALTY

If $X$ is a random variable that takes only *nonnegative* values, then for any value $a > 0$

$$\boxed{P\{X \geq a\} \leq \frac{E[X]}{a}}$$

# CHEBYSHEV'S INEQUALTY

If $X$ is a random variable with mean $\mu$ and variance $\sigma^2$, then, for any value $K > 0$

$$\boxed{P\left(|X-\mu| \geq K\right) \leq \frac{\sigma^2}{K^2}}$$

Note that for $k > 1$:

$$\boxed{P\left(\mu - k\sigma < X < \mu + k\sigma\right) \geq 1 - \frac{1}{k^2}} \qquad (K = k\sigma)$$

# EXAMPLE

Suppose that the number of items produced in a factory during a week is a random variable with mean 500.

a. What can be said about the probability that this week's production will be at least 1000?

$$P\{X \geq 1000\} \leq \frac{E[X]}{1000} = 0.5$$

b. If the variance of a week's production is 1000, then what can be said about the probability that this week's production will be between 400 and 600?

$$P(|X-500| \geq 100) \leq \frac{\sigma^2}{100^2} = 0.1 \Rightarrow P(400 \leq X \leq 600) \geq 0.9$$

# THE STRONG LAW OF LARGE NUMBER

Let $X_1, X_2, \ldots$ be independent random variables having a common distribution with mean $\mu$. Then, with probability 1

$$\boxed{\lim_{n\to\infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu}$$

# CENTRAL LIMIT THEOREM

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with mean $\mu$ and variance $\sigma^2$. Then the distribution of

$$\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$, i.e.,

$$P\left(\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx$$