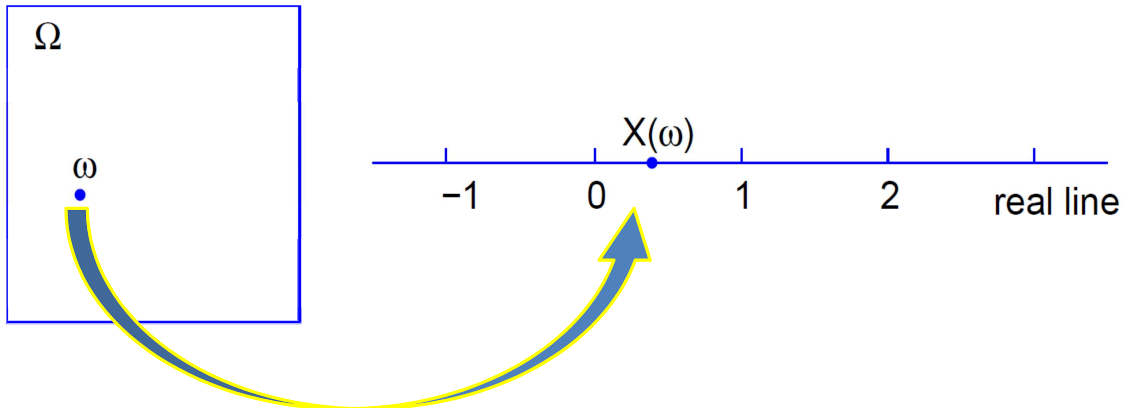
Chapter 4. Probability Distribution (A)

February 7, 2011

1 Random Variables

- A random variable is any function that assigns possible outcome a value. (or a random variable is a real-valued variable that takes on values randomly)
- Mathematically, a random variable (r.v.) X is a real-valued function $X(\omega)$ over the sample space of a random experiment, i.e., $X : \Omega \to R$



- Randomness comes from the fact that outcomes are random (X(ω) is a deterministic function of ω)

- Sometimes the random variable in random experiment are naturally defined. For example, measuring the height, r.v. X is the height; recorded the number of traffic accidents, Y is the number; ...

Notation:

- usually use upper case letters for random variables ($X(\omega), Y(\omega), ...$)
- Very often, people write $X(\omega), Y(\omega), ...$ as $X, Y, ...$
- usually use lower case letters for values of random variables: $X = x$ means that the random variable X takes on the value x

Examples:

1. Flip a coin 3 times. Here $\Omega = \{H, T\}$. Define the random variable $X \in \{0, 1, 2, ..., n\}$ to be the number of heads

2. Roll a 4-sided die twice.

(a) Define the random variable X as the maximum of the two rolls ($X \in \{1, 2, 3, 4\}$)

(b) Define the random variable Y to be the sum of the outcomes of the two rolls ($Y \in \{2, 3, ..., 8\}$)

(c) Define the random variable Z to be 0 if the sum of the two rolls is odd and 1 if it is even

(d) Flip coin until first heads shows up. Define the random variable $X \in \{1, 2....\}$ to be the number of flips until the first heads

