

1.017/1.010 Class 7

Random Variables and Probability Distributions

Random Variables

A **random variable** is a function (or rule) $x(\xi)$ that associates a real number 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: $x(\xi) = 1$ if outcome is heads, $x(\xi) = 0$ if outcome is tails

Event: $x(\xi)$ greater than 0

Probability Distributions

Random variables are characterized/defined by their probability distributions.

Cumulative distribution function (CDF)

Consider events:

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

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

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

$$F_x(x) = P[x(\xi) \leq x] = P(A)$$

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

$$P[x(\xi) > x] = 1 - F_x(x)$$

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

$$P[x_l < x(\xi) \leq x_u] = F_x(x_u) - F_x(x_l) = P(B)$$

Example: Discrete uniform CDF

$$\begin{aligned}
 F_x(x) &= 0.0 & x < 0 \\
 F_x(x) &= 0.3 & 0 < x \leq 1 \\
 F_x(x) &= 0.7 & 1 < x \leq 2 \\
 F_x(x) &= 1.0 & x > 2
 \end{aligned}$$

Example: Continuous uniform CDF

$$F_x(x) = 0.4x \quad 0 < x \leq 2.5, \quad 0 \text{ otherwise}$$

Probability mass function (PMF)

For a **discrete** x with possible outcomes $x_1 \dots x_N$, PMF is probability of x_i :

$$p_x(x_i) = P[\mathbf{x}(\xi) = x_i] = F_x(x_i) - F_x(x_i^-)$$

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

$$P[x_l < \mathbf{x}(\xi) \leq x_u] = \sum_{x_l < x_i \leq x_u} p_x(x_i) = P(B)$$

Example: Discrete uniform PMF:

$$\begin{aligned}
 p_x(0) &= 0.3 \\
 p_x(1) &= 0.4 \\
 p_x(2) &= 0.3
 \end{aligned}$$

Probability density function (PDF)

For a **continuous** $x \dots$ PDF is derivative of CDF:

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

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

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

Example: Continuous uniform PDF

$$f_x(x) = .4 \quad 0 < x \leq 2.5, \quad 0 \text{ otherwise}$$

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 $x = \text{sum of the 4 trial outcomes}$ (varies from 0 to 4).

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

Plot the PMF and CDF of 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 $x_i = 0, 1, 2, 3, \dots, 100$. Plot the CDF and PDF.

Confirm your results with a virtual experiment.



Copyright 2003 Massachusetts Institute of Technology
Last modified Oct. 8, 2003