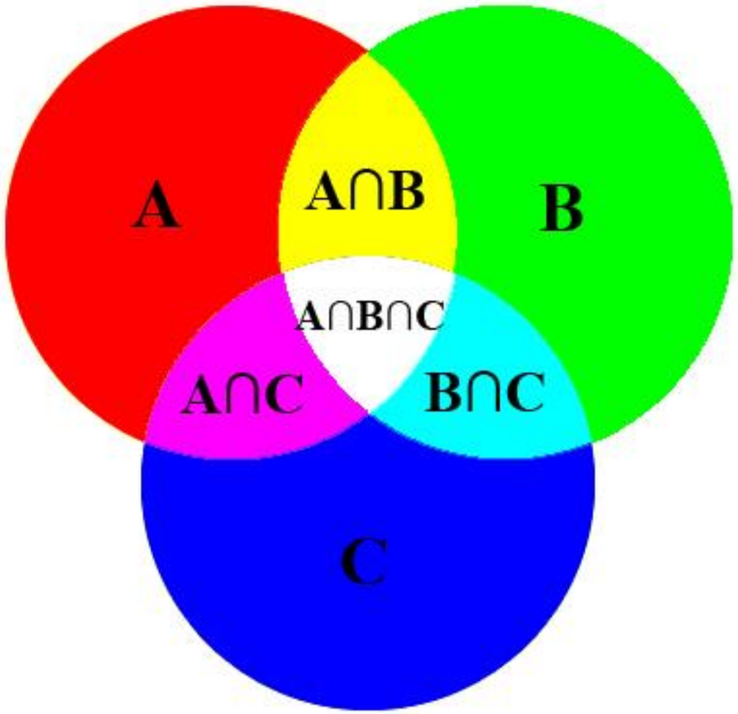
Chapter 2 Experiments, Outcomes and Events

- 2.1 Basic Definition
- 2.2 Set Theory
- 2.3 Venn Diagram
- 2.4 Summary

- May 28, 2018



2.1 Basic Definition

- In probability, any process of observation is an *experiment*.
- The results of an observation are the *outcomes* of the experiment.

Example 1

1) Roll of a die
2) Toss of a coin

are examples of an experiment.

2.1 Basic Definition

- A ***trial*** is a single occurrence of an experiment.
- If there are n trials, then we have a ***sample*** of size n consisting of n sample points.

Example 4

Where you are required to differentiate between a trial and an experiment, consider the experiment to be a larger entity formed by the combination of a number of trials.

i. In the experiment of tossing 4 coins, we may consider tossing each coin as a trial and therefore say that there are 4 trials in the experiment.

2.1 Basic Definition

- The set of all possible outcomes of an experiment is called the *sample space* S. So contains the results of all trials
- An element in S is a *sample point*.

Example 2

Find the sample space for the experiment of tossing a coin

i. Once (1 trial): $S = \{H, T\}$

ii. Twice (2 trials): $S = \{HH, HT, TH, TT\}$

2.2 Set Theory

Consider the events (subsets) $A, B, C, \cdots$ of a given sample space S.

- The ***union*** *(OR)*

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

- The ***intersection*** (AND)

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

We can generalize to

$$\bigcup_{j=1}^{m} A_j = A_1 \cup A_2 \cup \cdots \cup A_m$$

$$\bigcap_{j=1}^{m} A_j = A_1 \cap A_2 \cap \cdots \cap A_m$$

2.2 Set Theory

Example Union and intersection Roll a die

Event E = face up is even = {2, 4, 6}

Event G = face up is > than 3 = {4, 5, 6}

$E \cup G$ = {2, 4, 5, 6} (either in E or in G or in both)

note: 4 and 6 are in both events but show only once

$E \cap G$ = {4, 6} (in E and in G)

2.2 Set Theory

- If A and B *mutually exclusive* if $\boldsymbol{A \cap B = \emptyset}$

- If $\boldsymbol{A_i \cap A_j = \emptyset}$ for $i \neq j$ and $\boldsymbol{\bigcup_{i=1}^{k} A_i} = S$ then the collection $\{A_i : 1 \leq i \leq k\}$ forms a *partition* of S.

- The *complement* of A, denoted $\boldsymbol{\bar{A}}$, read "Not A", is

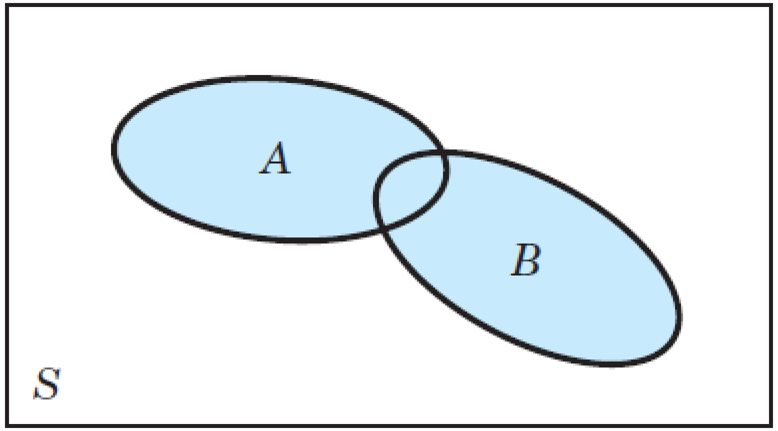
$$\bar{A} = \{x : x \in S \text{ and } x \notin A\}$$

2.3 Venn Diagram

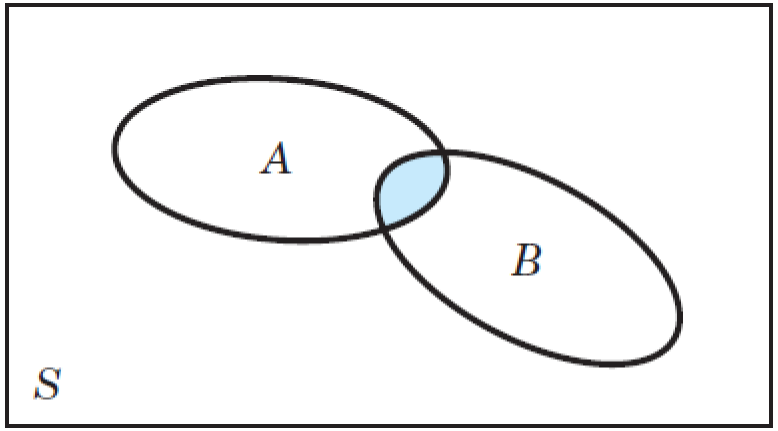
It is a graphical representation useful for illustrating the set operations.

