Lecture 10

The Normal Distributions

Readings: IntroStat Chapter 4; OpenIntro Chapter 3

STAT 8010 Statistical Methods I

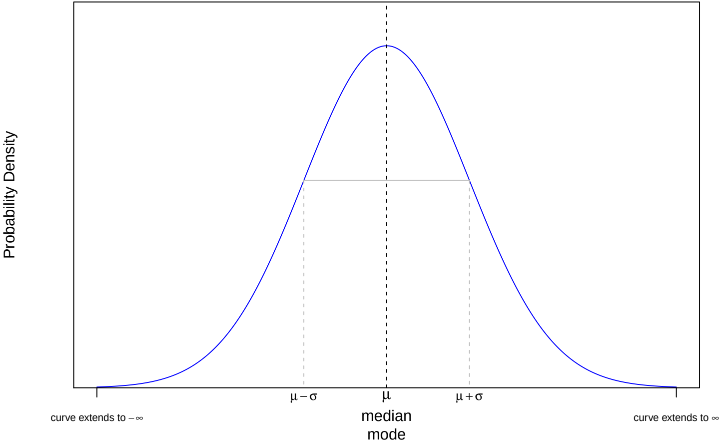
May 30, 2023

Whitney Huang
Clemson University

Agenda

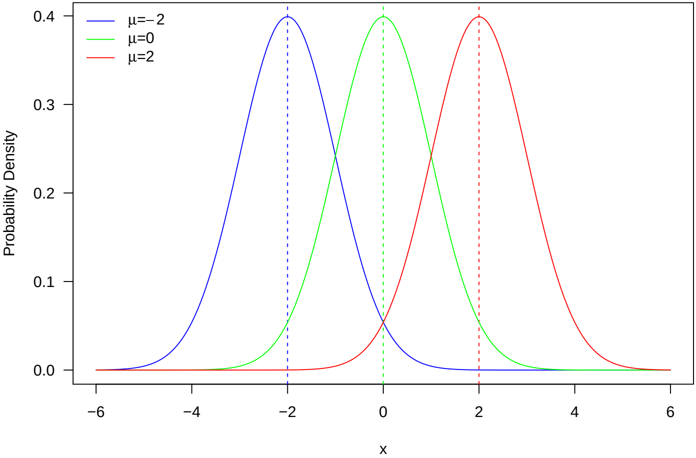
1. **Normal Distributions**
2. **Sums of Normal Random Variables**
3. **Normal approximation of Binomial Distribution**