3. Let $\Omega = R$ and the exmeriment is pick a number randomly from R. Define the two random variables

   (a) $X(\omega) = \omega$

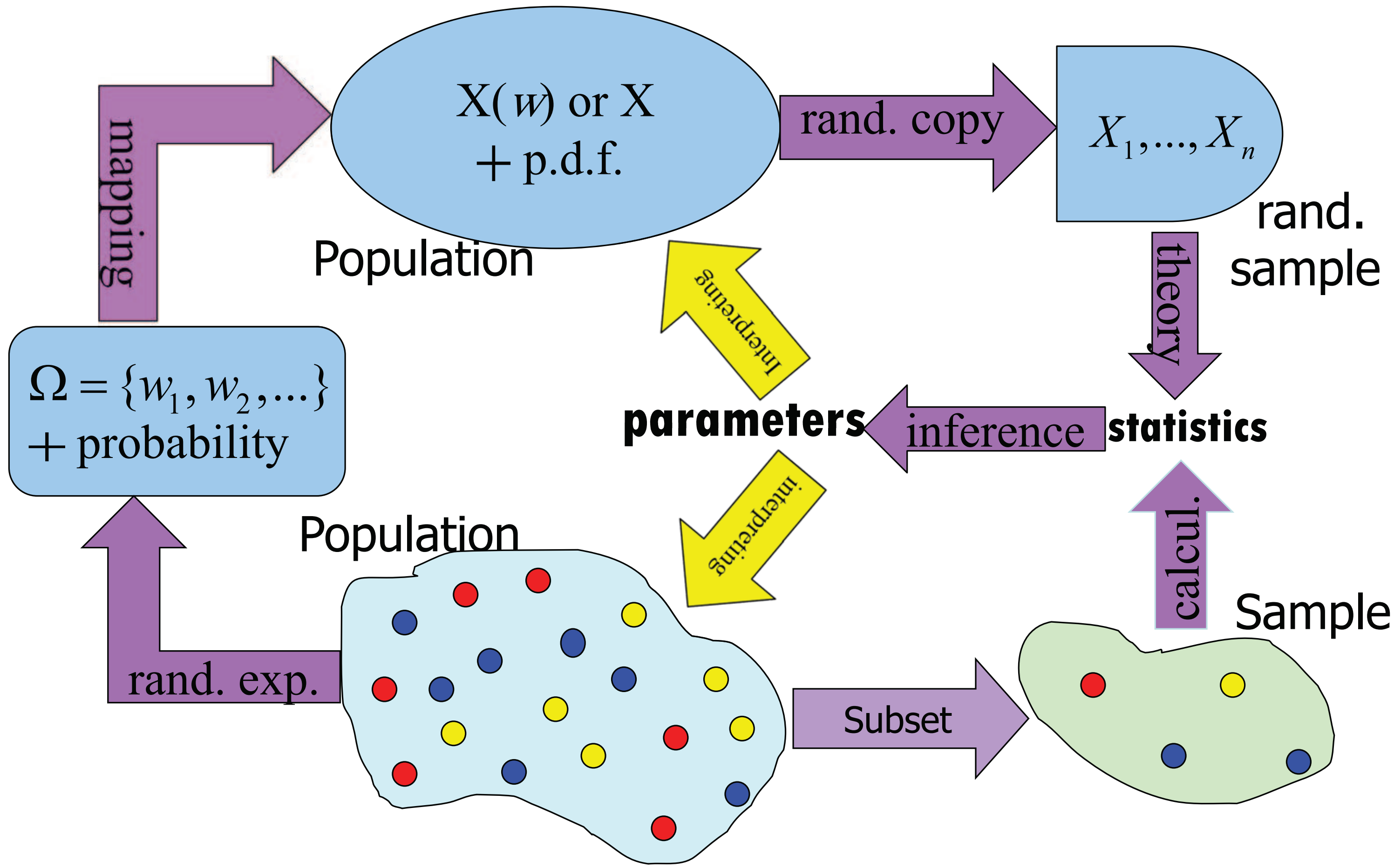
   (b)

$$Y = \begin{cases} +1 & \text{for } \omega > 0 \\ -1 & \text{otherwise} \end{cases}$$

4. Measure a person's height (X) and weight Y. Then $Y/X = Y(\omega)/X(\omega)$ make sense, it measure how fit the person ω is. Note that sometimes indicating ω is helpful!

Why do we need random variables?

1. To investigate the Population in a more convenient way.
2. In most applications we care more about these costs/measurements than the underlying probability space (i.e. Sample space + probability)
3. Very often we work directly with random variables without knowing (or caring to know) the underlying probability space



X(w) or X
+ p.d.f.
rand. copy
$X_1, \ldots, X_n$
rand. sample
mapping
Population
interpreting
theory
$\Omega = \{w_1, w_2, \ldots\}$
+ probability
parameters
inference
statistics
interpreting
calcul.
Population
Sample
rand. exp.
Subset

2 Specifying the distribution of a Random Variable

To determine the probability that $\{X \in A\}$ for any event A $\subset$ R. Note that

$$\{X \in A\} \Leftrightarrow \{\omega : X(\omega) \in A\}$$

thus

$$P(\{X \in A\}) = P(\{w : X(w) \in A\})$$

or in short

$$P(X \in A) = P(\{w : X(w) \in A\})$$

Example: Roll fair 4-sided die twice independently: Define the r.v. X to be the maximum of the two rolls. What is the $P(0.5 < X < 2)$?

