Probability Distributions

1. Discrete probability distributions
2. Continuous probability distributions
3. In-class exercise
4. Expectations and moments

Probability Distributions

Discrete Probability Distributions

Background

- Experimental measurements can be thought of as random samples from an underlying, unknown probability distribution
- The actual probability distribution can be reconstructed from an infinite number of samples
- More practically the probability distribution must be assumed or deduced from the available samples
- Key properties of the probability distribution can be estimated from a finite, usually small number of samples
- A few distributions are commonly used for statistical data analysis

Probability Distributions

- Random variables
  - » Experimental measurements are not reproducible in a deterministic fashion
  - » Each measurement can be viewed a random variable X
  - » Defined on sample space S of an experiment
- Probability distribution $f(x)$
  - » Determines probability of particular events
  - » Discrete distributions: random variables are discrete quantities
  - » Continuous distributions: random variables are continuous quantities
- Cumulative probability distribution function $F(x)$

$$F(x) = P(X \le x) \quad P(a < X \le b) = F(b) - F(a)$$

UMASS.

Discrete Probability Distributions

- Random variable X can only assume countably many discrete values: $x_1, x_2, x_3, \ldots$
- Probability distribution function $f(x)$

$$f(x) = \begin{cases} p_j \equiv P(X = x_j) & \text{if } x = x_j \\ 0 & \text{otherwise} \end{cases}$$

- Cumulative distribution function:

$$F(x) = \sum_{x_j \le x} f(x_j) = \sum_{x_j \le x} p_j$$

- Properties:

$$P(a < X \le b) = F(b) - F(a) = \sum_{a < x_j \le b} p_j$$

$$\sum_j p_j = 1$$

Discrete Distribution Example 1

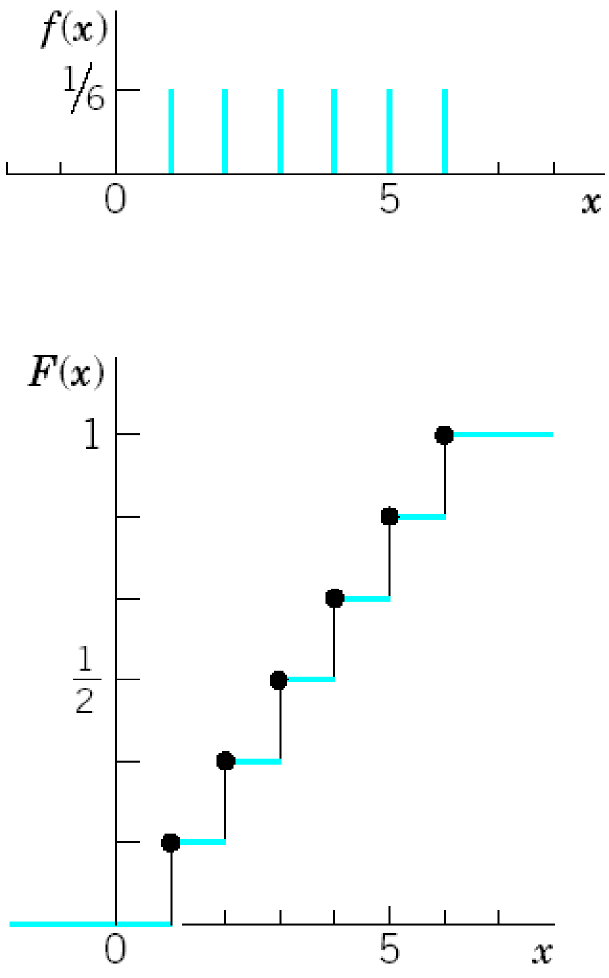

Fig. 512. Probability function $f(x)$ and distribution function $F(x)$ of the random variable X = *Number obtained in tossing a fair die once*

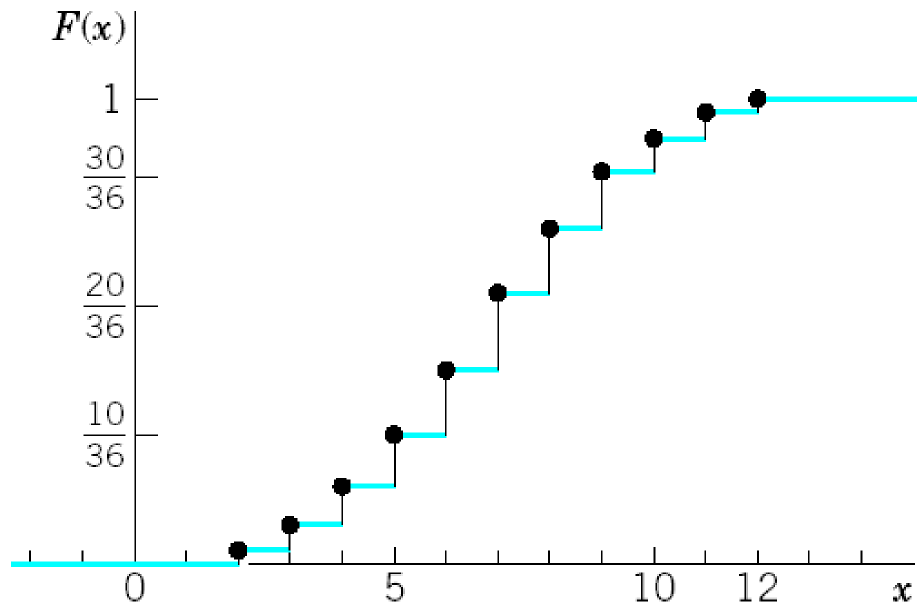

Fig. 513. Probability function $f(x)$ and distribution function $F(x)$ of the random variable X = *Sum of the two numbers obtained in tossing two fair dice once*