Example 6

For events A, B such that $A \cap B \neq \emptyset$,



Union $A \cup B$



Intersection $A \cap B$

Chapter 3 Probability

- 3.1 First Definition of Probability
- 3.2 Axioms of Probability
- 3.3 Basic Theorems of Probability
- 3.4 Conditional Probability
- 3.5 Independent Events
- 3.6 Summary

28 May 2018

3.2 Axioms of Probability

Given a sample space S, with each event A of S there is associated a number $P(S)$, called the *probability* of A, satisfying the following axioms.

1. For every A in S, $0 \leq P(A) \leq 1$.
2. For sample space S, $P(S) = 1$.
3. For mutually exclusive events A and B,
$$P(A \cup B) = P(A) + P(B).$$
4. If S has infinitely many points with partition $A_1, A_2, \cdots$
   then $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$.

3.3 Basic Theorems of Probability

Important Rules

$\bar{A}$ is the complement to A and

$$P(\bar{A}) = 1 - P(A).$$

Probability of a union

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

3.4 Conditional Probability

The *conditional probability* of B given A:

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Similarly, the *conditional probability* of A given B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

3.4 Conditional Probability

1) Multiplicative Rule

$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$.

Also

$$P(A \cap B \cap C) = P(A|B, C)P(B|C)P(C)$$

3.5 Independence

Two events are Independent if

If A and B are such that $P(A \cap B) = P(A)P(B)$

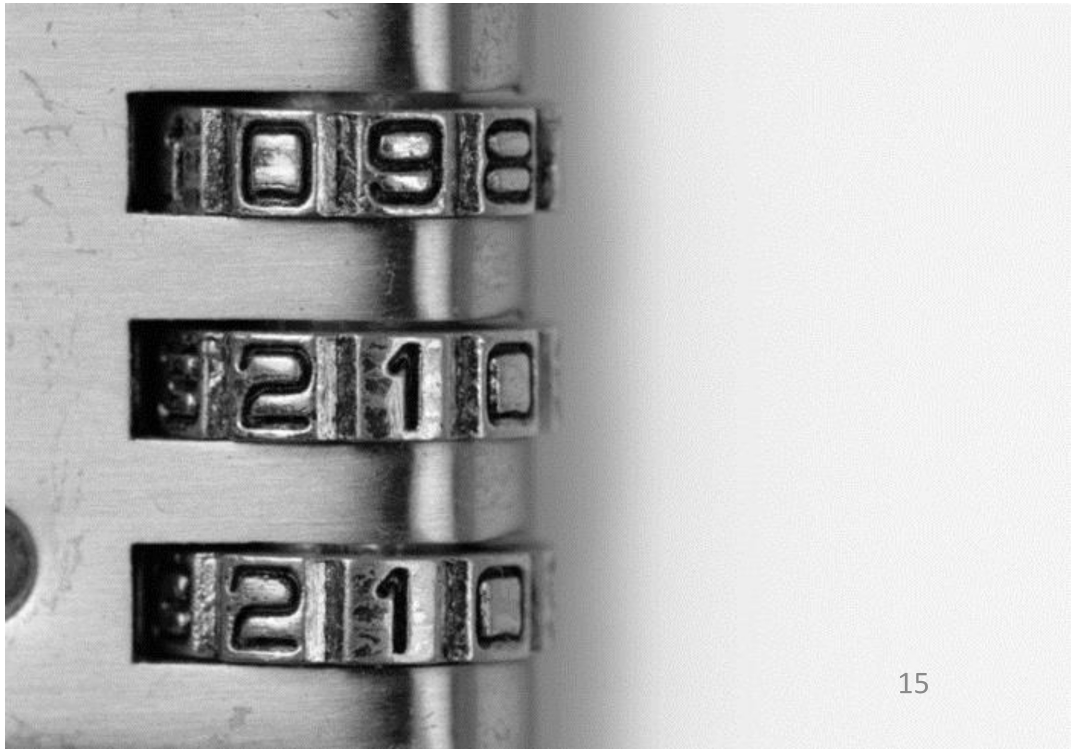
Or

$P(A|B) = P(A)$ and $P(B|A) = P(B)$

Permutations and Combinations

- 4.1 Permutations
- 4.2 Combinations

May 28, 2018



4.1 Permutations

Permutations **Order counts**.

We distinguish between choosing *k* objects from *n*

- *with repetition*: n^k $(n \; in \; n \Rightarrow n^n)$

$$\{a, b, c\} \rightarrow \{aa, ab, ac, ba, bb, bc, ca, cb, cc\}: 3^2 = 9$$

- *without repetition*: $\mathrm{P(n, r)} = \frac{n!}{n-k}$ $(n \; in \; n \Rightarrow n!)$

$$\{a, b, c\} \rightarrow \{ab \; ac, ba, bc, ca, cb, \}: \frac{n!}{(n-k)!} = \frac{3 * 2 * 1}{1} = 6$$

Binning *n* objects of *c* types : $\frac{n!}{n_1! n_2! \cdots n_c!}$

4.2 Combinations

Order does not count.

