Review

1 Linear Algebra Review

- Review definitions of eigenpairs
- Find eigenvalues.
- Conditions to diagonalize a matrix.

1.1 The matrix and its eigenpairs

Any $n \times m$ matrix A, written $A \in \mathbb{R}^{n \times m}$, is associated to a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. EG, the linear transformation T mapping $\mathbb{R}^2$ to $\mathbb{R}^3$ so that T maps $(0,1)^T$ to $(0,1,2)^T$ and $(1,0)^T$ to $(3,4,5)^T$ has associated matrix $A \in \mathbb{R}^{3\times 2}$,

$$A = \begin{pmatrix} 3 & 0 \\ 4 & 1 \\ 5 & 2 \end{pmatrix}$$

in this case we say the $2,1$ entry of A is $A_{2,1} = 4$

Let us write I_n to be the $n \times n$ identity matrix, that is $(I_n)_{i,j} = 0$ if $i \neq j$ and $(I_n)_{ij} = 1$ for $i = j$ for all $1 \leq i, j \leq n$. For $A \in \mathbb{R}^{n\times m}$, $I_n A = A$. The inverse matrix of A is the matrix (if it exists) A^{-1} so that $AA^{-1} = I_n$.

Given a matrices $A \in \mathbb{R}^{n\times n}$ we may ask if it is invertible, which is equivalent to the associated linear transformation being invertible. A is invertible if the equation $Ax = 0$ has only the trivial solution $x = 0$, if $Ax = 0$ has more than one solution we say it is singular.

The eigenpairs of $A \in \mathbb{R}^{n\times n}$ are pairs $(v, \lambda) \in \mathbb{R}^n \times \mathbb{R}$ so that $Av = \lambda v$. Here λ is an eigenvalue and v is an eigenvector. The matrix A is invertible if λ is not an eigenvalue for any v.

Given an eigenvalue λ, the geometric multiplicity is the number of linearly independent eigenvectors v that may be associated to λ.

The eigenvalues are exactly the solutions to the characteristic polynomial of the matrix. The characteristic polynomial of A is

$$p(t) = \det(tI_n - A)$$

The algebraic multiplicity for an eigenvalue λ is the largest q so that $(t-\lambda)^q$ divides $p(t)$. The geometric multiplicity of an eigenvalue is at least as great as its algebraic multiplicity.

1.2 Diagonalization

A diagonal matrix is a matrix $D \in \mathbb{R}^{n\times n}$ so that $D_{ij} = 0$ for $i \neq j$. Let us write $D = diag(d_1, ..., d_n)$ for the diagonal matrix D with $D_{ii} = d_i$.

We say the matrix A is diagonalizable if there exists a diagonal matrix similar to A. That is there is a matrix M so that

$$M^{-1}AM = D.$$

The matrix A is diagonalizable if and only if the algebraic multiplicity of each eigenvalue is equal to it's geometric multiplicity.

Let us give sufficient conditions for diagonalizability. The matrix $A \in \mathbb{R}^n$ is diagonalizable if there are n distinct eigenvalues - equivalently the algebraic multiplicity of each eigenvalue is 1.

1.3 Symmetric matrices

The transpose of a matrix A is the matrix A^T defined so that $(A^T)_{ij} = A_{ji}$. A real matrix is self symmetric if $A^T = A$. Real symmetric matrices are diagonalizable and their eigenvalues are real.

The diagonalization of a real symmetric matrix A can be written as $A = O^T DO$, where $D = diag(\lambda_1, .., \lambda_n)$ and O is orthogonal so that $O^T O = I_n$. Recall orthogonal matrices are matrices O so that the column (or row) vectors of O form an orthonormal basis of $\mathbb{R}^n$.

If v_1 and v_2 are eigenvectors of a real symmetric matrix with eigenvalues $\lambda_1 \neq \lambda_2$ then $(v_1, v_2) = 0$, that is, v_1 and v_2 are perpendicular.

positive / positive definite If all eigenvectors λ_i are nonnegative $\lambda_i \geq 0$ then for any $v \in \mathbb{R}^n$, $v^t Av \geq 0$. In this case we say that A is positive semidefinite.

If all eigenvectors λ_i are positive $\lambda_i > 0$ then for any $v \in \mathbb{R}^n$, $v^t Av > 0$. In this case we say that A is positive definite.

We see below that the covariance matrix of an n-tuple of random variables is a positive definite matrix.

