Continuous Distributions: Expected Value, Variance, Standard Deviation

BIOS 6611

CU Anschutz

Week 3

Random Variables

Random Variable

Random variables and probability distributions are the theoretical or mathematical representations of data values and frequency distributions.

- **Random Variable**: variable whose values depend on a random phenomenon. Possible values represent possible outcomes of a yet-to-be-performed experiment.
- We call these possible values the **sample space**
- Notation:
  - X, Y, Z denotes a random variable
  - x, y, z denotes a value of the random variable

Probability Distribution

The **probability distribution** describes the probabilities for each outcome for

- a discrete random variable (probability mass function, *pmf*)
- the probabilities of values in a range for a continuous random variable (probability density function, *pdf*)

The pmf/pdf is a useful tool for describing the frequency distributions of random variables for the entire population of interest and for providing probability statements about events involving random variables.

Continuous Random Variable

- **Continuous random variable**: sample space is an interval (or continuum) with infinite possible values
  - Ex: Let X denote the height of individuals in Denver. The possible outcomes may be anything in the interval from (0, 2.72) meters.
  - Characterized by a **probability density function (PDF)**: function that gives probability that X is a value that falls in some range of the sample space.
  - Notation:

$$\begin{aligned} P(X \in [c,d]) &= P(c \leq X \leq d) \\ P(X \in [c,d]) &= \text{*function that depends on interval [c,d]*}, \\ [c,d] &= \{\text{*interval in the sample space*}\} \end{aligned}$$

Continuous Random Variable

The probability at any specific value of a continuous random variable is 0: $P(X = x) = 0$.

PDFs and CDFs

PDF

Let X be a continuous random variable and let x represent the possible values that X can take on.

The PMF of discrete r.v. does *not* apply to continuous r.v., instead we denote the function that assigns probability values to the r.v. as $f(x)dx$, or the **probability density function**.

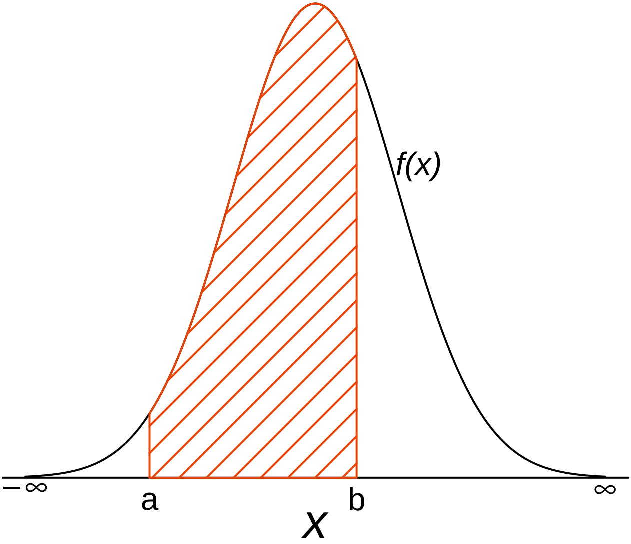
The probability distribution of X over the interval $[a, b]$ is therefore $P(a \leq X \leq b) = \int_a^b f(x)dx$.

Similar to discrete r.v., the following conditions must be met

- $f(x) \geq 0$
- $\int_a^b f(x)dx = 1$, where $[a, b]$ is the range of the entire possible sample space

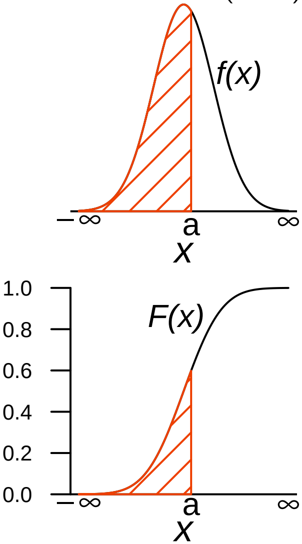
PDF and Intervals

$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a)$



Cumulative Distribution Function (CDF)

The **cumulative probability distribution** of X:
$F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$ (i.e., the area under the curve (AUC) from $-\infty$ to a)



- The CDF is a monotone increasing function: $F(-\infty) = 0, \ F(\infty) = 1$
- The CDF and PDF are directly related through the first derivative: $f(x) = \frac{d}{dx}F(x)$
- Every time we execute a statistical test or determine a p-value (level of significance), we will be using the cdf of a relevant continuous or discrete random variable.

PDF Example - Part 1

Assume that X is a r.v. with the PDF $f(x) = 2x$ where $0 \leq x \leq 1$. Plot the PDF and check that it is valid.

PDF Example - Part 2

Consider the PDF $f(x) = 2x$ where $0 \leq x \leq 1$. What is the probability that X falls between 0 and 0.5?

PDF Example - Part 3

Consider the PDF $f(x) = 2x$ where $0 \leq x \leq 1$. The CDF is $F(x) = x^2$, check the relationship with the PDF.

PDF Example - Part 4

Consider the PDF $f(x) = 2x$ where $0 \leq x \leq 1$. What is the probability that X is 0.5?

Expected Value (Mean), Variance, Standard Deviation (SD)

Formulas

Expected Value: $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \mu$

Variance: $Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \sigma^2$

Note, we also have the computational calculation of $Var(X) = E[X^2] - (E[X])^2$.

Standard Deviation: $s.d.(X) = \sqrt{Var(X)} = \sigma$

Expected Value Example

Consider the PDF $f(x) = 2x$ where $0 \leq x \leq 1$. What is $E(X)$?

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

Variance Example

Consider the PDF $f(x) = 2x$ where $0 \leq x \leq 1$. What is $Var(X)$?

$Var(X) = E[X^2] - (E[X])^2 =$

Formula Review for Random Variables

Discrete r.v.

$$E(X) = \mu = \sum_x xP(X = x)$$

$$Var(X) = \sigma^2 = \sum_x (x - \mu)^2 P(X = x)$$
$$= E[X^2] - (E[X])^2$$

$$SD(X) = \sigma = \sqrt{Var(X)}$$

Continuous r.v.

$$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)dx$$

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$
$$= E[X^2] - (E[X])^2$$

$$SD(X) = \sigma = \sqrt{Var(X)}$$