MATH 3339
Statistics for the Sciences
Sec 5.1-5.3

Wendy Wang
wwang60@central.uh.edu

Lecture 9 - 3339

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

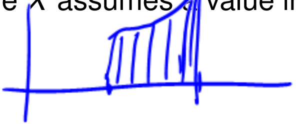
Outline

Types of Random Variables

- A random variable that may assume either a finite number of values or an infinite sequence of values such as $0, 1, \ldots$ is referred to as a **discrete random variable.**
- A random variable that may assume any numerical value in an interval or collection of intervals is called a **continuous random variable.**

Probability distributions

- A **probability distribution** for random variables describes how probabilities are distributed over the values of the random variable.
- For a discrete random variable X, the probability distribution is defined by **probability mass function**, denoted by $f(x)$. This provides the probability for each value of the random variable.
- For a continuous random variable, this is called the **probability density function** $f(x)$.The probability density function (pdf) $f(x)$ is a graph of an equation. The area under the graph of $f(x)$ corresponding to a given interval provides the probability that the random variable X assumes a value in that interval.

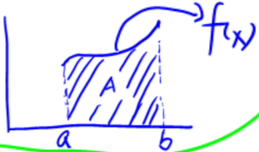


Discrete r.v : pmf : $f(x) = P(X = x)$

Continuous r.v : pdf : $f(x) \neq P(X = x)$

& $P(X = x) = 0$ for all x

for pdf



$\because P(X=a) = P(X=b) = 0$

$\therefore P(a \le X \le b) = P(a < X < b)$

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

$$\rightarrow P(X = a) = P(a \le X \le a) = \int_a^a f(x)\,dx = 0$$

$X = a$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

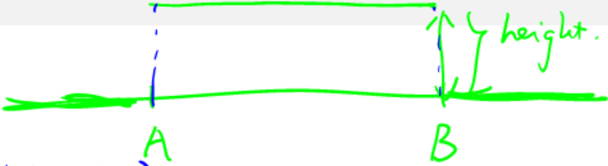
Probability Density Function

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

1. $f(x) \geq 0$ for all x.
2. The area under the entire graph of $f(x)$ must equal 1.

for discrete r.v. $\sum P(X=x) = 1$

Uniform Distribution



$X \sim \text{uniform}(A, B)$

A continuous random variable X is said to have a **uniform distribution** on the interval $[A, B]$ if the pdf of X is:

$$f(x) = \begin{cases} \frac{1}{B-A}, & A \le x \le B \\ 0, & \text{otherwise} \end{cases}$$

if $X < A$, or $X > B$

$\text{Area} = \text{height} \times \text{length}$

$= ? \times (B - A) = 1$

$\text{height} = \dfrac{1}{B-A}$

$$\int_{-\infty}^{\infty} f(x)\,dx$$

$$= \underbrace{\int_{-\infty}^{A} f(x)\,dx} + \int_{A}^{B} f(x)\,dx + \int_{B}^{\infty} f(x)\,dx$$

$$= \underbrace{\cancel{\int_{-\infty}^{A} 0\,dx} + \int_{A}^{B} \frac{1}{B-A}\,dx + \cancel{\int_{B}^{\infty} 0\,dx}}$$

$(X \leq A)$

$$= \int_{A}^{B} \frac{1}{B-A}\,dx = \frac{1}{B-A} \underbrace{\int_{A}^{B} 1\,dx}$$

$$= \frac{1}{B-A} \underbrace{x \Big|_{A}^{B}} = \frac{1}{B-A}(B-A)$$

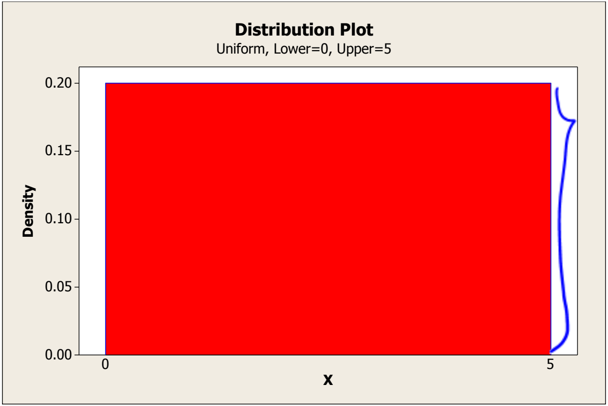
$$= 1$$

Density curve for waiting time

$X \sim \text{uniform}(0, 5)$

The rectangle ranges between 0 and 5. The height of the rectangle is: $\frac{1}{\text{highest value} - \text{lowest value}} = \frac{1}{5-0} = 0.2$.

