Lecture 13
probability and random variables

Review From Friday 2/16

Deriving Sampling Distributions

Population Distribution

Probability distribution for
a single roll of a 4-sided die

X	$P(X)$
1	0.25
2	0.25
3	0.25
4	0.25

Sampling Distribution

Probability distribution for
<u>the average of $n = 2$ rolls</u> of a 4-sided die

$\bar{x}$	Possible Outcomes	
1	(1,1)	1/16
1.5	(1,2), (2,1)	2/16
2	(1,3), (3,1),(2,2)	3/16
2.5	(3,2), (2,3), (4,1),(1,4)	4/16
3	(3,3), (4,2), (2,4)	3/16
3.5	(3,4),(4,3)	2/16
4	(4,4)	1/16

Mean and Standard Deviation of Discrete Random Variables

- The mean of a probability distribution is defined as

$$\mu = \sum_{x} xP(x)$$

- The variance and standard deviation of a probability distribution are defined as

$$\sigma^2 = \sum_{x} (x - \mu)^2 P(x)$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{x} (x - \mu)^2 P(x)}$$

Where x denotes an outcome of the random variable X and $P(x)$ denotes the probability of the outcome

The Bernoulli distribution

- The **probability mass function (PMF)** of a discrete random is a function that gives the probability that the variable is exactly equal to some value
- A Bernoulli random variable is on which there are two possible outcomes with probabilities p and $1\ - p$
- Whenever we assign the outcomes of a random variable to either "success" or "failure" (1 or 0) we are dealing with a Bernoulli random variable

$$\text{mean} = p$$

$$\text{variance} = p(1 - p)$$

$$\text{PMF: } P(X = x) = \begin{cases} p, & \text{if success} \\ (1 - p), & \text{else} \end{cases}$$

X	$P(X)$
1 (success)	p
0 (failure)	$1 - p$

The Binomial Distribution

- A discrete distribution which describes the probabilities for the number of successful outcomes in a given number of independent trials where each trial has the same probability of success

  It has two parameters:

  $n =$ the number of trials

  p = the probability of "success" or the probability of the outcome of interest.

  mean = np variance = $np(1-p)$

- It describes the proportion of trials in which a particular outcome of interest occurs

- It is a sum of n independent Bernoulli random variables

- There are many examples of binomial random variables
  - the number of heads observed in n flips of a coin where (each times heads has probability $p = \frac{1}{2}$ of occurring)
  - The proportion of deer with chronic wasting disease (CWD)
  - The number of patients who experience headaches as side of effect of taking a drug

The Binomial Distribution

- Probability Mass Function:

$$P(X = k) = \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

- n is the total number of trials (e.g flips of a coin)
- k successes occur with probability p^k
- $n - k$ failures occur with probability p^{n-k}
- $\binom{n}{k}$ is called the binomial coefficient – it represents the number of ways to arrange k successes in n trials

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Example:

- The tiger trout is a sterile hybridization of a brown trout with a brook trout produced in fish hatcheries and stocked into ponds and lakes. It is a prized catch among fly fisherman. Suppose that of the 10,000 trout stocked in spring valley reservoir, 500 of them are tiger trout. Assuming that tiger trout and other species are homogeneously mixed:
- what is the probability that I catch a tiger trout on my first cast?
- What is the probability that I catch a tiger trout in my first 10 casts?

The Poisson Distribution

- A Poisson distribution is a discrete probability distribution. It gives the probability of an event happening a certain number of times k within a given interval of time or space.

- The Poisson distribution has only one parameter, λ (lambda), which is the mean number of events. $\lambda > 0$

Probability Mass Function: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

Examples:

The number of traffic accidents at a particular intersection in a given day can be modeled using a Poisson distribution.

The number of defective items produced by a machine in a fixed period of time can be modeled with a Poisson distribution, assuming a constant defect rate.

Example:

- Consider a scenario in the Star Wars universe where the number of rebel attacks on an Imperial outpost follows a Poisson distribution.

- If the average rate of rebel attacks is 3.5 attacks per week, what is the probability of experiencing exactly 5 rebel attacks in a given week

- What is the probability of experiencing less than 2 attacks in a given week?