Probability Density Curve for Normal Random Variable



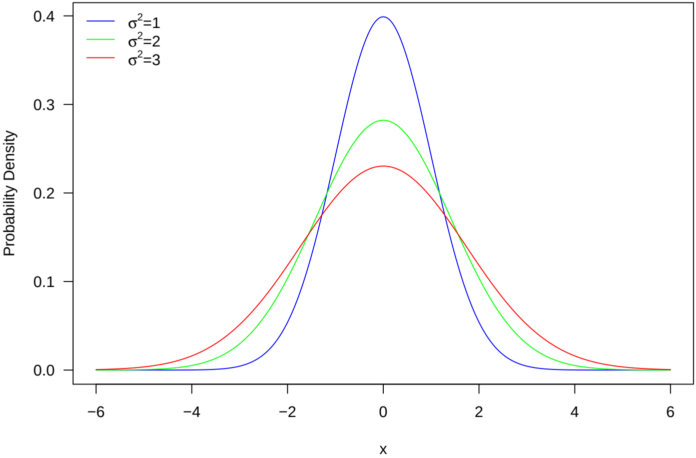
Normal Density Curves

Different μ but same σ^2

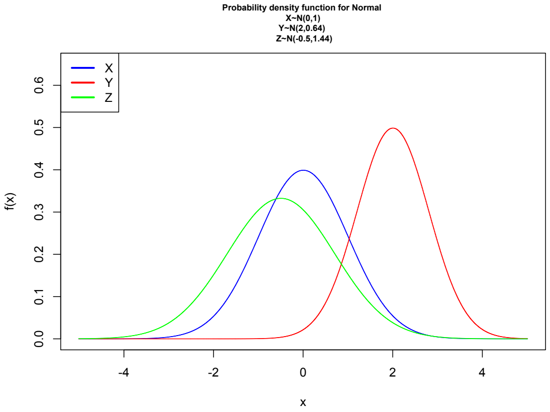


Normal Density Curves Cont'd

Same μ but different σ^2



Normal Density Curves



- The parameter μ determines the center of the distribution
- The parameter σ^2 determines the spread of the distribution
- Also called bell-shaped distribution

Characteristics of Normal Random Variables

Let X be a Normal r.v.

- The support for X: $(-\infty, \infty)$
- Parameters: μ : mean and σ^2 : variance
- The probability density function (pdf): $\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ for $-\infty < x < \infty$
- The cumulative distribution function (cdf): No explicit form, look at the value $\Phi(\frac{x-\mu}{\sigma})$ for $-\infty < x < \infty$ from **standard normal table**
- The expected value: $\mathrm{E}[X] = \mu$
- The variance: $\mathrm{Var}(X) = \sigma^2$

Standard Normal $Z \sim N(\mu = 0, \sigma^2 = 1)$

- Normal random variable X with mean μ and standard deviation σ can be converted to standard normal Z by the following :

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

Standard Normal $Z \sim N(\mu = 0, \sigma^2 = 1)$

- Normal random variable X with mean μ and standard deviation σ can be converted to standard normal Z by the following :

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

- The cdf of the standard normal, denoted by $\Phi(z)$, can be found from the **standard normal table**

Standard Normal $Z \sim N(\mu = 0, \sigma^2 = 1)$

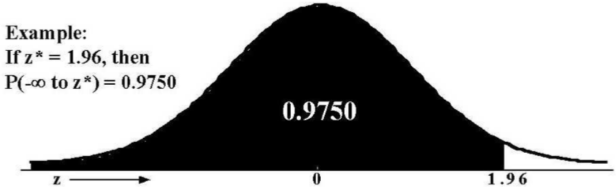
- Normal random variable X with mean μ and standard deviation σ can be converted to standard normal Z by the following :

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

- The cdf of the standard normal, denoted by $\Phi(z)$, can be found from the **standard normal table**
- The probability $\mathrm{P}(a \le X \le b)$ where $X \sim N(\mu, \sigma^2)$ can be computed

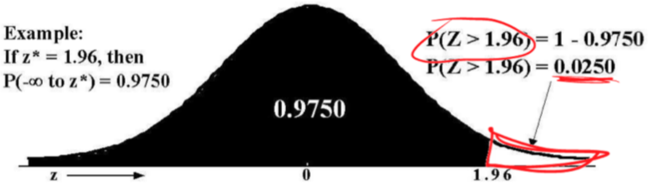
$$\mathrm{P}(a \le X \le b) = \mathrm{P}\left(\frac{a - \mu}{\sigma} \le Z \le \frac{b - \mu}{\sigma}\right)$$
$$= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

Standard Normal Table



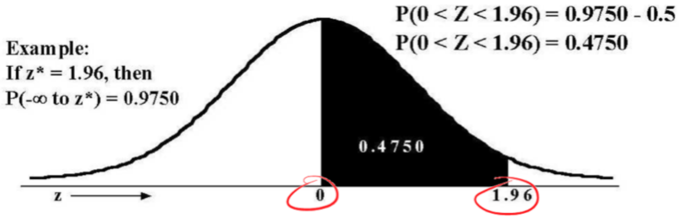
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767

Standard Normal Table Cont'd



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767

Standard Normal Table Cont'd



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767

Properties of Φ

- $\Phi(0) = .50 \Rightarrow$ Mean and Median (50_{th} percentile) for standard normal are both 0

