CEE 260/MIE 273: Probability and Statistics in Civil Engineering
Lecture 3a: Introduction: Random variables

Prof. Oke

UMass Amherst

College of Engineering

September 23, 2025

M2c Recap: Conditional Probability and Bayes' Theorem

- Total probability:

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_n)P(E_n) \tag{1}$$

- Bayes' Theorem for two events:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{2}$$

For multiple events:

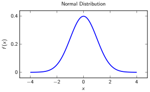
$$P(E_1|A) = \frac{P(A|E_1)P(E_1)}{P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_n)P(E_n)} \tag{3}$$

$$= \frac{P(A|E_1)P(E_1)}{P(A)} \tag{4}$$

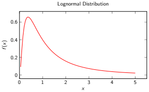
Overview of Module 3

Overview

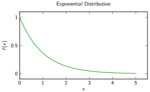
- Lecture 3a: Introduction: Random Variables

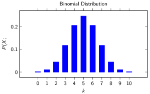


- Lecture 3b: Normal Distribution

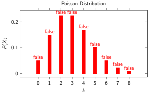


- Lecture 3c: Lognormal and Exponential Distributions





- Lecture 3d: Binomial Distribution



- Lecture 3e: Poisson Distribution
- Lecture 3f: Joint Distributions and further topics

Objectives and outline

- Understand random variables
- Distinguish between discrete and continuous random variables
- Compute measures of centrality and dispersion, as well as sketch PMFs, PDFs and CDFs

Random variables

Definitions

- A random variable (r.v.) represents the values of the outcomes in a sample space (e.g. the outcome of die roll: $X = \{1, 2, 3, 4, 5, 6\}$
- A random variable is a function that uniquely maps events in a sample space to the set of real numbers.

A random variable X may be:

- *Discrete*
- *Continuous*

Describing random variables

Central values

- Mean
- Median
- Mode

Measures of dispersion

- Variance
- Standard deviation
- Coefficient of variation (COV)

Probability distribution

A probability distribution governs the values of a random variable.
It can be described by the following functions:

- probability mass function, PMF discrete random variable
- probability density function, PDF continuous random variable
- cumulative distribution function, CDF discrete/continuous random variable

Cumulative distribution function (CDF)

The CDF (F_X) of a random variable X is given by

$$F_X \equiv P(X \leq x) \quad \text{for all } x \tag{5}$$

The CDF satisfies the basic axioms of probability:

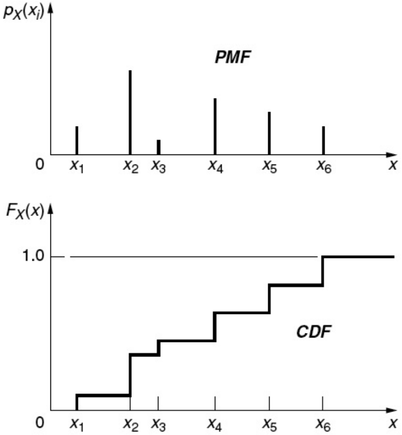
1. $F_X(-\infty) = 0$ and $F_X(\infty) = 1$
2. $F_X(x) \geq 0 \quad \forall x$ and is nondecreasing with x.[1]
3. $F_X(x)$ is continuous to the right with x.

[1] Note that the symbol $\forall$ means "for all"

Probability mass function (PMF)

The PMF is given by

$$p_X(x_i) \equiv P(X = x_i) \quad \forall x \qquad (6)$$



CDF of discrete random variable

$$\begin{aligned} F_X(x) &= \sum_{x_i \leq x} P(X = x_i) \\ &= \sum_{x_i \leq x} p_X(x_i) \end{aligned}$$

The probability masses in a PMF sum up to 1.

Probability density function (PDF)

The PDF is denoted $f_X(x)$ such that the probability of X in the interval $(a, b]$ is:

$$P(a < X \leq b) = \int_a^b f_X(x)dx \tag{7}$$

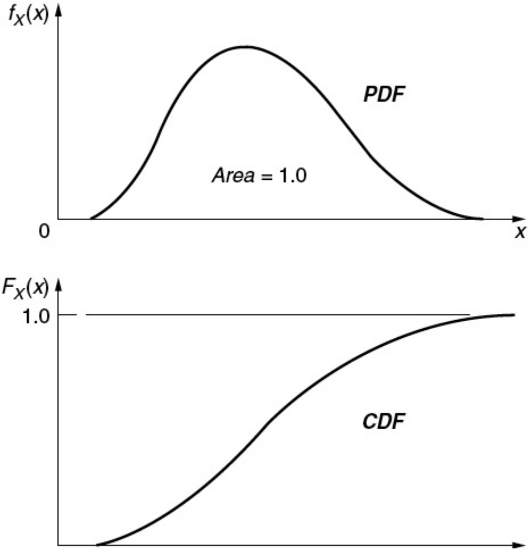
CDF of continuous random variable

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f_X(\tau)d\tau \end{aligned}$$

It follows that the PDF is the derivative of the CDF:

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{8}$$

PDF (cont.)



