Probability and Statistics

Kristel Van Steen, PhD[2]

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

ULg

kristel.vansteen@ulg.ac.be

CHAPTER 2: RANDOM VARIABLES AND ASSOCIATED FUNCTIONS

1 Random variables

1.1 Introduction

1.2 Formal definition

1.3 The numbers game

2 Functions of one variable

2.1 Probability distribution functions

2.2 The discrete case: probability mass functions

2.3 The binomial distribution

2.4 The continuous case: density functions

2.5 The normal distribution

2.6 The inverse cumulative distribution function

5.3 Moment generating functions

6 Functions of random variables

6.1 Functions of one random variable

6.2 Functions of two or more random variables

6.3 Two or more random variables: multivariate moments

7 Inequalities

7.1 Jensen inequality

7.2 Markov's inequality

7.3 Chebyshev's inequality

7.4 Cantelli's inequality

7.5 The law of large numbers

1 Random variables

1.1 Introduction

- We have introduced before the concept of a probability model to describe random experiments.
- Such a model consists of
  - a universal space of events Ω
  - a sample space of events ω; S_e
  - a probability P

1.2 Formal definition

- So, thus far we have focused on *probabilities of events*.
- For example, we computed the probability that you win the Monty Hall game or that you have a rare medical condition given that you tested positive.
- But, in many cases we would like to know more.
  - For example, how many contestants must play the Monty Hall game until one of them finally wins?
  - How long will this condition last?
  - How much will I lose gambling with strange dice all night?
- To answer such questions, we need to work with *random variables*

Often random experiments have associated numerical values, i.e. for each elementary event (outcome) ω there is a number $X(\omega) = x$.

Example: Random draw of a playing card
Define the value function X by:

$$\begin{aligned}
\omega = \text{ Ace} &\mapsto X(\omega) = 11 \\
\omega = \text{ King} &\mapsto X(\omega) = 4 \\
\omega = \text{ Queen} &\mapsto X(\omega) = 3 \\
\omega = \text{ Jack} &\mapsto X(\omega) = 2 \\
\omega = \text{ Ten} &\mapsto X(\omega) = 10 \\
\omega = \text{ Nine} &\mapsto X(\omega) = 0 \\
\vdots \quad & \quad \vdots \\
\omega = \text{ Six} &\mapsto X(\omega) = 0
\end{aligned}$$

Thus in the example above, $X(\cdot)$ is a function. In general we define:

A random variable X is a function:

$$X: \quad \Omega \rightarrow \mathbb{R}$$
$$\omega \mapsto X(\omega)$$

The function $X(\cdot)$ is not random, but its argument ω is.

- While it is rather unusual to denote a function by X (or Y, Z, …) we shall see that random variables sometimes admit calculations like those with ordinary variables such as X (or Y, Z,…).
- The outcomes of the random experiment (i.e., ω) yield different possible values of $x = X(\omega)$: the value of x is a **realization** of the random variable X. Thus a realization of a random variable is the result of a random experiment (which may be described by a number)

- We call a **random variable discrete** if its range $W = W_X$ (the set of potential values of X) is discrete, i.e. countable (its potential values can be numbered).
  - o $W = \{0, 1, ..., 10\}$ is finite and thus discrete, while
  - o $W = \{0, 1, 2, 3, ...\}$ (natural numbers including zero) is infinite but still discrete, and while
  - o the set of real numbers is not discrete (but continuous)
- In case of a sample space having an uncountably infinite number of sample points, the associated random variable is called a **continuous random variable**, with its values distributed over one or more continuous intervals on the real line.
- We need to make this distinction because they require different probability assignment considerations…

A special random variable

An *indicator random variable* is a random variable that maps every outcome to either 0 or 1. Indicator random variables are also called *Bernoulli variables*.

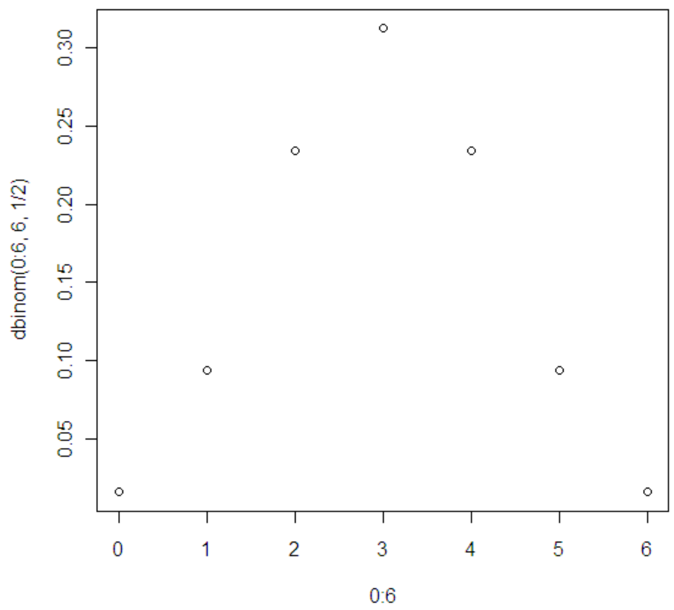
Indicator random variables are closely related to events. In particular, an indicator random variable partitions the sample space into those outcomes mapped to 1 and those outcomes mapped to 0. For example, the indicator M partitions the sample space into two blocks as follows:

$$\underbrace{HHH \quad TTT}_{M=1} \quad \underbrace{HHT \quad HTH \quad HTT \quad THH \quad THT \quad TTH}_{M=0}.$$

In the same way, an event E partitions the sample space into those outcomes in E and those not in E. So E is naturally associated with an indicator random variable, I_E, where $I_E(\omega) = 1$ for outcomes $\omega \in E$ and $I_E(\omega) = 0$ for outcomes $\omega \notin E$. Thus, $M = I_E$ where E is the event that all three coins match.

If I were to repeat an experiment

- Suppose M=1 when the result of me tossing a coin is head (a success); suppose M=0 when the result is tails



plot(0:6, dbinom(0:6, 6, 1/2))
(http://www.r-project.org/)

1.3 The numbers game

Enough definitions —let's play a game! We have two envelopes. Each contains an integer in the range $0, 1, \ldots, 100$, and the numbers are distinct. To win the game, you must determine which envelope contains the larger number. To give you a fighting chance, we'll let you peek at the number in one envelope selected at random. Can you devise a strategy that gives you a better than 50% chance of winning?

For example, you could just pick an envelope at random and guess that it contains the larger number. But this strategy wins only 50% of the time. Your challenge is to do better.

So you might try to be more clever. Suppose you peek in one envelope and see the number 12. Since 12 is a small number, you might guess that the number in the other envelope is larger. But perhaps we've been tricky and put small numbers in *both* envelopes. Then your guess might not be so good!

An important point here is that the numbers in the envelopes may *not* be random. We're picking the numbers and we're choosing them in a way that we think will defeat your guessing strategy. We'll only use randomization to choose the numbers if that serves our purpose, which is making you lose!

Intuition Behind the Winning Strategy

Amazingly, there is a strategy that wins more than 50% of the time, regardless of what numbers we put in the envelopes!

Suppose that you somehow knew a number x that was in between the numbers in the envelopes. Now you peek in one envelope and see a number. If it is bigger than x, then you know you're peeking at the higher number. If it is smaller than x, then you're peeking at the lower number. In other words, if you know a number x between the numbers in the envelopes, then you are certain to win the game.

The only flaw with this brilliant strategy is that you do *not* know such an x. Oh well.

But what if you try to *guess* x? There is some probability that you guess correctly. In this case, you win 100% of the time. On the other hand, if you guess incorrectly, then you're no worse off than before; your chance of winning is still 50%. Combining these two cases, your overall chance of winning is better than 50%!

Informal arguments about probability, like this one, often sound plausible, but do not hold up under close scrutiny. In contrast, this argument sounds completely implausible —but is actually correct!

Analysis of the Winning Strategy

For generality, suppose that we can choose numbers from the set $\{0, 1, \ldots, n\}$. Call the lower number L and the higher number H.

Your goal is to guess a number x between L and H. To avoid confusing equality cases, you select x at random from among the half-integers:

$$\left\{\frac{1}{2}, 1\frac{1}{2}, 2\frac{1}{2}, \ldots, n - \frac{1}{2}\right\}$$

But what probability distribution should you use?

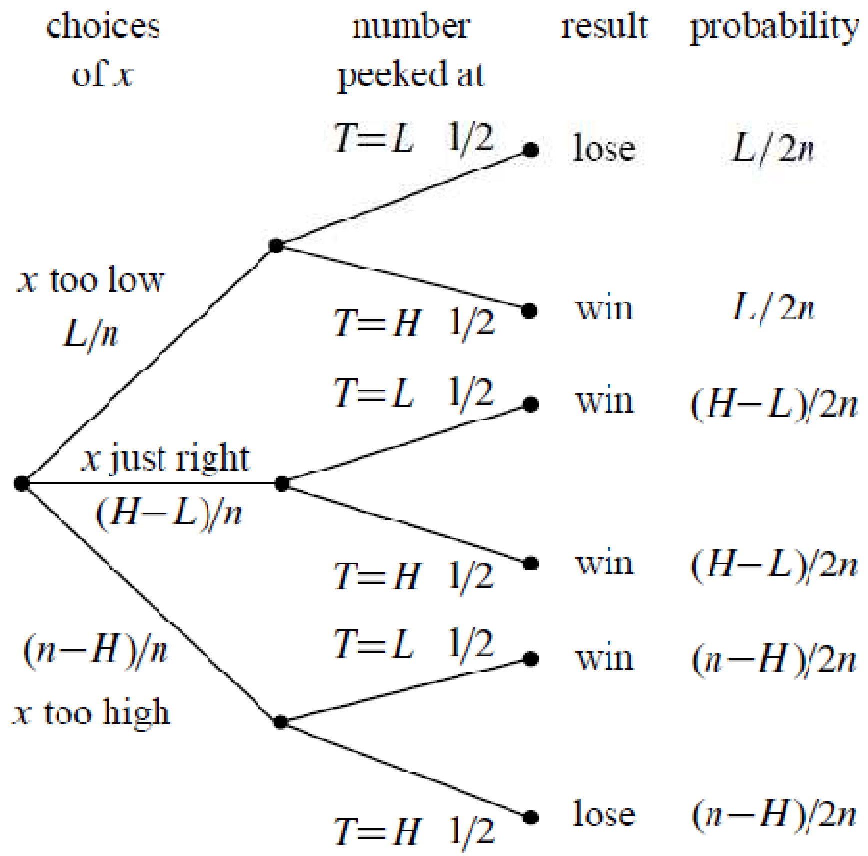
The uniform distribution turns out to be your best bet. An informal justification is that if we figured out that you were unlikely to pick some number —say $50\frac{1}{2}$ —then we'd always put 50 and 51 in the envelopes. Then you'd be unlikely to pick an x between L and H and would have less chance of winning.

After you've selected the number x, you peek into an envelope and see some number T. If $T > x$, then you guess that you're looking at the larger number. If $T < x$, then you guess that the other number is larger.

All that remains is to determine the probability that this strategy succeeds. We can do this with the usual four step method and a tree diagram.

Step 1: Find the sample space.

You either choose x too low ($< L$), too high ($> H$), or just right ($L < x < H$). Then you either peek at the lower number ($T = L$) or the higher number ($T = H$). This gives a total of six possible outcomes .



Step 2: Define events of interest.

The four outcomes in the event that you win are marked in the tree diagram.

Step 3: Assign outcome probabilities.

First, we assign edge probabilities. Your guess x is too low with probability L/n, too high with probability $(n - H)/n$, and just right with probability $(H - L)/n$. Next, you peek at either the lower or higher number with equal probability. Multiplying along root-to-leaf paths gives the outcome probabilities.

Step 4: Compute event probabilities.

The probability of the event that you win is the sum of the probabilities of the four outcomes in that event:

$$\begin{aligned}\Pr[\text{win}] &= \frac{L}{2n} + \frac{H-L}{2n} + \frac{H-L}{2n} + \frac{n-H}{2n} \\ &= \frac{1}{2} + \frac{H-L}{2n} \\ &\geq \frac{1}{2} + \frac{1}{2n}\end{aligned}$$

The final inequality relies on the fact that the higher number H is at least 1 greater than the lower number L since they are required to be distinct.

Sure enough, you win with this strategy more than half the time, regardless of the numbers in the envelopes!

So choosing numbers in the range 0,…,100, will make you win with prob at least 1/2+1/200 = 50.5%. Even better, if you are allowed only numbers in the range 0,…,10, then your probability of winning rises to 55%! Not bad he ….

Randomized algorithms

The best strategy to win the numbers game is an example of a *randomized algorithm* —it uses random numbers to influence decisions. Protocols and algorithms that make use of random numbers are very important in computer science. There are many problems for which the best known solutions are based on a random number generator.

For example, the most commonly-used protocol for deciding when to send a broadcast on a shared bus or Ethernet is a randomized algorithm known as *exponential backoff*. One of the most commonly-used sorting algorithms used in practice, called *quicksort*, uses random numbers. You'll see many more examples if you take an algorithms course. In each case, randomness is used to improve the probability that the algorithm runs quickly or otherwise performs well.

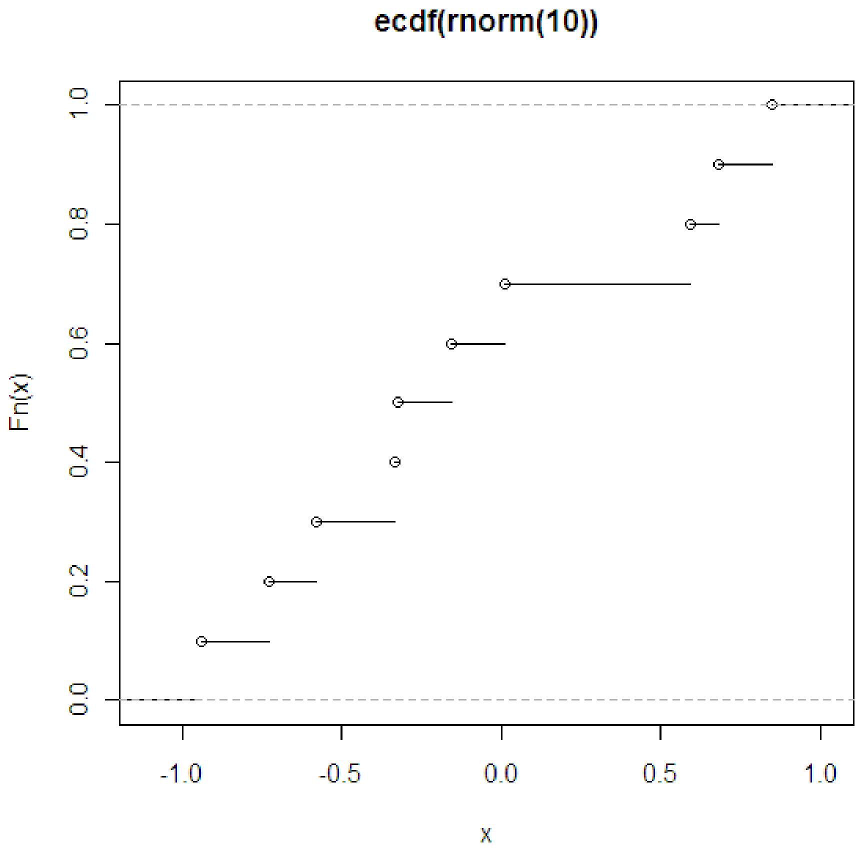
2 Functions of one random variable

2.1 Probability distribution functions

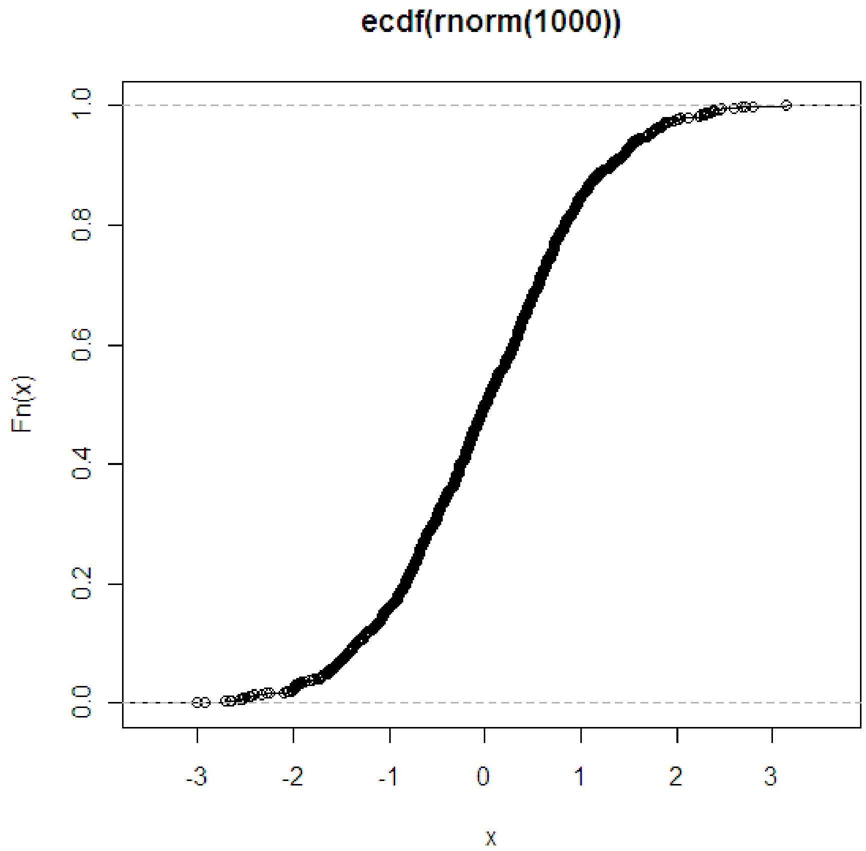
- Given a random experiment with associated random variable X, and given a real number x, the function

  $$F_X(x) = P(X \leq x)$$

  is defined as the **cumulative distribution function** (CDF).
- As x increases, the value of the CDF will increase as well, until it reaches 1 (which explains its name)
- Note that $F_X(x)$ is simply a P(A), the probability of an event A occurring, the event here being $X \leq x$
- This function is sometimes referred to in the literature as a **probability distribution function** (PDF) or a distribution function (omitting cumulative), which may cause some confusion…



plot(ecdf(rnorm(10)))



plot(ecdf(rnorm(1000)))