Classification of Random variables

- Discrete: X can assume only one of a countable number of values. Such r.v. can be specified by a probability mass function (pmf). Examples 1, 2, 3(b) are discrete r.v.s

- Continuous: X can assume one of a continuum of values and the probability of each value is 0. Such r.v. can be specified by a probability density function (pdf). Examples 3(a) and 4 are of continuous r.v.s.

- Mixed: X is neither discrete nor continuous. Such r.v. (as well as discrete and continuous r.v.s) can be specified by a cumulative distribution function (cdf)

Example Toss a coin. $\Omega = \{H, T\}$. If the coin is even, then $P(H) = 0.5, P(T) = 0.5$. Define a r.v. $X(H) = 1, X(T) = 0$. Then we have

$$\{X = 1\} = \{H\}, \quad \{X = 0\} = \{T\}$$

Thus

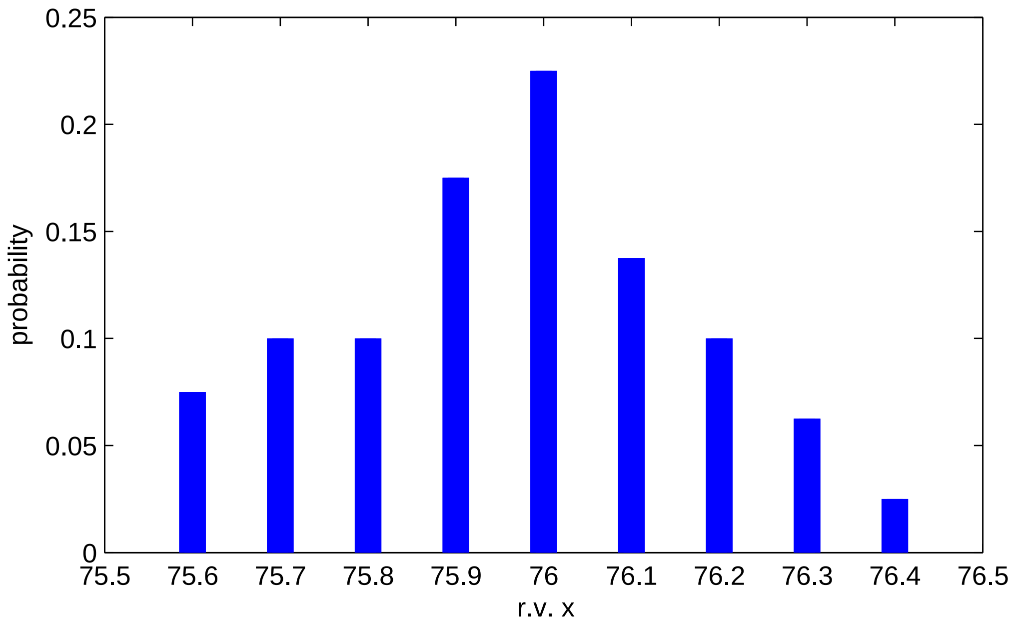
$$P(X = 1) = P(\{H\}) = 0.5, \quad P(X = 0) = P(\{T\}) = 0.5.$$

or sometimes

$$P(1) = 0.5, \quad P(0) = 0.5.$$

Example The male adults in the Village (a very special case).

x_k	75.6	75.7	75.8	75.9	76.0	76.1	76.2	76.3	76.4
$P(X = x_k)$	0.075	0.100	0.100	0.175	0.225	0.1375	0.100	0.0625	0.025



Example Two fair dice are thrown. Let A_1 be the event that the first die shows an odd number. Let A_2 be the event that the second die shows an odd number. Denote A_1 and A_2 using random variables.