Continuous Random Variables

- Recall that a **continuous random variable** is a random variable with an uncountable number of values

- Finding probabilities for continuous random variables requires a different mathematical treatment.

  - we cannot simply list the possible values of a continuous random variable and their probabilities

- The <u>probability distribution</u> of a continuous random variable is typically represented by a *function*, which we will usually consider graphically

  - This function relates the value of the random variable to its <u>probability density</u>

Probability Density Functions

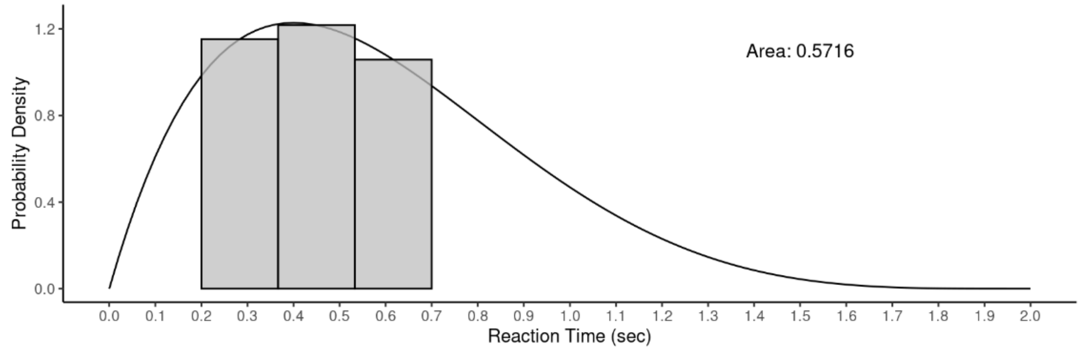
- A probability density function is a function $f(x)$ which gives the probability density of a random variable x. It has the following 3 properties

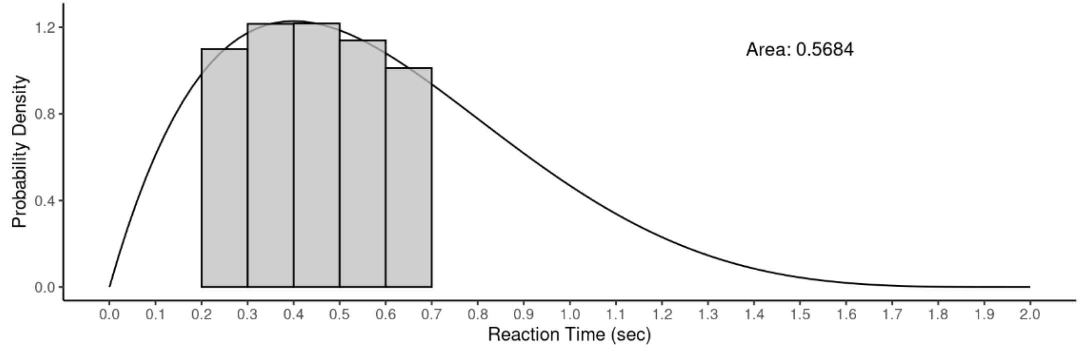
1. The function $f(x)$ is non-negative: $f(x) \geq 0$

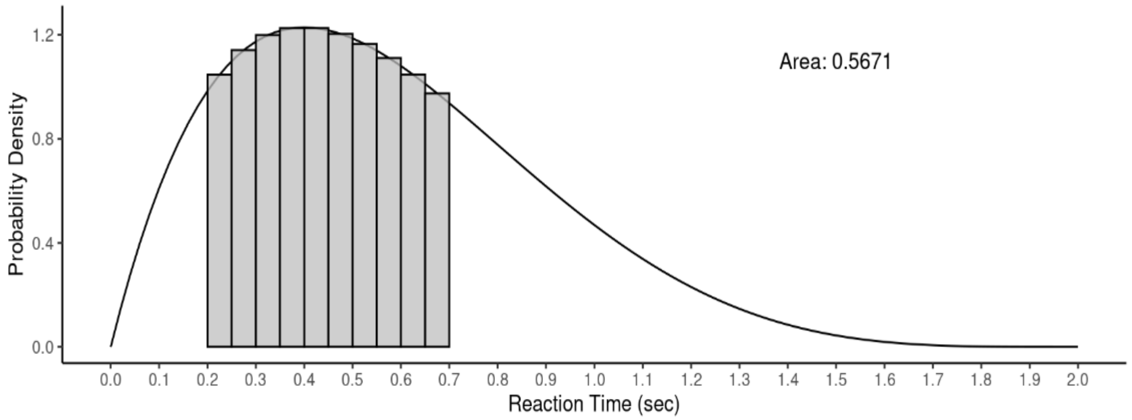
2. The area under the curve of $f(x)$ and above zero equals 1

