Probability Theory

Applications for Data Science
Module 4 Continuous Random Variables

Anne Dougherty

February 27, 2021

TABLE OF CONTENTS

Random Variables

At the end of this module, students should be able to

- **Define a continuous random variable and give examples of a probability density function and a cumulative distribution function.**
- Identify and discuss the properties of a **uniform**, exponential, and normal random variable.
- Calculate the expectation and variance of a continuous rv.

Continuous Random Variables

Definition: A random variable is **continuous** if possible values comprise either a single interval on the number line or a union of disjoint intervals.

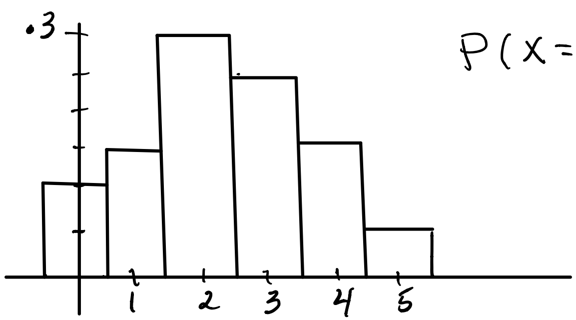
Examples:

- In the study of the ecology of a lake, a rv X could be the depth measurements at randomly chosen locations. $X \in [0, \text{maximum depth of lake}]$.
- In a study of a chemical reaction, Y could be the concentration level of a particular chemical in solution.
- In a study of customer service, W could be the time a customer waits for service.

Note: If X is continuous, $P(X = x) = 0$ for any x! Why?

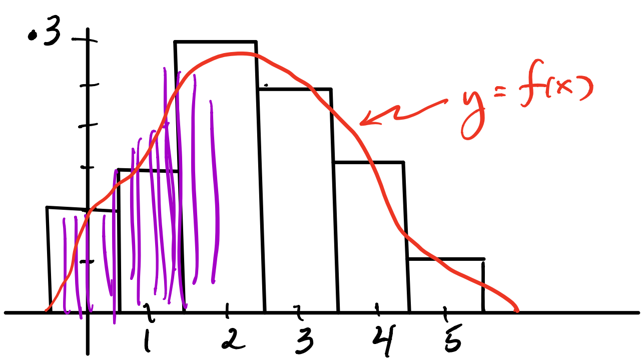
Motivating example: Suppose a train is scheduled to arrive at 1 pm. Let X be the minutes past the hour that it arrives and $X \in \{0, 1, 2, 3, 4, 5\}$.

x	0	1	2	3	4	5
p(x)	.1	.15	.3	.25	.15	.05



$P(X=2) = P(1.5 \le X \le 2.5) = .3$

$$P(1.5 \le X \le 2.5) = \int_{1.5}^{2.5} f(x)\,dx$$



$y = f(x)$ is called the probability density function for the continuous r.v. X if

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

Properties of the probability density function

For any continuous rv X with probability density function (pdf) f we have:

- The probability density function $f : (-\infty, \infty) \to [0, \infty)$, so $f(x) \geq 0$.
- $P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)\, dx = 1$ $\leftarrow$ $P(S) = 1$
- $P(a \leq X \leq b) = \int_a^b f(x)\, dx$

Note: $P(X = a) = \int_a^a f(x)\, dx = 0$ for all real numbers a.

Cumulative Distribution Function

Definition The cumulative distribution function (cdf) for a continuous rv X is given by $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$

1. $0 \leq F(x) \leq 1$
2. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
3. $F'(x) = f(x)$ by Fundamental Thm of Calculus

Uniform Random Variable

Definition A random variable X has the **uniform distribution** on the interval $[a, b]$ if it's density function is given by

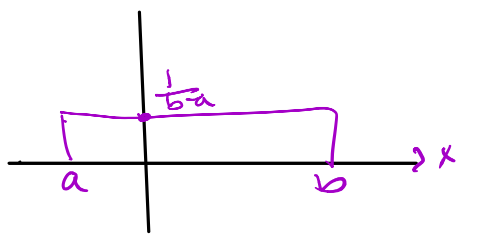
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{else} \end{cases}$$

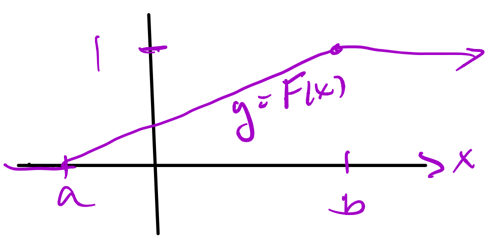
Notation: $X \sim U[a, b]$

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

$$= \int_a^x \frac{1}{b-a}\,dt \quad \text{for} \quad a \le x \le b$$

$$= \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } b < x \end{cases}$$





Example: Random number generators select numbers uniformly from a specific interval, usually $[0, 1]$.

Example: Suppose the diameter of aerosol particles in a particular application is uniformly distributed between 2 and 6 nanometers. Find the probability that a randomly measured particle has diameter greater than 3 nanometers.

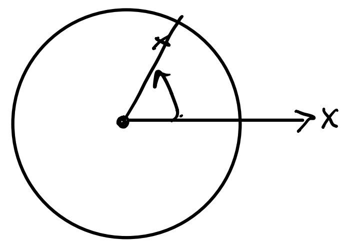
$X \sim U[2,6]$

$$f(x) = \begin{cases} \frac{1}{4} & \text{for } 2 \le x \le 6 \\ 0 & \text{else} \end{cases}$$

$$P(X \geq 3) = 1 - P(X \leq 3) = 1 - \int_2^3 \frac{1}{4}\,dx = \frac{3}{4}$$

$$= \int_3^6 \frac{1}{4}\,dx$$

Example: You throw a dart at a dartboard. The radial distance of the dart from the x-axis can be modeled by a uniform random variable.



$Y \sim U[0, 360]$

Expectation and variance for a continuous rv X:

Recall: if Y is discrete rv

$$E(Y) = \sum_k k\, P(Y=k) \quad \text{and} \quad V(Y) = \sum_{k=1}^{\infty} (k-\mu_Y)^2 P(Y=k)$$

If X is continuous then

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

and

$$\begin{aligned} V(X) &\overset{\text{def}}{=} \int_{-\infty}^{\infty} (x-\mu_X)^2 f(x)\, dx \\ &= \int_{-\infty}^{\infty} (x^2 - 2\mu_X x + \mu_X^2)\, f(x)\, dx \\ &= \underbrace{\int_{-\infty}^{\infty} x^2 f(x)\, dx}_{E(X^2)} - 2\mu_X \underbrace{\int_{-\infty}^{\infty} x\, f(x)\, dx}_{E(X)} + \mu_X^2 \underbrace{\int_{-\infty}^{\infty} f(x)\, dx}_{1} \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

Compute expectation and variance for $X \sim U[a, b]$

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{else} \end{cases}$$

$$E(X) = \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \underbrace{\frac{b+a}{2}}_{\text{this is what we expect since it is midway between } a \,\&\, b.}$$

We saw a similar result for discrete uniform r.v.

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a}\,dx = \frac{1}{b-a} \frac{x^3}{3}\Big|_a^b$$

$$= \frac{b^3-a^3}{3(b-a)} = \frac{1}{3}(b^2+ab+a^2)$$

$$V(X) = E(X^2) - (E(X))^2$$

$$= \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2$$

$$= \frac{4(b^2+ab+a^2) - 3(b^2+2ab+a^2)}{12}$$

$$= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}$$