2 Probability Review

- Definition - Probability space + Random variables + Examples
- PDF and CDF
- Expectation
- Joint distribution
- Marginal distribution
- Independent Random Variables
- Moments, Variance, Covariance
- Conditional Random Variables / Distribution
- Large numbers + CLT

2.1 Definitions

Let us carry out an experiment. On day zero, say the price of a stock is \$100, (set $S_0 = 100$) let us suppose that every day the price of the stock will change by \$1. On the first day, flip a coin, if it is heads the stock rises it's price by \$1, ($S_1 = 101$) if it is tails the stock lowers it's price by \$1 ($S_1 = 99$). Of course in 'real life' the stock doesn't change value according to a coin flip, but to volume of trading and updated market for forecasts etc but we can simplify to a coin flip. Repeat the experiment up to day 10.

We can expect the experiment to generate a 'word' $w_1, .., w_{10}$ where $w_i = H$ if the coin flip comes up 'heads' on the i^{th} day and $w_i = T$ if the coin flip comes up 'tails' on the i^{th} day. We have a sequence of stock prices $S_0, .., S_{10}$ on each day and price changes $X_i = S_i - S_{i-1}$; equivalently, $S_i = X_i + \cdots + X_1 + S_0$

Notice the different variables depend on different sets of information. X_i depends only on the outcome of the i^{th} flip w_i; in other words X_i is a function of w_i, written $X_i = X_i(w_i)$. S_i depends on the first i flips, thus $S_i = S_i(w_1, .., w_i)$.

We formalize this setup in the definition.

2.1.1 Probability space

The definition of probability space requires several parts

(1.) **Sample space** Ω containing all outcomes of the experiment.

In the above example, Ω is the set of all outcomes of experiments,

$$\Omega = \{(w_1, .., w_{10}) : w_i = T, H\}$$

- A **State space** is an image of a function on Ω.

  In the above example: We may use a common state space $V = \{-1, 1\}$ for all the X_i so that X_i is a function on Ω taking values in V. We have a state space $W = \mathbb{N}$, for all the S_i so that S_i is a function on Ω taking values in W.

(2.) A **σ-algebra** is a collection $\mathcal{F}$ of the subsets of Ω, which obeys the following,

- If $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$
- If $A \in \mathcal{F}$ then $\Omega \backslash A \in \mathcal{F}$.
- $\mathcal{F}$ must contain both Ω and $\emptyset$.

If we only were interested in a single (discrete) random variable, it would be sufficient for us to simply take $\mathcal{F}$ to be the set of all subsets of Ω. There is a practical reason to do more with σ-algebras which we will discuss shortly.

(3.) A **Probability measure** is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ so that

- For all $A \in \mathcal{F}$, $0 \le \mathbb{P}(A) \le 1$.
- For $(E_i)_{i=1}^{\infty}$ a subcollection of disjoint sets of $\mathcal{F}$ we have $\mathbb{P}(\cup_i E_i) = \sum_i \mathbb{P}(E_i)$.
- $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(S) = 1$.

The σ algebra $\mathcal{F}$ may simply be understood as the desired refinement of information. This point of view is helpful when we introduce sequences (a filtration) of σ algebras which model the increasing information over time.

In simple cases it is often sufficient to identify the sample space and the state space, for instance for a random variable X so that $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$. Then we may define the state and sample space to be $S = \{0, 1\}$, then $X(0) = 0$ and $X(1) = 1$ - so that we have $\mathbb{P}(\{i\}) = \mathbb{P}(X = i)$. As we have seen it may be useful to make a distinction between the sample space and the state spaces in the case that several random variables depend on the same outcomes.

2.1.2 Random Variables

As discussed above, a random variable X is a function on the sample space Ω, which takes values in the state space S, for example in $S = \mathbb{R}^n$ for $n \geq 1$. Let us write $X : \Omega \to S$ in this case, which means X is a function with Domain Ω and Range S. That is, we need a probability space defined by a triple $(\Omega, \mathcal{F}, \mathbb{P})$ and a function $X : \Omega \to S$.

Let us recall standard examples of random variables and construct the associated probability space.

Example: (Bernoulli Trial)

■

Recall a Bernoulli trial is a model of a (weighted) coin flip, which returns a one if the flip turns up head and returns a zero if the flip turns up tails. Lets go through the full construction. Lets call $\Omega = \{H, T\}$, and $\mathcal{F} = \{\emptyset, \{T\}, \{H\}, \{H, T\}\}$, which in this case is the full power set of S. Define $\mathbb{P}(T) = q$ and $\mathbb{P}(H) = p$.

From 2.1.1 rule (3), we have $1 \geq p, q \geq 0$ and $p + q = 1$.