$$f(x) = \begin{cases} \frac{1}{5-0} = 0.2, & 0 \le x \le 5 \\ 0 & x > 5, \text{ or } x < 0 \end{cases}$$

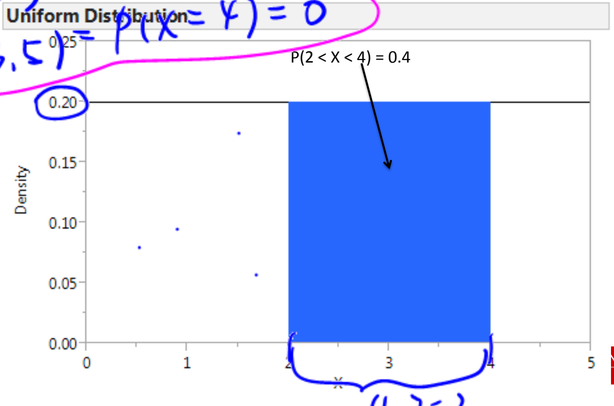


The probability of any event between a range of values is the same as the area between the range under the density curve.

$$\text{Area of rectangle} = \text{height} \times \text{width} = \underbrace{0.2}_{\text{height}} \times \underbrace{2}_{(4-2)} = 0.4$$

$P(X=3) = 0$

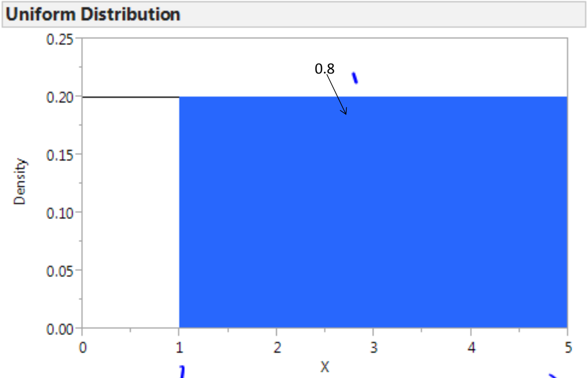
$P(X=3.5) = P(X=4) = 0$



Example continued

What is the probability that a person waits for at least one minute?

$P(X \geq 1) = 0.2 * (5-1) = 0.8$



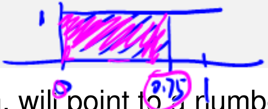
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example continued

What is the probability that a person waits for at least one minute?

$$\begin{aligned} P(X \geq 1) &= \text{area above } 1 \\ &= \text{height} \times \text{width} \end{aligned}$$

Example



Consider a spinner that, after a spin, will point to a number between zero and 1 with "uniform probability."

$X \sim \text{uniform}(0,1)$

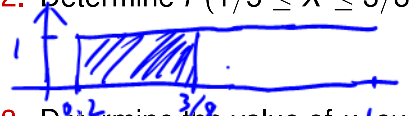
$$f(x) = \begin{cases} \frac{1}{1-0} = 1, & 0 < x \le 1 \\ 0, & \text{o.w.} \end{cases}$$

1. What is the probability that the spinner will land on something less than 0.75?

$P(X < 0.75)$

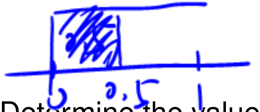
$= \text{area} = 0.75 * 1 = 0.75$

2. Determine $P(1/5 \le X \le 3/8)$.



$\text{Area} = \left(\frac{3}{8} - \frac{1}{5}\right) * 1$

3. Determine the value of x_0 such that $P(X \le x_0) = 0.5$



4. Determine the value of X_0 such that $P(X \ge X_0) = 0.35$.

$\text{Area} = (1 - ?) * 1 = 0.35$

$? = 0.65$

$x_0 = 0.65$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Definition of a Density Function

- A **density function** is a nonnegative function f defined of the set of real numbers such that:

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

- If f is a density function, then its integral $F(x) = \int_{-\infty}^{x} f(u)du$ is a continuous cumulative distribution function (cdf), that is $P(X \leq x) = F(x)$. $= \int_{-\infty}^{x} f(t)\,dt$

- If X is a random variable with this density function, then for any two real numbers, a and b