The total area under a PDF is 1.

Further derivations

1. Continuous case:

$$P(a < X \leq b) = \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx \tag{9}$$

2. Discrete case:

$$P(a < X \leq b) = \sum_{x_i \leq b} p_X(x_i) - \sum_{x_i \leq a} p_X(x_i) \tag{10}$$

3. For all random variables:

$$P(a < X \leq b) = F_X(b) - F_X(a) \tag{11}$$

Example 1: Operating condition of bulldozers

Each of 3 bulldozers equally likely to operational or nonoperational after 6 months. Plot the PMF and CDF of the random variable X which represents the operating condition of the bulldozers after 6 months.

- Let the outcomes be O (operational) and N (nonoperational)
- There are $2 \times 2 \times 2 = 8$ possibilities:
  1. *OON*
  2. *OOO*
  3. *ONO*
  4. *ONN*
  5. *NOO*
  6. *NON*
  7. *NNO*
  8. *NNN*

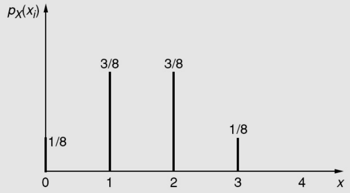


Figure: PMF

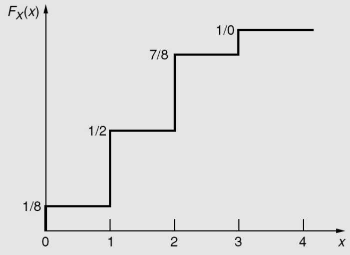


Figure: CDF

Mean and variance

Mean

Weighted average or expected value

$$E(X) \quad = \quad \sum_i x_i p_X(x_i) \quad \text{discrete case} \tag{12}$$

Variance

In the discrete case:

$$Var(X) = \sum_i (x_i - \mu_X)^2 p_X(x_i) \tag{13}$$

Expanding results in:

$$Var(X) = E(X^2) - \mu_X^2 \tag{14}$$

Measures of dispersion (cont.)

Standard deviation

The standard deviation is convenient as it has the same unit as the random variable:

$$\sigma_X = \sqrt{Var(X)} \tag{15}$$

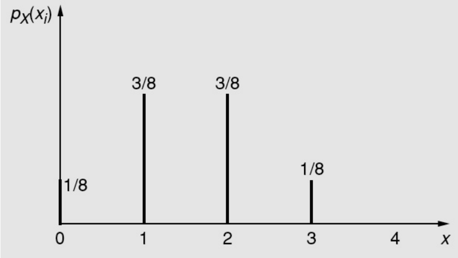
Coefficient of variation

The COV gives the deviation relative to the mean. It is unitless.

$$\delta_X = \frac{\sigma_X}{\mu_X} \tag{16}$$

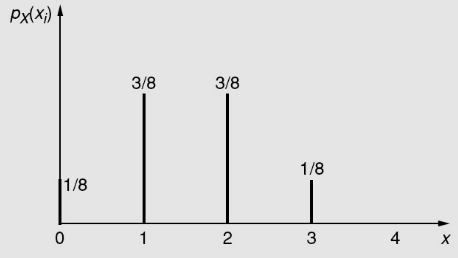
Example 2: Bulldozers revisited

You are given the PMF (probability mass function) of the operating condition X of bulldozers after 6 months.



Find the mean, variance, standard deviation and coefficient of variation of X.

Example 2: Bulldozers revisited (cont.)



(a) $\mu_X = E(X) = 0\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = 1.5.$

(b) $Var(X) = [0^2\left(\frac{1}{8}\right) + 1^2\left(\frac{3}{8}\right) + 2^2\left(\frac{3}{8}\right) + 3^2\left(\frac{1}{8}\right)] - (1.5)^2 = 0.75$

(c) $\sigma_X = \sqrt{0.75} = 0.866$

(d) $\delta_X = \frac{0.866}{1.50} = 0.577$

Mean and variance

These include the mean, median and mode.