Example If a, b and c are constants, then $a + bX$, $(X - c)^2$ and $X + Y$, $f(X)$ (for any function f) are random variables defined by

$$
\begin{aligned}
(a + bX)(w) &= a + bX(w) \qquad \text{and} \\
(X - c)^2(w) &= (X(w) - c)^2 \\
(X + Y)(w) &= X(w) + Y(w) \\
f(X)(w) &= f(X(w)).
\end{aligned}
$$

Definition. The probability mass function (pmf) of a discrete random variable X is given by

$$f(x_i)[\text{or}\ \ p(x_i)\ \text{or}\ \ p_i] = P(\{X = x_i\})$$

Properties of any probability mass function

1. $p(x_i) \geq 0$ for every x_i
2. $\sum_{\text{all } x_i} p(x_i) = 1$
3. $P(X \in E) = \sum_{x_i \in E} p(x_i)$

   A special case

$$P(a < X \leq b) = \sum_{a < x_i \leq b} p(x_i)$$

3 Bernoulli Trials

A Bernoulli trial is an experiment with TWO outcomes (e.g., "success" vs. "failure", "head" vs. "tail", +/-, "yes" vs. "no", etc.).

Examples of Bernoulli Trials. A Coin Toss: We can obverse H="heads", conventionally denoted success, or T="tails" denoted as failure; Rolling a Die: The outcome space is binarized to "success"=$\{6\}$ and "failure" = $\{1, 2, 3, 4, 5\}$ if someone only cares about "rolling a 6"; Polls: Choosing a voter at random to ascertain whether that voter will vote "yes" in an upcoming referendum.

The Bernoulli random variable (r.v.): Mathematically, a Bernoulli trial is modeled by a random variable

$$X = \begin{cases} 1, & \text{success} \\ 0, & \text{failure} \end{cases}$$

Writing the distribution in table

x	0	1
$P(X = x)$	p	1-p

A Bernoulli Process consists of repeatedly performing independent but identical Bernoulli trials.

3.1 Binomial Random Variables

Suppose we conduct an experiment observing an n-trial (fixed) Bernoulli process. If we are interested in the r.v. X=Number of successes in the n trials, then X is called a Binomial r.v. and its distribution is called Binomial Distribution.

Examples: Roll a standard die ten times. Let X be the number of times 6 turned up; If a student randomly guesses at 5 multiple-choice questions, find the probability that the student gets exactly three correct; A family with 4 kids, what is the probability of 2 boys and 2 girls?

Binomial Distribution

If the random variable X follows the Binomial distribution with (fixed) parameters n (sample-size) and p (probability of success at one trial), we write $X \sim B(n,p)$. The probability of getting exactly x successes is given by the Binomial probability (or mass) function:

$$P(X=x) = C_n^x p^x (1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, ..., n$$

x	0	1	...	k	...	n
$P(X=x)$	$(1-p)^n$	$n(1-p)^{n-1}p$	...	$C_n^k p^k (1-p)^{n-k}$	...	p^n

also denoted as $b(x;n,p) = C_n^x p^x (1-p)^{n-x}$

This probability expression has an easy and intuitive interpretation. The probability of the x successes in the n trials is (p^x). Similarly, the probability of the n-x failures is $(1-p)^{n-x}$. However, the x successes can be arranged anywhere among the n trials, and there are different ways of arranging the x successes in a sequence of n trials: C_n^x.

R command

dbinom(x, n, p)   # which is $b(x; n, p)$

pbinom(x, n, p)   # which is $P(X \leq x)$ for $X \sim B(n, p)$.

Example Pat Statsdud failed to study for the next stat exam. Pat's exam strategy is to rely on luck for the next quiz. The quiz consists of 10 multiple-choice questions (n=10). Each question has five possible answers, only one of which is correct (p=0.2). Pat plans to guess the answer to each question.

What is the probability that Pat gets two answers correct?

$$P(X = 2) = P(2) = 0.3019899$$

What is the probability that Pat fails the quiz? (suppose it is considered a failed quiz if a grade on the quiz is less than 50% , i.e. 5 questions out of 10)

$$P(\text{fail quiz}) = P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4) = 0.9672065$$

Example It has been claimed that in 60% of all solar-heat installations the utility bill is reduced by at least one-third. Accordingly, what are the probabilities that the utility bill will be reduced by at least one-third in

(a) four of five installations?

$$b(4; 5, 0.60) = C_5^4 (0.60)^4 (1 - 0.60)^{5-4} = 0.259$$

(b) at least four of five installations?

$$b(5; 5, 0.60) = C_5^5 (0.60)^5 (1 - 0.60)^{5-5} = 0.078$$

and the answer is b(4; 5, 0.60) $+$ b(5; 5, 0.60) $=$ 0.259 $+$ 0.078 $=$ 0.337.