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx.$$

pdf, not cdf

Example of a Density Function

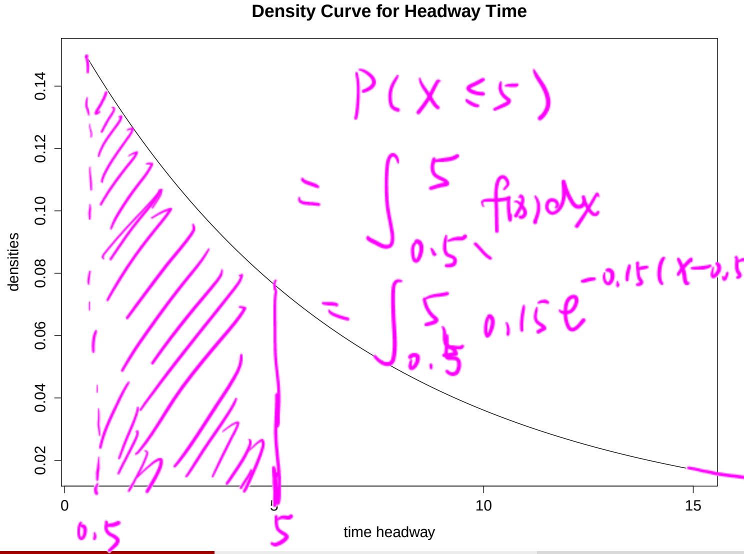
"Time headway" in traffic flow is the elapsed time between the time that one car finishes passing a point and the instant that the next car begins to pass that point. Let X = the time headway (in sec) for two randomly chosen consecutive cars on a freeway during a period of heavy flow. The following pdf of X is essentially the one suggested in "The Statistical Properties of Freeway Traffic" (*Transp. Res.*, vol. 11: 221 - 228):

$$f(x) = \begin{cases} 0.15e^{-0.15(x-0.5)} & x \geq 0.5 \\ 0 & otherwise \end{cases}$$

Density Function

CASA

This is the graph of the density function.



$$P(X \le 5) = \int_{0.5}^{5} f(x)\,dx = \int_{0.5}^{5} 0.15\, e^{-0.15(x-0.5)}\,dx$$

Determine Probability

What is the probability that headway time is at most 5 seconds.

$$P(X \le 5) = 1 - e^{-\frac{27}{40}}$$

Uniform Distribution

A continuous random variable X is said to have a **uniform distribution** on the interval $[A, B]$ if the pdf of X is:

$$f(x) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B \\ 0, & \text{otherwise} \end{cases}$$

$f(x)$: pdf

$F(x)$: cdf

Determine the cdf of a Uniform Distribution

$$F(x) = P(X \le x)$$

$$= \int_{-\infty}^{x} f(t)\,dt$$

$$= \begin{cases} 0 & x < A \\ ? & A \le x \le B \\ 1 & x > B \end{cases}$$

if $A \le X \le B$,

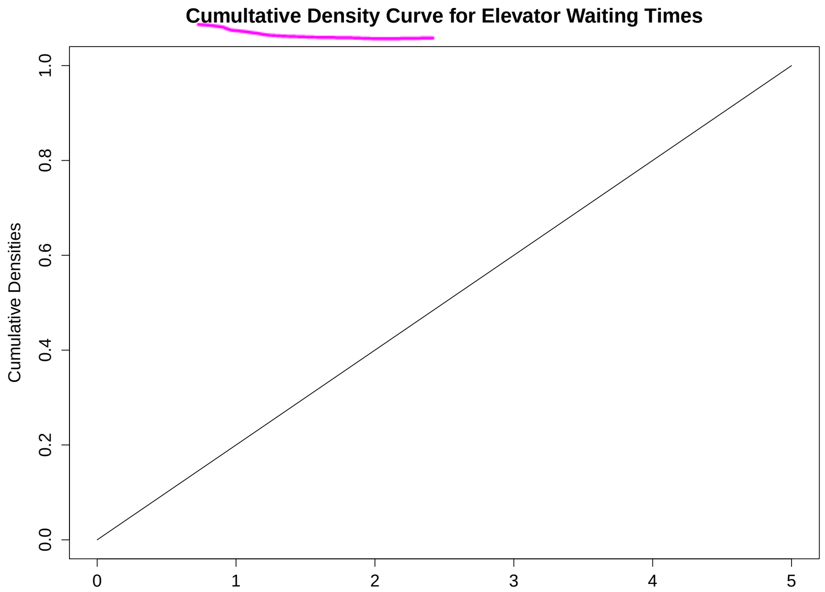
$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

$$= \cancel{\int_{-\infty}^{A} 0\,dt} + \int_{A}^{x} \frac{1}{B-A}\,dt$$

$$= \frac{1}{B-A}\, t \Big|_{A}^{x} = \frac{x-A}{B-A}$$

$$F(x) = \begin{cases} 0 & x < A \\ \dfrac{x-A}{B-A} & A \le x \le B \\ 1 & x > B \end{cases}$$