3. The area under $f(x)$ and between a and b equals $P(a < x < b)$

Example: Consider computing the probability $P(0.2 < X < 0.7)$.

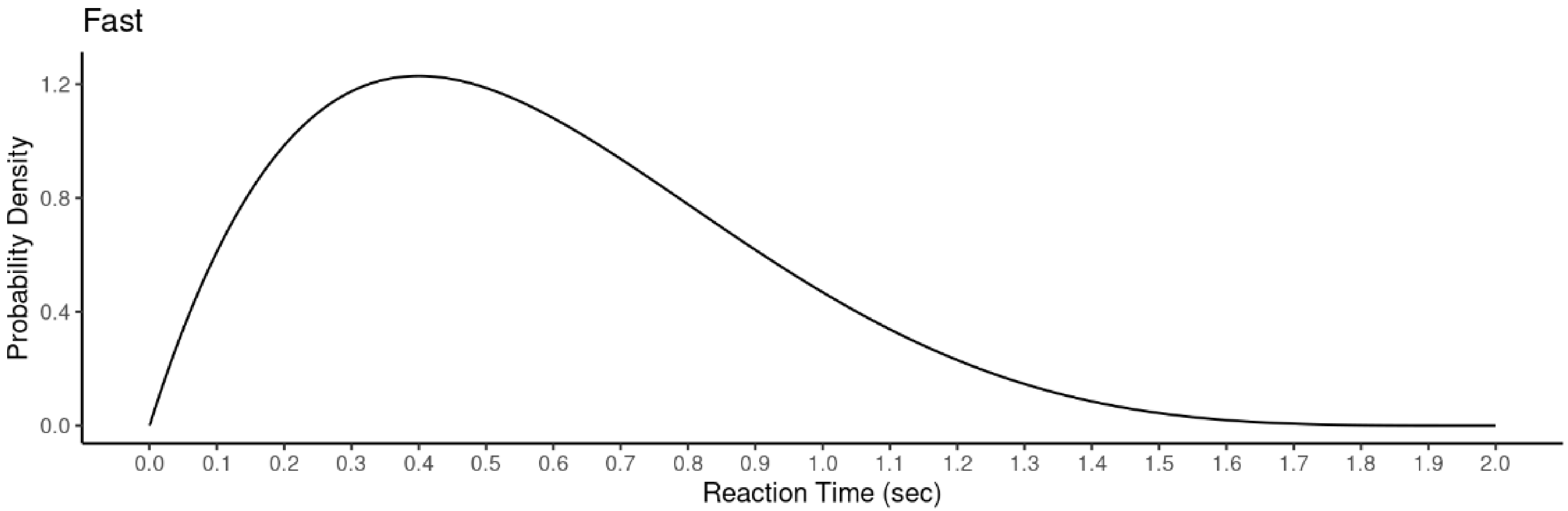