Let $X : S \to \{0, 1\}$, defined by $X(T) = 0$ and $X(H) = 1$. Thus the probability X is 1 is given by

$$\mathbb{P}(X = 1) = P(X^{-1}(1)) = P(\{H\}) = p$$

We use the notation $X \sim B(p)$ to say X is a Bernoulli variable with probability p of returning 1. ■

Example: (Binomial)

■

A Binomial random variable is a model of flipping n identical independent weighted coins and counting the total number of heads.

The set of outcomes of n coin flips is the set

$$\Omega = \{\omega = (\omega_1, .., \omega_n) : \omega_i = H, T \text{ for } i = 1, .., n\}$$

The σ algebra $\mathcal{F}$ can be taken to be the set of all subsets of Ω, ie the power set of Ω.

We will assume each coin flip is independent of the others and each flip is head with probability p. Then for a word $(w_1, .., w_n)$ with $w_i = H, T$ we have

$$\mathbb{P}(w_1, ..., w_n) = \mathbb{P}(w_1) \cdots \mathbb{P}(w_n)$$

where $\mathbb{P}(w_i) = p$ for $w_i = H$ and $\mathbb{P}(w_i) = 1 - p$ for $w_i = T$.

Similar to the example in the definition let us define random variables $X_i : \Omega \to \{0, 1\}$ by

$$X_i(\omega) = \begin{cases} 1 & \text{for } \omega_i = H \\ 0 & \text{for } \omega_i = T \end{cases}.$$

Let $\mathbb{P}(X_i = 1) = p$, for $0 < p < 1$.

Then define Y by

$$Y = X_1 + \cdots + X_n$$

which counts the number of total heads. The probability distribution of Y is given by

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

■

Example: (Geometric distribution)

■

The geometric distribution is an experiment in flipping a series of coins. The value of the random variable Y is flip number of the first head.

The set of outcomes is the set,

$$\Omega = \{\omega = (\omega_1, \omega_2, ..) : \omega_i = H, T \text{ for } i = 1, 2, ...\}.$$

Notice it is possible that the experiment does not end in a fixed amount of time, that is it might not end by flip 100. Let us construct σ - algebras that model the amount of information we have after a given amount of time. Let $\mathcal{F}_i$ be the σ algebra resolving the information up to the i^{th} flip. Notice $\mathcal{F}_i \subset \mathcal{F}_{i+1}$, that is, the information we get after every flip is increasing. It suffices to consider the σ algebra $\mathcal{F} = \cup_i \mathcal{F}_i$. Note that $\mathcal{F}$ resolves the information of the experiment up to any finite time.

For any finite word $(w_1, .., w_n)$ with $w_i = H, T$ for $i = 1, .., n$ let us define the set

$$A_{w_1,..,w_n} = \{\omega \in \Omega : \omega_i = w_i \text{ for } i = 1, .., n\}.$$

Let us again assume the probability of a head is p, and each trial is independent, so that

$$\mathbb{P}(A_{w_1,..,w_n}) = \mathbb{P}(w_1) \cdots \mathbb{P}(w_n)$$

and $\mathbb{P}(w_i) = p$ if $w_i = H$ and $\mathbb{P}(w_i) = 1 - p$ if $w_i = T$.

The event $\{Y = k\}$ is then the event of the first $k - 1$ flips being tails and the k^{th} being head,

$$\{Y = k\} = \underbrace{A_{T...TH}}_{\text{with } k-1 \ Ts}$$

Thus

$$\mathbb{P}(\{Y = k\}) = \mathbb{P}\left(\underbrace{A_{T...TH}}_{\text{with } k-1 \ Ts}\right) = (1-p)^{k-1}p$$

■

Mass function and cumulative distribution - Discrete A more common - often more useful - way to introduce a random variable is by it's density function, in this case the above probability space is implicit (or the Sample space is identified with the State space).

Ie if the Sample space is identified with the state space $\Omega = S$, $\mathcal{F}$ is the power set of S.

In the discrete case, with random variable X and state space $S \subset \mathbb{R}$, for any $x \in S$ define mass function p

$$p(x) = \mathbb{P}(\{x\}) = \mathbb{P}(X = x)$$

We can define the cumulative distribution,

$$F_X(x) = \mathbb{P}(\{-\infty, x\}) = \mathbb{P}(X \leq x).$$

Example: (Bernoulli Trial - part 2)

■

We can define the cumulative and density distribution of X. The density, $f_X(0) = q$ and $f_X(1) = p$. The cumulative distribution is,

$$F_X(x) = \begin{cases} 0 & x < 0, \\ q & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

■

Density function and cumulative distribution - Continuous Let us take $S = \Omega = \mathbb{R}$. The σ algebra is a bit technical but it is enough to know it contains intervals $[a, b]$, unions of intervals $\cup_i [a_i, b_i]$ etc.