$C(n, r)$ or $\binom{n}{r}$. This is also called <u>binomial coefficient</u>

$$C(n, r) = \frac{n!}{r!\,(n-r)!}$$

Combinations are applied when

1. order is not important, and
2. repetitions are not allowed, and

$\binom{0}{0}$ is defined as 1, $\binom{n}{0} = 1$, $\binom{n}{1} = n$, $\binom{n}{k} = \binom{n}{n-k}$

Tutorial 1

- Look at page 3 and 4

Random Variables. Probability Distributions

- 5.1 Random Variables
- 5.2 Discrete Random Variables and Distributions
- 5.3 Continuous Random Variables and Distributions
- 5.4 Summary

28 May 2018

5.2 Discrete Random Variables and Distributions

The *probability mass function* (**pmf**) of X is, for $j = 1,2,\cdots,$

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

the cumulative *distribution function* (**cdf**)

$$F(x) = P\left(X \le x_j\right) = \sum_{x \le x_j} p_j$$

Note: we often refer to the *cdf* simply as *the distribution*

5.2 Discrete Random Variables and Distributions

Properties of the Probability Mass Function (pmf)
1. $f(x) \geq 0$
2. $\sum f(x) = 1$
Properties of the (Discrete) Cumulative Distribution Function
1. $\lim_{x \to -\infty} F(x) = 0$
2. $\lim_{x \to \infty} F(x) = 1$
3. $F(x) = P(X \leq x) = \sum_{x_j \leq x} p_j$
4. F is non-decreasing function
5. $0 \leq F(x) \leq 1$

5.2 Discrete Random Variables and Distributions

One useful formula for discrete distribution is

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

This is the sum of all probabilities p_j for which x_j satisfies the condition $a < x_j \leq b$.

6.1 Mean of a Discrete Random Variable

The ***mean*** (or ***expectation*** or ***expected value***) is

$$\mu = \mathrm{E}[X] = \sum_{j=1}^{n} x_j \, p(x_j)$$

The variance is

$\sigma^2 = \mathrm{V}(X) = \mathrm{E}\left[\left(X - \mathrm{E}(X)\right)^2\right] = \sum_j \left(x_j - \mathrm{E}(X)\right)^2 p(x_j) =$ $\mathrm{E}(\mathrm{X}^2) - \mathrm{E}(X)^2$.

The standard deviation is $\sigma = \sqrt{V(X)}$

7.4 Important Discrete Distributions

7.1 Binomial Distribution $P(X = k|n) = \binom{n}{k} p^k (1-p)^{n-k}$

$$E(X) = np, \qquad V(X) = np(1-p)$$

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

$$E(X) = \frac{1}{p}, \qquad V(X) = \frac{1-p}{p^2}$$

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

$$E(X) = \lambda, \qquad V(X) = \lambda$$

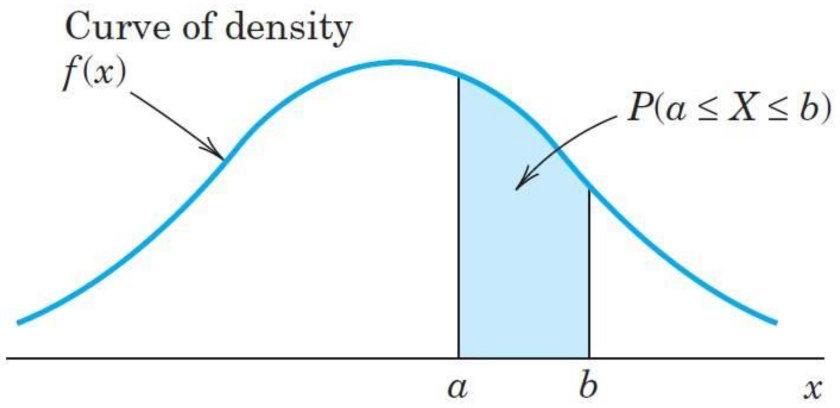
- Tutorial 2 problem 2.1
- Tutorial 3 problem 1.3
- Tutorial 4 problem 2.3

5.3 Continuous Random Variables and Distributions

We shall now consider continuous random variables which may take any value on $\mathbb{R}$.

Instead of the pmf, we now have a ***probability density function*** (pdf). This is a continuous function f such that $P(a < X \leq b)$ is equal to the area under the graph of f between $x = a$ and $x = b$.

The probability associated with any particular value $X = a$ is zero. So we need to find the probability of an interval [*X*, *X* + *dX*]



5.3 Continuous Random Variables and Distributions

Properties of the Probability Density Function (pdf)
1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
Properties of the (Continuous) Distribution Function
1. $\lim_{x \to -\infty} F(x) = 0$
2. $\lim_{x \to \infty} F(x) = 1$
3. $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(v) dv$
4. F is differentiable [under special conditions], non-decreasing function
5. $0 \leq F(x) \leq 1$
6. $f(x) = F'(x)$

5.3 Continuous Random Variables and Distributions: example

Consider a variable X that takes values in $[a, b]$ with constant density function: $f(x) = k \; if \; a < x \le b$, $\quad f(x) = 0 \;\; otherwise$.

This is called Uniform variable.

The cdf is $P(X \le x) = \int_a^x k \, du = [u]_a^x = k(x-a)$ for $a < x \le b$

Since $\mathrm{P}(\mathrm{X} \le b) = F(b) = k(b-a) = 1$, we have $k = 1/(b-a)$. So,

$$f(x) = \begin{cases} \frac{1}{b-a}; \; 0 < x \le 1 \\ 0; otherwise \end{cases} \quad \text{and } F(x) = \begin{cases} 0 & ; x \le a \\ x & ; a < x \le b \\ 1 & ; x > b \end{cases}$$

6.3 Mean and variance of a Continuous Variable

The *mean* is

$\mu = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$ where f is the pdf of X.

Remember that $\mathrm{E}(\mathrm{Y}) = \mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$

$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}\left[\left(X - \mathrm{E}(X)\right)^2\right] = \int_{-\infty}^{\infty} \left(x - \mathrm{E}(X)\right)^2 f(x)\, dx =$
$\int_{-\infty}^{\infty} x^2 f(x) dx - \mathrm{E}(X)^2$.