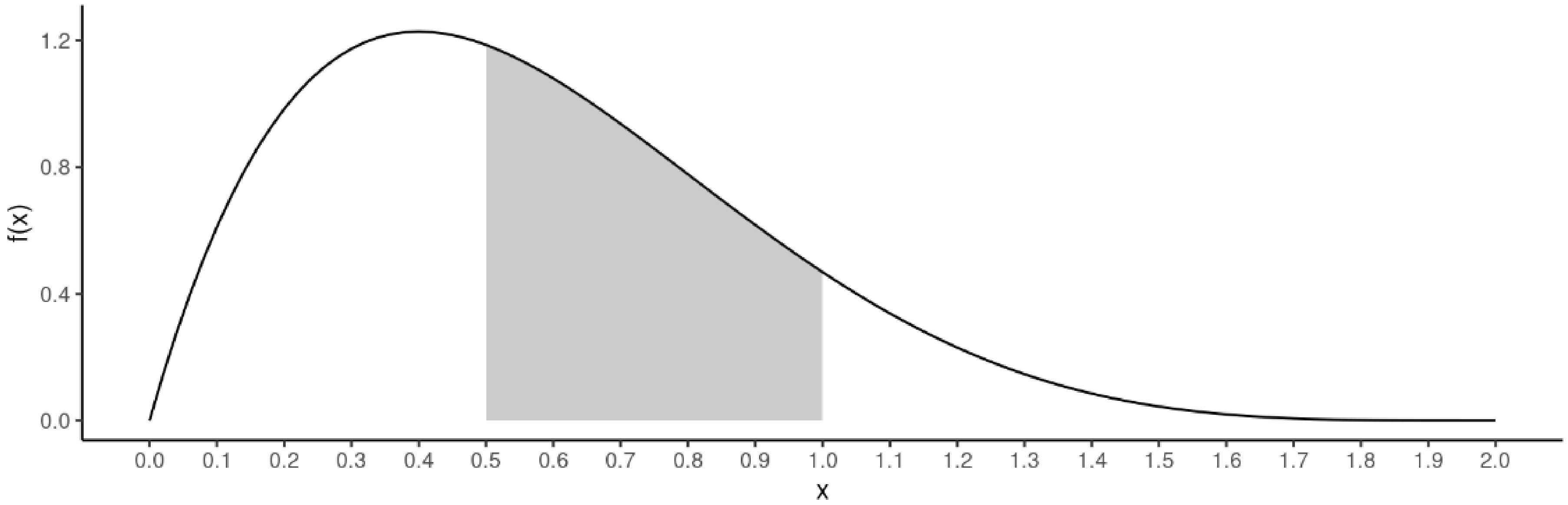


The actual probability is about 0.5667.

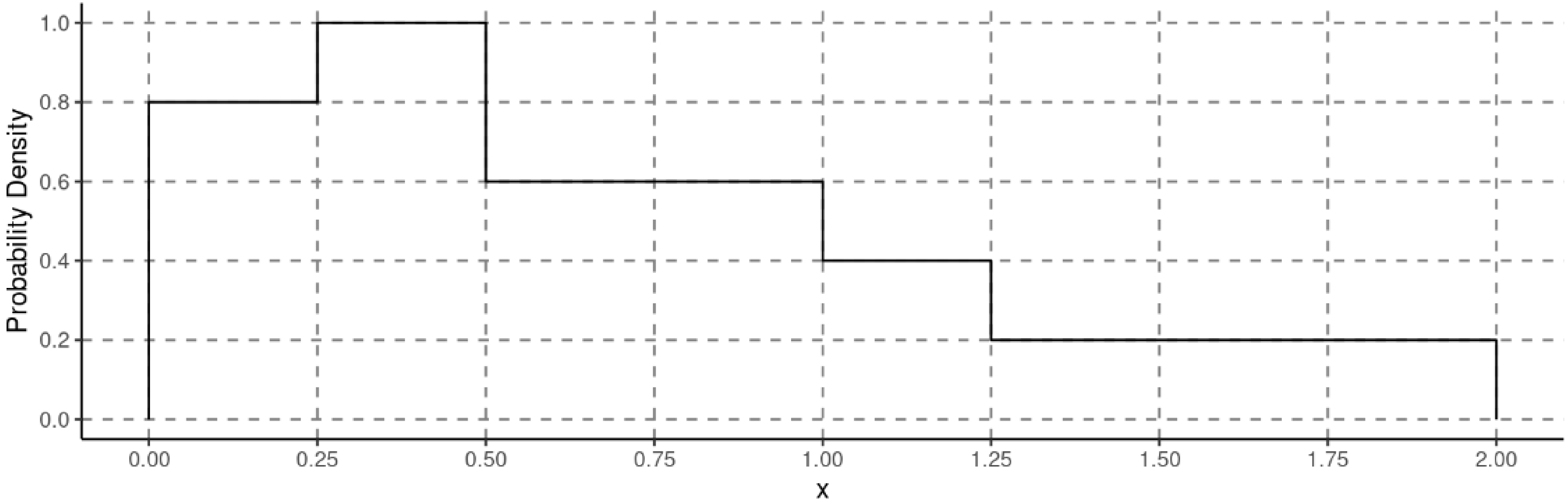
Example: Consider the following hypothetical probability distributions of reaction time.



