Lecture 8

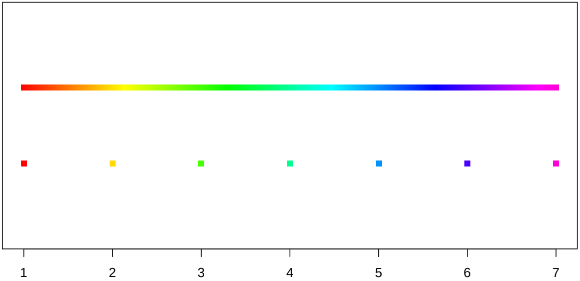
Probability IV

Readings: IntroStat Chapter 4; OpenIntro Chapter 3

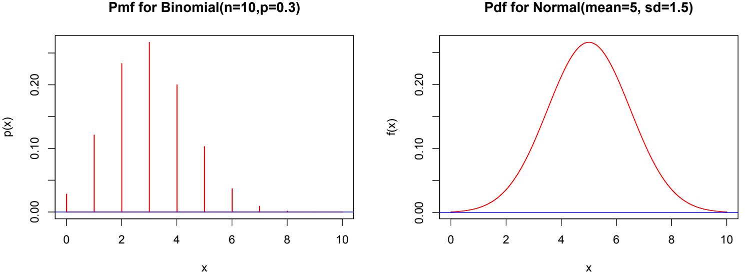
STAT 8010 Statistical Methods I
May 25, 2023

Whitney Huang
Clemson University

From Discrete to Continuous Random Variables



Probability Mass Functions vs. Probability Density Functions



Remarks:

- pmf assigns probabilities to each possible values of a discrete random variable
- pdf describes the relative likelihood for a continuous 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):

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

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

- $f_X(x) \ge 0$ 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 r.v.s

- The cdf $F_X(x)$ is defined as $F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(x)\,dx$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Example 1

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 1

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 1

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \le x \le 4$ 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 1

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \le x \le 4$ 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 $F_X(x)$

Example 1 Cont'd

1. $\int_1^4 f_X(x)\, dx = \int_1^4 c(4x - x^2)\, dx = 9c = 1 \Rightarrow c = \frac{1}{9}$

2. $E[X] = \int_1^4 x f_X(x)\, dx = 2.25$
   $E[X^2] = \int_1^4 x^2 f_X(x)\, dx = 5.6$
   $\Rightarrow Var(X) = E[X^2] - (E[X])^2 = .5375$

3. $P(1.75 \le X \le 2.75) = \int_{1.75}^{2.75} f_X(x)\, dx = .4282$

4. $F_X(x) = \int f_X(x)\, dx = \frac{1}{9}(2x^2 - \frac{x^3}{3})$ for $1 \le X \le 4$ and 0 if $x < 1$ and 1 if $x > 4$

Example 2

Let $f_X(x) = .25x$ for $1 \leq x \leq 3$ and 0 otherwise:

1. Is X more likely to be within $[1, 2]$ or within $[2, 3]$? First answer this question using logic. Check your answer by calculating the probabilities
2. What is the probability that X is more than 2.2?
3. Find the mean and standard deviation of X
4. Find $F_X(x)$
5. What value of X represents the top 15% of the distribution?

CLEMSON UNIVERSITY

Example 2 Cont'd

Solution.

1. X is more likely be within $[2, 3]$ because $f_X(x)$ is increases with x
   $\mathbb{P}(1 \le X \le 2) = \int_1^2 f_X(x)\, dx = .375$ $\mathbb{P}(2 \le X \le 3) = \int_2^3 f_X(x)\, dx = .625$

CLEMSON UNIVERSITY

Example 2 Cont'd

Solution.

1. X is more likely be within $[2, 3]$ because $f_X(x)$ is increases with x
$\mathbb{P}(1 \leq X \leq 2) = \int_1^2 f_X(x)\, dx = .375$ $\mathbb{P}(2 \leq X \leq 3) = \int_2^3 f_X(x)\, dx = .625$

2. $\mathbb{P}(X \geq 2.2) = \int_{2.2}^3 f_X(x)\, dx = .52$

Example 2 Cont'd

Solution.

1. X is more likely be within $[2,3]$ because $f_X(x)$ is increases with x
$\mathbb{P}(1 \le X \le 2) = \int_1^2 f_X(x)\,dx = .375$ $\mathbb{P}(2 \le X \le 3) = \int_2^3 f_X(x)\,dx = .625$

2. $\mathbb{P}(X \ge 2.2) = \int_{2.2}^3 f_X(x)\,dx = .52$

3. $\mathbb{E}[X] = \int_1^3 x f_X(x)\,dx = 2.1667$
$\mathbb{E}[X^2] = \int_1^3 x^2 f_X(x)\,dx = 5$
$\Rightarrow Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = .3056 \Rightarrow Sd(X) = .5528$

Example 2 Cont'd

Solution.

1. X is more likely be within $[2,3]$ because $f_X(x)$ is increases with x
$\mathbb{P}(1 \leq X \leq 2) = \int_1^2 f_X(x)\,dx = .375$ $\mathbb{P}(2 \leq X \leq 3) = \int_2^3 f_X(x)\,dx = .625$

2. $\mathbb{P}(X \geq 2.2) = \int_{2.2}^3 f_X(x)\,dx = .52$

3. $\mathbb{E}[X] = \int_1^3 x f_X(x)\,dx = 2.1667$
$\mathbb{E}[X^2] = \int_1^3 x^2 f_X(x)\,dx = 5$
$\Rightarrow Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = .3056 \Rightarrow Sd(X) = .5528$

4. $F_X(x) = \int f_X(x)\,dx = \frac{1}{8}(x^2 - 1)$ for $1 \leq X \leq 3$ and 0 if $x < 1$ and 1 if $x > 3$

Example 2 Cont'd

Solution.

1. X is more likely be within $[2,3]$ because $f_X(x)$ is increases with x
$\mathbb{P}(1 \le X \le 2) = \int_1^2 f_X(x)\, dx = .375$ $\mathbb{P}(2 \le X \le 3) = \int_2^3 f_X(x)\, dx = .625$

2. $\mathbb{P}(X \ge 2.2) = \int_{2.2}^3 f_X(x)\, dx = .52$

3. $\mathbb{E}[X] = \int_1^3 x f_X(x)\, dx = 2.1667$
$\mathbb{E}[X^2] = \int_1^3 x^2 f_X(x)\, dx = 5$
$\Rightarrow Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = .3056 \Rightarrow Sd(X) = .5528$

4. $F_X(x) = \int f_X(x)\, dx = \frac{1}{8}(x^2 - 1)$ for $1 \le X \le 3$ and 0 if $x < 1$ and 1 if $x > 3$

5. $X_{.85}$ is the x value such that $\int_1^x f_X(x)\, dx = .85 \Rightarrow X_{.85} = 2.7928$

Summary

In this lecture, we learned

- Continuous Random Variables
- Probability Density Function / Cumulative Distribition Function
- Calculating Expected Value, Variance, and Percentiles