- Mean: weighted average or expected value

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad \text{continuous case} \tag{17}$$

Variance

In the continuous case:

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \tag{18}$$

Expanding both equations results in:

$$Var(X) = E(X^2) - \mu_X^2 \tag{19}$$

Example 3: Loaded beam

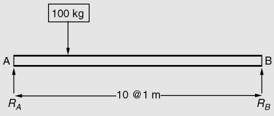


Figure E2.5a Beam AB.

Consider the beam under a 100-kg load. If the load is equally likely to be placed anywhere along the 10m span of the beam,then the PDF of the load position X is uniformly distributed in $0 < x \leq 10$, i.e.:

$$f_X(c) = \begin{cases} c & 0 < x \leq 10 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

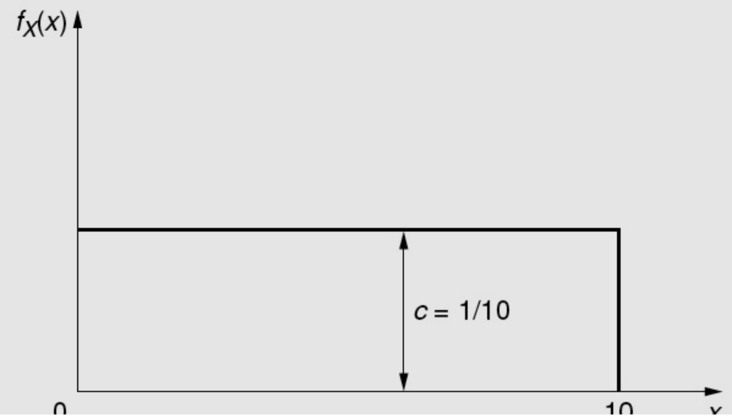
(a) Plot the PDF of X.

(b) Solve the integral for the CDF and plot.

(c) Find $P(2 < X \leq 5)$.

Example 3: Loaded beam (cont.)

(a) The area under the PDF must be 1. Thus, c must be $\frac{1}{10}$.



Example 3: Loaded beam (cont.)

(b) The CDF is given by:

$$F_X = \int_0^x c dx = cx = \frac{x}{10} \qquad 0 < x \leq 10$$

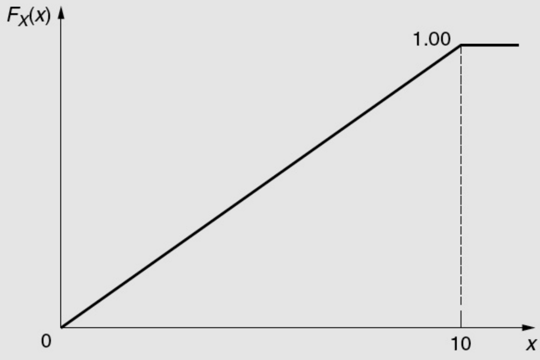


Figure E3.2b CDF of X.

Example 3: Loaded beam (cont.)

(c) To compute $P(2 < X \leq 5)$, we use the CDF:

$$\begin{aligned} P(2 < X \leq 5) &= F_X(5) - F_X(2) \\ &= \frac{5-2}{10} = 0.3 \end{aligned}$$

Example 4: Useful life of machines

The useful life T of welding machines is a random variable with an exponential distribution. The PDF and CDF are:

$$
\begin{aligned}
f_T(t) &= \lambda e^{-\lambda t} \qquad t \geq 0 \\
F_T(t) &= 1 - e^{-\lambda t} \qquad t \geq 0
\end{aligned}
$$

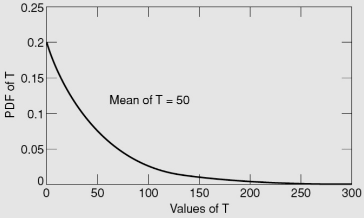


Figure E3.3a Exponential PDF of useful life T.

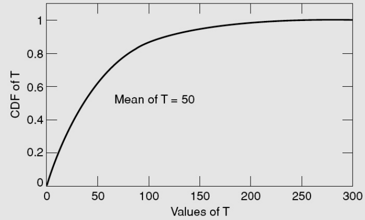


Figure E3.3b CDF of useful life T.