Properties of Φ

- $\Phi(0) = .50 \Rightarrow$ Mean and Median (50_{th} percentile) for standard normal are both 0
- $\Phi(-z) = 1 - \Phi(z)$

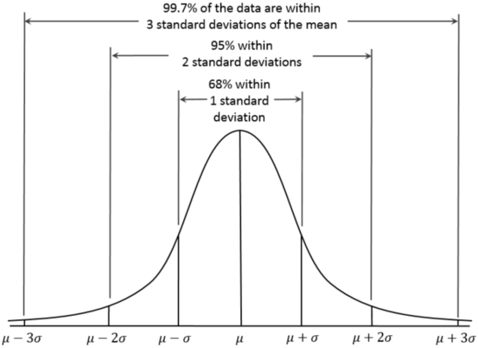
Properties of Φ

- $\Phi(0) = .50 \Rightarrow$ Mean and Median (50_{th} percentile) for standard normal are both 0
- $\Phi(-z) = 1 - \Phi(z)$
- $\mathbb{P}(Z > z) = 1 - \Phi(z) = \Phi(-z)$

The Empirical Rules

The Empirical Rules provide a quick way to approximate certain probabilities for the Normal Distribution as the following table:

Interval	Percentage with interval
$\mu \pm \sigma$	68%
$\mu \pm 2\sigma$	95%
$\mu \pm 3\sigma$	99.7%



Example

Let us find the following probabilities with respect to Z:

1. Z is at most -1.75
2. Z is between -2 and 2 inclusive
3. Z is less than .5

Example Cont'd

Solution.

1. $P(Z \leq -1.75) = \Phi(-1.75) = .0401$

Example Cont'd

Solution.

1. $P(Z \leq -1.75) = \Phi(-1.75) = .0401$
2. $P(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2) = .9772 - .0228 = .9544$

Example Cont'd

Solution.

1. $P(Z \leq -1.75) = \Phi(-1.75) = .0401$
2. $P(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2) = .9772 - .0228 = .9544$
3. $P(Z < .5) = \Phi(.5) = .6915$

Example

Suppose a STAT-8020 exam score follows a normal distribution with mean 78 and variance 36. Let X to denote the exam score, answer the following questions:

1. What is the probability that a randomly chosen test taker got a score greater than 84?
2. Suppose the passing score for this exam is 75. What is the probability that a randomly chosen test taker got a score greater than 84 given that she/he pass the exam?
3. Using the empirical rule to find the 84_{th} percentile.

Example

Find the following percentile with respect to Z

1. 10_{th} percentile
2. 55_{th} percentile
3. 90_{th} percentile

Example Cont'd

Solution.

1. $Z_{10} = -1.28$
2. $Z_{55} = 0.13$
3. $Z_{90} = 1.28$

```
> qnorm(0.1)
[1] -1.281552
> qnorm(0.55)
[1] 0.1256613
> qnorm(0.9)
[1] 1.281552
```

Example

Let X be Normal with a mean of 20 and a variance of 49. Find the following probabilities and percentile:

1. X is between 15 and 23
2. X is more than 30
3. X is more than 12 knowing it is less than 20
4. What is the value that is smaller than 20% of the distribution?

Example Cont'd

Solution.

1. $P(15 \le X \le 23) = \Phi(\frac{23-20}{7}) - \Phi(\frac{15-20}{7}) = \Phi(0.43) - \Phi(-0.71) = .6664 - .2389 = .4275$
2. $P(X > 30) = 1 - P(X \le 30) = 1 - \Phi(\frac{30-20}{7}) = 1 - .9236 = .0764$
3. $P(X > 12 | X < 20) = \frac{P(12<X<20)}{\mathbb{P}(X<20)} = \frac{\Phi(0)-\Phi(-1.14)}{\Phi(0)} = .7458$
4. $Z_{80} = 0.84 \Rightarrow X_{80} = \mu + Z_{80} \times \sigma = 20 + 0.84 \times \sqrt{49} = 25.88$

Sums of Normal Random Variables

If X_i $1 \leq i \leq n$ are independent normal random variables with mean μ_i are variance σ_i^2, respectively.

