Introductory Statistics Lectures

Random variables

Theoretical distributions for populations versus histograms for samples.

ANTHONY TANBAKUCHI
DEPARTMENT OF MATHEMATICS
PIMA COMMUNITY COLLEGE

(Compile date: Tue May 19 14:49:16 2009)

Contents

1 Random variables

1.1 Random variables

RANDOM VARIABLE x. DEFINITION 1.1
is a variable determined by chance for each outcome of a procedure.

Example 1. Examples of random variables:
- Height
- Weight
- Number of males in a class

Types of random variables

discrete are countable. Have an associated **probability distribution**. (ex. Number of males)

continuous have infinitely many values. Have an associated **probability density function**. (ex. Height)

Types of questions we want to answer
"Find the probability that the mean is ..."

english statement	mathematical notation
"equal to fifty"	$P(\mu = 50)$
"not equal to fifty"	$P(\mu \neq 50)$
"not fifty"	$P(\mu \neq 50)$
"greater than fifty"	$P(\mu > 50)$
"at least fifty"	$P(\mu \geq 50)$
"less than fifty"	$P(\mu < 50)$
"no more than fifty"	$P(\mu \leq 50)$
"between forty and fifty" (inclusive)	$P(40 \leq \mu \leq 50)$
"between forty and fifty"	$P(40 < \mu < 50)$

1.2 Probability distributions

DEFINITION 1.2

PROBABILITY DISTRIBUTION $P(x_i)$.
describes the probability of a discrete random variable x. $P(x_i)$ is the probability of observing x_i. Describes **population**.

Two key properties of probability distributions

$$\boxed{\sum_{i=1}^{k} P(x_i) = 1} \tag{1}$$

with k possible values of x.

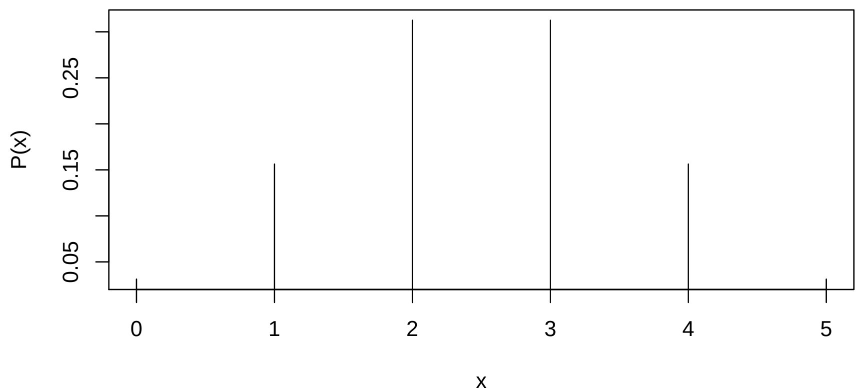
$$\boxed{0 \leq P(x_i) \leq 1, \quad \text{for } i = 1, 2, \ldots, k} \tag{2}$$

Ways we can represent $P(x_i)$
Three basic ways we can represent $P(x_i)$.

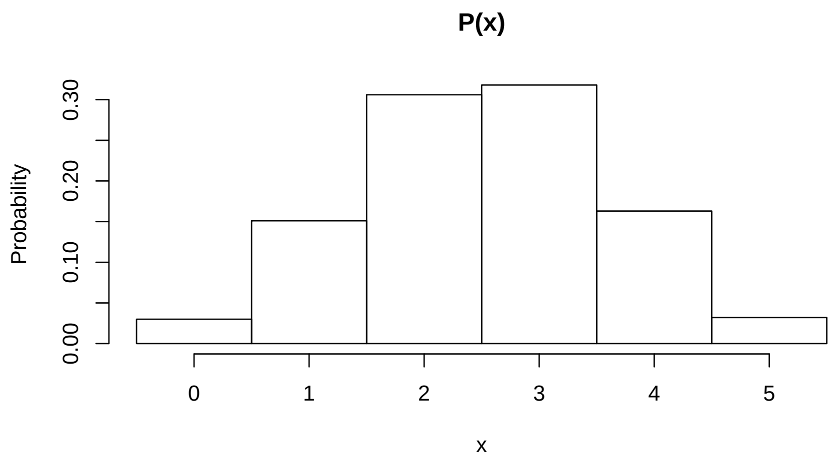
1. Table

x	P(x)
0	0.03125
1	0.15625
2	0.31250
3	0.31250
4	0.15625
5	0.03125

2. Probability distribution plot



3. Probability histogram



MEAN AND STANDARD DEVIATION OF $P(X_I)$

Given $P(x_i)$, we can find the mean and standard deviation:

EXPECTED VALUE. DEFINITION 1.3

of a random variable is the mean value μ. The expected value E in terms of the probability distribution is:

$$E = \mu = \sum_{i}^{k} x_i \cdot P(x_i) \tag{3}$$

Definition 1.4

Standard deviation.
of a probability distribution is given by:

$$\boxed{\sigma = \sqrt{\sum_i^k (x_i - \mu)^2 \cdot P(x_i)}} \tag{4}$$

Example 2. Find the expected value of x using the previous table giving $P(x_i)$.

x	P(x)	x P(x)
0	0.03125	0.00000
1	0.15625	0.15625
2	0.31250	0.62500
3	0.31250	0.93750
4	0.15625	0.62500
5	0.03125	0.15625

Finally, summing up the last column: $E = 2.5$

Example 3. Find the σ of x using the previous table giving $P(x_i)$.

x	P(x)	(x-2.5)^2 P(x)
0	0.03125	0.19531
1	0.15625	0.35156
2	0.31250	0.07812
3	0.31250	0.07812
4	0.15625	0.35156
5	0.03125	0.19531

Finally, summing up the last column: $\sigma^2 = 1.25$, taking the square root: $\sigma = 1.12$

UNUSUAL VALUES

Definition 1.5

Unusual value.
if $P(x_i) \leq 0.05$

Given that $P(x = 5) = 0.0312$.