Let X be a continuous random variable with probability density function, f_X, then

$$\mathbb{P}(a < X \le b) = \int_a^b f_X(x)dx$$

Notice f_X must obey

$$\int_{-\infty}^{\infty} f_X(x)dx = \mathbb{P}(X \in \mathbb{R}) = 1$$

Let us define the cumulative distribution function,

$$F_X(x) = \int_{-\infty}^{x} f_X(s)ds$$

Example: (Gaussian/Normal)

■

Let $Z \sim N(\mu, \sigma^2)$ be a Normal random variable with mean μ and variance σ^2. The standard Normal is defined as $Z \sim N(0, 1)$.

The density function of the Gaussian/Normal $Z \sim N(\mu, \sigma^2)$ is

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

so that

$$\mathbb{P}(Z \le z) = \int_{-\infty}^{z} e^{-\left(\frac{s-\mu}{\sigma}\right)^2} \frac{ds}{\sigma\sqrt{2\pi}}$$

■

2.2 Expectation

The most important quantity to measure a random variable is the expectation. We'll define this in the case of $S \subset \mathbb{R}$.

Lets use the probability space definition to define the expectation,

$$\mathbb{E}(X) = \sum_{x \in S} x\mathbb{P}(\{x\}) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}).$$

On the other hand, if we use a mass function,

$$\mathbb{E}(X) = \sum_{x \in S} xp(x)$$

In the continuous case:

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx$$

Example: (Continuous 'triangular' distribution)

■

Consider X given by density $f_X(x) = 2x$ for $0 \leq x \leq 1$. Find the expectation,

$$\mathbb{E}(X) = \int_0^1 x f_X(x) dx = \int_0^1 x(2x) dx = \left. \frac{2x^3}{3} \right|_0^1 = \frac{2}{3}$$

■

2.2.1 Tail sum formula

In this section we demonstrate a formula for the expectation of a nonnegative random variable

If X is a nonnegative random variable with cumulative distribution F_X then

$$F_X(x) = 0$$

for all $x < 0$. If X is continuously distributed with density f_X then $f_X(x) = 0$ for all $x < 0$.

If X is a non negative random variable we can calculate the expectation

$$\mathbb{E}(X) = \int_{x=0}^{\infty} x f_X(x) dx = \int_{x=0}^{\infty} \int_{t=0}^{x} f_X(x) dt dx = \int_{t=0}^{\infty} \int_{x=t}^{\infty} f_X(x) dx dt = \int_{t=0}^{\infty} [1 - F_X(t)] dt$$

we therefore have the formula,

$$\mathbb{E}(X) = \int_{t=0}^{\infty} [1 - F_X(t)] dt = \int_{t=0}^{\infty} \mathbb{P}(X > t) dt$$

This formula is valid for discrete nonnegative random variables as well.

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k) = \sum_{k=0}^{\infty} k p(k)$$

Example: (Geometric distribution)

■

Suppose X is a geometric random variable for coin flips with probability of success p, $p + q = 1$ and $\mathbb{P}(X = k) = q^{k-1} p$. Let's calculate the expectation using the tail sum formula.

So

$$\mathbb{P}(X \leq k) = p + qp + ... + q^{k-1} p = \frac{1 - q^k}{1 - q} p = 1 - q^k$$

So $\mathbb{P}(X > k) = q^k$ Now we have,

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k) = 1 + q + q^2 + \cdots = \frac{1}{1 - q} = \frac{1}{p}.$$

■

3 Random Variables with joint distribution

Suppose Z takes on values in $S = \mathbb{R}^2$, then we can write $Z = (X, Y)$ where there is some joint density function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}^+$ where $\mathbb{R}^+ = [0, \infty)$. Then for $A \subset \mathbb{R}^2$, we have

$$\mathbb{P}(Z \in A) = \int\int_{(x,y)\in A} f_{X,Y}(x, y)$$

Let us use the notation $\chi(z) = \chi_{P(z)}$ where $P(z)$ is some truth/false statement, such as $P(z)$ is is the statement $|z| < 1$. χ is a function so that $\chi_{P(z)} = 1$ when condition $P(z)$ is true and $\chi_{P(z)} = 0$ otherwise.