Properties of cumulative distribution functions

$$(i) \qquad F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0,$$

$$F_X(+\infty) = \lim_{x \to +\infty} F_X(x) = 1$$

$$(ii) \quad F_X(.) \text{ is a montone, nondecreasing function; that is,}$$

$$F_X(a) \le F_X(b) \text{ for } a < b$$

$$(iii) \qquad F_X(.) \text{ is continuous from the right; that is,}$$

$$\lim_{0<h\to 0} F_X(x+h) = F_X(x)$$

- Actually, ANY function F(.) with domain the real line and counterdomain [0,1] satisifying the above three properties is defined to be a cumulative distribution function.

2.2 The discrete case: probability mass functions

The random variable X takes its values (its potential realizations) with certain probabilities. These are defined as follows:

$$\begin{aligned}
&\text{Probability of } X \text{ taking the value } x \\
&= P(X = x) = P(\{\omega;\ X(\omega) = x\}) \\
&= \sum_{\omega; X(\omega)=x} P(\omega).
\end{aligned}$$

Example (cont.): $X =$ Value of a playing card drawn at random

$$\begin{aligned}
&\text{Probability of } 4 = P(X = 4) \\
= \quad & P(\{\omega;\ \omega = \text{ a king}\}) \\
= \quad & P(\text{King of diamonds}) + P(\text{King of hearts}) + P(\text{King of clubs}) + P(\text{King of spades}) \\
= \quad & 4/36 = 1/9.
\end{aligned}$$

The "list" of probabilities $P(X = x)$ for all possible values of x is called the (discrete) **(probability) distribution** of the (discrete) random variable X. Each random variable X has a corresponding (probability) distribution, and vice versa:

$$\text{Random variable } X \;\Leftrightarrow\; \text{(probability) distribution}$$

Each (discrete) probability distribution satisfies the equality

$$\sum_{\text{all } x \text{ possible}} P(X = x) = 1.$$

The function

$$p_X(x) = P(X = x)$$

is called **the probability mass function** of the discrete random variable X, or **discrete density function** or probability function of X, amongst others.

Example (cont.): $X =$ Value of a playing card drawn at random
The probability distribution of X is

$$P(X = 11) = 1/9$$

$$\begin{aligned} P(X = 10) &= 1/9 \\ P(X = 4) &= 1/9 \\ P(X = 3) &= 1/9 \\ P(X = 2) &= 1/9 \\ P(X = 0) &= 4/9 \end{aligned}$$

Relation between density function and cumulative distribution function

- The cumulative distribution function and probability mass function of a discrete random variable contain the same information; each is recoverable from the other:

$$P_X(x_i) = F_X(x_i) - F_X(x_{i-1}),$$

$$F_X(x) = \sum_{i=1}^{i:x_i \leq x} P_X(x_i),$$

assuming that $x_1 < x_2 < \ldots$

- The discrete random variable X is completely characterized by these functions

Simplified definition for discrete density functions

Any function $f(\cdot)$ with domain the real line and counterdomain [0, 1] is defined to be a *discrete density function* if for some countable set $x_1, x_2, \ldots, x_n, \ldots,$

(i) $f(x_j) > 0$ for $j = 1, 2, \ldots.$

(ii) $f(x) = 0$ for $x \neq x_j; j = 1, 2, \ldots.$

(iii) $\sum f(x_j) = 1$, where the summation is over the points $x_1, x_2, \ldots, x_n, \ldots.$ ////

- This definition allows us to speak about discrete density functions without reference to some random variable.
- We can therefore talk about properties of discrete density functions without referring to a random variable

2.3 The binomial distribution

Regard the situation where the quantity of interest is the number of successes (or failures) at something. Examples of this include quality control, success or failure of (medical or biological) treatments, or gambling.

Example: Coin toss
A coin is tossed and randomly comes up heads (K) or tails (Z).
Regard the random variable X with values in $W = \{0, 1\}$ describing the following:

$$X = 0 \qquad \text{if the outcome is tails,}$$
$$X = 1 \qquad \text{if the outcome is heads.}$$

The probability distribution of X can be described by a single parameter π:

$$P(X = 1) = \pi, \quad P(X = 0) = 1 - \pi, \quad 0 \leq \pi \leq 1.$$

A fair coin has the parameter $\pi = 1/2$.

Bernoulli distribution

A random variable X with range $W = \{0, 1\}$ has a Bernoulli(π) distribution if

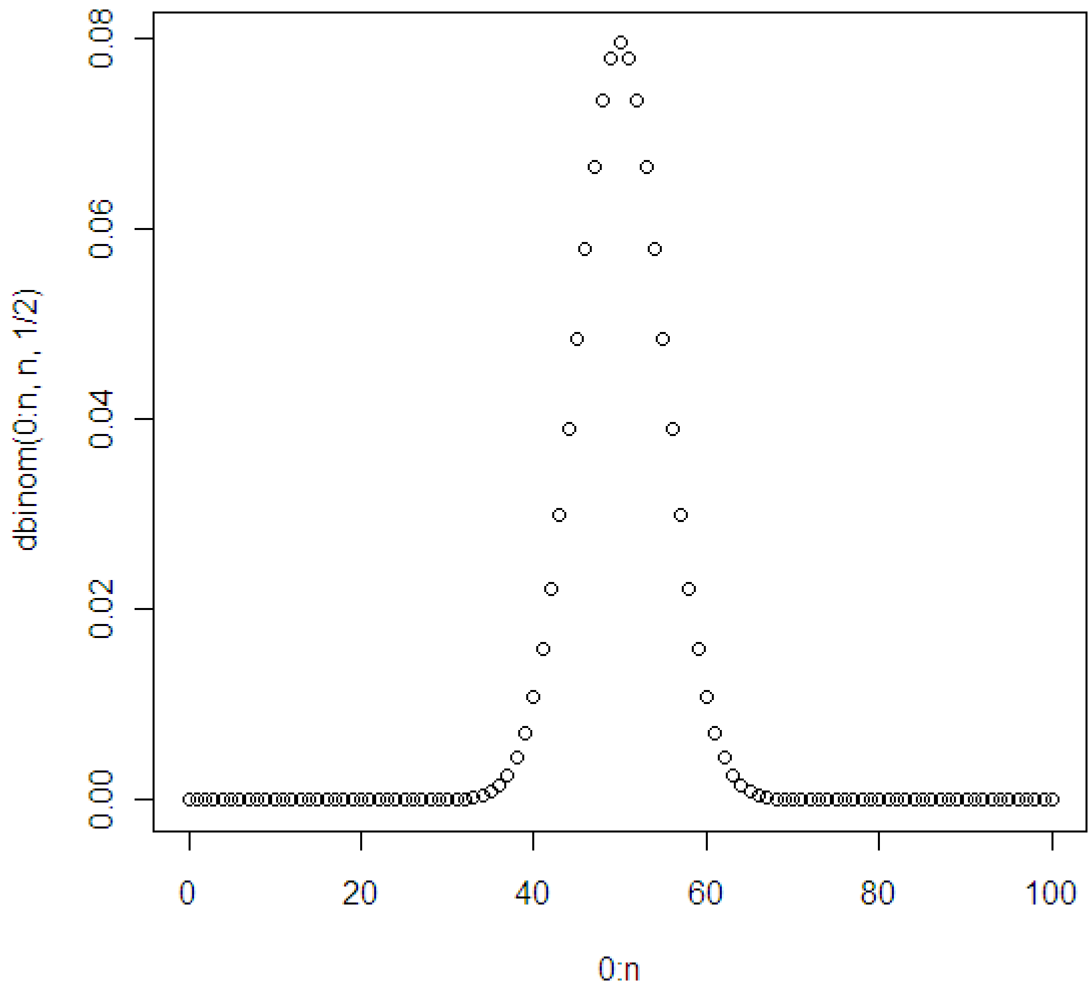
$$P(X = 1) = \pi, \quad P(X = 0) = 1 - \pi, \quad 0 \leq \pi \leq 1.$$

The Bernoulli distribution is a trivial mathematical description of the (non-)occurence of an event.

Example (cont.): n-fold coin toss
Regard $X =$ Number of heads from n independent coin tosses. Obviously the range of X is the set $W = \{0, 1, \ldots, n\}$. X can also be written as the sum of independent Bernoulli-distributed random variables:

$$X = \sum_{i=1}^{n} X_i,$$

$$X_i = \begin{cases} 1 & \text{i-th toss comes up heads} \\ 0 & \text{i-th toss comes up tails.} \end{cases}$$




```
binom <- function(n) {
  plot(0:n, dbinom(0:n, n,
1/2))
  Sys.sleep(0.1)
}
ignore <- sapply(1:100,
binom)
```

Binomial distribution

A random variable X with range $W = \{0, 1, \ldots, n\}$ has a Binomial(n, π) distribution if

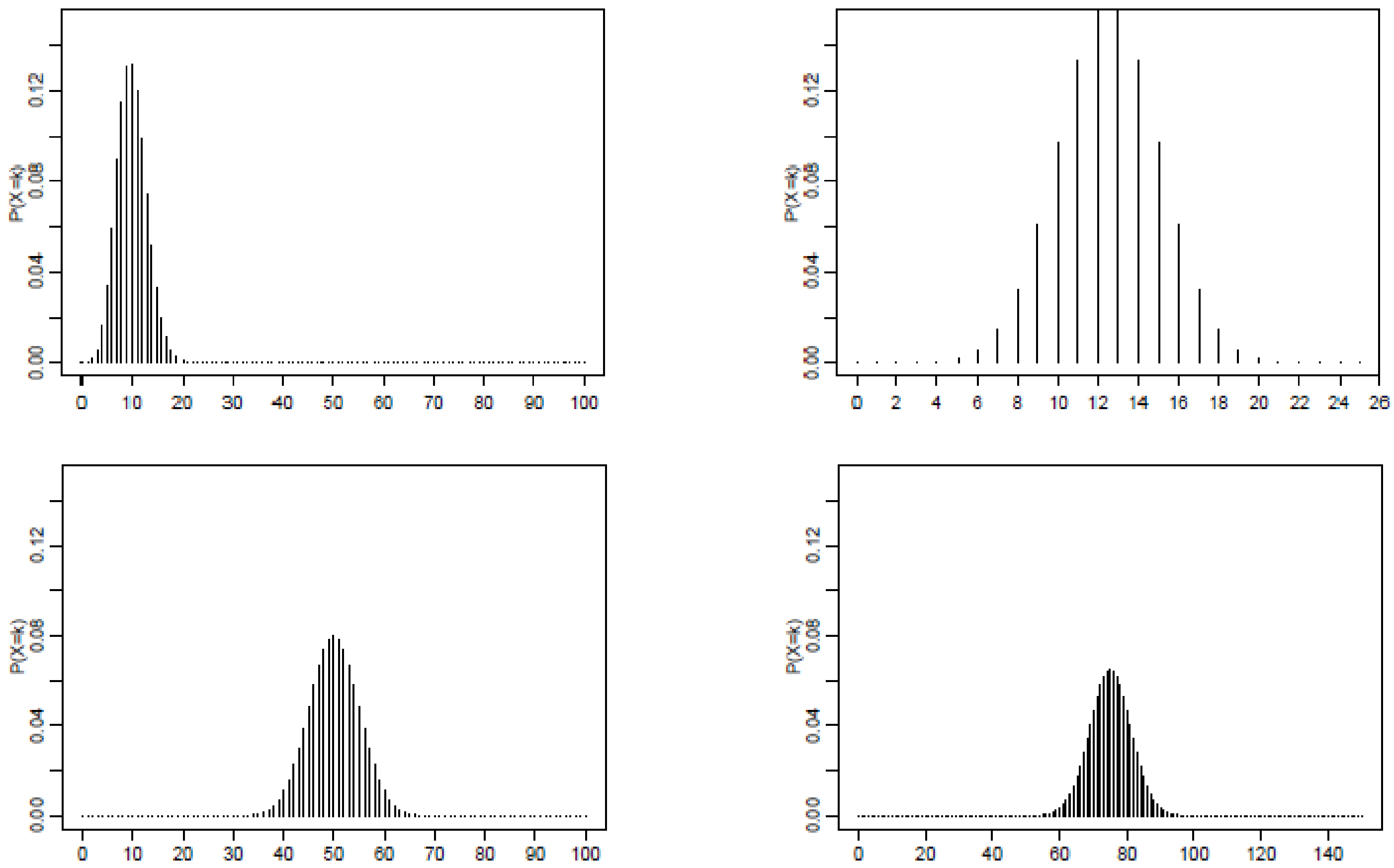
$$P(X = x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \; x = 0, 1, \ldots, n$$

where $0 \leq \pi \leq 1$ is the success rate associated to the distribution.

(Here $\binom{n}{x}$ is the binomial coefficient, which denotes the number of possible arrangements of x successes and $n - x$ failures).

As in the previous example, X denotes the number of successes/failures (occurence of a particular event) out of n **independent** experiments. The independence of these experiments is crucial if the binomial distribution is to apply.

Binomial probabilities P(X = x) as a function of x for various choices of n and π. On the left, n=100 and π=0.1,0.5. On the right, π=0.5 and n=25,150



2.4 The continuous case: density functions

- For a continuous random variable X, the CDF $F_X(.)$ is a continuous function and the derivative

$$f_X(x) = \frac{dF_X(x)}{dx},$$

  exists for all x.
- The function $f_X(.)$ is called the **density function** of X or the probability density function of X

Relation between density function and cumulative distribution function

- The cumulative distribution function and probability mass function of a continuous random variable contain the same information; each is recoverable from the other:

$$f_X(x) = \frac{dF_X(x)}{dx},$$

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

- The continuous random variable X is completely characterized by these functions

Simplified definition for density functions

Any function $f(\cdot)$ with domain the real line and counterdomain $[0, \infty)$ is defined to be a *probability density function* if and only if

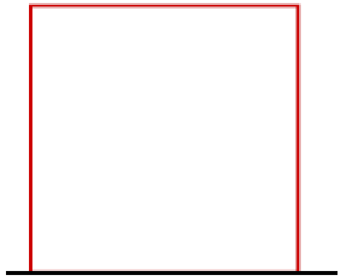
(i) $f(x) \geq 0$ for all x.