Sums of Normal Random Variables

If X_i $1 \leq i \leq n$ are independent normal random variables with mean μ_i are variance σ_i^2, respectively.

- Let $S_n = \sum_{i=1}^{n} X_i$ then $S_n \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$

Sums of Normal Random Variables

If X_i $1 \le i \le n$ are independent normal random variables with mean μ_i are variance σ_i^2, respectively.

- Let $S_n = \sum_{i=1}^{n} X_i$ then $S_n \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$

Sums of Normal Random Variables

If X_i $1 \leq i \leq n$ are independent normal random variables with mean μ_i are variance σ_i^2, respectively.

- Let $S_n = \sum_{i=1}^n X_i$ then $S_n \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$
- This can be applied for any integer n

Example

Let X_1, X_2, and X_3 be mutually independent, Normal random variables. Let their means and standard deviations be $3k$ and k for $k = 1$, 2, and 3 respectively. Find the following distributions:

1. $\sum_{i=1}^{3} X_i$
2. $X_1 + 2X_2 - 3X_3$
3. $X_1 + 5X_3$

Example Cont'd

Solution.

1. $\sum_{i=1}^{3} X_i \sim N(\mu = 3 + 6 + 9 = 18, \sigma^2 = 1^2 + 2^2 + 3^2 = 14)$
2. $X_1 + 2X_2 - 3X_3 \sim N(\mu = 3 + 12 - 27 = -12, \sigma^2 = 1^2 + 4 \times 2^2 + 9 \times 3^2 = 98)$
3. $X_1 + 5X_3 \sim N(\mu = 3 + 45 = 48, \sigma^2 = 1^2 + 25 \times 3^2 = 226)$

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large
- Rule of thumb for this approximation to be valid (in this class) is $np > 5$ and $n(1-p) > 5$

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large

- Rule of thumb for this approximation to be valid (in this class) is $np > 5$ and $n(1-p) > 5$

- If $X \sim \text{Bin}(n, p)$ with $np > 5$ and $n(1-p) > 5$ then we can use $X^* \sim \text{N}(\mu = np, \sigma^2 = np(1-p))$ to approximate X

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large
- Rule of thumb for this approximation to be valid (in this class) is $np > 5$ and $n(1-p) > 5$
- If $X \sim \text{Bin}(n, p)$ with $np > 5$ and $n(1-p) > 5$ then we can use $X^* \sim \mathrm{N}(\mu = np, \sigma^2 = np(1-p))$ to approximate X
- Notice that Binomial is a discrete distribution but normal is a continuous distribution so that $\mathbb{P}(X^* = x) = 0 \ \forall x$

Normal approximation of Binomial Distribution

- We can use a Normal Distribution to approximate a Binomial Distribution if n is large
- Rule of thumb for this approximation to be valid (in this class) is $np > 5$ and $n(1-p) > 5$
- If $X \sim \text{Bin}(n, p)$ with $np > 5$ and $n(1-p) > 5$ then we can use $X^* \sim \mathrm{N}(\mu = np, \sigma^2 = np(1-p))$ to approximate X
- Notice that Binomial is a discrete distribution but normal is a continuous distribution so that $\mathbb{P}(X^* = x) = 0 \ \forall x$
- **Continuity correction**: we use $\mathbb{P}(x - 0.5 \le X^* \le x + 0.5)$ to approximate $\mathbb{P}(X = x)$

Example

Suppose a class has 400 students (to begin with), that each student drops independently of any other student with a probability of .07. Let X be the number of students that finish this course

1. Find the probability that X is between 370 and 373 inclusive
2. Is an approximation appropriate for the number of students that finish the course?
3. If so, what is this distribution and what are the parameter(s)?
4. Find the probability that is between 370 and 373 inclusive by using the approximation

Summary

In this lecture, we learned

- Normal Distributions
- Sum of Normal Random Variables
- Normal approximation of Binomial Distribution