Example: What is the probability that reaction time is between 0.5 and 1.0 seconds – i.e., $P(0.5 < X < 1)$?



Example: Suppose reaction time had the following probability distribution.

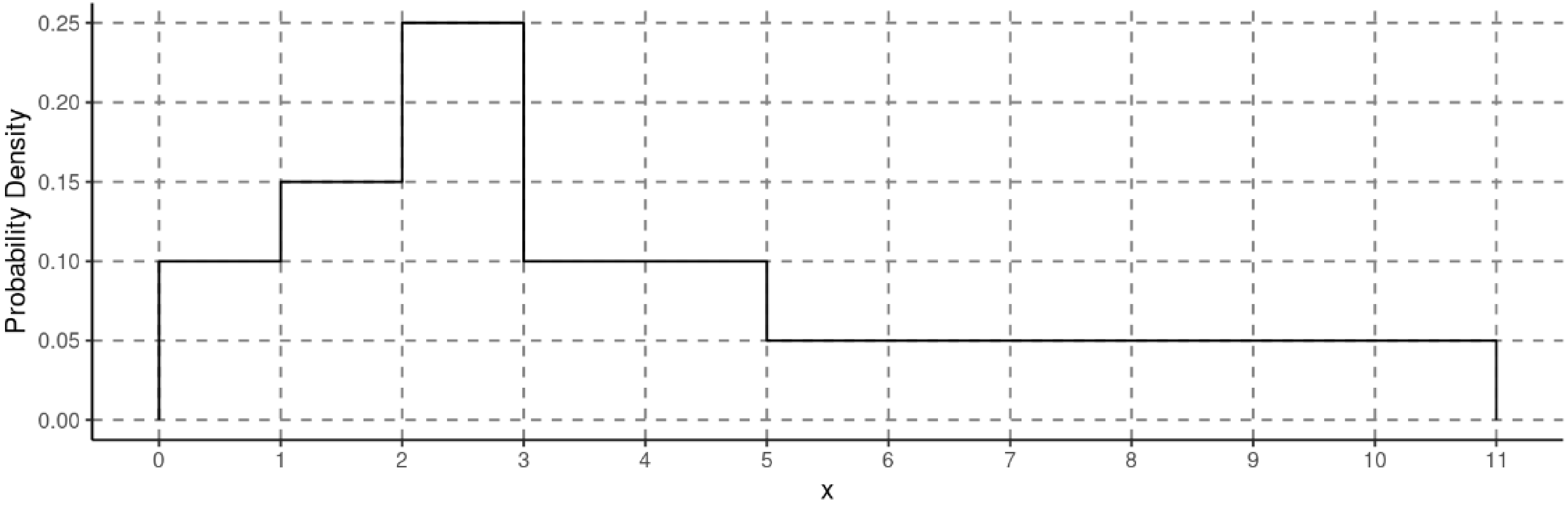


$$P(X < 0.5) = ?$$

$$P(0.25 < X < 0.5) = ?$$

Quartiles

Example: Consider the following probability distribution of a continuous random variable x.



The three *quartiles* have the property that

$$P(X < Q_1) = 0.25,$$
$$P(X < Q_2) = 0.5,$$
$$P(X < Q_3) = 0.75.$$

Mean and variance of Continuous Random Variables

- The mean μ of continuous random variable **cannot** (usually) be defined or computed without calculus, but it is the "balance point" of its probability distribution.