8.1 Uniform Distribution

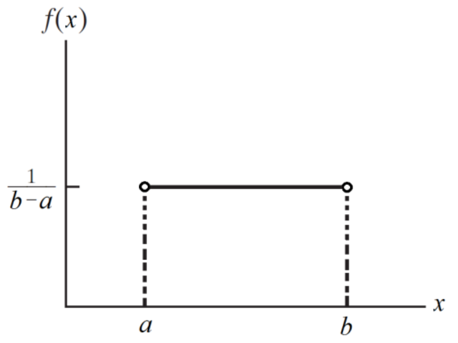
The random variable X follows a uniform distribution on the interval (a, b) if its pdf is

$$f(x) = \begin{cases} \frac{1}{b-a} & ; \quad \text{if } a < x \leq b \\ 0 & ; \quad \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0\,; if\ x \leq a \\ \frac{x-a}{b-a} \quad ; \quad \text{if } a < x \leq b \\ 1 \quad ; f \quad x > b \end{cases}$$

The mean and variance of X are $E(X) = \frac{a+b}{2}$

$$Var(X) = \frac{(b-a)^2}{12}.$$



Percentiles

- To find the value of X for which

$$P(X \leq k) = F(k) = p$$

The percentile k is found by inverting the cdf.

Example Uniform:

$$P(X \leq k) = 0.95 \Rightarrow \frac{k-a}{b-a} = 0.95.$$

Therefore,

$$k = a + 0.95(b-a)$$

6.5 Properties of Mean and Variance:

Proof that if X is a random variable then the mean and variance of $Y = aX + b$ with constants $a, b \in \mathbb{R}$ are equal to

$$\mathrm{E}(Y) = a\mathrm{E}(X) + b$$

$$\mathrm{Var}(Y) = a^2\mathrm{Var}(X)$$

5.4 Summary

- Discrete: probability mass function (pmf) $P(X = x_j)$
- Continuous: probability density function (pdf) $f(x)$
  - not a probability!
- Cumulative distribution functions (cdf) $F(X) = P(X \leq x)$
  - discrete $F(x_j) = \sum_{i \leq j} P(X = x_i)$
  - continuous $F(x) = \int_{-\infty}^{x} f(u)du$
- $P(a < X \leq b) = F(b) - F(a)$

6.6 Summary

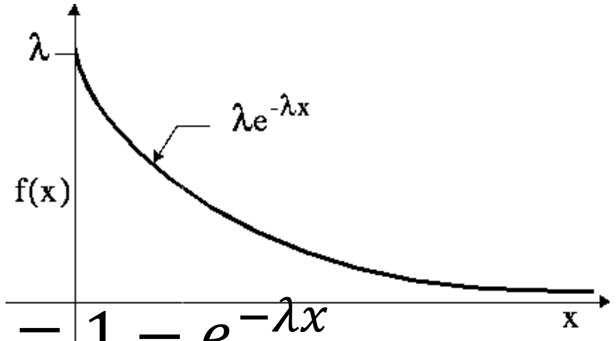
- Mean of a random variable
  - $\mu = \mathrm{E}[X] = \sum_{j=1}^{n} x_j\, p(x_j)$ if discrete
  - $\mu = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$ if continuous
- Variance of a random variable
  - $\sigma^2 = \sum_j \left(x_j - \mathrm{E}(X)\right)^2 p(x_j) = \mathrm{E}(\mathrm{X}^2) - \mathrm{E}(X)^2$
  - $\sigma^2 = \displaystyle\int_{-\infty}^{\infty} \left(x - \mathrm{E}(X)\right)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mathrm{E}(X)^2$

- $\mathrm{E}(ax + b) = a\mathrm{E}(X) + b;\quad \mathrm{Var}(ax + b) = a^2\mathrm{Var}(X)$

8.2 Exponential Distribution

The random variable $X \sim \mathrm{Exp}(\lambda)$ follows an exponential distribution if its pdf is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; \quad x \geq 0 \\ 0 \quad ; & \text{otherwise} \end{cases}.$$



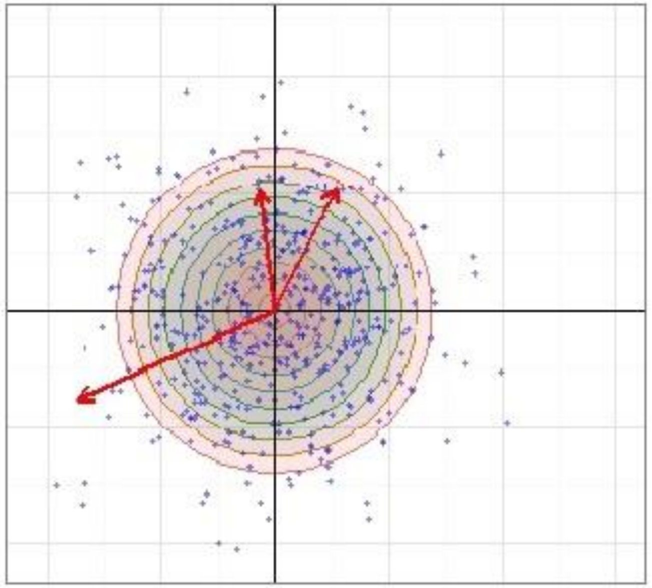
$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}$$

$E(X) = \frac{1}{\lambda}$ and $\mathrm{Var(X)} = \frac{1}{\lambda^2}$.

