1.017/1.010 Class 7
Random Variables and Probability Distributions

Random Variables

A **random variable** is a function (or rule) $\boldsymbol{x}(\xi)$ that associates a real number $\boldsymbol{x}$ with each outcome ξ in the sample space S of an experiment. Assignment of such rules enables us to quantify a wide range of real-world experimental outcomes.

Example:

Experiment: Toss of a coin
Outcome: Heads or tails
Random Variable: $\boldsymbol{x}(\xi) = 1$ if outcome is heads, $\boldsymbol{x}(\xi) = 0$ if outcome is tails
Event: $\boldsymbol{x}(\xi)$ greater than 0

Probability Distributions

Random variables are characterized/defined by their probability distributions.

Cumulative distribution function (CDF)

Consider events:

$\boldsymbol{x}(\xi)$ less than x : $A = \{\boldsymbol{x}(\xi) \leq x\}$

$\boldsymbol{x}(\xi)$ lies in the interval $[x_l, x_u]$: $B = \{x_l < \boldsymbol{x}(\xi) \leq x_u\}$

For **any** random variable $\boldsymbol{x}$ . . . the **cumulative distribution function** (CDF) gives the probability that $\boldsymbol{x}(\xi)$ is **less than** a specified value x:

$$F_{\boldsymbol{x}}(x) = P[\boldsymbol{x}(\xi) \leq x] = P(A)$$

The probability that $\boldsymbol{x}(\xi)$ is **greater than** x:

$$P[\boldsymbol{x}(\xi) > x] = 1 - F_{\boldsymbol{x}}(x)$$

The probability that $\boldsymbol{x}(\xi)$ **lies in interval** $[x_l, x_u]$ is:

$$P[x_l < \boldsymbol{x}(\xi) \leq x_u] = F_{\boldsymbol{x}}(x_u) - F_{\boldsymbol{x}}(x_l) = P(B)$$

Example: Discrete uniform CDF

$F_x(x) = 0.0 \qquad x < 0$
$F_x(x) = 0.3 \qquad 0 < x \le 1$
$F_x(x) = 0.7 \qquad 1 < x \le 2$
$F_x(x) = 1.0 \qquad x > 2$

Example: Continuous uniform CDF

$F_x(x) = 0.4x \qquad 0 < x \le 2.5,\ \ 0$ otherwise

Probability mass function (PMF)

For a **discrete** $\boldsymbol{x}$ with possible outcomes $x_1 \ldots x_N$, PMF is probability of x_i:

$$p_{\boldsymbol{x}}(x_i) = P[\boldsymbol{x}(\xi) = x_i] = F_{\boldsymbol{x}}(x_i) - F_{\boldsymbol{x}}(x_i^-)$$

The probability that $\boldsymbol{x}(\xi)$ lies in the interval $[x_l, x_u]$ is:

$$P[x_l < \boldsymbol{x}(\xi) \le x_u] = \sum_{x_l < \boldsymbol{x}_i \le x_u} p_{\boldsymbol{x}}(x_i) = P(B)$$

Example: Discrete uniform PMF:

$p_x(0) = 0.3$
$p_x(1) = 0.4$
$p_x(2) = 0.3$

Probability density function (PDF)

For a **continuous** $\boldsymbol{x}$ . . . PDF is derivative of CDF:

$$f_x(x) = \frac{dF_x(x)}{dx}$$

The probability that $\boldsymbol{x}(\xi)$ lies in the interval $[x_l, x_u]$ is:

$$P[x_l < \boldsymbol{x}(\xi) \le x_u] = \int_{x_l}^{x_u} f_x(x)dx = P(B)$$

Example: Continuous uniform PDF

$f_x(x) = .4 \qquad 0 < x \le 2.5,\ \ 0$ otherwise

Exercise: Constructing probability distributions from virtual experiments

Consider a sequence of 4 repeated independent trials, each with the outcome 0 or 1. Suppose that $P(0) = 0.3$ and $P(1) = 0.7$. Define the discrete random variable $\boldsymbol{x}$ = sum of the 4 trial outcomes (varies from 0 to 4).

There are 2^4=16 possible experimental outcomes, each giving a particular value of $\boldsymbol{x}$. In some cases, several different outcomes give the same value of $\boldsymbol{x}$ (e.g. $C_{1,4} = 4$ outcomes give $\boldsymbol{x}$ = 1). This experiment yields a **binomial probability distribution**.

Plot the PMF and CDF of $\boldsymbol{x}$, using the rules of probability to evaluate $F_x(x)$ and $p_x(x_i)$ for $x_i = 0, 1, 2, 3, 4$.

Duplicate these results with a virtual experiment that generates many sequences of 4 trials each. Derive $F_x(x)$ and $p_x(x_i)$ by evaluating the fraction of replicates that yield $x_i = 0, 1, 2, 3, 4$.

Generalize your pencil and paper analysis to give a general expression for $p_x(x_i)$ when there are 100 rather than 4 repeated independent trials and $x_i = 0, 1, 2, 3, \ldots 100$. Plot the CDF and PDF.

Confirm your results with a virtual experiment.