For example for $A \subset \mathbb{R}^2$

$$\chi_{z\in A} = \begin{cases} 1 & z \in A \\ 0 & z \notin A \end{cases}$$

Example: (Uniform distribution on a triangle)

■

Consider $Z = (X, Y) \sim \mathcal{U}(A)$ where A is the triangle $A = \{(x, y) : 0 \leq y \leq x \leq 1\}$. The area of the triangle is $|A| = 1/2$. Thus Z has the density $f_Z(z) = 2\chi_{z\in A}$. Equivalently, (X, Y) has joint density $f_{X,Y}(x, y) = 2\chi_{(x,y)\in A}$ ■

Remark Suppose X is uniformly distributed on the set $A \subset \mathbb{R}^2$. Suppose A has area $|A|$. Then for any $B \subset A$ it holds that $\mathbb{P}(X \in B) = |B|/|A|$.

Conversly if X is a random variable taking values in a set $A \subset \mathbb{R}^2$ so that there is a $c > 0$ so that the probability X is in B for any $B \subset A$ is $\mathbb{P}(X \in B) = c|B|$ then X is uniform and $c = 1/|A|$.

Cumulative density We can define cumulative distribution as usual
$F_{X,Y}(x, y) = \mathbb{P}(X \leq x;\ Y \leq y)$.

If the density $f_{X,Y}$ exists (the variable $Z = (X, Y)$ is continuously distributed) then,

$$F_{X,Y}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u, v)dudv$$

so

$$\frac{d^2}{dxdy}F_{X,Y} = f_{X,Y}$$

Example: (Two day stock price)

■

Suppose the value of a stock on day zero (today) is $S_0 = 3/2$ on day one the stock value S_1 is uniformly distributed on an interval of length one centered at $3/2$ ie uniformly distributed on $[1, 2]$. Suppose the value on the second day S_2 is uniformly distributed over an interval of length one centered at the value of day 1.

We can write independent variables $U_i \sim \mathcal{U}[-.5, .5]$ for $i = 1, 2$. Then $S_1 = 1.5 + U_1$, and $S_2 = S_1 + U_2$.

The support of the distribution of (S_1, S_2) in $\mathbb{R}^2$ is a parallelogram with vertices $(1, 0.5), (1, 1.5), (2, 1.5)$ and $(2, 2.5)$. $f_{X_1,X_2}(x_1, x_2) = 1$ for (x_1, x_2) on the interior of the parallelogram and 0 otherwise.

Is the distribution of (S_1, S_2) uniform in $\mathbb{R}^2$?

Notice

$$\begin{pmatrix} S_1 \\ S_2 \end{pmatrix} = \begin{pmatrix} 1.5 + U_1 \\ 1.5 + U_1 + U_2 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$$

A linear transformation of a uniform distribution is uniform. ✓

Thus if A is the parallelogram described above, then $f(s_1, s_2) = \chi_{(s_1,s_2)\in A}$ is the joint distribution. (The area of the parallelogram is 1 so we don't need a normalizing factor in front.)

Notice S_2 however is distributed from 0.5 to 2.5. Is it uniformly distributed?

■

Day to day prices of stock are marginal distributions of the process of stock prices.

Ie if stock prices over 10 days is a random variable $(S_1, S_2, ...S_{10}) \in \mathbb{R}^{10}$ then the price on day 9, S_9 is a marginal random variable, what is it's distribution?

3.1 Marginals

In the discrete case, for bivariate random variable $Z = (X, Y)$ with probability mass function $f_{X,Y}$, the formula for marginal density is

$$f_X(x) = \sum_{y:(x,y)\in S} \mathbb{P}[(X, Y) = (x, y)].$$

Similarly for the continuous case, where $f_{X,Y}$ is joint density, the marginal density is

$$f_X(x) = \int_{y\in\mathbb{R}} f_{X,Y}(x, y)dy$$

Example: (Continuous example)

■

Let $A = \{(x, y) : 0 \leq y \leq x \leq 1\}$.
Define the density of the variable (X, Y),

$$f_{(X,Y)}(x, y) = 6y\chi_{(x,y)\in A}$$

Find the marginal distributions. Find the expectation of each variable.

marginals Compute the X marginal,

$$\int_{\mathbb{R}} f_{(X,Y)}(x, y)dy = \int_0^x 6ydy = 3x^2$$

thus

$$f_X(x) = \begin{cases} 3x^2 & \text{for } x \in (0, 1) \\ 0 & \text{for } x \notin (0, 1) \end{cases}$$

Compute the Y marginal,

$$\int_{\mathbb{R}} f_{(X,Y)}(x, y)dx = \int_y^1 6ydx = 6y(1 - y)$$

thus

$$f_Y(y) = \begin{cases} 6y(1 - y) & \text{for } x \in (0, 1) \\ 0 & \text{for } x \notin (0, 1) \end{cases}$$

expectations Let us calculate the expectation of X using the marginal distribution,

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx = 3/4$$

Let us calculate the expectation of Y using the joint distribution,

$$\mathbb{E}(Y) = \int\int_{\mathbb{R}^2} y f_{XY}(x,y) dx dy = \int_0^1 \int_0^x 6y^2 dy dx = \int_0^1 2x^3 dy dx = 1/2$$

■

3.2 Sums

We wish to consider the special examples of sums of random variables.

