Probability Distributions
Mike Veatch

Probability distributions

support of r.v. set S of possible values of the r.v.

discrete r.v. takes on discrete values, e.g., integers. Support S is a finite (or countable) set

probability mass function (pmf): $f(x) = P(X = x)$

continuous r.v. takes on any numerical value in an interval. Support S is an interval

probability density function (pdf): $P(a \le X \le b) = \int_a^b f(x)dx$

(cumulative) distribution function (df): $F(x) = P(X \le x)$

continuous r.v.: $F(x) = \int_{-\infty}^{x} f(y)dy$ **discrete r.v.:** $F(x) = \sum_{y:y \le x} f(y)$

mean (expectation) continuous r.v. $E(X) = \int_{-\infty}^{\infty} xf(x)dx$ **discrete r.v.** $E(X) = \sum_x xf(x)$

expectation of function of r.v.: $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$ or $E[g(X)] = \sum_x g(x)f(x)$

mean of a r.v. $\mu = E(X)$

variance of a r.v. $\sigma^2 = V(X) = E[(X - \mu)^2]$

standard deviation of a r.v. $\sigma = \sqrt{V(X)}$

Histograms are closely related to probability distributions. If we had a histogram of the *population*, the probability of a randomly selected individual falling in a class is the relative frequency. We call this probability distribution the **population distribution**. A sample histogram approximates it.

Populations and samples

population collection of individuals of interest.
sample subset of the population that is selected by some sampling method and observed

(sample) statistic Anything computed from the sample. Includes **descriptive statistics**, $\bar{x}$, s

Sampling distributions

Variables have distributions over the population. The values of the variable in the sample are **random variables**; call them $X_1, \ldots, X_n$. *After* choosing a sample, the observed values $x_1, \ldots, x_n$ are known.

Sample statistics, such as $\bar{X}$, are random variables (computed values, e.g., $\bar{x}$, are not). Their probability distribution is called a **sampling distribution**.

For a random sample from a large population, the X_i are **independent** and their probability distribution is the same as the population distribution. A key fact is that the sampling distribution of $\bar{X}$ has

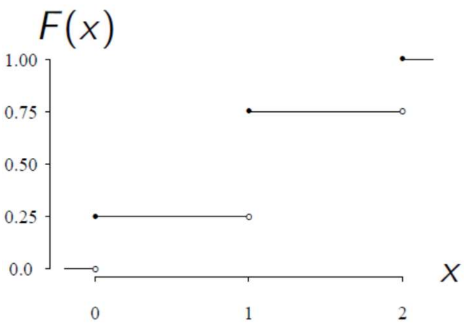
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

	Sample statistic	Population/distribution parameter
mean	$\bar{x}$	μ
variance	s^2	σ^2
standard deviation	$s = \sqrt{s^2}$	σ

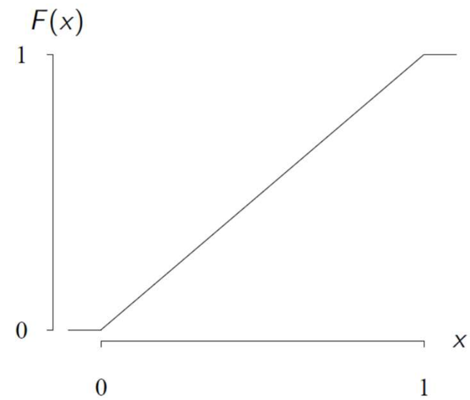
Exercises

1. Find the pmf or pdf $f(x)$ for each distribution function.

a)



b)



2. Consider the following probability distribution of the number of lesions observed by a physician on subjects who had regularly used tanning salons.

Lesions	0	1	2	3	4
Probability	0.05	0.1	0.25	0.30	0.30

a) What is the support of this r.v.?
b) Find the probability that a subject has three or more lesions.
c) Find the mean number of lesions.
d) Graph the probability mass function.

3. For the probability density function $f(x) = 1 - x/2,\ 0 \le x \le 2$
a) Find $P(X \le 1)$. Note: you may use integrals or geometry.
b) Optional: Find $E(X)$.
c) Optional: Find and graph the distribution function $F(x)$.

4. Suppose X is a normally distributed r.v.
a) What is the support of X?
b) What is the support of X^2? That is, square the values of the normal r.v.

5. Suppose Y is the number of times a person voted for the winner in the last 12 presidential elections. What is the support of Y?

6. What happens to the standard deviation when a variable is rescaled? Let X be the amount of an ATM withdrawal in Euros and suppose its standard deviation is $\sigma_X = 50$.
a) Let $Y = 1.2X$ be the withdrawal in dollars. Find the standard deviation of Y.
b) There is a \$3 withdrawal fee, so the charge to the account is $T = 1.2X - 3$ dollars. Find the standard deviation of T. Hint: remember that standard deviation measures *spread* of a distribution.
c) Optional: Generalizing (a) and (b), what is a formula for the standard deviation of $aX + b$, where a and b are constants? The variance of $aX + b$?