Lecture 13

Continuous Random Variables

Text: Chapter 4

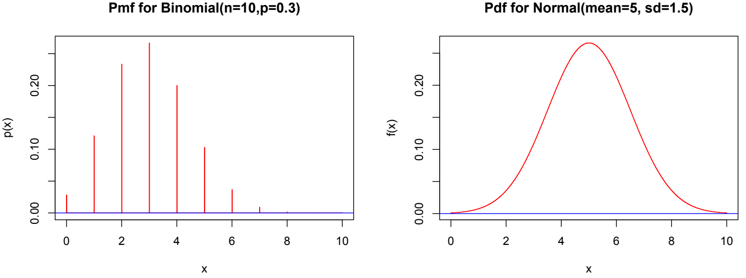
STAT 8010 Statistical Methods I
September 18, 2019

Whitney Huang
Clemson University

Agenda

1. From Discrete to Continuous Random Variables
2. Cumulative Distribution Functions
3. Expected Value and Variance
4. Normal Distributions

Probability Mass Functions vs. Probability Density Functions



Remarks:

- pmf assigns probabilities to each possible values of a discrete distribution
- pdf describes the relative likelihood for this random variable to take on a given interval

Probability Mass Functions v.s. Probability Density Functions cont'd

Recall the properties of discrete probability mass functions (Pmfs):

- $0 \leq p_X(x) \leq 1$ for all possible values of x

Probability Mass Functions v.s. Probability Density Functions cont'd

Recall the properties of discrete probability mass functions (Pmfs):

- $0 \leq p_X(x) \leq 1$ for all possible values of x
- $\sum_x p_X(x) = 1$

Probability Mass Functions v.s. Probability Density Functions cont'd

Recall the properties of discrete probability mass functions (Pmfs):

- $0 \leq p_X(x) \leq 1$ for all possible values of x
- $\sum_x p_X(x) = 1$
- $\mathbb{P}(a \leq X \leq b) = \sum_{x=a}^{x=b} p_X(x)$

Probability Mass Functions v.s. Probability Density Functions cont'd

Recall the properties of discrete probability mass functions (Pmfs):

- $0 \leq p_X(x) \leq 1$ for all possible values of x
- $\sum_x p_X(x) = 1$
- $\mathbb{P}(a \leq X \leq b) = \sum_{x=a}^{x=b} p_X(x)$

Probability Mass Functions v.s. Probability Density Functions cont'd

Recall the properties of discrete probability mass functions (Pmfs):

- $0 \leq p_X(x) \leq 1$ for all possible values of x
- $\sum_x p_X(x) = 1$
- $\mathbb{P}(a \leq X \leq b) = \sum_{x=a}^{x=b} p_X(x)$

For continuous distributions, the properties for probability density functions (Pdfs) are similar:

- $f_X(x) \geq 0$ for all possible values of x

Probability Mass Functions v.s. Probability Density Functions cont'd

Recall the properties of discrete probability mass functions (Pmfs):

- $0 \leq p_X(x) \leq 1$ for all possible values of x
- $\sum_x p_X(x) = 1$
- $\mathbb{P}(a \leq X \leq b) = \sum_{x=a}^{x=b} p_X(x)$

For continuous distributions, the properties for probability density functions (Pdfs) are similar:

- $f_X(x) \geq 0$ for all possible values of x
- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

Probability Mass Functions v.s. Probability Density Functions cont'd

Recall the properties of discrete probability mass functions (Pmfs):

- $0 \leq p_X(x) \leq 1$ for all possible values of x
- $\sum_x p_X(x) = 1$
- $\mathbb{P}(a \leq X \leq b) = \sum_{x=a}^{x=b} p_X(x)$

For continuous distributions, the properties for probability density functions (Pdfs) are similar:

- $f_X(x) \geq 0$ for all possible values of x
- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx$

Cumulative Distribution Functions (cdfs) for Continuous Distribution

- The cdf $F_X(x)$ is defined as $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(x)\,dx$
- we use cdf to calculate probabilities of a continuous random variable within an interval, i.e. $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx = \int_{-\infty}^{b} f_X(x)\,dx - \int_{-\infty}^{a} f_X(x)\,dx = \boxed{F_X(b) - F_X(a)}$

Remark: $\mathbb{P}(X = x) = \int_x^x f_X(x)\,dx = 0$ for all possible values of x

Expected Value and Variance

Recall the expected value formula for the discrete random variable: $\mathbb{E}[X] = \sum_x x p_X(x)$
For continuous random variables, we have similar formulas:
Let a, b, and c are constant real numbers

- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$

Expected Value and Variance

Recall the expected value formula for the discrete random variable: $\mathbb{E}[X] = \sum_x x p_X(x)$
For continuous random variables, we have similar formulas:
Let a, b, and c are constant real numbers

- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$
- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$

Expected Value and Variance

Recall the expected value formula for the discrete random variable: $\mathbb{E}[X] = \sum_x x p_X(x)$
For continuous random variables, we have similar formulas:
Let a, b, and c are constant real numbers

- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$
- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Expected Value and Variance

Recall the expected value formula for the discrete random variable: $\mathbb{E}[X] = \sum_x x p_X(x)$
For continuous random variables, we have similar formulas:
Let a, b, and c are constant real numbers

- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$
- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$
- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[cX] = c\mathbb{E}[X]$

Expected Value and Variance

Recall the expected value formula for the discrete random variable: $\mathbb{E}[X] = \sum_x x p_X(x)$
For continuous random variables, we have similar formulas:
Let a, b, and c are constant real numbers

- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$
- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[cX] = c\mathbb{E}[X]$
- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$

Expected Value and Variance

Recall the expected value formula for the discrete random variable: $\mathbb{E}[X] = \sum_x x p_X(x)$
For continuous random variables, we have similar formulas:
Let a, b, and c are constant real numbers

- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$
- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$
- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[cX] = c\mathbb{E}[X]$
- $\mathbb{E}[aX+bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \left(\int_{-\infty}^{\infty} x f_X(x)\, dx\right)^2$

Expected Value and Variance

Recall the expected value formula for the discrete random variable: $\mathbb{E}[X] = \sum_x x p_X(x)$
For continuous random variables, we have similar formulas:
Let a, b, and c are constant real numbers

- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$
- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$
- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[cX] = c\mathbb{E}[X]$
- $\mathbb{E}[aX+bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$

- $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right)^2$
- $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$

Expected Value and Variance

Recall the expected value formula for the discrete random variable: $\mathbb{E}[X] = \sum_x x p_X(x)$
For continuous random variables, we have similar formulas:
Let a, b, and c are constant real numbers

- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$
- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$
- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[cX] = c\mathbb{E}[X]$
- $\mathbb{E}[aX+bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$

- $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \left(\int_{-\infty}^{\infty} x f_X(x)\, dx\right)^2$
- $\text{Var}(cX) = c^2 \text{Var}(X)$
- $\text{Var}(X-c) = \text{Var}(X)$

Example

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \leq x \leq 4$ and be 0 otherwise. Answer the following questions:

Example

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \leq x \leq 4$ and be 0 otherwise. Answer the following questions:

1. Find the value of c that makes this a valid pdf

Example

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \leq x \leq 4$ and be 0 otherwise. Answer the following questions:

1. Find the value of c that makes this a valid pdf

Example

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \leq x \leq 4$ and be 0 otherwise. Answer the following questions:

1. Find the value of c that makes this a valid pdf
2. Find the expected value and variance of X

Example

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \leq x \leq 4$ and be 0 otherwise. Answer the following questions:

1. Find the value of c that makes this a valid pdf
2. Find the expected value and variance of X

Example

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \leq x \leq 4$ and be 0 otherwise. Answer the following questions:

1. Find the value of c that makes this a valid pdf
2. Find the expected value and variance of X
3. What is the probability that X is within $.5$ inches of the expected diameter?

Example

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \leq x \leq 4$ and be 0 otherwise. Answer the following questions:

1. Find the value of c that makes this a valid pdf
2. Find the expected value and variance of X
3. What is the probability that X is within $.5$ inches of the expected diameter?

Example

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \leq x \leq 4$ and be 0 otherwise. Answer the following questions:

1. Find the value of c that makes this a valid pdf
2. Find the expected value and variance of X
3. What is the probability that X is within $.5$ inches of the expected diameter?
4. Find $F_X(x)$

Normal Distribution

Characteristics of the Normal random variable:
Let X be a Normal r.v.

- The support for X: $(-\infty, \infty)$

Normal Distribution

Characteristics of the Normal random variable:
Let X be a Normal r.v.

- The support for X: $(-\infty, \infty)$
- Its parameter(s) and definition(s): μ : mean and σ^2 : variance

Normal Distribution

Characteristics of the Normal random variable:
Let X be a Normal r.v.