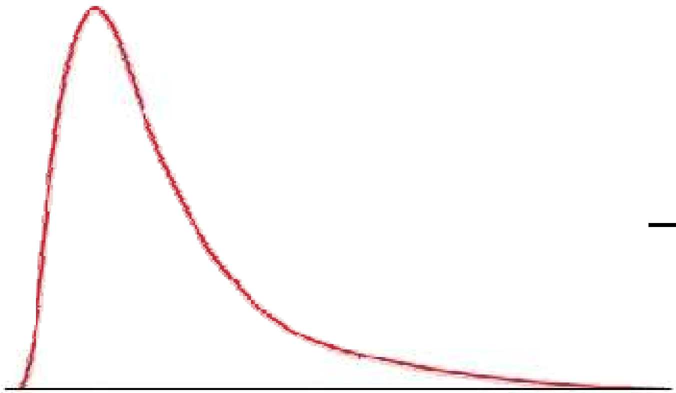
(ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$. ////

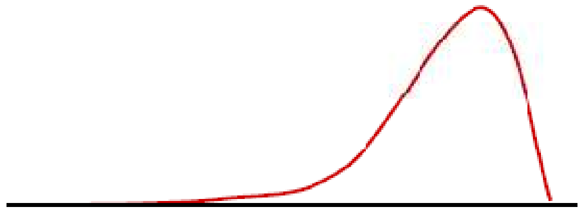
- With this definition , we can speak of density functions without reference to random variables

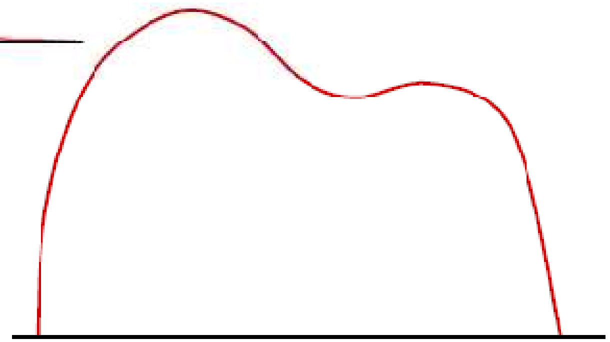
Density curves as a mathematical model of a distribution

Density curves come in any imaginable shape.









Some are well known mathematically and others aren't.

2.5 The normal distribution

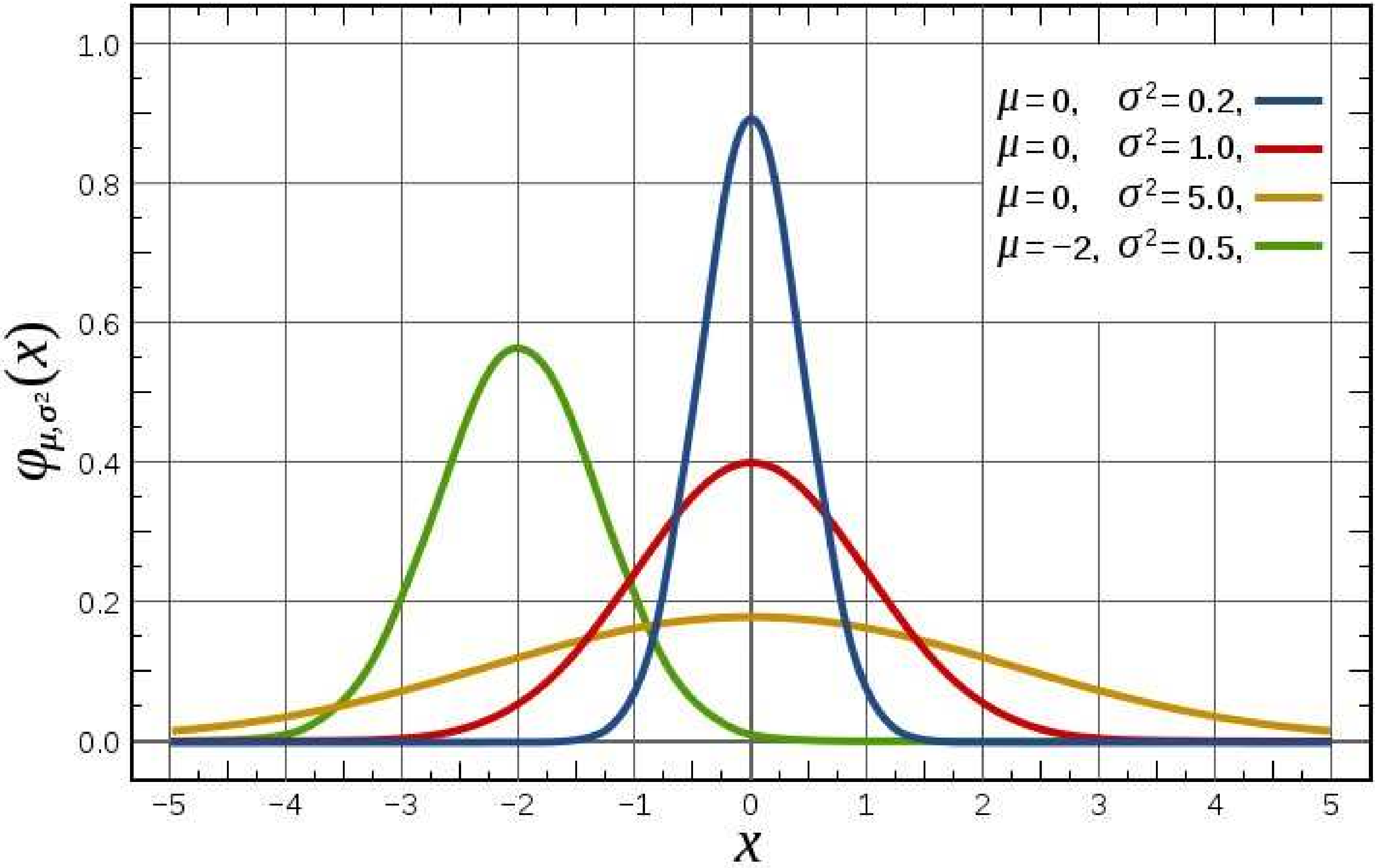
- In probability theory, the **normal** (or **Gaussian**) **distribution** is a continuous probability distribution that is often used as a first approximation to describe real-valued random variables that tend to cluster around a single mean value.
- The graph of the associated probability density function is "bell"-shaped, and is known as the Gaussian function or bell curve:

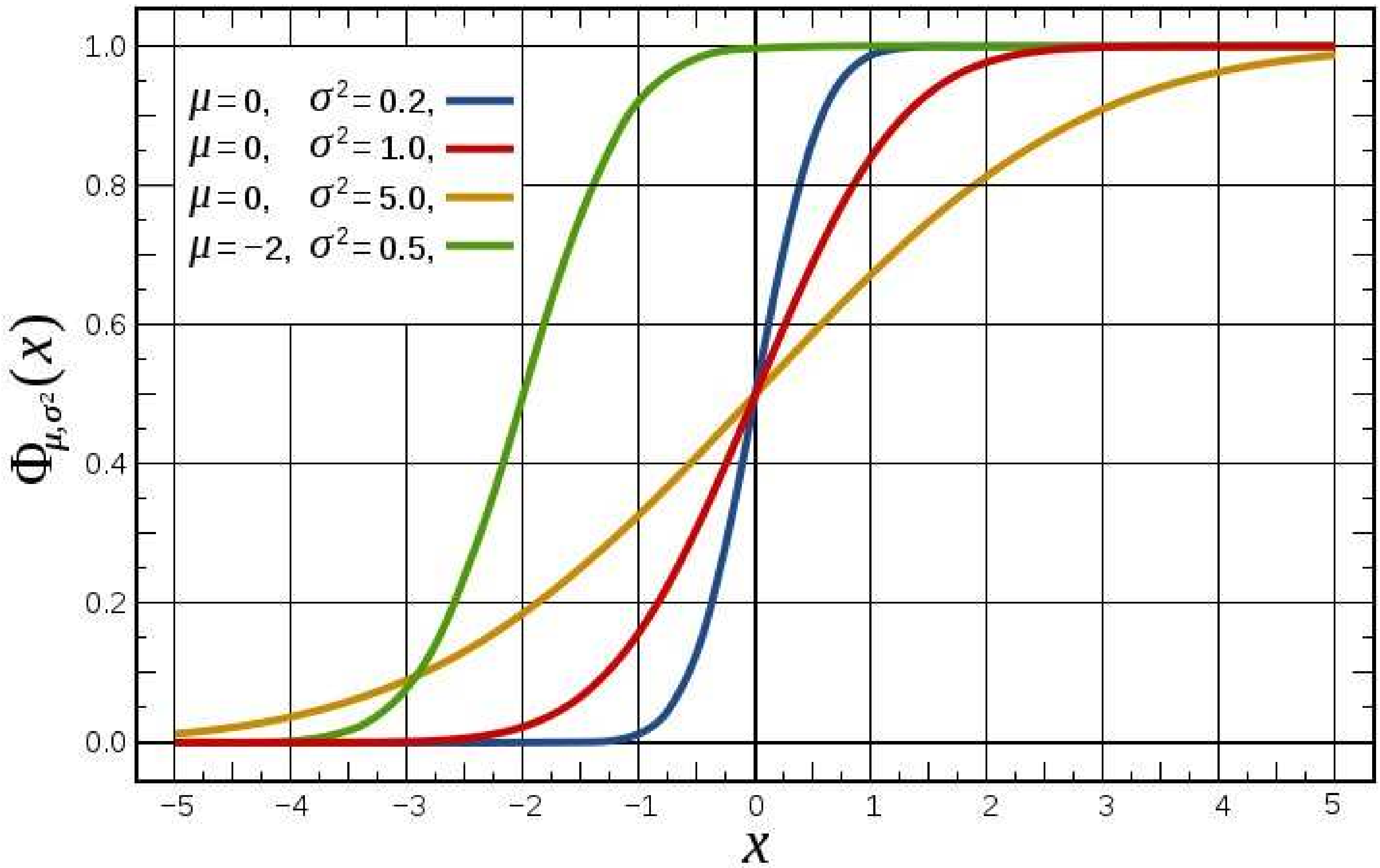
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where parameter μ is the *mean* (location of the peak) and σ^2 is the *variance* (the measure of the width of the distribution).

- The distribution with $\mu = 0$ and $\sigma^2 = 1$ is called the **standard normal**.

- Density and cumulative distribution function for several normal distributions. The red curve refers to the standard normal distribution.





1.0
0.8
0.6
0.4
0.2
0.0
$\mu=0, \quad \sigma^2=0.2,$
$\mu=0, \quad \sigma^2=1.0,$
$\mu=0, \quad \sigma^2=5.0,$
$\mu=-2, \quad \sigma^2=0.5,$
$\Phi_{\mu,\sigma^2}(x)$
−5 −4 −3 −2 −1 0 1 2 3 4 5
x

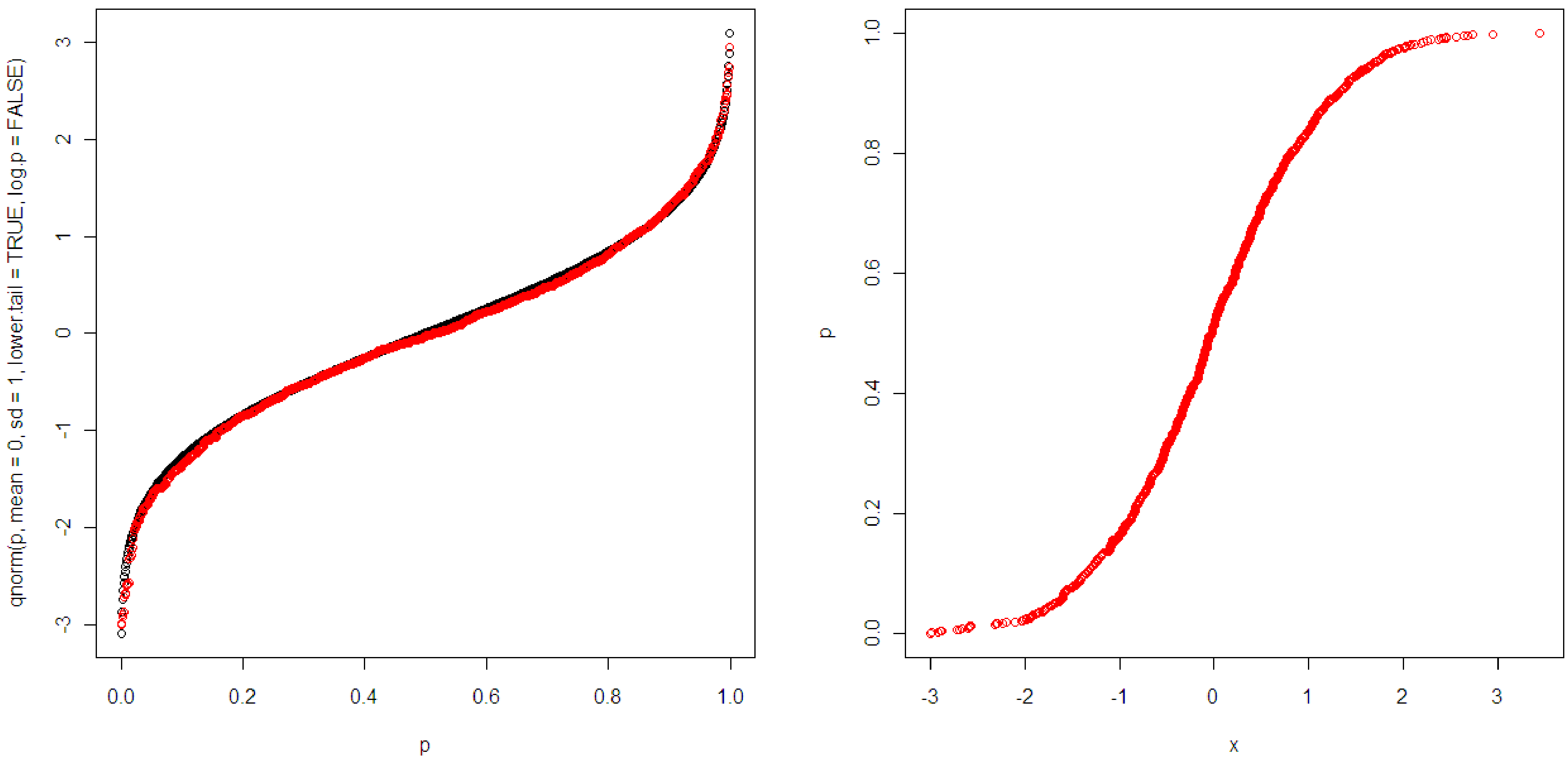
2.6 The inverse cumulative distribution function (=quantile function)

- If the CDF $F_X(.)$ is strictly increasing and continuous then $F_X^{-1}(y), y \in [0,1]$ is the unique real number x such that $F_X(x) = y$

- Unfortunately, the distribution does not, in general, have an inverse. One may define, for $y \in [0,1]$, the **generalized inverse distribution function**:

$$F_X^{-1}(y) = \inf_{x \in \mathbb{R}} \{F_X(x) \geq y\}$$

  (infimum = greatest lower bound)
- The inverse of the CDF is called the **quantile** function (evaluated at 0.5 it gives rise to the median – see later).
- The inverse of the CDF can be used to translate results obtained for the uniform distribution to other distributions (see later).