- The standard deviation σ of a continuous random variable **cannot** (usually) be defined or computed without calculus, but it measures the "spread" of the probability distribution.

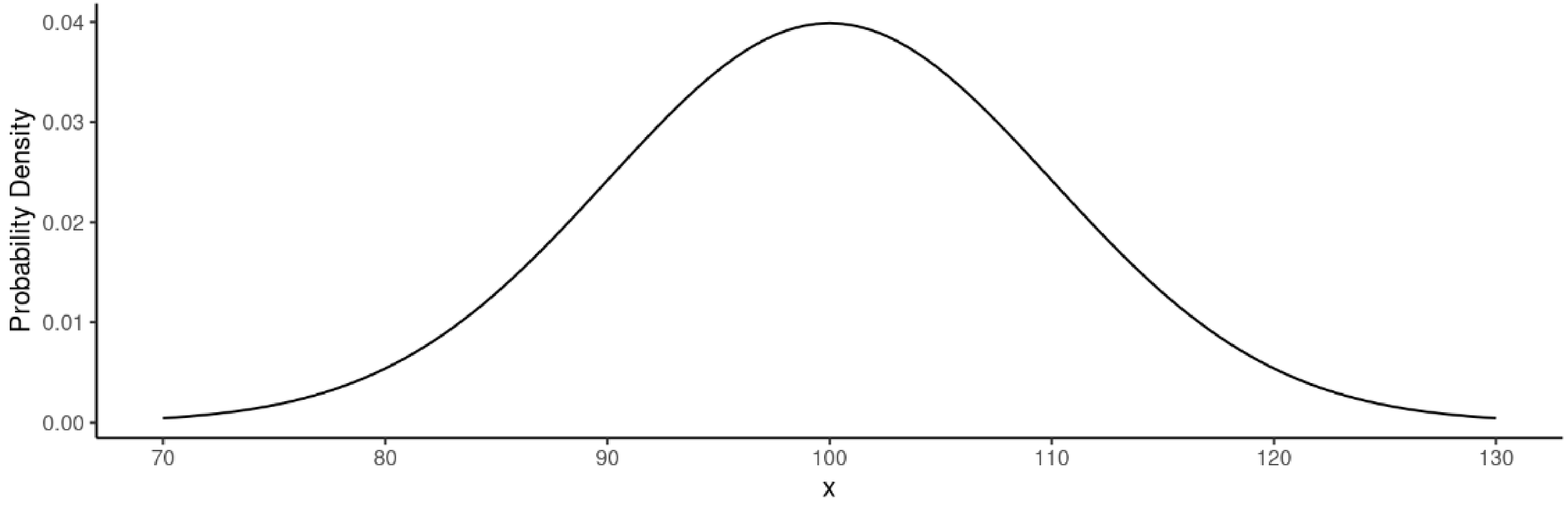
The Normal Distribution

- One important family of continuous probability distributions is the **normal distribution**.

- PDF normal

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Example: Normal probability distribution with mean $\mu = 100$ and standard deviation $\sigma = 10$.



Computing Probabilities From A Normal Distribution

- If we want to find the probability of a given value that we know follows a normal probability distribution we must first find its z-score

$$z = \frac{x - \mu}{\sigma} \sim N(0,1) \quad \text{if } X \sim N(\mu, \sigma)$$

- We can use a probability table for the standard normal distribution or use software such as http://www.statdistributions.com/normal/ to compute the probabilities based on z-scores.

Examples:

Using StatDistributions.com find the following probabilities for a random variable (remember to first convert to z- scores)

$X \sim N(\mu = 100, \sigma = 10)$:

- $P(X < 90)$

- $P(X > 90)$

- $P(90 < X < 110)$

Examples:

Using the z table on Canvas, find the following probabilities for a random variable (remember to first convert to z- scores)

$X \sim N(\mu = 100, \sigma = 5)$:

- $P(X < 90)$ $z = \frac{90 - 100}{5} = -2$
- $P(X > 80)$ $z = \frac{85 - 100}{5} = -3$
- $P(90 < X < 110) = P(-2 \leq z \leq 2)$