Memoryless: $P(X > x + t | X > t) = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x} = P(X > x)$

Rayleigh distribution

- It’s the distribution of the square root of the sum of two squared normal variables
- $X \sim N(0, \sigma^2)$ and $\mathrm{Y} \sim N(0, \sigma^2)$ and independent then
- $r = \sqrt{X^2 + Y^2} \sim Rai(\sigma^2)$
- $f(x) = \frac{x}{\sigma^2} \exp\left(\frac{x^2}{2\sigma^2}\right)$
- $F(x) = 1 \ - \exp\left(\frac{x^2}{2\sigma^2}\right)$
- $E(X) = \sigma\sqrt{\frac{\pi}{2}}, V(X) = \sigma^2\,\frac{4-\pi}{2}$



8.3 Normal Distribution

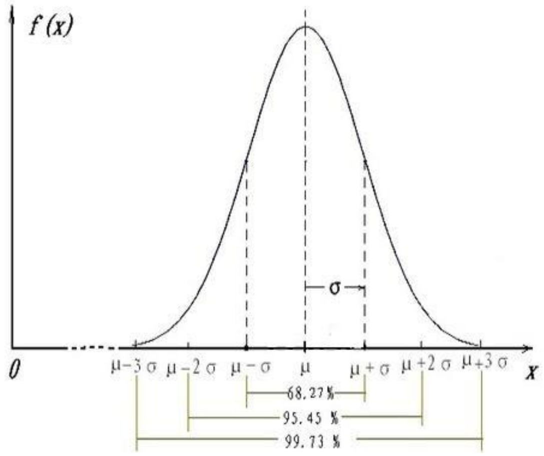
The random variable X follows a normal distribution if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad ; \quad -\infty < x < \infty.$$

We write $X \sim \mathrm{N}(\mu, \sigma^2)$ for a random variable that has normal distribution. The mean and variance of X are respectively

$\mathrm{E(X)} = \mu$ and $\mathrm{Var(X)} = \sigma^2$.

The pdf is a bell-shaped curve symmetric about μ.



8.3 Normal Distribution

If $X \sim \mathrm{N}(\mu, \sigma^2)$,

then $\mathrm{z} = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1^2)$ is a standard normal distribution.

If $\mathrm{z} \sim \mathrm{N}(0,1^2)$ then $\mathrm{X} = \mu + z\sigma \sim \mathrm{N}(\mu, \sigma^2)$

So we find $P(X \leq x) = P\left(z \leq \frac{x-\mu}{\sigma}\right)$ using the tables

8.4 Summary

- Uniform $P(x \leq c) = \frac{c-a}{b-a}, a < x \leq b$
- Exponential $P(X \leq x) = 1 - \ e^{-\lambda x}, x > 0$
- Raiyleigh $P(X \leq x) = 1 - e^{-\frac{x^2}{2\sigma^2}}, x > 0$
- Normal $X \sim N(\mu, \sigma^2) \rightarrow Z = \frac{x-\mu}{\sigma} \sim N(0,1)$
  - $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

- Tutorial 5 Q1.3

Chapter 9 Distributions of Several Random Variables

9.1 Discrete 2D Distributions

9.2 Continuous 2D Distributions

9.3.1 Sum of Means

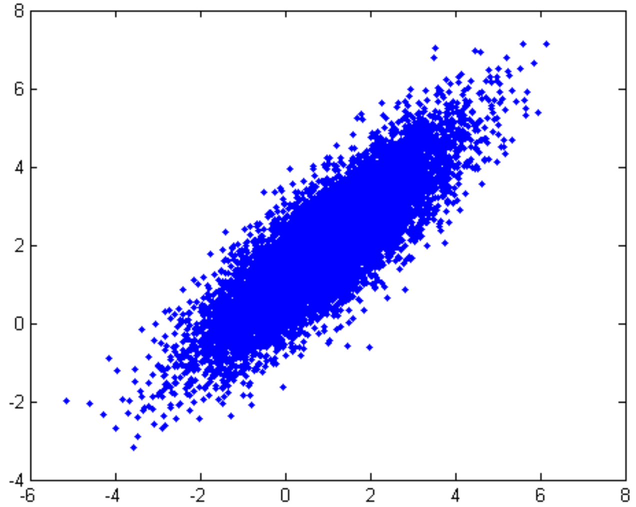
9.3.2 Multiplication of Means

9.4 Independence and Uncorrelatedness

9.5 Covariance

9.6 Correlation

9.7 Summary



28 May 2018

9.1 Discrete 2D Distributions

We have the following results:

1. Probability mass function
$$f(x,y) = P(X = x, Y = y)$$
2. Cumulative distribution function
$$F(x,y) = P(X \le x, Y \le y) = \sum_{x_i \le x} \sum_{y_j \le y} f\left(x_i, y_j\right)$$
3. $\sum_i \sum_j f\left(x_i, y_j\right) = 1$

4. Marginal distribution of Y
$$f(x) = P(X = x) = \sum_{all\, y_j} f\left(x, y_j\right)$$
$$f(y) = P(Y = y) = \sum_{all\, x_i} f(x_i, y)$$

1.9.1 Discrete 2D Distributions, example (2)

Example 1

$f_{Y_1,Y_2}(y_1, y_2)$ $f_{Y_2}(y_2)$ $f_{Y_1}(y_1)$

		Y_1			
		0	1	2	Tot
Y_2	0	0	0.1	0.2	0.3
	1	0.1	0.2	0	0.3
	2	0.4	0	0	0.4
	Tot	0.5	0.3	0.2	1

i. Find the probability $\mathrm{F}(1,1) = P(Y_1 \leq 1, Y_2 \leq 1)$;

ii. Find the marginal probability $f_{Y_1}(y_1) = P(Y_1 = y_1)$;

iii. Find the marginal cdf of Y_1, $F_{Y_1}(y_1) = P(Y_1 \leq y_1)$.

9.1 Discrete 2D Distributions

5. If X and Y are independent, then

   $f(x_i|y_j) = f(x_i)$ or $f(x_i, y_j) = f(x_i)f(y_j)$ for **all** x_i's and y_j's.

   [Strategy:

   X and Y are independent if each entry in the joint distribution table is the product of the marginal entries.]

6. The expected value

   $\mathrm{E}(g(X,Y)) = \sum_i \sum_j g(x_i, y_j)f(x_i, y_j)$ where g is some function of X and Y.