Example: (Two day stock price – part 2)

■

We continue the example above. Let us find the density of S_2.

$$f_{S_2}(s_2) = \int_{y \in \mathbb{R}} f_{S_1,S_2}(y, s_2) dy = \begin{cases} 0 & 2.5 < s_2 \\ -s_2 + 2.5 & 1.5 < s_2 \leq 2.5 \\ s_2 - 0.5 & 0.5 < s_2 \leq 1.5 \\ 0 & s_2 \leq 0.5 \end{cases}$$

Thus, S_1 and (S_1, S_2) are uniform but S_2 is not uniform.

By symmetry it is easy to see that $\mathbb{E}(S_2) = 3/2$. ■

3.3 Independent Random Variables

Variables X and Y taking values in S are independent if, for all $A, B \subset S$, we have

$$\mathbb{P}(X \in A; Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

For $S = \mathbb{R}$, in terms of cumulative distributions, $x, y \in \mathbb{R}$ we have

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x; Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) \leq F_X(x) F_Y(y)$$

If F is differentiable we can define the density. If X, Y are independent, the density is multiplicative, ie,

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

the same holds for probability mass.

A corollary of this representation is that for X, Y independent random variables we have

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Indeed,

$$\mathbb{E}(XY) = \int_{\mathbb{R}}\int_{\mathbb{R}} xy f_{XY}(x,y) dx dy = \int_{\mathbb{R}} x f_X(x) dx \int_{\mathbb{R}} y f_X(y) dy = \mathbb{E}(X)\mathbb{E}(Y)$$

4 Derived random Variables

Given random variable X, define probability for a random variable $Y = g(X)$.

The cumulative here is defined as

$$F_Y(y) = \int_{x:g(x)\leq y} f_X(x)dx$$

Suppose for every y an interval of $(y-\epsilon, y+\epsilon)$ exists so that there are local functions (x_i) so that $g(x_i(y')) = y'$ for $y' \in (y-\epsilon, y+\epsilon)$. Then the derivative of the cumulative density,

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \sum_i f_X(x_i(y)) \left|\frac{d}{dy}[x_i(y)]\right|$$

Example: (Square of uniform)

■

Let $X \sim \mathcal{U}([-2,2])$ and $Y = X^2$. Calculate the CDF and PDF.

Then for $0 \leq y \leq 4$,

$$F_Y(y) = \mathbb{P}(X : X^2 \leq y) = \int_{-\sqrt{y}}^{\sqrt{y}} (1/4)dx = \frac{2\sqrt{y}}{4}.$$

Let $x_{\pm}(y) = \pm\sqrt{y}$, then

$$f_Y(y) = f_X(+\sqrt{y})\,|(+\sqrt{y})'| + f_X(-\sqrt{y})\,|(-\sqrt{y})'| = \frac{1}{4\sqrt{y}}$$

for $0 < y < 4$, and 0 otherwise. ■

4.1 Moments

The moments of the random variables are the expectations of powers of the random variables The k^{th} moment of X is

$$\mathbb{E}(X^k) = \int_{\mathcal{R}} x^k f_X(x)dx$$

The k^{th} *central* moment is

$$\mathbb{E}([X - \mathbb{E}(X)]^k) = \sum_i (-1)^i \binom{k}{i} \mathbb{E}(X^i)[\mathbb{E}(X)]^{k-i}$$

4.1.1 Variance

Of course the most important central moment is the second. It is known as the variance,

$$Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 =: \sigma_X^2$$

The square root of the variance is the Standard deviation

$$\sigma_X = \sqrt{var(X)}$$

Recall these helpful properties of variance:

1 For $a, b \in \mathbb{R}$, we have $var(aX+b) = a^2 var(X)$.

2 For X_1, X_2 independent random variables $var(X_1+X_2) = var(X_1) + var(X_2)$

3 For X_1, X_2 independent random variables $var(X_1X_2) = var(X_1)var(X_2) - \mu_{X_1}^2 var(X_2) - \mu_{X_2}^2 var(X_1)$. As a corollary, if $\mu_{X_1} = \mu_{X_2} = 0$ then $var(X_1X_2) = var(X_1)var(X_2)$.