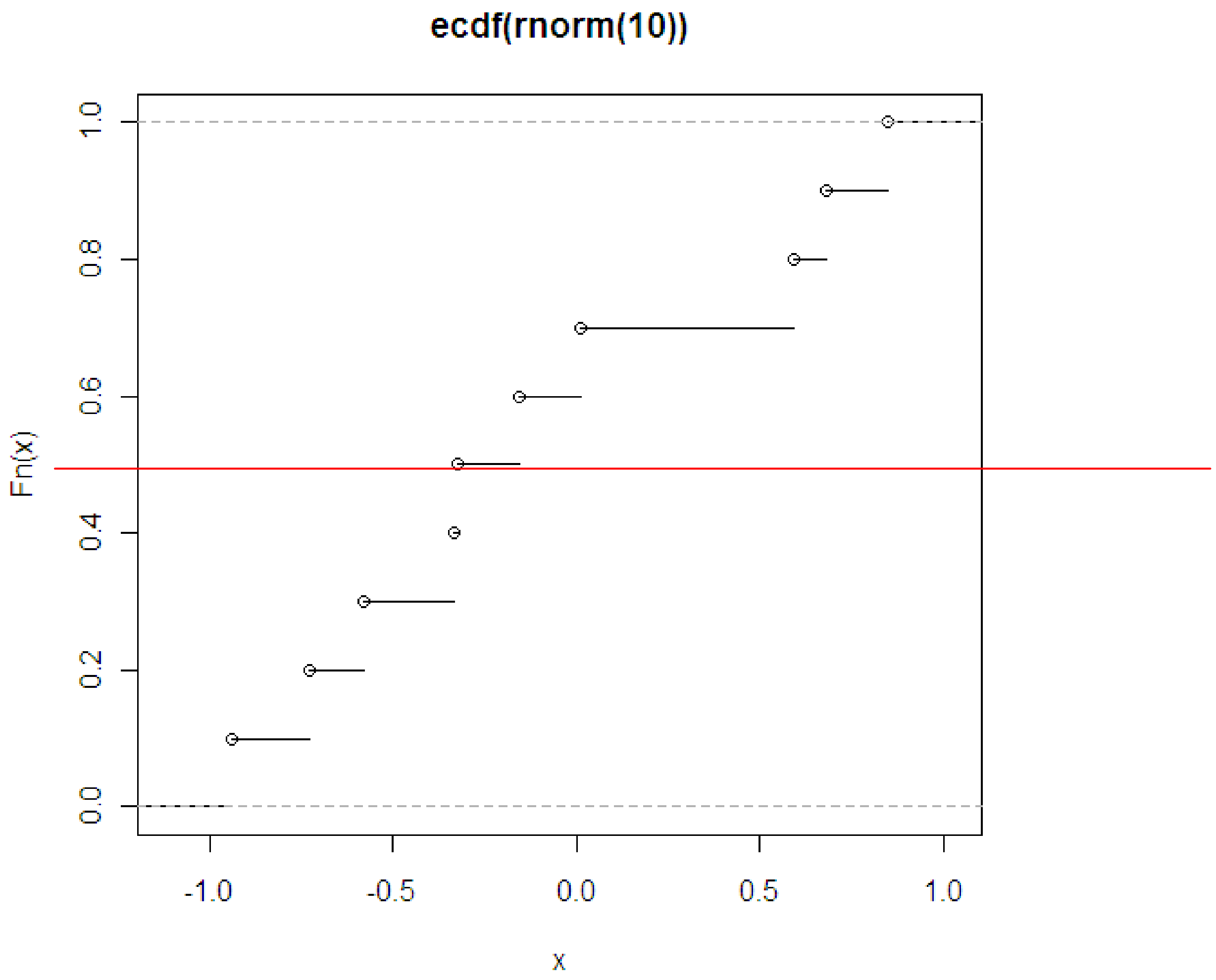

qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
-3
-2
-1
0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
p
p
0.0
0.2
0.4
0.6
0.8
1.0
-3
-2
-1
0
1
2
3
x

This is how I generated the plots on the previous page in the free software package R

(http://www.r-project.org/)

```
par(mfrow=c(1,2))

p <- seq(0,1,length=1000)

plot(p, qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE))

x <- rnorm(1000)

F<- ecdf(x)

points(F(sort(x)),sort(x),col="red")

plot(sort(x),F(sort(x)),xlab="x",ylab="p",col="red")
```

$$F_X^{-1}(y) = \inf_{x \in \mathbb{R}} \{F_X(x) \geq y\}$$



Some useful properties of the inverse CDF

- F^{-1} is non-decreasing
- $F^{-1}(F(x)) \leq x$
- $F(F^{-1}(y)) \geq y$
- $F^{-1}(y) \leq x \text{ if and only if } y \leq F(x)$
- If Y has a uniform distribution on [0,1] then $F^{-1}(Y)$ is distributed as F. This is used in random number generation using what is called "*the inverse transform sampling method*"

Quantiles

- By a **quantile**, we mean the fraction (or percent) of points below the given value. That is, the 0.3 (or 30%) quantile is the point at which 30% percent of the data fall below and 70% fall above that value.
- More formally, the qth quantile of a random variable X or of its corresponding distribution is defined as the smallest number x_q satisfying $F_X(x_q) \geq q$
- The **generalized** inverse cumulative distribution function is always a well defined function that gives the limiting value of the sample at qth quantile of the distribution of X: $F_X^{-1}(q) = \inf_{x \in \mathbb{R}} \{F_X(x) \geq q\}$

- Furthermore, if $F^{-1}(q)$, and hence X, is a continuous function, then it can be inverted to give us the cumulative distribution function for X (i.e., the unique quantile at which a particular value of x would appear in an ordered set of samples in the limit as N grows very large).
- Several approaches in statistics and visualization tools in statistics exist that are based on quantiles: box plots, qq-plots, quantile regression, … We refer to more details about these in subsequent chapters.

2.7 Mixed-type distributions

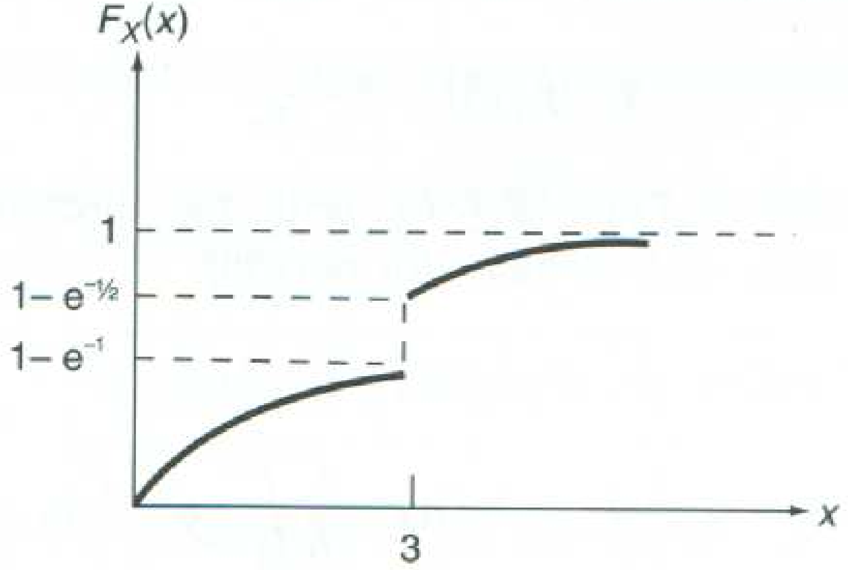
- There are situations in which one encounters a random variable that is partially discrete and partially continuous
- <u>Problem</u>: Since it is more economical to limit long-distance telephone calls to 3 minutes or less, the cumulative distribution of X – the duration in minutes of long-distance calls – may be of the form

$$F_X(x) = \begin{cases} 0, & \text{for } x < 0 \\ 1 - e^{-x/3}, & \text{for } 0 \leq x < 3 \\ 1 - \dfrac{e^{-x/3}}{2}, & \text{for } x \geq 3. \end{cases}$$

Determine the probability that X is (a) more than 2 minutes and (b) between 2 and 6 minutes.

A plot of the CDF is given below. Furthermore for part (a):

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F_X(2)$$
$$= 1 - (1 - e^{-2/3}) = e^{-2/3}.$$



For part (b):

$$P(2 < X \leq 6) = F_X(6) - F_X(2)$$
$$= \left(1 - \frac{e^{-2}}{2}\right) - (1 - e^{-2/3}) = e^{-2/3} - \frac{e^{-2}}{2}.$$

The partial probability density function of X is given by

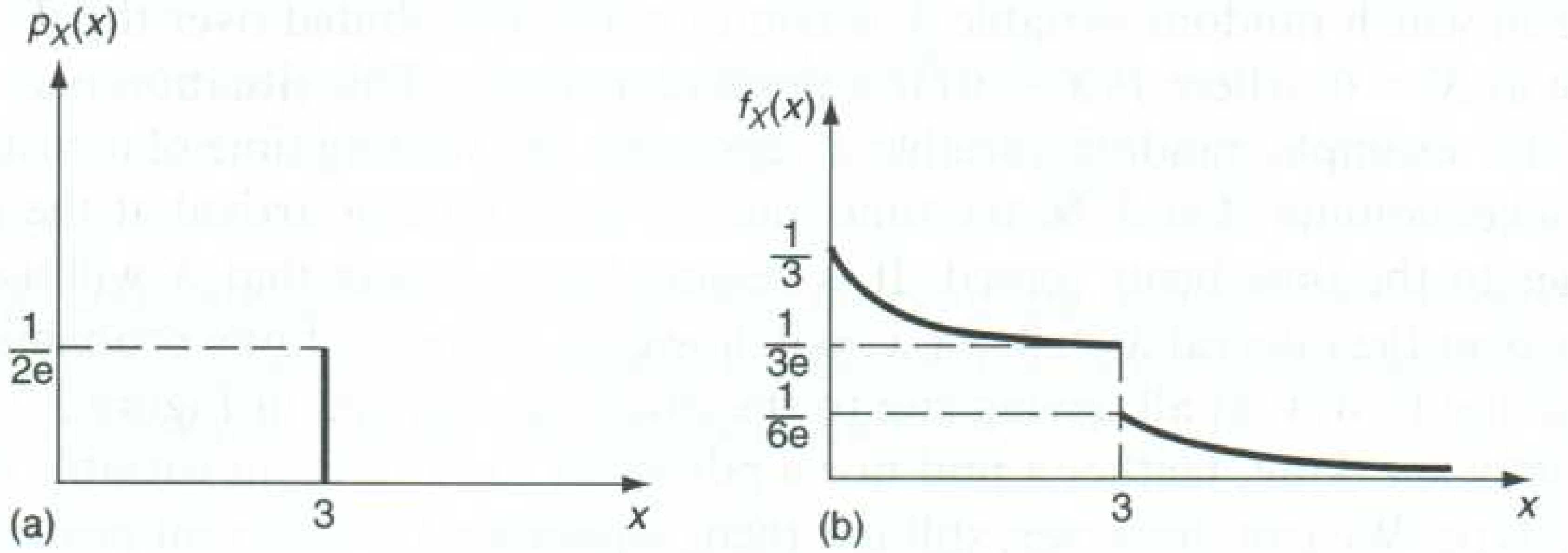
$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = \begin{cases} 0, & \text{for } x < 0; \\ \frac{1}{3}\,\mathrm{e}^{-x/3}, & \text{for } 0 \le x < 3; \\ \frac{1}{6}\,\mathrm{e}^{-x/3}, & \text{for } x \ge 3. \end{cases}$$

Note that the area under $f_X(x)$ is no longer 1 but

$$1 - p_X(3) = 1 - \frac{1}{2\mathrm{e}}.$$

Hence the partial probability mass function of X is given by

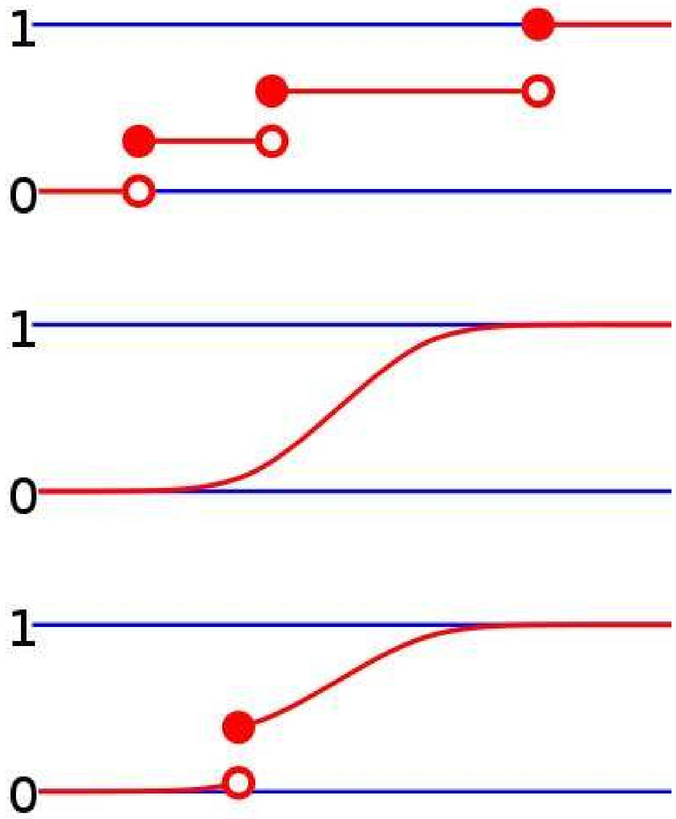
$$p_X(x) = \begin{cases} \frac{1}{2\,\mathrm{e}}, & \text{at } x = 3; \\ 0, & \text{elsewhere;} \end{cases}$$



To obtain $P(X > 2)$ and $P(2 < X \leq 6)$, both the discrete and continuous portions come into play, and we have, for part (a),

$$
\begin{aligned}
P(X > 2) &= \int_2^{\infty} f_X(x)\,\mathrm{d}x + p_X(3) \\
&= \frac{1}{3}\int_2^{3} \mathrm{e}^{-x/3}\,\mathrm{d}x + \frac{1}{6}\int_3^{\infty} \mathrm{e}^{-x/3}\,\mathrm{d}x + \frac{1}{2\mathrm{e}} \\
&= \mathrm{e}^{-2/3}
\end{aligned}
$$

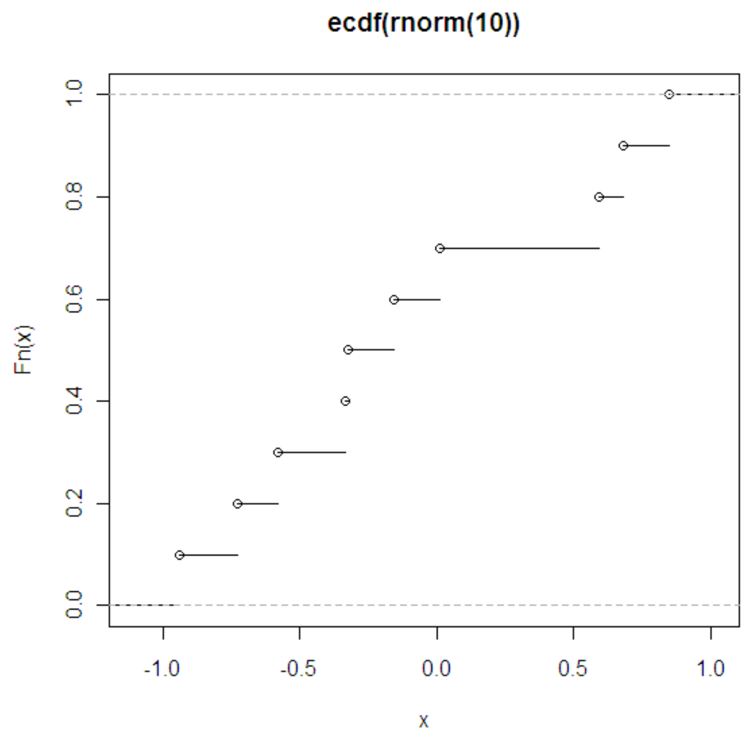
In conclusion:



From top to bottom, the cumulative distribution function of a discrete probability distribution, continuous probability distribution,
and a distribution which has both a continuous part and a discrete part.

Even though the underlying model that generated these date come from a normal (hence non-discrete) distribution, I only have a discrete nr of possibilities because of the small size (here: 10) of my sample.

- When drafting a cumulative distribution function from the 10 generated sample points, the CDF graph will look as if generated for a discrete random variable.



2.8 Comparing cumulative distribution functions

- In statistics, the **empirical distribution function**, or empirical CDF, is the cumulative distribution function associated with the empirical measure of the sample at hand. This CDF is a step function that jumps for 1/n at each of the n data points. The empirical distribution function *estimates* the true underlying CDF of the points in the sample.
- The Kolmogorov–Smirnov (KS) test is based on quantifying a distance between cumulative distribution functions and can be used to test to see whether two empirical distributions are different or whether an empirical distribution is different from an ideal distribution (i.e. a reference distribution).
- The 2-sample KS test is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples.

3 Two or more random variables

- In many cases it is more natural to describe the outcome of a random experiment by two or more numerical numbers simultaneously, such as when characterizing both weight and height in a given population
- When for instance two random variables X and Y are in play, we can also consider these as components of a two-dimensional random vector, say **Z**
- Joint probability distributions, for X and Y jointly, are sometimes referred to as **bivariate distributions**.
- Although most of the time we will give examples for two-variable scenarios, the definitions, theorems and properties can easily be extended to **multivariate** scenarios (dimensions > 2)

3.1 Joint probability distribution functions

- The joint probability distribution function of random variables X and Y, denoted by $F_{XY}(x, y)$, is defined by

$$F_{XY}(x, y) = P(X \leq x \cap Y \leq y),$$

for all x, y

- As before, some obvious properties follow from this definition of joint **cumulative distribution function**:

$$\left.\begin{aligned} &F_{XY}(-\infty, -\infty) = F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0, \\ &F_{XY}(+\infty, +\infty) = 1, \\ &\quad F_{XY}(x, +\infty) = F_X(x), \\ &\quad F_{XY}(+\infty, y) = F_Y(y). \end{aligned}\right\}$$

- $F_X(x)$ and $F_Y(y)$ are called **marginal distribution functions** of X and Y, resp.