1.9.2 Continuous 2D Distributions

4. Marginal densities are

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$f(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

5. The marginal cdfs

$$F_X(x) = \int_{-\infty}^{x} f_X(u) du = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, v) dv du = \lim_{y \to \infty} F_{XY}(x, y)$$

$$F_Y(y) = \int_{-\infty}^{y} f_Y(v) dv = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(u, v) du dv = \lim_{x \to \infty} F_{XY}(x, y)$$

1.9.2 Continuous distributions, example 2 (1)

Consider two variables, X, Y, with joint pdf

$$f_{XY}(x,y) = \begin{cases} 4xy, if\ 0 < x \leq 1, 0 < y \leq 1 \\ 0, otherwise \end{cases}$$

Find the joint cdf, marginal pdfs and cdfs. Also $P(X \leq 0.3 | Y \leq 0.5)$,

i) The joint cdf is simply

$$F_{XY}(x,y) = \int_0^x \int_0^y 4uv\ dvdu = x^2y^2, 0 < x \leq 1, 0 < y \leq 1$$

$F_{XY}(x,y) = 0, x, y \leq 0$ and $F_{XY}(x,y) = 1, x, y > 1.$

1.9.2 Continuous distributions, example 2 (2)

$$f_{XY}(x,y) = \begin{cases} 4xy, if\ 0 < x \leq 1, 0 < y \leq 1 \\ 0, otherwise \end{cases}$$

ii) The marginal pdf's are

$$f_X(x) = \int_0^1 4xv\,dv = 4x\left[\frac{y^2}{2}\right]_0^1 = 2x, 0 < x \leq 1$$

$$f_Y(y) = \int_0^1 4uy\,du = 4y\left[\frac{x^2}{2}\right]_0^1 = 2y, 0 < y \leq 1$$

Remember for discrete variables we took the sums

1.9.2 Continuous distributions, example 2 (3)

$$f_{XY}(x,y) = \begin{cases} 4xy, if\ 0 < x \leq 1, 0 < y \leq 1 \\ 0, otherwise \end{cases}$$

iii. The marginal cdf's can be simply computed from the joint cdf:

$$F_X(x) = F_{XY}(x,1) = [x^2y^2]_{y=1} = x^2, 0 < x \leq 1;$$

$$F_Y(y) = F_{XY}(1,y) = [x^2y^2]_{x=1} = y^2, 0 < y \leq 1$$

In fact, $F_X(x) = \int_0^x \int_0^1 4uv\, dv\, du = \int_0^x f_X(u) du = [u^2]_0^x = x^2$;

Naturally we can show the same for Y.

1.9.2 Continuous distributions, example 2 (3.1)

Recall that, $F_{XY}(x, y) = P(X \leq x, Y \leq y)$. So the marginal cdfs are given by

$$\mathrm{F_X(x)} = P(X \leq x) = \lim_{y \to \infty} P(X \leq x, Y \leq y)$$
$$= P(X \leq x, Y \leq 1) = F_{XY}(x, 1)$$

because Y must be ≤ 1. In general substitute the maximum value of the other variable or take the limit to ∞.

$$\mathrm{F}_Y(\mathrm{y}) = P(Y \leq y) = F_{XY}(1, y)$$

1.9.2 Continuous 2D Distributions, example 3

To solve this we need to use integration over a general region

Example 3

For joint probability density function

$$f(x,y) = 2; \quad 0 < x \le y, 0 < y \le 1$$

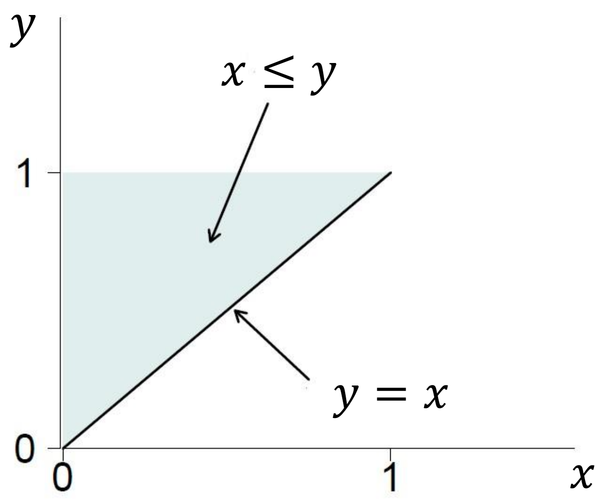
find its distribution function $F(x,y)$.

Solution

We can write the domain R as

$0 < x \le 1, x < y \le 1$

This is more useful because we integrate from below



1.9.3 Covariance and correlation

The covariance between two variables is a measure of their association. It is defined as

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y).$$

the correlation of X and Y is given by

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

$-\infty < Cov(X,Y) < \infty$, $-1 \le \rho \le 1$.

1.9.4 Mean of a sum of variables

Recall that

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

and

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

These results are always true (for continuous/discrete, and Independent/dependent random variables).

For the variance it is different.

1.9.4 Variance of a sum of variables

$$Var(aX + bY + c) \stackrel{\text{def}}{=}$$

$$a^2 Var(X) + 2abCov(X, Y) + b^2 Var(Y) =$$

$$\mathrm{a}^2 V(X) + 2ab\rho\sigma_{\mathrm{X}}\sigma_Y + b^2 V(Y)$$

1.9.5 Summary

- Know how to obtain marginal distributions from joint pdf or joint pmf and pdf
- Know how to find covariance $\mathrm{Cov}(X, Y)$ and correlation $\rho(X, Y)$
- Understand the concept of correlation for discrete/continuous 2D distributions
- Know how to compute the mean and the variance of a sum of variables

9.4 Independence and Uncorrelatedness

For <u>two</u> random variables X and Y

1) They are **independent** if $f(xy) = f(x)f(y)$,
2) therefore $E(XY) = E(X)E(Y)$.
3) And $Cov(X, Y) = E(XY) - E(X)E(Y) = 0$
   so $\rho = 0$

4) Independence⇒Uncorrelated but

but uncorrelatedness does not imply independence

9.3.2 Product of Means

Theorem 2

The mean of the product of **independent** random variables equals the product of means, i.e.