- The support for X: $(-\infty, \infty)$
- Its parameter(s) and definition(s): 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 $-\infty < x < \infty$

Normal Distribution

Characteristics of the Normal random variable:
Let X be a Normal r.v.

- The support for X: $(-\infty, \infty)$
- Its parameter(s) and definition(s): 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

Normal Distribution

Characteristics of the Normal random variable:
Let X be a Normal r.v.

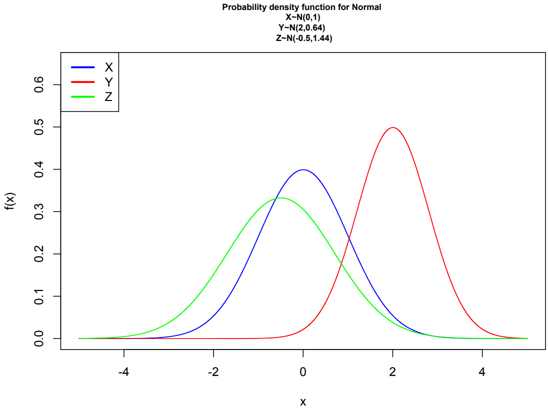
- The support for X: $(-\infty, \infty)$
- Its parameter(s) and definition(s): μ : mean and σ^2 : variance
- The probability density function (pdf): $\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ for $-\infty < x < \infty$
- The cumulative distribution function (cdf): No explicit form,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

Normal Distribution

Characteristics of the Normal random variable:
Let X be a Normal r.v.

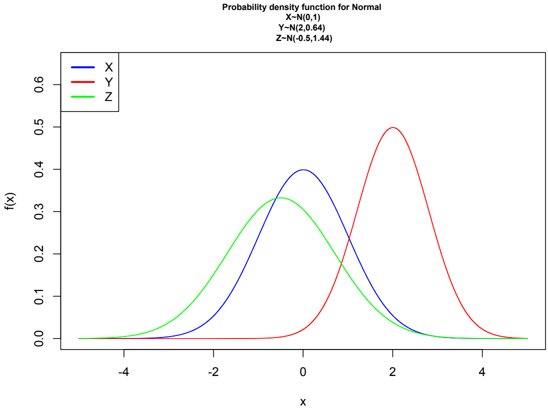
- The support for X: $(-\infty, \infty)$
- Its parameter(s) and definition(s): μ : mean and σ^2 : variance
- The probability density function (pdf): $\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ for $-\infty < x < \infty$
- The cumulative distribution function (cdf): No explicit form,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

Normal Density Curves



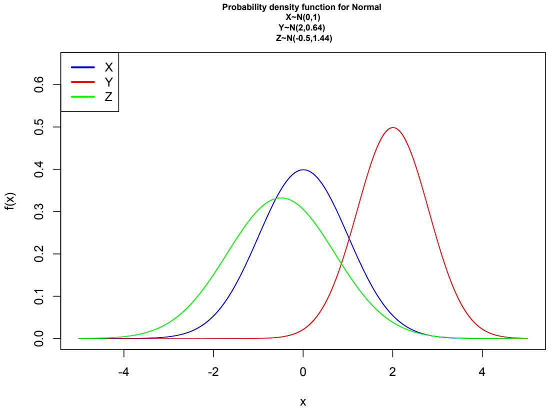
- The parameter μ determines the center of the distribution

Normal Density Curves



- The parameter μ determines the center of the distribution
- The parameter σ^2 determines the spread of the distribution

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

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

- Normal random variable X with mean μ and standard deviation σ can convert 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 convert 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 convert 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 convert 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 $\mathbb{P}(a \leq X \leq b)$ where $X \sim N(\mu, \sigma^2)$ can be compute

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma})$$
$$= \Phi(\frac{b - \mu}{\sigma}) - \Phi(\frac{a - \mu}{\sigma})$$

Properties of Φ

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

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
- $\Phi(-z) = 1 - \Phi(z)$

Properties of Φ

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

Example

Let us examine Z. 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. $\mathbb{P}(Z \leq -1.75) = \Phi(-1.75) = .0401$

Example Cont'd

Solution.

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

Example Cont'd

Solution.

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

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.

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 $1 \leq i \leq n$ 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 $1 \leq i \leq n$ 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. $\sum_{i=1}^{3} X_i \sim N(\mu = 3+6+9 = 15, \sigma^2 = 1^2+2^2+3^2 = 14)$
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. $X_1 + 5X_3 \sim N(\mu = 3+45 = 48, \sigma^2 = 1^2 + 25 \times 3^2 = 226)$