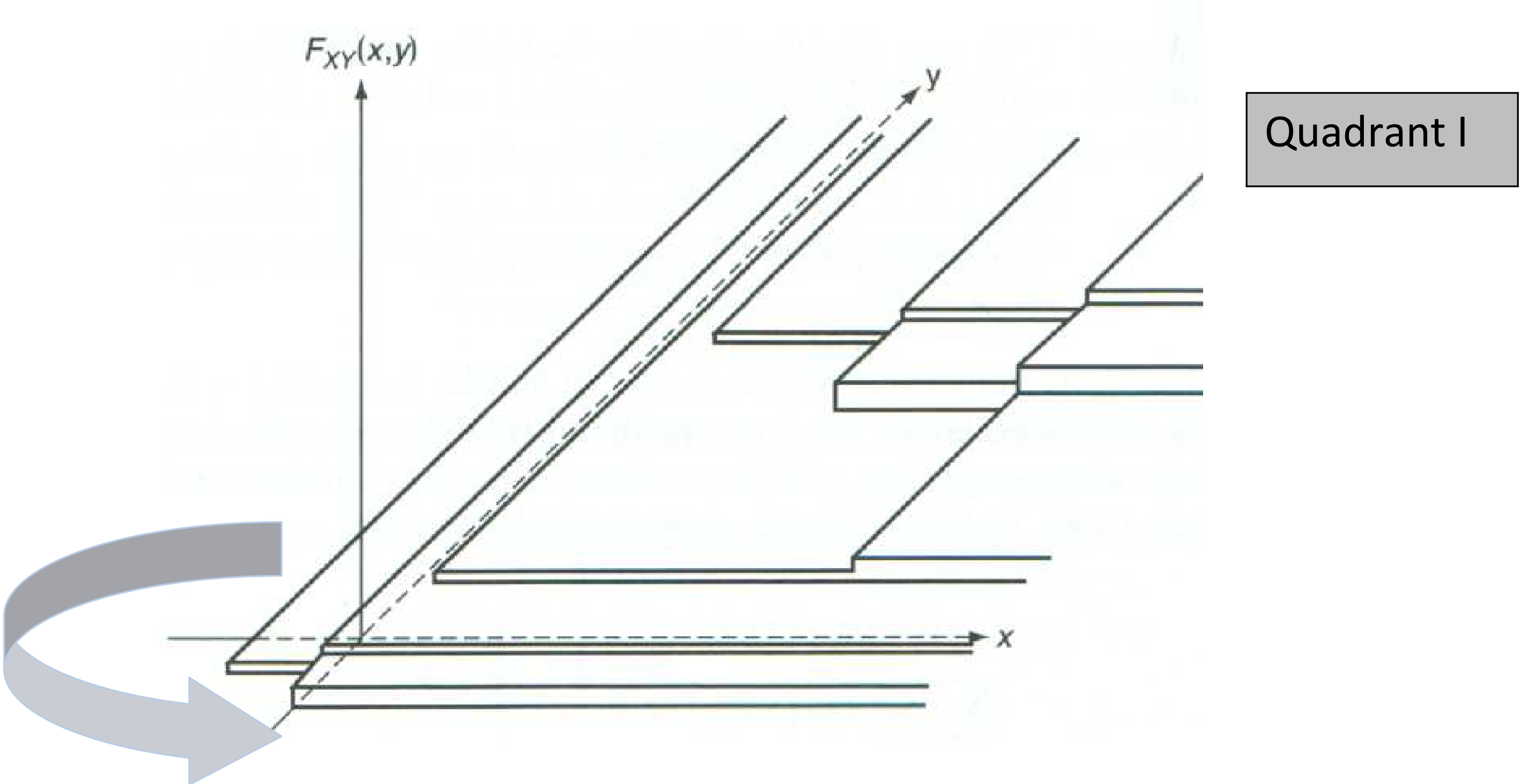
- Also, it can be shown that the probability $P(x_1 < X \leq x_2 \cap y_1 < Y \leq y_2)$ is given in terms of $F_{XY}(x, y)$ by

$$F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1),$$

indicating that all probability calculations involving random variables X and Y can be made with the knowledge of their joint cumulative distribution function

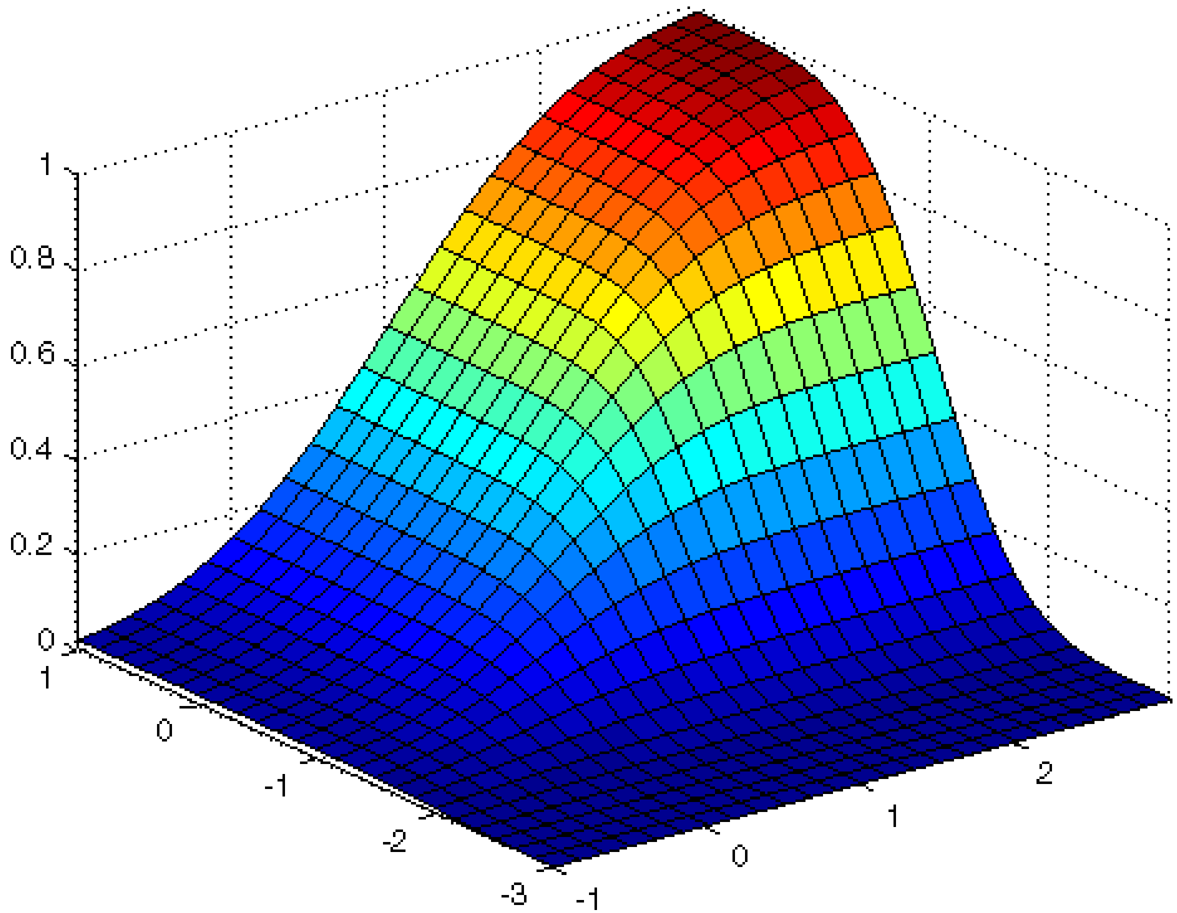
- The general shape of $F_{XY}(x, y)$ can be visualized from the properties given before:

- In case of X and Y being discrete, their joint probability distribution function has the appearance of a corner of an irregular staircase, as shown below.

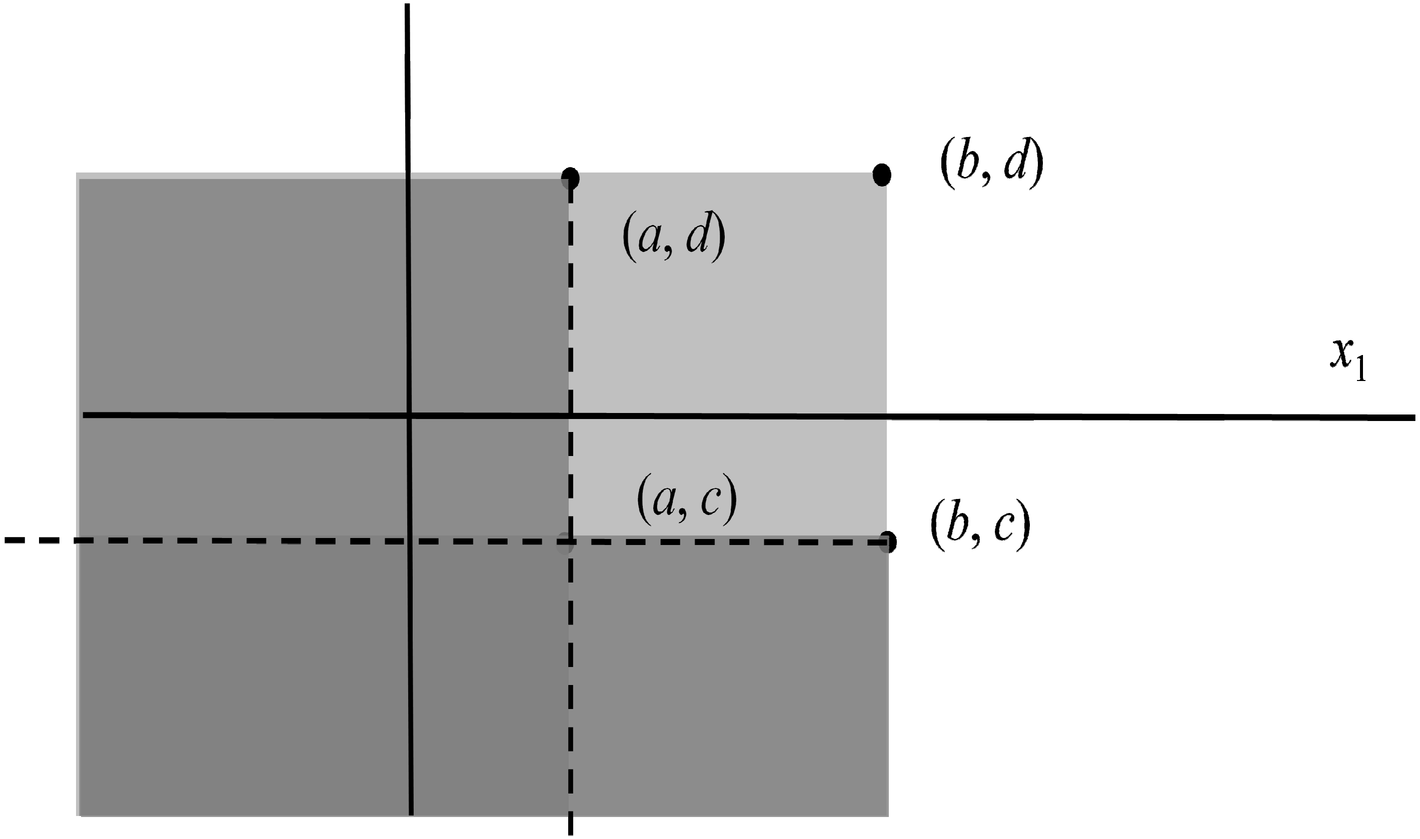


- It rises from 0 to height 1 in moving from quadrant III to quadrant I

- In case of X and Y being continuous, their joint probability distribution function becomes a smooth surface with the same features as in the discrete case:



- Intuition behind the computation of $P(a < X \leq b \cap c < Y \leq d)$ as $F_{XY}(b,d) - F_{XY}(a,d) - F_{XY}(b,c) + F_{XY}(a,c)$



Copulas

- In probability theory and statistics, a **copula** can be used to describe the dependence between random variables. Copulas derive their name from linguistics.
- The cumulative distribution function of a random vector can be written in terms of marginal distribution functions and a copula. The marginal distribution functions describe the marginal distribution of each component of the random vector and the copula describes the dependence structure between the components.
- Copulas are popular in statistical applications as they allow one to easily model and estimate the distribution of random vectors by estimating marginals and copula separately. There are many parametric copula families available, which usually have parameters that control the strength of dependence

- Consider a random vector (X_1, X_2) and suppose that its margins F_1 and F_2 are continuous. By applying the **probability integral transformation** to each component, the random vector

$$(U_1, U_2) = (F_1(X_1), F_2(X_2))$$

has uniform margins. The **copula** of (X_1, X_2) is defined as the joint cumulative distribution function of (U_1, U_2):

$$C(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2)$$

- Note that it is also possible to write

$$(X_1, X_2) = (F_1^{-1}(U_1), F_2^{-1}(U_2)),$$

Where the inverse functions are unproblematic as the marginal distribution functions were assumed to be continuous. The analogous identity for the copula is

$$C(u_1, u_2) = P(X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2))$$

- Sklar's theorem provides the theoretical foundation for the application of copulas. Sklar's theorem states that a multivariate cumulative distribution function

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$$

of a random vector (x_1, x_2) with margins $F_1(x_1) = F_{X_1}(x_1)$ and $F_2(x_2) = F_{X_2}(x_2)$ can be written as $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$, where C is a copula.

- The theorem also states that given $F(x_1, x_2)$, the copula is unique on $\mathrm{Range}(F_1) \times \mathrm{Range}(F_2)$, which is the Cartesian product of the ranges of the marginal CDF’s.
- The converse is also true: given a copula $C : [0,1]^2 \to [0,1]$ and margins $F_i(x)$ then $C(F_1(x_1), F_2(x_2))$ defines a 2-dimensional cumulative distribution function.

3.2 The discrete case: joint probability mass functions

- Let X and Y be two discrete random variables that assume at most a countable infinite number of value pairs (x_i, y_j), i,j = 1,2, …, with nonzero probabilities. Then the **joint probability mass function** of X and Y is defined by

$$P_{XY}(x, y) = P(X = x \cap Y = y),$$

for all x and y. It is zero everywhere except at the points (x_i, y_j), i,j = 1,2, …,

where it takes values equal to the joint probability $P(X = x_i \cap Y = y_j)$.