$E(X_1X_2\cdots X_n) = E(X_1)E(X_2)\cdots E(X_n)$.

The result is true for continuous/discrete random variables. ■

11.3 Bivariate Normal Random Variables

Theorem 1

Two random variables X and Y are said to have a **bivariate normal distribution** with parameters μ_X, σ_X^2, μ_Y, σ_Y^2 and ρ, if their joint pdf is given by

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

where ρ is the correlation between X and Y. ■

11.3 Bivariate Normal Random Variables

Theorem 2

Let X and Y be bivariate normal random variables and uncorrelated, $(\rho = 0)$. Then they are independent.

$$f(x,y)$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]} =$$

$$\left[\frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]}\right]\left[\frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left[\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}\right] = f(x)f(y) \blacksquare$$

11.3 Bivariate Normal Random Variables

Theorem 3

Let X and Y be bivariate normal random variables then X and Y are normal. For U, V standard Normal

$$E(v|u) = \rho u, \qquad V(v|u) = (1-\rho^2)$$

$$(X,Y) \sim MN\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2\end{pmatrix}\right)$$

$E(X) = \mu_X$ and $V(X) = \sigma_X^2$ $\qquad$ $E(Y) = \mu_Y$ and $V(Y) = \sigma_Y^2$

$$E(Y|X) = \mu_Y + \rho\,\frac{\sigma_Y}{\sigma_X}(X - \mu_X) \qquad Var(Y|X) = \sigma_Y^2(1-\rho^2)$$

11.3 Bivariate Normal Random Variables

Theorem 4 important

Two random variables X and Y are said to be *bivariate (jointly) normal*, if
$\mathrm{Z} = aX + bY$ has a normal distribution for all $a, b \in \mathbb{R}$, with
$E(Z) = a\mu_X + b\mu_Y$ and
$var(Z) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\ cov(XY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\ \rho\sigma_X\sigma_Y$

Note

1) If X and Y are independent normal random variables, then they are bivariate normal.

2.3.1 Inequalities

If X is a nonnegative random variable, $X > 0$, then, for any value $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a} \quad MARKOV's$$

For any X

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2} \quad CHEBYCHEV's$$

2.3.2 Convergence

Convergence *almost sure* (*a.s.*)

$$P(\lim_{n\to\infty} X_n = X) = 1 \Leftrightarrow X_n \xrightarrow{a.s.} X$$

Convergence in *probability* (*p*)

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1, \forall\epsilon > 0 \Leftrightarrow X_n \xrightarrow{p} X$$

Convergence in *mean square*

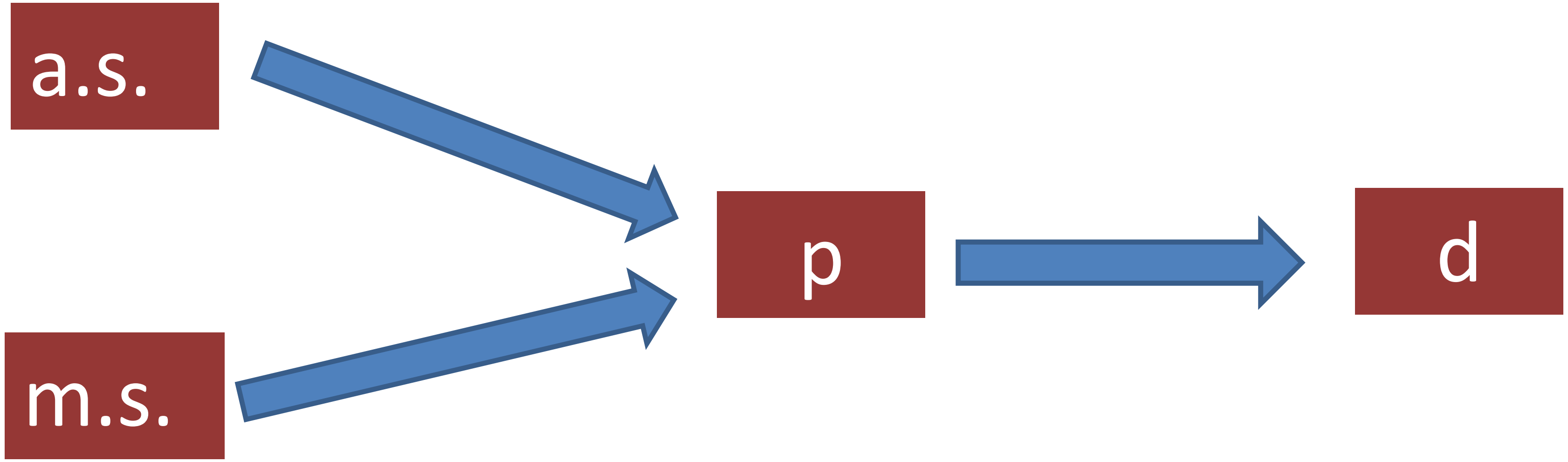
$$\lim_{n\to\infty} E\ [(X_n - X)^2]\ = 0 \Leftrightarrow X_n \xrightarrow{m.s.} X$$

- Convergence *in distribution* (*d*).

$$\lim_{n\to\infty} F(X_n) = F(X) \Leftrightarrow X_n \xrightarrow{d} X$$

2.3.2 Relationship between convergences

- Convergence *almost sure* implies convergence *in p.*
- Convergence *in mean square* implies convergence *in p.*
- Convergence *in probability* implies convergence *in distribution.*