Example: (Two day stock price)

■

Let us find the variance of S_1 and S_2. Define $U_i \sim \mathcal{U}[-.5, .5]$ for $i = 1, 2$. Let $S_1 = 3/2 + U_1$ and $S_2 = S_1 + U_2$.

Thus

$$var(S_1) = var(U_1) = 1/12$$

having used property 1 and *that for* $U \sim \mathcal{U}[a, b]$ *that* $var(U) = [b-a]/12$.
On the other hand the variance for S_2 is,

$$var(S_2) = var(\frac{1}{2} + U_1 + U_2) = var(U_1) + var(U_2) = 1/6.$$

having used property 1 and 2. ■

5 Covariance

Let X and Y be random variables. We may assume there is a joint density $f_{X,Y}$ (or joint mass $p_{X,Y}$) define the covariance

$$cov(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$$

for $\mu_X = \mathbb{E}X$ and $\mu_Y = \mathbb{E}Y$.

We can easily derive $cov(X,Y) = \mathbb{E}(XY) - \mu_X\mu_Y$. Notice as well $cov(X,Y) = cov(Y,X)$ and $var(X) = cov(X,X)$.

Define the covariance matrix

$$\Sigma = \Sigma_{X,Y} = \begin{pmatrix} var(X) & cov(X,Y) \\ cov(X,Y) & var(Y) \end{pmatrix}$$

Let $Z = aX + bY$ then (*verify this*)

$$var(Z) = cov(aX+bY, aX+bY) = a^2 var X + 2ab\, cov(X,Y) + b^2 var(Y) = \begin{pmatrix} a & b \end{pmatrix} \Sigma \begin{pmatrix} a \\ b \end{pmatrix}$$

In particular, for $Z = X+Y$ *we have* $var(Z) = var(X) + 2cov(X,Y) + var(Y)$

n-variables This discussion generalizes to a set of n jointly distributed random variables $X_1, ..., X_n$. Thus for the covariance matrix Σ defined by $\Sigma_{ij} = cov(X_i, X_j)$. The matrix Σ is positive semidefinite.

For $Z = v_1X_1 + \cdots + v_nX_n$ the variance of Z is $var(Z) = v^T\Sigma v \geq 0$.

It follows that Σ has the property that for all $v \in \mathbb{R}^n$ we have $v^T\Sigma v \geq 0$. This property is known as positive semidefinite.

Moreover, Σ is real and symmetric so that $\Sigma = ODO^T$ where the matrices O are orthogonal and D is diagonal - with entries being the eigenvalues of Σ. As Σ is positive definite, the eigenvalues are nonnegative.

Example: (Two day stock price – part 3)

■

We continue the example above. Let us find the covariance matrix and the correlation of X_1 and X_2.

First find the covariance of X_1 and X_2. Find the expectation of the product,

$$\begin{aligned}\mathbb{E}(X_1X_2) &= \mathbb{E}((1+U_1)(1/2+U_1+U_2)) \\ &= \mathbb{E}\{(3/2+[U_1-1/2])(3/2+[U_1-1/2]+[U_2-1/2])\} \\ &= (3/2)^2+(3/2)\mathbb{E}([U_1-1/2]+[U_2-1/2])+(3/2)\mathbb{E}([U_1-1/2])+ \\ &+\mathbb{E}((U_1-1/2)^2)+\mathbb{E}(U_1-1/2)(U_2-1/2) \\ &= (3/2)^2+(1/12)+0=(3/2)^2+(1/12)\end{aligned}$$

But $\mathbb{E}(X_1)\mathbb{E}(X_2)=(3/2)^2$ so

$$cov(X_1,X_2)=\mathbb{E}(X_1X_2)-\mathbb{E}(X_1)\mathbb{E}(X_2)=1/12$$

■

5.1 Gaussians

One dimensional Gaussians $X\sim N(\mu,\sigma^2)$ have the PDF

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

Recall Gaussians are stable, that is for $X_i\sim N(\mu_i,\sigma_i^2)$ we have $X_1+X_2\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

For any covariance matrix Σ (that is a positive definite real symmetric matrix) in $\mathbb{R}^{n\times n}$ (real $n\times n$ matrices) and vector $\overline{\mu}\in\mathbb{R}^n$ we can define a Multivariate Gaussian distribution X with PDF

$$f_X(x)=\frac{1}{\sqrt{(2\pi)^n\det(\Sigma)}}e^{-\frac{1}{2}(x-\overline{\mu})^T\Sigma^{-1}(x-\overline{\mu})} \tag{5.1}$$

Notice if $\Sigma=ODO^T$ for D diagonal $D=diag(\lambda_1,..,\lambda_n)$ then $\Sigma^{-1}=OD^{-1}O^T$ and $D=diag(\lambda_1^{-1},..,\lambda_n^{-1})$.