- As a direct consequence of this definition:

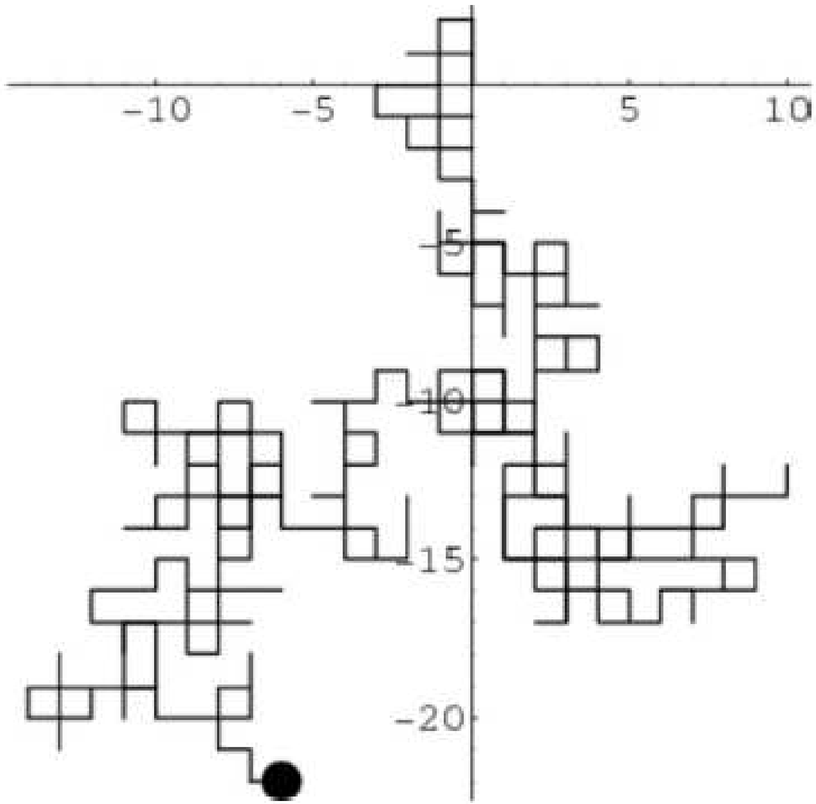
$$\left.\begin{aligned} 0 < p_{XY}(x_i, y_j) \le 1, \\ \sum_i \sum_j p_{XY}(x_i, y_j) = 1, \\ \sum_i p_{XY}(x_i, y) = p_Y(y), \\ \sum_j p_{XY}(x, y_j) = p_X(x), \end{aligned}\right\}$$

  Here, $P_X(x) \,\text{and}\, P_Y(y)$ are called **marginal probability mass functions.**

- Also,

$$F_{XY}(x, y) = \sum_{i=1}^{i:x_i \le x} \sum_{j=1}^{j:y_j \le y} P_{XY}(x_i, y_j)$$

A two-dimensional simplified random walk

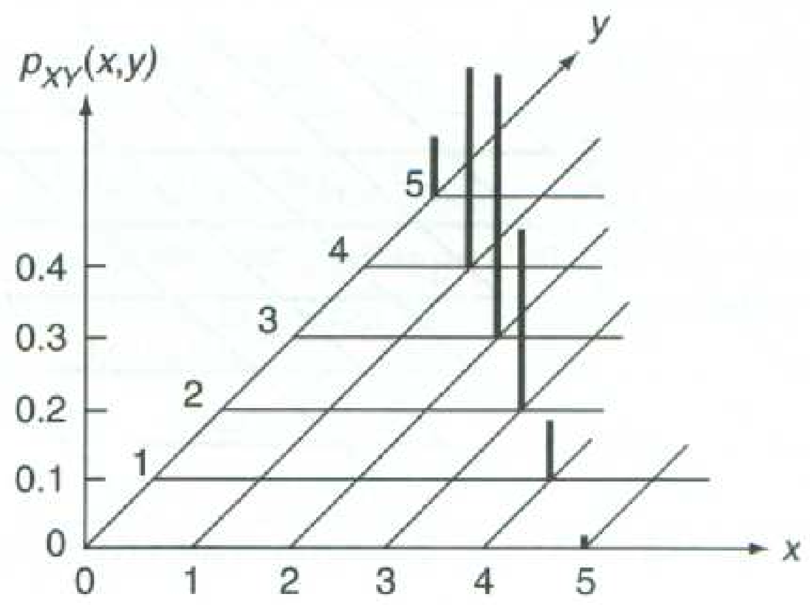


- It has been proven that on a two-dimensional lattice, a *random walk* like this has unity probability of reaching any point (including the starting point) as the number of steps approaches infinity.

- Now, we imagine a particle that moves in a plane in unit steps starting from the origin. Each step is one unit in the positive direction, with probability p along the x axis and probability q (p+q=1) along the y axis. We assume that each step is taken independently of the others.
- Question: What is the probability distribution of the position of this particle after 5 steps?
- Answer: We are interested in $P_{XY}(x,y)$ with the random variable X representing the x coordinate and the random variable Y representing the y-coordinate of the particle position after 5 steps.
  - Clearly, $P_{XY}(x,y) = 0$ except at $x + y = 5 \, \text{and} \, x, y \geq 0$
  - Because of the independence assumption $P_{XY}(5,0) = p^5$
  - Similarly, $P_{XY}(4,1) = 5p^4q$

- In all other settings:

$$p_{XY}(x,y) = \begin{cases} 10p^3q^2, & \text{for } (x,y) = (3,2); \\ 10p^2q^3, & \text{for } (x,y) = (2,3); \\ 5pq^4, & \text{for } (x,y) = (1,4); \\ q^5, & \text{for } (x,y) = (0,5). \end{cases}$$



Check whether the sum over all x, y equals 1 (as should be!)

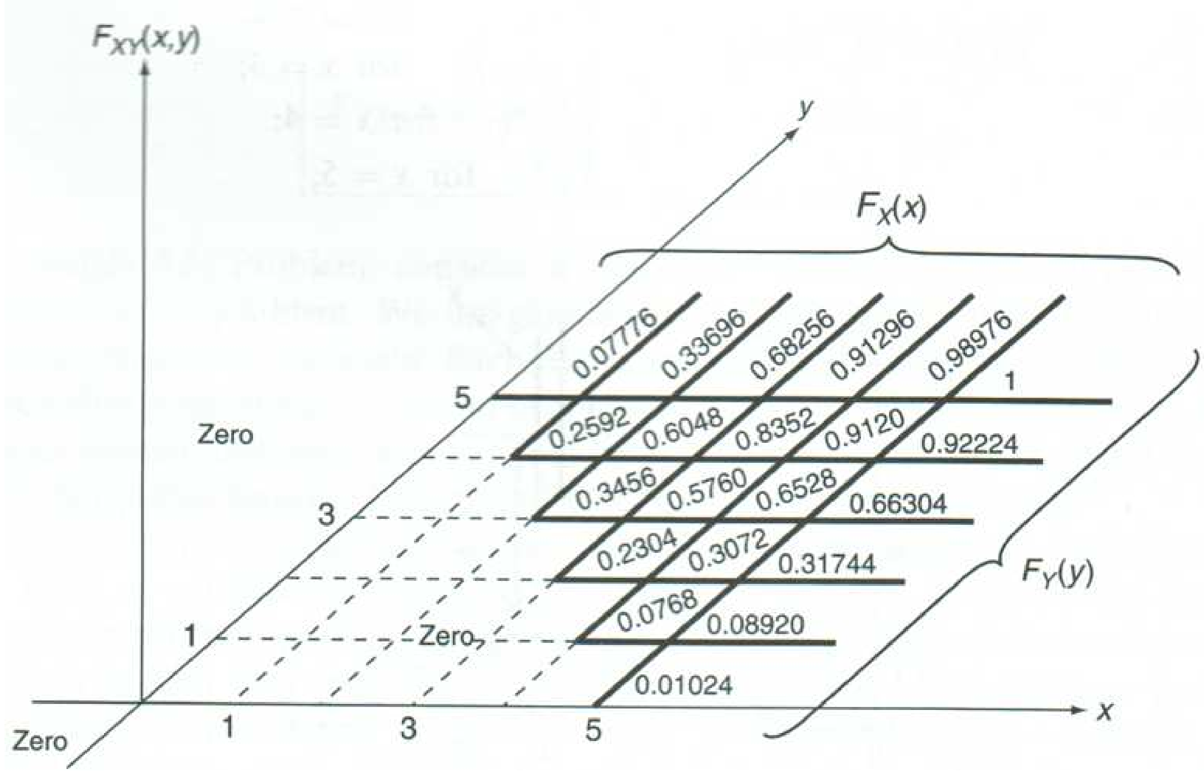
- From this, the marginal probability mass functions can be derived:

$$p_X(x) = \sum_j p_{XY}(x, y_j) = \begin{cases} q^5, & \text{for } x = 0; \\ 5pq^4, & \text{for } x = 1; \\ 10p^2q^3, & \text{for } x = 2; \\ 10p^3q^2, & \text{for } x = 3; \\ 5p^4q, & \text{for } x = 4; \\ p^5, & \text{for } x = 5; \end{cases}$$

$$p_Y(y) = \sum_i p_{XY}(x_i, y) = \begin{cases} p^5, & \text{for } y = 0; \\ 5p^4q, & \text{for } y = 1; \\ 10p^3q^2, & \text{for } y = 2; \\ 10p^2q^3, & \text{for } y = 3; \\ 5pq^4, & \text{for } y = 4; \\ q^5, & \text{for } y = 5. \end{cases}$$

- The joint probability distribution function can also be derived using the formulae seen before (example shown for p=0.4 and q=0.6):

$$F_{XY}(x,y) = \sum_{i=1}^{i:x_i \leq x} \sum_{j=1}^{j:y_j \leq y} P_{XY}(x_i, y_j)$$



3.3 The continuous case: joint probability density functions

- The joint probability density function $f_{XY}(x,y)$ of 2 continuous random variables X and Y is defined by the partial derivative

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

- Since $F_{XY}(x,y)$ is monotone non-decreasing in both x and y, the associated joint probability density function is nonnegative for all x and y.
- As a direct consequence:

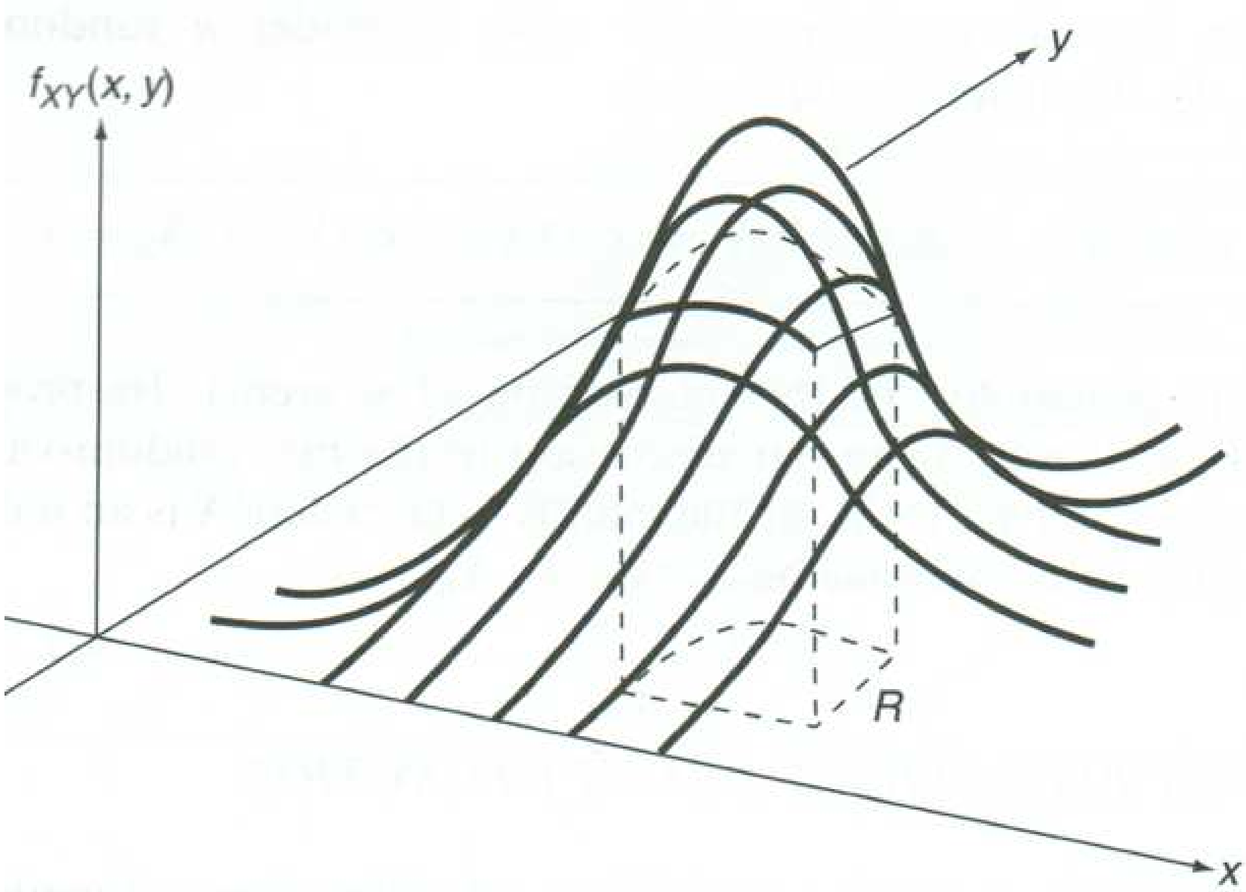
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x\mathrm{d}y = 1,$$

$$\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}y = f_X(x),$$

$$\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x = f_Y(y).$$

where $f_X(x) \text{ and } f_Y(y)$ are now called the marginal density functions of X and Y respectively

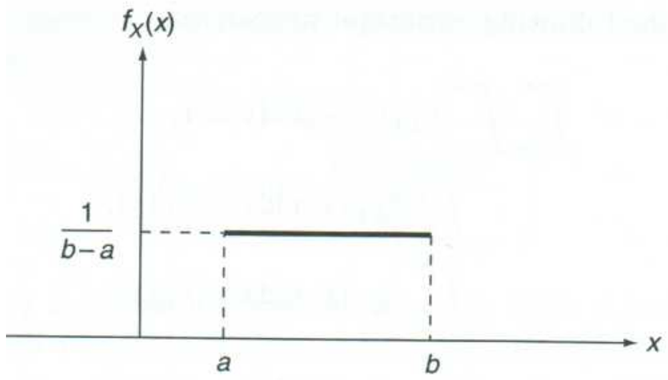
- Also, $F_{XY}(x,y) = P(X \leq x \cap Y \leq y) = \int_{<\infty}^{y} \int_{-\infty}^{x} f_{XY}(u,v)dudv$ and $P(x_1 < X \leq x_2 \cap y_1 < Y \leq y2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x,y)dxdy$ for $x_1 < x_2 \, \text{and} \, y_1 < y_2$

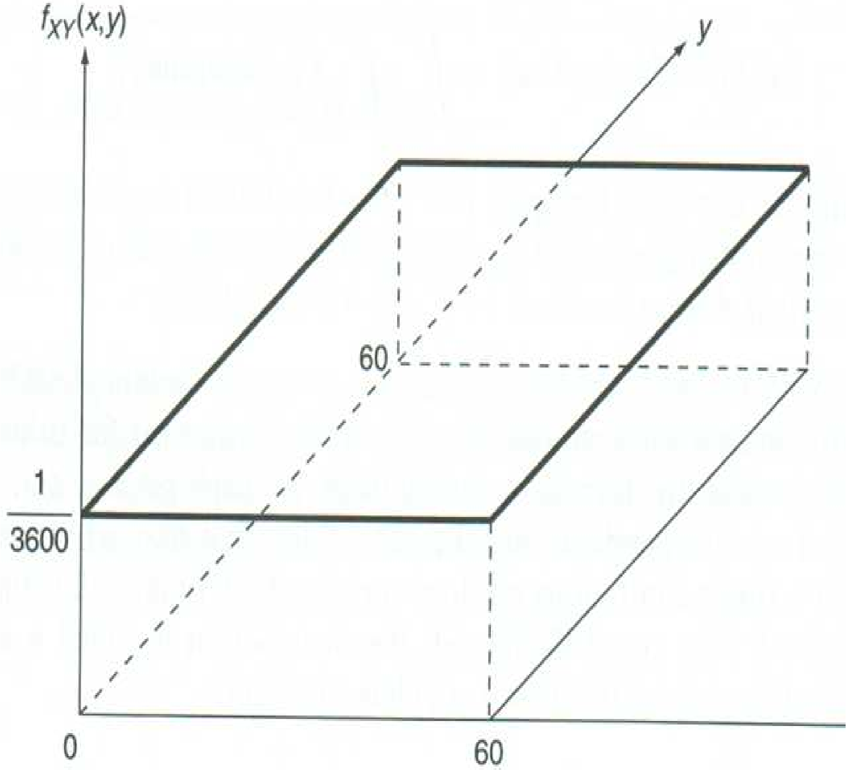


Meeting times

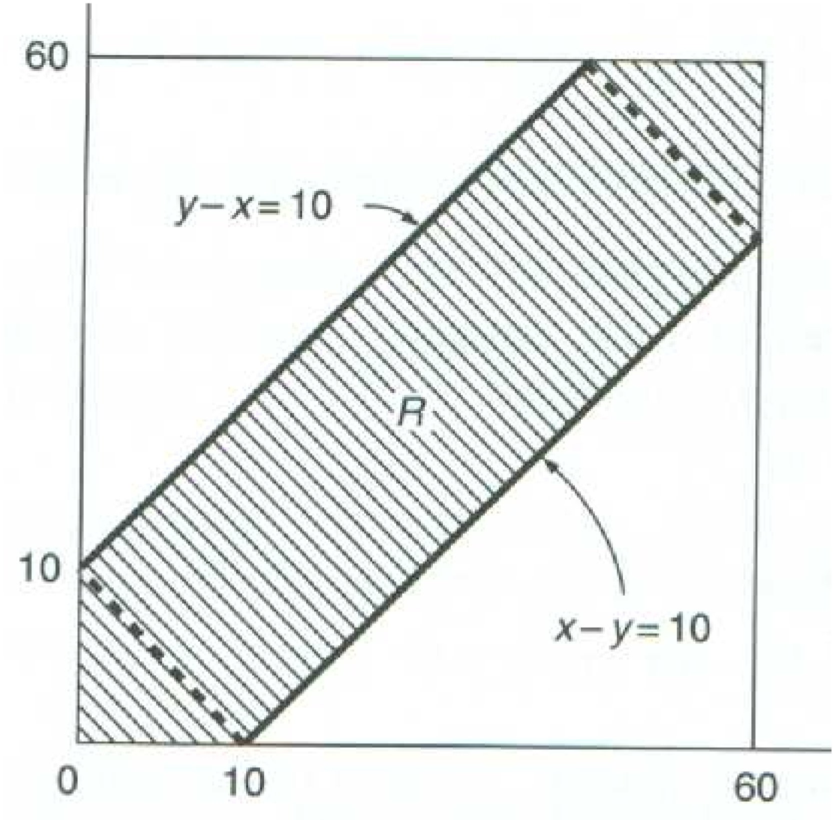
- A boy and a girl plan to meet at a certain place between 9am and 10am, each not wanting to wait more than 10 minutes for the other. If all times of arrival within the hour are equally likely for each person, and if their times of arrival are independent, find the probability that they will meet.
- Answer: for a single continuous random variable X that takes all values over an interval a to b with equal likelihood, the distribution is called a uniform distribution and its density function has the form

$$f_X(x) = \begin{cases} \frac{1}{b-a} & , \text{for } a \leq x \leq b \\ 0 & , \text{otherwise} \end{cases}$$