3.2 Poisson Random Variables and Experiments

Poisson Distribution is a discrete probability distribution that expresses the probability of the number of events occurring in a fixed period of time (or fixed region). Denoted by $Poi(\lambda)$, where parameter $\lambda > 0$, being the expected number of occurrences that occur during the given period/region.

Mass function: For $X \sim Poi(\lambda)$, the Poisson mass function is given by

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

where k is the number of occurrences of an event

x	0	1	...	k	...
$P(X = x)$	$e^{-\lambda}$	$\frac{\lambda e^{-\lambda}}{1!}$	...	$\frac{\lambda^k e^{-\lambda}}{k!}$	...

Poisson Distribution and Binomial Distribution[1]

For $X \sim B(n,p)$, when n is very large and p is very small, then the distribution may be approximated by the Poisson distribution $X \sim Poi(np)$

More precisely, let

$$X_n \sim B(n, \lambda/n), \quad Y \sim Poi(\lambda)$$

Then for all $0 \leq k \leq n$

$$\lim_{n\to\infty} P(X_n = k) = P(Y = k)$$

This is sometimes known as the law of rare events, since each of the n individual Bernoulli events rarely occurs.

[1]you can skip the details of this page

Examples of events that may be modeled by Poisson Distribution include:

- The number of cars that pass through a certain point on a road (sufficiently distant from traffic lights) during a given period of time.
- The number of spelling mistakes one makes while typing a single page.
- The number of phone calls at a call center per minute.
- The number of times a web server is accessed per minute.
- The number of road kill (animals killed) found per unit length of road.
- The number of mutations in a given stretch of DNA after a certain amount of radiation exposure.

- The number of unstable atomic nuclei that decayed within a given period of time in a piece of radioactive substance.
- The number of pine trees per unit area of mixed forest.
- The number of stars in a given volume of space.
- The distribution of visual receptor cells in the retina of the human eye.
- The number of light bulbs that burn out in a certain amount of time.
- The number of viruses that can infect a cell in cell culture.
- The number of inventions of an inventor over their career.

Example The number of infections [X] in a hospital each week has been shown to follow a poisson distribution with mean 3.0 infections per week.

- $P(X = 0) =$
- $P(X < 4) =$
- $P(X > 9) =$
- If you found 9 infections next week, what would you say??

`R command`

`dpois(x, lambda)  # which is` $P(X = x)$ `for` $X \sim Poi(\lambda)$.

`ppois(q, lambda)  # which is` $P(X \leq x)$ `for` $X \sim Poi(\lambda)$.

3.3 The geometric distribution

The geometric distribution is either of two discrete probability distributions:

- The probability distribution of the number X of Bernoulli trials needed to get one success, supported on the set $\{ 1, 2, 3, ... \}$
- The probability distribution of the number Y = X - 1 of failures before the first success, supported on the set $\{ 0, 1, 2, 3, ... \}$

Geometric distribution is named for the fact that the sequence of probabilities is a geometric sequence.

To avoid ambiguity, it is considered wise to indicate which is intended, by mentioning the range explicitly.

If the probability of success on each trial is p, then the probability that the kth trial (out of k trials) is the first success is

$$P(X = k) = (1 - p)^{k-1} p$$

for k = 1, 2, 3, ...

Equivalently, the probability that there are k failures before the first success is[2]

$$P(Y = k) = (1 - p)^k p$$

for k = 0, 1, 2, 3, ...

x	0	1	...	k	...
$P(X = x)$	p	$(1 - p)p$	...	$(1 - p)^k p$	...

[2]In our book, it is also denoted as $g(k, p)$

Example suppose an ordinary die is thrown repeatedly until the first time a "1" appears. The probability distribution of the number of times it is thrown is supported on the infinite set $\{1, 2, 3, \ldots\}$ and is a geometric distribution with $p = 1/6$.

Example If the probability is 0.05 that a certain kind of measuring device will show excessive drift. what is the probability that the sixth measuring device tested will be the first to show excessive drift?

$$g(6; 0.05) = (0.05)(1 - 0.05)^{6-1} = 0.039$$