(a) Find the mean of this distribution

(b) Find the median

(c) Show that the variance is $\frac{1}{\lambda^2}$

Example 4: Useful life of machines

$$\begin{aligned} \text{PDF} : f_T(t) &= \lambda e^{-\lambda t} \qquad t \geq 0 \\ \text{CDF} : F_T(t) &= 1 - e^{-\lambda t} \qquad t \geq 0 \end{aligned}$$

(a) The mean is given by $\mu_T = E(T) = \int_0^{\infty} t\lambda e^{-\lambda t} dt$.We use integration by parts: $\int udv = uv - \int vdu$.

$$\begin{aligned} \mu_T &= \int_0^{\infty} t\lambda e^{-\lambda t} dt \\ &= \lambda \int_0^{\infty} \lambda e^{-\lambda t} dt \\ &= \lambda \left[t \left(-\frac{1}{\lambda} e^{-\lambda t} \right) \right]_0^{\infty} - \left[-\frac{1}{\lambda} e^{-\lambda t} dt \right] \\ &= \lambda \left(0 + \frac{1}{\lambda} \frac{-e^{-\lambda t}}{\lambda} \Big|_0^{\infty} \right) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} \end{aligned}$$

Recap

- Random variables
- Probability mass/density function (PMF/PDF) and cumulative distribution function (CDF)
- Measures of centrality
- Measures of dispersion

Reading

- Open Intro Statistics Section 3.4 (Random variables)
- Open Intro Statistics Section 3.5 (Continuous distributions)

Skewness

The skewness or symmetry of a distribution is measured by the third central moment:
In the discrete case:

$$E(X - \mu_X)^3 = \sum_i (x_i - \mu_X)^3 p_X(x_i) \tag{21}$$

In the continuous case:

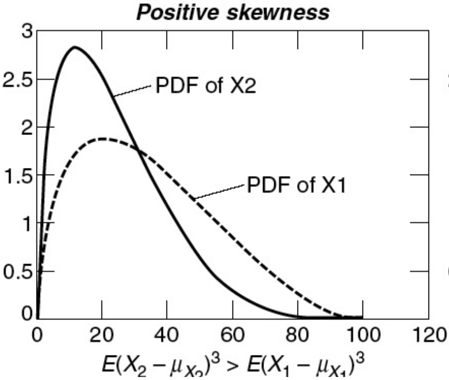
$$E(X - \mu_X)^3 = \int_{-\infty}^{\infty} (x - \mu_X)^3 f_X(x) dx \tag{22}$$

For convenience, the skewness coefficient is also used (unitless):

$$\theta = \frac{E(X - \mu_X)^3}{\sigma^3} \tag{23}$$

Skewness (cont.)

- Positive skewness is characterized by a long right tail (right-skewed)
- Negative skewness is characterized by a long left tail (left-skewed)



Kurtosis

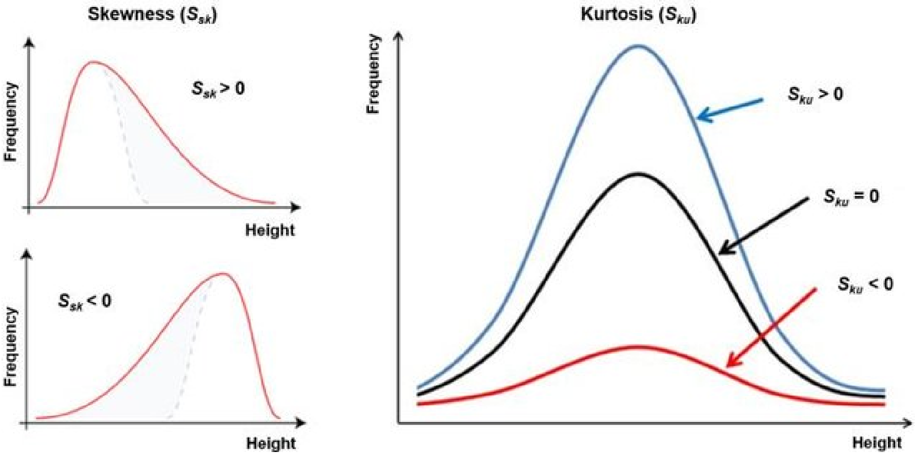
This is the measure of peakedness in a distribution. It is the fourth central moment:
In the discrete case:

$$E(X-\mu_X)^4 = \sum_i (x_i - \mu_X)^4 p_X(x_i) \tag{24}$$

In the continuous case:

$$E(X-\mu_X)^4 = \int_{-\infty}^{\infty} (x-\mu_X)^4 f_X(x)dx \tag{25}$$

Skewness vs. kurtosis



Source: Bonyar, A (2015) "Application of localization factor for the detection of tin oxidation with AFM" DOI: 10.1109/SIITME.2015.7342289

Generalized expectation

The mathematical expectation can be defined for a function g of random variable X:

$$
\begin{aligned}
E[g(X)] &= \sum_i g(x_i) p_X(x_i) \quad \text{discrete case} && (26)\\
E[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad \text{continuous case} && (27)
\end{aligned}
$$