The joint density function of two independent uniformly distributed random variables is a flat surface within prescribed bounds. The volume under the surface is unity.



$$P(\text{they will meet}) = P(|X - Y| \leq 10)$$
$$= [2(5)(10) + 10\sqrt{2}(50\sqrt{2})]/3600 = \frac{11}{36}$$

- We can derive from the joint probability, the joint probability distribution function, as usual

$$F_{XY}(x, y) = \begin{cases} 0, & \text{for } (x, y) < (0, 0); \\ 1, & \text{for } (x, y) > (60, 60). \end{cases}$$

$$F_{XY}(x, y) = \int_0^y \int_0^x \left(\frac{1}{3600}\right) \mathrm{d}x\mathrm{d}y = \frac{xy}{3600}.$$

- From this we can again derive the marginal probability density functions, which clearly satisfy the earlier definition for 2 random variables that are uniformly distributed over the interval [0,60]

4 Conditional distribution and independence

- The concepts of conditional probability and independence introduced before also play an important role in the context of random variables
- The **conditional distribution** of a random variable X, given that another random variable Y has taken a value y, is defined by

$$F_{XY}(x|y) = P(X \leq x|Y = y)$$

- When a random variable X is discrete, the definition of **conditional mass function** of X given Y=y is

$$p_{XY}(x|y) = P(X = x|Y = y)$$

- For a continuous random variable X, the **conditional density function** of X given Y=y is

$$f_{XY}(x|y) = \frac{dF_{XY}(x|y)}{dx}$$

- In the discrete case, using the definition of conditional probability, we have

$$p_{XY}(x|y) = P(X = x|Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)},$$

$$p_{XY}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}, \text{ if } p_Y(y) \neq 0,$$

an expression which is very useful in practice when wishing to derive joint probability mass functions …

- Using the definition of independent events in probability theory, when the random variables X and Y are assumed to be independent,

$$P_{XY}(x|y) = P_X(x)$$

so that $P_{XY}(x,y) = P_X(x)P_Y(y)$

- The definition of a conditional density function for a random continuous variable X, given Y=y, entirely agrees with intuition …:

$$P(x_1 < X \leq x_2 | y_1 < Y \leq y_2) = \frac{P(x_1 < X \leq x_2 \cap y_1 < Y \leq y_2)}{P(y_1 < Y \leq y_2)}.$$

In terms of jpdf $f_{XY}(x, y)$, it is given by

$$P(x_1 < X \leq x_2 | y_1 < Y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \Big/ \int_{y_1}^{y_2} \int_{-\infty}^{\infty} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \Big/ \int_{y_1}^{y_2} f_Y(y)\,\mathrm{d}y.$$

By setting $x_1 = -\infty, x_2 = x, y_1 = y, y_2 = y + \Delta y$ and by taking the limit $\Delta y \to 0$, this reduced to

$$F_{XY}(x|y) = \frac{\int_{-\infty}^{x} f_{XY}(u, y)\,\mathrm{d}u}{f_Y(y)},$$

provided $f_Y(y) \neq 0$

From

$$f_{XY}(x|y) = \frac{dF_{XY}(x|y)}{dx}$$

and

$$F_{XY}(x|y) = \frac{\int_{-\infty}^{x} f_{XY}(u, y)\, \mathrm{d}u}{f_Y(y)},$$

we can derive that

$$f_{XY}(x|y) = \frac{\mathrm{d}F_{XY}(x|y)}{\mathrm{d}x} = \frac{f_{XY}(x, y)}{f_Y(y)}, \quad f_Y(y) \neq 0,$$

a form that is identical to the discrete case. But note that

$$F_{XY}(x|y) \neq \frac{F_{XY}(x, y)}{F_Y(y)}.$$

- When random variables X and Y are independent, however, $F_{XY}(x|y) = F_X(x)$ (using the definition for $F_{XY}(x|y)$) and (using the expression

  $$f_{XY}(x|y) = \frac{\mathrm{d}F_{XY}(x|y)}{\mathrm{d}x} = \frac{f_{XY}(x,y)}{f_Y(y)}, \quad f_Y(y) \neq 0,$$

  it follows that

  $$f_{XY}(x|y) = f_X(x),$$

  $$f_{XY}(x,y) = f_X(x) f_Y(y),$$

- Finally, when random variables X and Y are discrete,

$$F_{XY}(x|y) = \sum_{i=1}^{i:x_i \leq x} p_{XY}(x_i|y),$$

and in the case of a continuous random variable,

$$F_{XY}(x|y) = \int_{-\infty}^{x} f_{XY}(u|y)\mathrm{d}u.$$

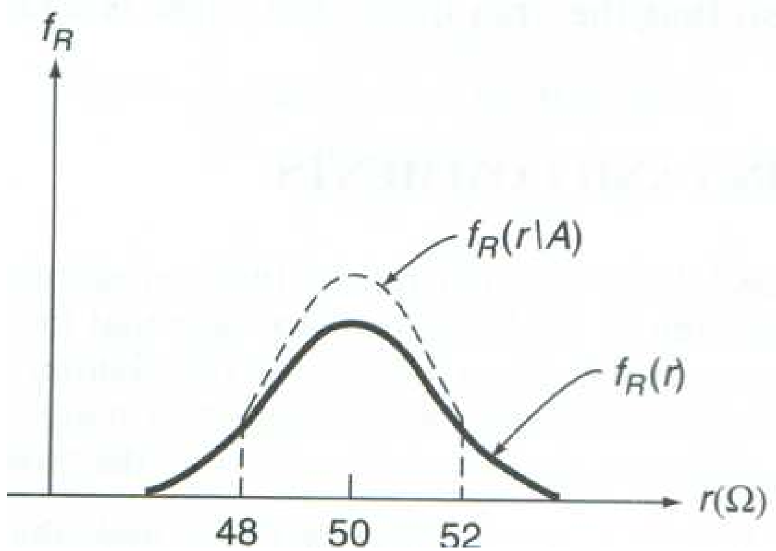
Note that these are very similar to those relating the distribution and density functions in the univariate case.

- Generalization to more than two variables should now be straightforward, starting from the probability expression

$$P(ABC) = P(A|BC)P(B|C)P(C)$$

Resistor problem

- Resistors are designed to have a resistance of R of $50 \pm 2\Omega$. Owing to some imprecision in the manufacturing process, the actual density function of R has the form shown (right), by the solid curve.
- Determine the density function of R after screening (that is: after all the resistors with resistances beyond the 48-52 Ω range are rejected.
- Answer: we are interested in the conditional density function $f_R(r|A)$ where A is the event $\{48 \leq R \leq 52\}$



We start by considering

$$F_R(r|A) = P(R \leq r | 48 \leq R \leq 52) = \frac{P(R \leq r \cap 48 \leq R \leq 52)}{P(48 \leq R \leq 52)}.$$

However,

$$R \leq r \cap 48 \leq R \leq 52 = \begin{cases} \emptyset, & \text{for } r < 48; \\ 48 \leq R \leq r, & \text{for } 48 \leq r \leq 52; \\ 48 \leq R \leq 52, & \text{for } r > 52. \end{cases}$$

Hence,

$$F_R(r|A) = \begin{cases} 0, & \text{for } r < 48; \\ \dfrac{P(48 \leq R \leq r)}{P(48 \leq R \leq 52)} = \dfrac{\int_{48}^{r} f_R(r)\mathrm{d}r}{c}, & \text{for } 48 \leq r \leq 52; \\ 1, & \text{for } r > 52; \end{cases}$$

Where

$$c = \int_{48}^{52} f_R(r)dr$$

is a constant.

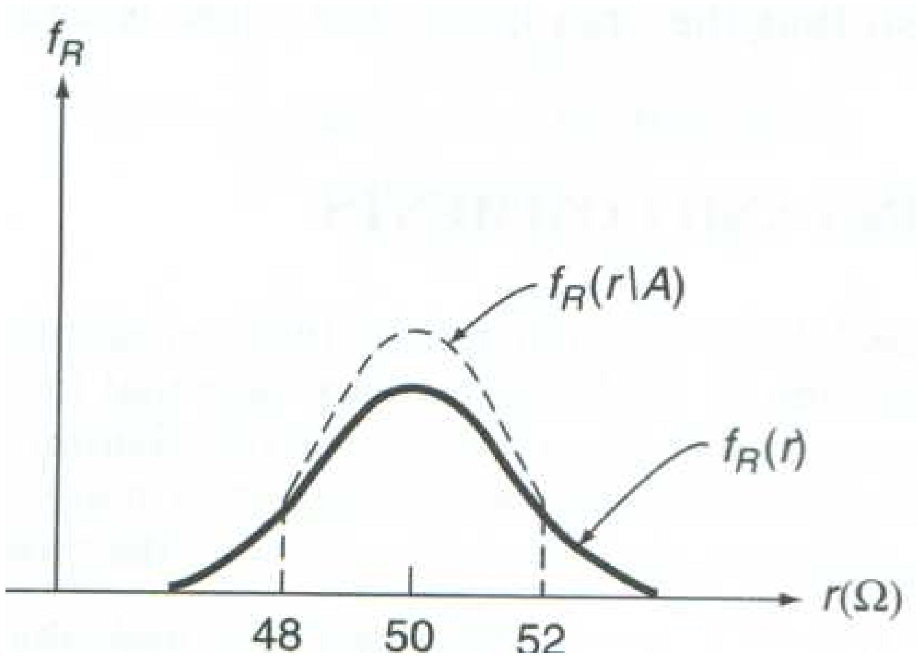
The desired function is then obtained by differentiation. We thus obtain

$$f_R(r|A) = \frac{dF_R(r|A)}{dr} = \begin{cases} \frac{f_R(r)}{c} & \text{for } 48 \leq r \leq 52 \\ 0 & \text{otherwise} \end{cases}$$

Now, look again at a graphical representation of this function. What do you observe?

Answer:

The effect of screening is essentially a truncation of the tails of the distribution beyond the allowable limits. This is accompanied by an adjustment within the limits by a multiplicative factor 1/c so that the area under the curve is again equal to 1.



5 Expectations and moments

5.1 Mean, median and mode

Expectations

- Let g(X) be a real-valued function of a random variable X. The mathematical expectation or simply expectation of g(X) is denoted by E(g(X)) and defined as

$$E(g(X)) = \sum_i g(x_i) P_X(x_i)$$

if X is discrete where $x_1, x_2, \ldots$are possible values assumed by X.

- When the range of i extends from 1 to infinity, the sum above exists if it converges absolutely; that is, $\sum_{i=1}^{\infty} |g(x_i)| P_X(x_i) < \infty$

- If the random variable X is continuous, then

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx,$$

if the improper integral is absolutely convergent, that is,

$$\int_{-\infty}^{+\infty} |g(x)| f_X(x) dx < \infty$$

then this number will exist.

- Basic properties of the expectation operator E(.), for any constant c and any functions g(X) and h(X) for which expectations exist include:

$$\left.\begin{array}{l} E\{c\} = c, \\ E\{cg(X)\} = cE\{g(X)\}, \\ E\{g(X) + h(X)\} = E\{g(X)\} + E\{h(X)\}, \\ E\{g(X)\} \leq E\{h(X)\}, \quad \text{if } g(X) \leq h(X) \text{ for all values of } X. \end{array}\right\}$$

Proofs are easy. For example, in the 3rd scenario and continuous case :

$$\begin{aligned} E\{g(X) + h(X)\} &= \int_{-\infty}^{\infty} [g(x) + h(x)] f_X(x) \mathrm{d}x \\ &= \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x + \int_{-\infty}^{\infty} h(x) f_X(x) \mathrm{d}x \\ &= E\{g(X)\} + E\{h(X)\}, \end{aligned}$$

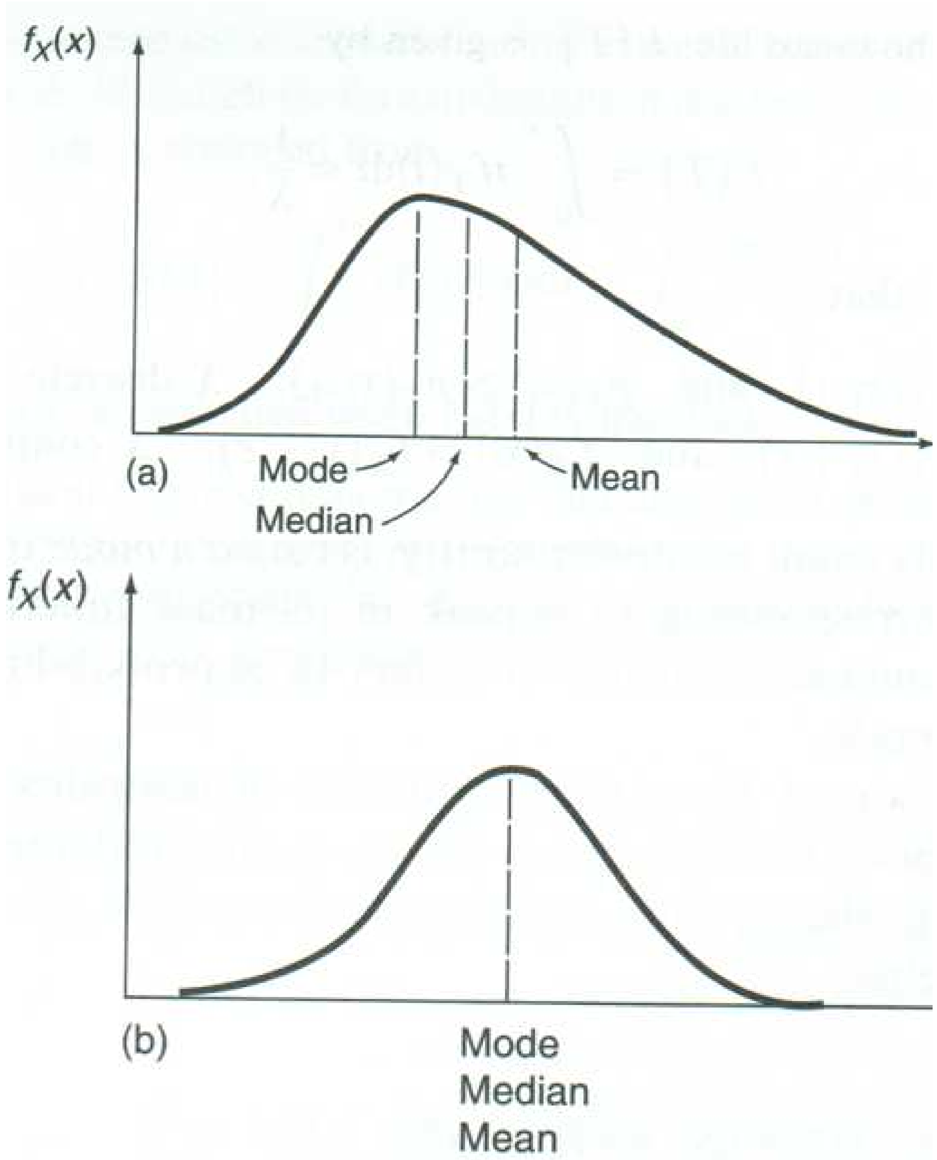
Moments of a single random variable

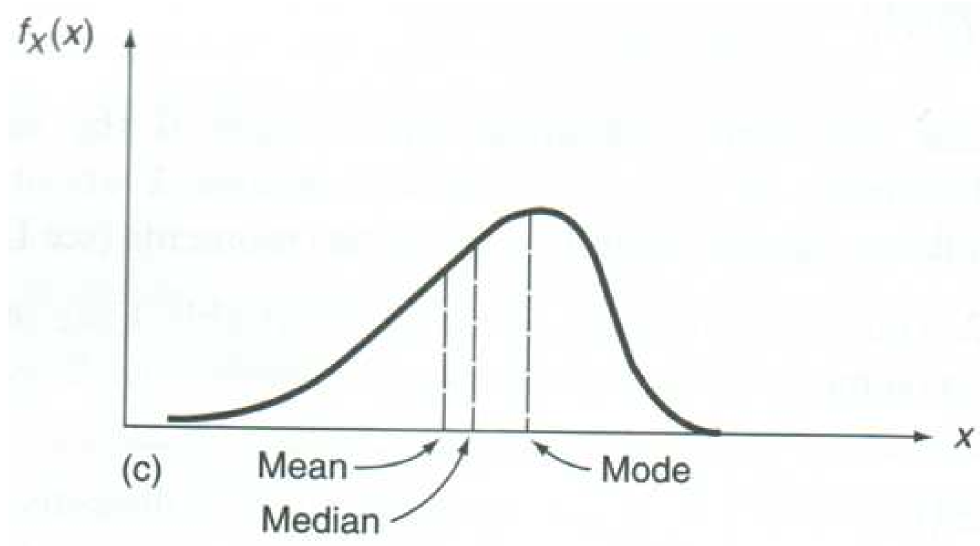
- Let $g(X) = X^n, n = 1, 2, ...$; the expectation $E(X^n))$, when it exists, is called the **nth moment** of X and denoted by μ'_n:

$$E\{X^n\} = \sum_i x_i^n p_X(x_i), \text{ for } X \text{ discrete;}$$

$$E\{X^n\} = \int_{-\infty}^{\infty} x^n f_X(x)\mathrm{d}x, \text{ for } X \text{ continuous.}$$

- The **first moment of X** is also called the **mean**, expectation, average value of X and is a measure of centrality
- Two other *measures of centrality* of a random variable:
  - A **median** of X is any point that divides the mass of its distribution into two equal parts → think about our quantile discussion
  - A **mode** is any value of X corresponding to a peak in its mass function or density function





From left to right: positively skewed, negatively skewed, symmetrical distributions

Time between emissions of particles

- Let T be the time between emissions of particles by a radio-active atom. It is well-established that T is a random variable and that it obeys what is called an exponential distribution (λ a positive constant):

$$f_\tau(t) = \begin{cases} e^{-\lambda t} & \text{for } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

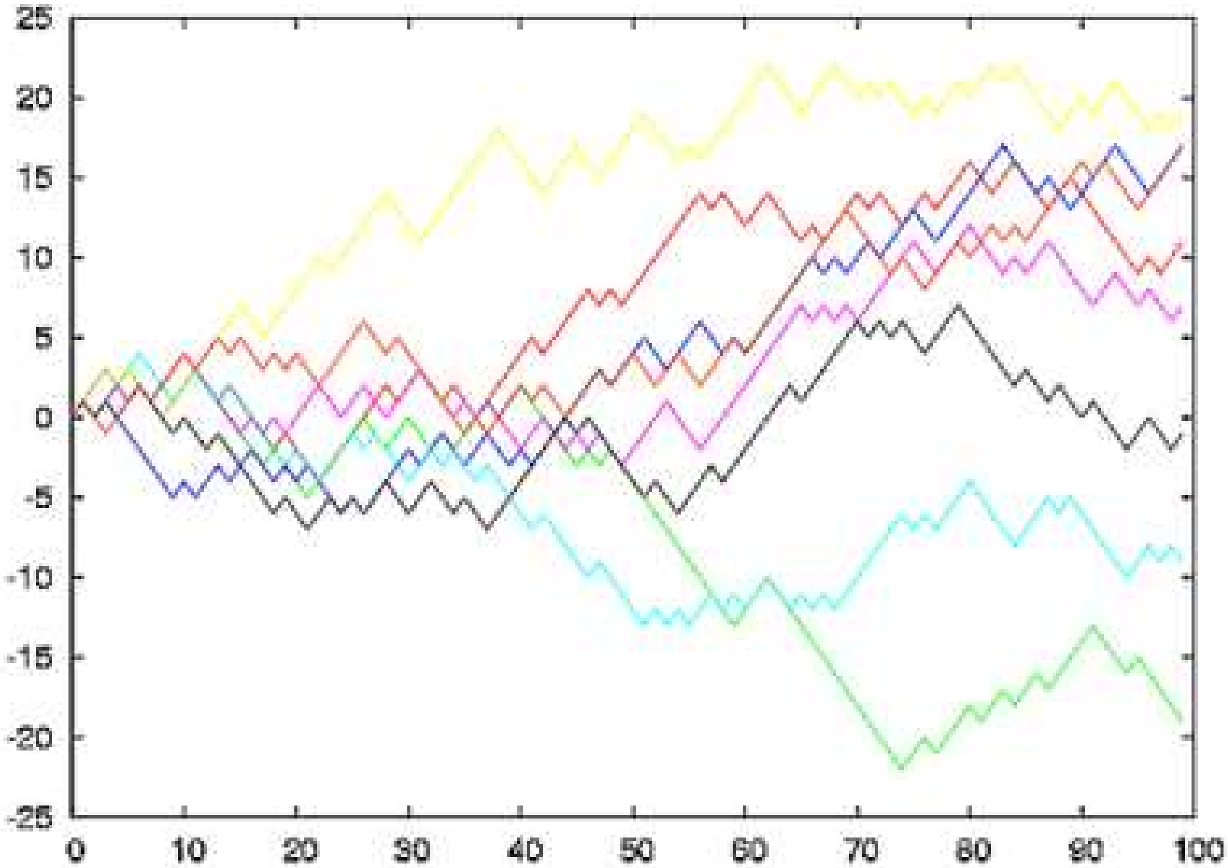
- The random variable T is called the lifetime of the atom, and a common average measure of this lifetime is called the half-life which is defined as the median of T. Thus the half-life, τ is found from

$$\int_0^\tau f_\tau(t)dt = 1/2, \text{ or } \tau = \ln(2/\lambda)$$

- The mean life time E(T) is given by $E(T) = \int_0^\infty t f_\tau(t)dt = 1/\lambda$

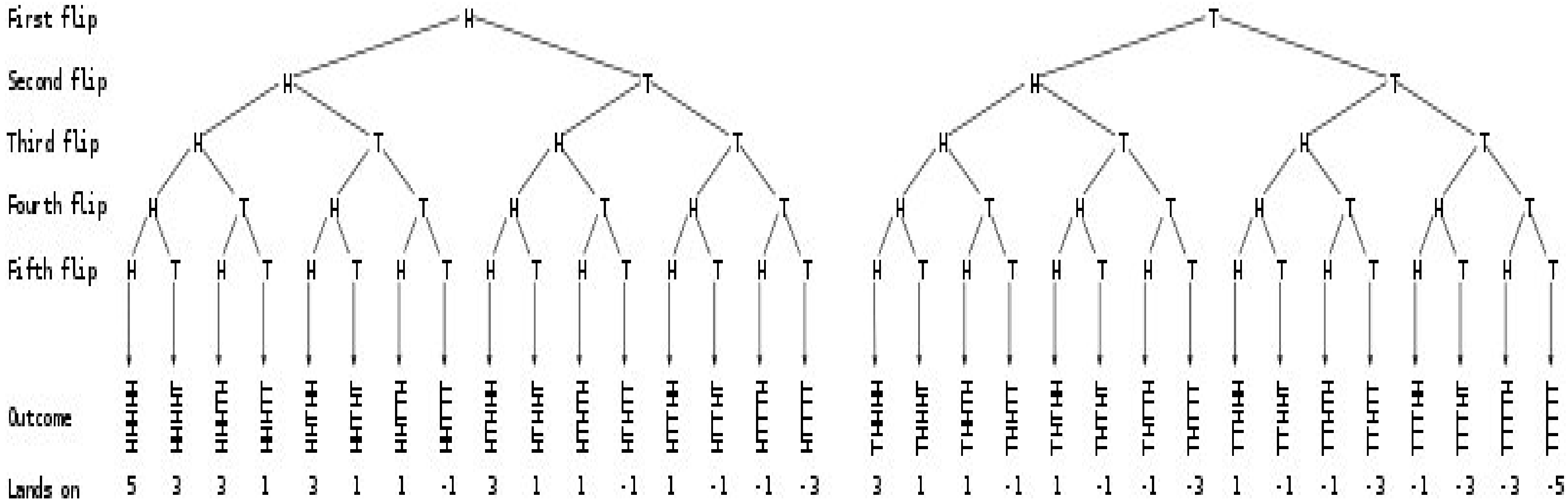
A one-dimensional random walk

- An elementary example of a random walk is the random walk on the integer number line, which starts at 0 and at each step moves +1 or −1 with equal probability.
- This walk can be illustrated as follows: A marker is placed at zero on the number line and a fair coin is flipped. If it lands on heads, the marker is moved one unit to the right. If it lands on tails, the marker is moved one unit to the left. After five flips, it is possible to have landed on 1, −1, 3, −3, 5, or −5. With five flips, three heads and two tails, in any order, will land on 1. There are 10 ways of landing on 1 or −1 (by flipping three tails and two heads), 5 ways of landing on 3 (by flipping four heads and one tail), 5 ways of landing on −3 (by flipping four tails and one head), 1 way of landing on 5 (by flipping five heads), and 1 way of landing on −5 (by flipping five tails).



- Example of eight random walks in one dimension starting at 0. The plot shows the current position on the line (vertical axis) versus the time steps (horizontal axis).

- See the figure below for an illustration of the possible outcomes of 5 flips.



- To define this walk formerly, take independent random variables $Z_1, Z_2, \ldots$ where each variable is either 1 or -1 with a 50% probability for either value, and set $S_0 = 0$ and $S_n = \sum_{j=1}^{n} Z_j$. The series is called the simple random walk on $\mathbb{Z}$. This series of 1's and -1's gives the distance walked, if each part of the walk is of length 1.

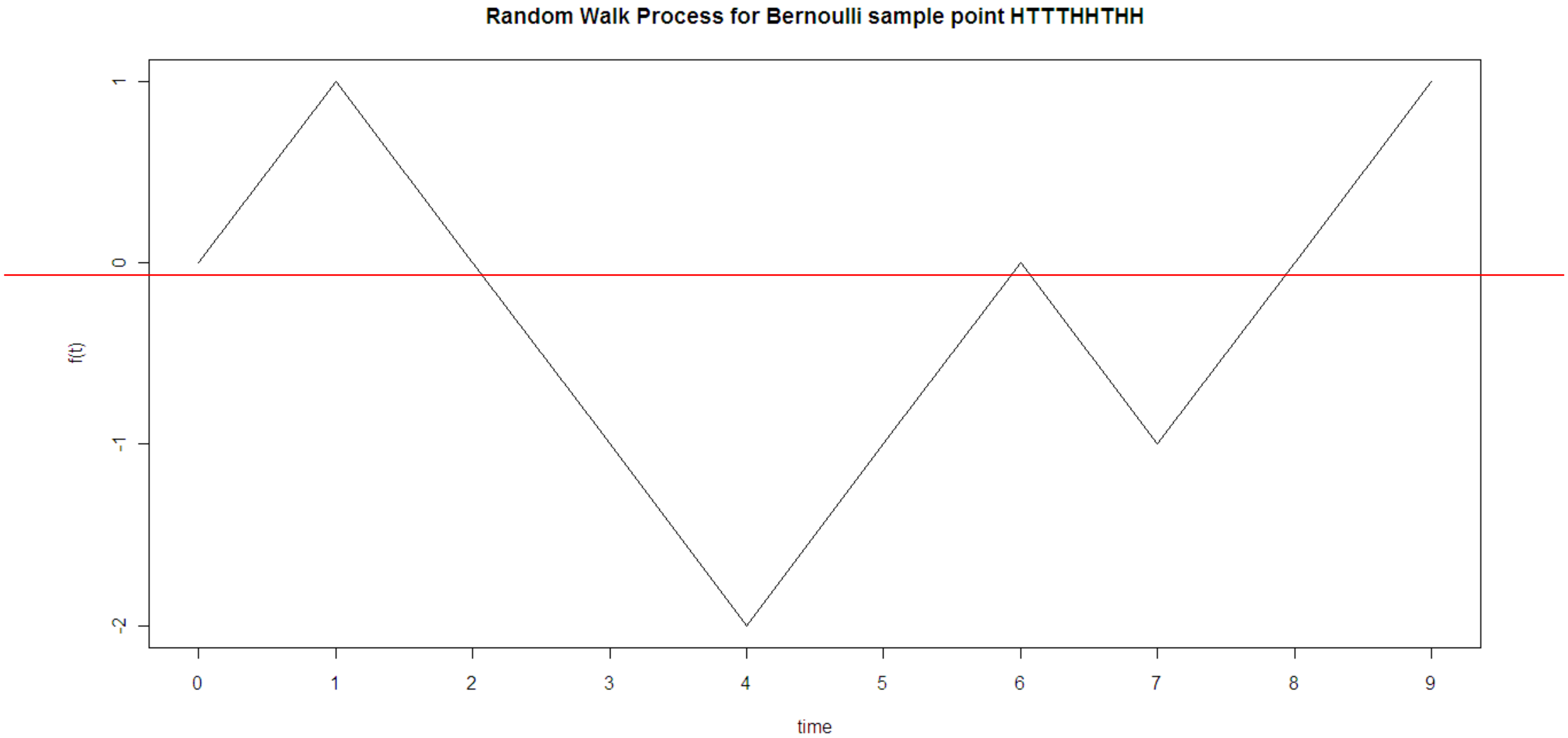
- The expectation $E(S_n)$ of S_n is 0. That is, the mean of all coin flips approaches zero as the number of flips increase. This also follows by the finite additivity property of expectations:

$$E(S_n) = \sum_{j=1}^{n} E(Z_j) = 0.$$

- A similar calculation, using independence of random variables and the fact that $E(Z_n^2) = 1,$ shows that

$$E(S_n^2) = \sum_{j=1}^{n} E(Z_j^2) = n.$$

- This hints that $E(|S_n|)$, the expected translation distance after n steps, should be of the order of $\sqrt{n}$.



- Suppose we draw a line some distance from the origin of the walk. How many times will the random walk cross the line?

- The following, perhaps surprising, theorem is the answer: for any random walk in one dimension, every point in the domain will almost surely be crossed an infinite number of times. [In two dimensions, this is equivalent to the statement that any line will be crossed an infinite number of times.] This problem has many names: the *level-crossing problem*, the *recurrence* problem or the *gambler's ruin* problem.
- The source of the last name is as follows: if you are a gambler with a finite amount of money playing *a fair game* against a bank with an infinite amount of money, you will surely lose. The amount of money you have will perform a random walk, and it will almost surely, at some time, reach 0 and the game will be over.

- At zero flips, the only possibility will be to remain at zero. At one turn, you can move either to the left or the right of zero: there is one chance of landing on -1 or one chance of landing on 1. At two turns, you examine the turns from before. If you had been at 1, you could move to 2 or back to zero. If you had been at -1, you could move to -2 or back to zero. So, f.i. there are two chances of landing on zero, and one chance of landing on 2. If you continue the analysis of probabilities, you can see *Pascal's triangle*

n	−5	−4	−3	−2	−1	0	1	2	3	4
$P[S_0 = k]$						1				
$2P[S_1 = k]$					1		1			
$2^2P[S_2 = k]$				1		2		1		
$2^3P[S_3 = k]$			1		3		3		1	
$2^4P[S_4 = k]$		1		4		6		4		1
$2^5P[S_5 = k]$	1		5		10		10		5	

$$P[S_n = k] = \frac{1}{2^n}\binom{n}{(n+k)/2}$$