Discrete Distribution Example 2

- Assume that the probability of performing exactly x successful experiments in a row is governed by:

$$f(x) = \begin{cases} p_1 = 0.1 & x = 1 \\ p_2 = 0.3 & x = 2 \\ p_3 = 0.4 & x = 3 \\ p_4 = 0.2 & x = 4 \\ 0 & \text{otherwise} \end{cases} \Rightarrow \quad F(x) = \begin{cases} 0.1 & x \le 1 \\ 0.4 & x \le 2 \\ 0.8 & x \le 3 \\ 1.0 & x \le \infty \end{cases}$$

- What is the probability that you will perform either 2 or 3 successful experiments in a row?

$$P(1 < X \le 3) = f(x_2) + f(x_3) = 0.3 + 0.4 = 0.7$$

$$P(1 < X \le 3) = F(3) - F(1) = 0.8 - 0.1 = 0.7$$

Probability Distributions

Continuous Probability Distributions

Continuous Probability Distributions

- Random variable *X* can assume infinitely many real values
- Cumulative distribution function:

$$F(x) = \int_{-\infty}^{x} f(v)dv$$

- Probability distribution function:

$$f(x) = \frac{dF(x)}{dx}$$

- Properties

$$P(a < X \le b) = F(b) - F(a) = \int_{a}^{b} f(v)dv$$

$$\int_{-\infty}^{\infty} f(v)dv = 1$$

Continuous Distribution Example 1

- Probability distribution function:

$$f(x) = \begin{cases} 0.75(1-x^2) & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad \int_{-\infty}^{\infty} f(v)dv = \int_{-1}^{1} 0.75(1-v^2)dv = 1$$

- Cumulative distribution function:

$$F(x) = \int_{-\infty}^{x} f(v)dv = \int_{-1}^{x} 0.75(1-v^2)dv = 0.5 + 0.75x - 0.25x^3$$

$$F(x) = \begin{cases} 0 & x \le -1 \\ 0.5 + 0.75x - 0.25x^3 & -1 < x \le 1 \\ 1 & x > 1 \end{cases}$$

- Probability of events:

$$P(-0.5 \le x \le 0.5) = F(0.5) - F(-0.5) = \int_{-0.5}^{0.5} 0.75(1-v^2)dv = 0.6875$$

$$P(X \le x) = F(x) = 0.95 = 0.5 + 0.75x - 0.25x^3 \Rightarrow x = 0.73$$

Continuous Distribution Example 2

- Assume that the probability of obtaining a thin film thickness measurement of *x* microns is governed by:

$$f(x) = \begin{cases} cx(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- Determine the constant *c* such that *f*(*x*) is a legitimate probability distribution function

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^1 f(x)dx = \int_0^1 cx(1-x)dx = 1$$

$$c\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Bigg|_{x=0}^{x=1} = c\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{c}{6} = 1 \quad \Rightarrow \quad c = 6$$

Continuous Distribution Example 2

- Compute the cumulative probability function $F(x)$

$$F(x) = \int_{-\infty}^{\infty} f(v)dv = \int_{0}^{x} 6v(1-v)dv = 6\left(\frac{x^2}{2} - \frac{x^3}{3}\right)$$

- Compute the probability that the measured thickness x will be 0.5 microns or less

$$P(x \leq 0.5) = F(0.5) = 6\left(\frac{0.5^2}{2} - \frac{0.5^3}{3}\right) = 0.5$$

Probability Distributions

In-class Exercise

Probability Distributions

Expectations and Moments

Mean and Variance of a Distribution

- Discrete distribution

Mean $$\mu = \sum_j x_j f(x_j)$$

Variance $$\sigma^2 = \sum_j (x_j - \mu)^2 f(x_j)$$

- Continuous distribution

Mean $$\mu = \int_{-\infty}^{\infty} x f(x) dx$$

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

- Symmetric distribution
  - » If $f(c-x) = f(c+x)$, then $f(x)$ is symmetric with respect to the mean $\mu = c$

Expectations and Moments

- Moments for continuous distributions

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

*k*th moment $$E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx$$

*k*th central moment $$E([X-\mu]^k) = \int_{-\infty}^{\infty} (x-\mu)^k f(x)dx$$

- Interpretation
  - » The mean is the first moment
  - » The variance is the second central moment
  - » The mean and variance calculated from samples are called the sample mean and variance

Sample mean $$\bar{x} = \tfrac{1}{n}\sum_{i=1}^{n} x_i$$ Sample variance $$s^2 = \tfrac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})^2$$

Expectations and Moments Example 1

- Given the following probability distribution, compute the true mean and variance

$$f(x) = \begin{cases} 0.75(1-x^2) & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mu = E(X) = \int_{-1}^{1} 0.75x(1-x^2)dx = \left[\tfrac{1}{2}(0.75)x^2 - \tfrac{1}{4}(0.75)x^4\right]_{-1}^{+1} = 0$$

$$\sigma^2 = E([\mu - X]^2) = \int_{-1}^{1} (x-0)^2 0.75(1-x^2)dx = \left[\tfrac{1}{3}(0.75)x^3 - \tfrac{1}{5}(0.75)x^5\right]_{-1}^{+1} = 0.2$$

Expectations and Moments Example 2

- Given n samples randomly drawn from a distribution with true mean $\mu = 1$ and variance $\sigma^2 = 4$, compute the sample mean $\bar{x}$ and variance s^2
- Effect of sample size n:

Size	Sample Mean	Sample Variance
5	1.619	4.630
10	1.220	4.827
50	1.005	3.965
100	0.836	4.626
1000	1.050	3.884
10000	0.995	3.962