Question 1. Is $P(x = 5)$ unusual?

Question 2. If we observe $x = 5$, what does the rare event rule tell us?

1.3 Probability densities

Definition 1.6

Probability density function (pdf) $f(x)$.
describes the probability of a continuous random variable x. Describes **population**.

Two key properties of probability densities

$$1 = \int_{-\infty}^{\infty} f(x)\,dx, \quad \text{(area under curve is 1)} \tag{5}$$

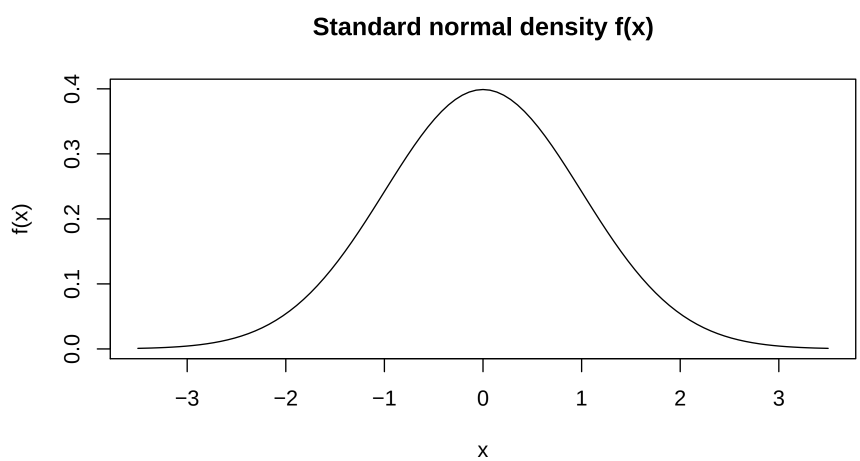
$$0 \le f(x) \quad \text{for all } x \tag{6}$$

Probability is area

$$P(A) = \int_A f(x)\,dx \tag{7}$$

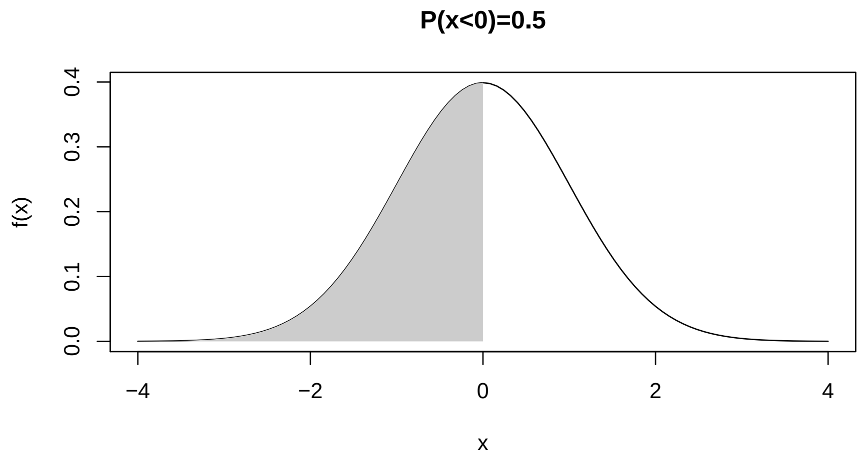
Unusual values: $P(A) \le 0.05$

Representing a probability density



Probability is represented by AREA. Height is not meaningful.

Representation of probability on a pdf



The probability $P(x < 0)$ is the area to the left of 0.

Given $f(x)$, we can generally find the mean and standard deviation:

DEFINITION 1.7 EXPECTED VALUE.
of a random variable is the mean value μ. The expected value E in terms of the probability density is:

$$E = \mu = \int_{-\infty}^{\infty} x \cdot f(x)dx \tag{8}$$

DEFINITION 1.8 STANDARD DEVIATION.
of a probability density is given by:

$$\sigma = \sqrt{\int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx} \tag{9}$$

1.4 Summary

1. Random variable x
2. $P(x_i)$ describes probability of observing x_i (discrete).
   - Key properties (for discrete):
     a) $0 \leq P(x_i) \leq 1$
     b) $\sum_{i=1}^{k} P(x_i) = 1$
   - Unusual value: $P(x_i) \leq 0.05$
   - Expected value (mean): $E = \mu = \sum x_i \cdot P(x_i)$
   - Standard deviation: $\sigma = \sqrt{\sum (x_i - \mu)^2 \cdot P(x_i)}$
3. $f(x)$ is probability density function (pdf), for **continuous** variables.

1.5 Additional Problems

Given:

x	P(x)
0	0.12500
1	0.37500
2	0.37500
3	0.12500

Question 3. Is the above a valid probability distribution?

Question 4. Find the expected value.

Question 5. Find the standard deviation.