Law of large Numbers and central convergence theorem

- Weak law of large numbers $\lim_{\mathrm{n}\to\infty} P\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n} - \mu\right| \geq \epsilon\right\} = 0;$
- Strong law of large numbers $\bar{X}_n = \frac{X_1+X_2+\cdots+X_n}{n} \to \mu$ as $n \to \infty$
  - $\lim_{n\to\infty} P(\bar{X}_n = \mu) = 1$ **weak**, $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$ **strong**
- Central convergence theorem $\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \underset{n\to\infty}{\sim} N(0,1)$

$$\bar{X}_n \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ for } n \to \infty$$

12.1 Stochastic processes: definition

- A random process is a collection of random variables defined on a set of indices T as
$$\{X(t), t \in T\}$$
- $X(t)$ and T can be either discrete or continuous;
- The process is defined by the collection of joint cumulative distributions
$$F_{X(1),X(2),\dots,X(k)}(x_1, x_2, \dots, x_k) \\ = P(X(1) \leq x_1, X(2) \leq x_2, \dots, X(k) \leq x_k)$$

12.1 Stochastic processes: mean and variance

- The mean, $\mu_X(t)$, of a random process is defined as

$$\mu_X(t) = E[X(t)] = \int_{-\infty}^{\infty} X(t) f_{X(t)}(x) dx.$$

In general, $\mu_X(t)$ is a function of time;

- Analogously, the variance is defined as

$$\sigma_X(t) = \int_{-\infty}^{\infty} [X(t) - \mu_X(t)]^2 f_{X(t)}(x) dx.$$

12.1 Stochastic processes: autocovariance and autocorrelation

- The autocovariance (function), $C_{XX}(t,s)$, is the covariance of the variables in two different times, (t,s)

$$C_{XX}(t,s) = E[(X(t) - \mu_X(t))(X(s) - \mu_X(s)]$$

- The autocorrelation (function), $R_{XX}(t,s)$, is defined as

$$R_{XX}(t,s) = E[(X(t)X(s)] = C_{XX}(t,s) + \mu_X(t)\mu_X(s)$$

12.1.1 Stochastic processes: strong stationarity

- A process is strongly (or strictly) stationary if its distribution does not change over time. That is, if

$$F_{X(t_{1+\tau}),X(t_{2+\tau}),\ldots,X(t_{k+\tau})}(x_1, x_2, \ldots, x_k) \\ = F_{X(t_1),X(t_2),\ldots,X(t_k)}(x_1, x_2, \ldots, x_k)$$

This means that the distribution is time invariant.

12.1.1 Stochastic processes: weak stationarity

- A process is weakly stationary if

$$\mu_X(t+\tau) = \mu_X(t) = \mu_X$$

And

$$C_{XX}(t, t+\tau) = C_{XX}(s, s+\tau) = C_{XX}(\tau)$$

Mean is constant and the covariance is only a functions of τ, or $C(t_1, t_2)$ is a function of only $(t_2 - t_1)$. Also $R_{XX}(t, t+\tau) = R_{XX}(s, s+\tau)$

12.1.3 Counting processes

A stochastic process $\{X(t), t \geq 0\}$ is said to be a counting process if X(t)represents the total number of “events” that occur by time t.

Some examples of counting processes are the following:

i. Number of people entering a store;

ii. Number of earthquakes;

iii. Numbers of calls to an emergency centre;

iv. Number of goals in a soccer game;

12.1 Counting processes

If $s < t$, the number of events that occur in $[s, t]$ is called increment and is equal to $X(t) - X(s)$;

Definition: <u>Increments are independent</u> if the number of events that occur in an interval is independent of the number of events occurred in a non-overlapping interval. If $s_1 < t_1 < s_2 < t_2$, $\left(X(t_1) - X(s_1)\right)$ is independent of $\left(X(t_2) - X(s_2)\right)$

Example, number of goals scored in the two halves of a soccer game

12.2 Poisson Process

A continuous-time $\{X(t): t \geq 0\}$ is a Poisson process with rate $\lambda > 0$ if

i. $X(0) = 0$

ii. It has stationary and independent increments.

iii. The number of events occurring in an interval, $X(t)$, is Poisson with mean λt, that is

$$P(X(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad ; \quad k = 0,1,2, \ldots$$

12.3 Waiting time for the first event

One important question regarding a Poisson process is: how long will it take for the first event to happen?

Let T_1 be the time at which the first event happens, then

$$P(T_1 \leq t) = P(X(t) > 0) = 1 - P(X(t) = 0) = 1 - e^{-\lambda t}$$

If you remember, this is the Exponential distribution with rate λ and mean $1/\lambda$ and pdf

$$f(t) = \lambda e^{-\lambda t}, t \geq 0; 0 \; if \; t < 0$$

12.3 Waiting time for the first event

The exponential distribution is memoryless

Let Y$\sim Exp(\lambda\}$

$$P(T_1 > t + s | T_1 > t) = \frac{P(T_1 > t + s)}{P(T_1 > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s}$$

Therefore, $P(T_1 \leq t + s | T_1 > t) = P(T_1 \leq s) = 1 - e^{-\lambda s}$

which does not depend on t.

12.4 Distribution of arrival times

Time of arrival is uniformly distributed in $[0, t]$ so that

$P(T_1 \leq s | X(t) = 1) = \frac{s}{t}$.

This is easily checked since, for $s \leq t$,

$$P(T_1 \leq s | X(t) = 1) = \frac{P(X(s) = 1)P(X(t-s) = 0)}{P(X(t) = 1)} = \frac{e^{-\lambda s}\lambda s e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}}$$

$$= \frac{s}{t}$$

12.3 Summary

- Increments are stationary and independent

$$\mathrm{P}(X(t+s)-X(s)=k)=P(X(t)=k)$$

- For a Poisson Process

$$P(X(t)=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t} \quad ; \quad k=0,1,2,\ldots$$

- The waiting time for the 1^{st} event is exponential with mean $\frac{1}{\lambda}$

$$P(T_1 \le t) = \ 1-e^{-\lambda t}$$

- The conditional arrival time of T_1 is uniform

$$P(T_1 \le s | N(t)=1)=\frac{s}{t}$$