Cumulative Density Function



Using the cdf $F(X)$ to Compute Probabilities

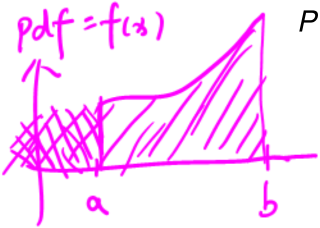
$P(X \le b)$

$P(X \le a)$

Let X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$. Then for any number a,

$$P(X > a) = 1 - F(a) = 1 - P(X \le a)$$

and for any two numbers a and b with $a < b$,



$$P(a \le X \le b) = F(b) - F(a)$$

$$= P(X \le b) - \underline{P(X < a)}$$

$$= P(X \le b) - P(X \le a)$$

$$= F(b) - F(a)$$

Example

The cdf for X = measurement error is

$$P(X \le x) \leftarrow F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{2} + \frac{3}{32}\left(4x - \frac{x^3}{3}\right) & -2 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

1. Compute $P(X < 0)$.

$$P(X < 0) = P(X \le 0) = F(0) = \frac{1}{2} + 0 = \frac{1}{2}$$

2. Compute $P(-1 < X < 1)$

$$P(-1 < X < 1) = F(1) - F(-1)$$
$$= \left(\frac{1}{2} + \frac{3}{32}\left(4 * 1 - \frac{1^3}{3}\right)\right) - \left(\frac{1}{2} + \frac{3}{32}\left(4 * (-1) - \frac{(-1)^3}{3}\right)\right)$$
$$= \frac{11}{16}$$

3. Compute $P(X > 0.5)$

$$P(X > 0.5) = 1 - P(X \le 0.5) = 1 - F(0.5) = 1 - \left(\frac{1}{2} + \frac{3}{32}(4 \ldots)\right)$$

4. $P(x > 2.2) = 1 - P(x \leq 2.2) = 1 - F(2.2)$

$= 1 - 1 = 0$

5. $P(x < -3) = F(-3) = 0$

6. $P(x > -5) = 1 - F(-5) = 1 - 0 = 1$

Going from CDF to PDF

$(CDF)' = PDF$

The cdf for X = measurement error is

$\int^x pdf = CDF$

$$F(x) = \begin{cases} 0' & x < -2 \\ \left(\frac{1}{2} + \frac{3}{32}\left(4x - \frac{x^3}{3}\right)\right)' & -2 \leq x < 2 \\ 1' & 2 \leq x \end{cases}$$

Determine the PDF $f(x)$.

$$f(x) = F'(x) = \begin{cases} 0' = 0 & x < -2 \\ \frac{3}{8} - \frac{3}{32}x^2 & -2 \leq x < 2 \\ 1' = 0 & x \geq 2 \end{cases}$$

$$= \begin{cases} \frac{3}{8} - \frac{3}{32}x^2 & -2 \leq x < 2 \\ 0 & \text{o.w} \end{cases}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad \left(\frac{x^3}{8}\right)' = \frac{3x^2}{8}$$

Suppose we have a pdf of

$$f(x) = \begin{cases} \frac{3}{8}x^2 & 0 \leq X \leq k \\ 0 & otherwise \end{cases}$$

a) Determine k.

$$\int_0^k \left(\frac{3}{8}x^2\right)dx = \frac{x^3}{8}\Big|_0^k = \frac{k^3}{8} - \frac{0}{8} = \frac{k^3}{8} = 1$$

$$\Rightarrow k = 2.$$

b) Give the cdf of this distribution.

$$F(x) = \begin{cases} 0 & x < 0 \\ \int_0^x \frac{3t^2}{8}\,dt & 0 \leq x \leq 2 \quad \Leftarrow \boxed{\frac{x^3}{8}} \\ 1 & x > 2 \end{cases}$$

c) Determine x_0 such that $P(X \leq x_0) = 0.125 = \frac{1}{8}$

$$F(x_0) = \frac{x_0^3}{8} = \frac{1}{8} \Rightarrow x_0 = 1.$$

Quantiles popper12: choose d for 1-5

Let F be a given cumulative distribution and let p be any real number between 0 and 1. The **(100p)th percentile** of the distribution of a continuous random variable X is defined as

$$F^{-1}(p) = min\{x|F(x) \geq p\}.$$

For continuous distributions, $F^{-1}(p)$ is the smallest number x such that $F(x) = p$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Determine the Percentiles

Given a cdf,

$$F(x) = \begin{cases} 0 & X < 0 \\ \frac{1}{8}x^3 & 0 \leq X \leq 2 \\ 1 & X > 2 \end{cases}$$

1. Determine the 90^{th} percentile.

2. Determine the 50^{th} percentile.

3. Find the value of c such that $P(X \leq c) = 0.75$.