5.2 Correlation

The covariance of two random variables may be positive or negative. Random variables which have positive covariance have the property that as one increases then *on average* the second increases.

On the other hand if the covariance is negative then as one increases on average the second decreases.

It is useful to normalize the covariance to a number ρ between -1 and 1. Where $|\rho|=1$ indicates 'perfect' correlation between the two random variables. Let

$$\rho=\rho_{XY}=corr(X,Y)=\frac{cov(X,Y)}{\sigma_X\sigma_Y}$$

Example: (positive correlation)

■

In the definition of multidimensional Gaussian, let $n=2$, $\mu=(0,0)^T$ and $\Sigma=\begin{pmatrix}1 & 1/2\\ 1/2 & 2\end{pmatrix}$. Then

$\rho_{X,Y} = \frac{1/2}{\sqrt{2}}$. ■

Example: (negative correlation)

■

In the definition of multidimensional Gaussian, let $n = 2$, $\mu = (0,0)^T$ and $\Sigma = \begin{pmatrix} 2 & -1/2 \\ -1/2 & 3 \end{pmatrix}$. Then $\rho_{X,Y} = \frac{-1/2}{\sqrt{2}\sqrt{3}}$. ■

If X an Y are independent random variables then $\rho_{XY} = 0$, but the converse does not hold

Example: (Zero correlation - independent)

■

Let $a < b$ and $c < d$. Let $X \sim \mathcal{U}[a,b]$ and $Y \sim \mathcal{U}[c,d]$, then $\rho_{(X,Y)} = 0$. ■

Example: (Zero correlation - not independent)

■

Let $Z = (X,Y)$ be uniformly distributed on the shape $A = \{(x,y) \in \mathbb{R}^2 : |x+y| \leq 1\}$.
Then $\rho_{XY} = 0$. ■

6 Conditionals

Given a probability experiment, one often would like to *condition* on some partial knowledge of the outcome.

Example: (Die roll)

■

Suppose we roll two die, one red and one green. The sample space is the set of ordered pairs $S = \{(i,j) : i,j \in \{1,..,6\}\}$. Consider the probability that one of the faces shows a 1.

$$\mathbb{P}(\text{At least one dice shows 1}) = 1 - \mathbb{P}(\text{Neither dice shows a 1}) = 1 - (5/6)^2 = 11/36$$

But if we have partial knowledge of the outcome the probability may change, suppose we know the sum is 4.

$$\mathbb{P}(\text{At least one dice shows 1}|\text{The sum of the die is 4}) = 2/3$$

This can be seen by deduction, by considering the three possible outcomes where the sum of the roll is 4: $\{(1,3),(2,2),(3,1)\}$, two of three of these have a face with 1 showing. ■

6.1 Conditional probabilities

We will define the conditional probabilities, recall Bayes rule: let $A, B \subset S$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{6.1}$$

this gives the probability of event A *given* event B.

Example: (Die roll 2)

■

Suppose we roll two die, one red and one green. The sample space is the set of ordered pairs $\Omega = \{(i,j) : i,j \in \{1,..,6\}\}$. Let $X(i,j) = i + j$ the sum of the faces of the die. Suppose we know that the sum is greater than 7, ie $X > 7$. What is the probability X is greater than or equal to 10?

Let $A = \{X \geq 10\}$ and $B = \{X > 7\}$.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) = 6/36.$$

on the other hand $\mathbb{P}(B) = 15/36$ so

$$\mathbb{P}(A|B) = \frac{6/36}{15/36} = \frac{2}{5}$$

Of course, we notice $\frac{2}{5} \neq \frac{6}{36} = \mathbb{P}(X \geq 10)$ ■

Example: (Die roll 3)

■

Let us continue the previous example. Let C be the event that the green dice is even. Find

$$\mathbb{P}(A|C)$$

Notice the elements of $A \cap C$ have the green dice is either 4 or 6. If the green dice is 4 the red dice is 6. If the green dice is 6 the red dice is 4,5, or 6.

$$\mathbb{P}(A \cap C) = 4/36 = 1/9$$

Of course $\mathbb{P}(C) = 1/2$. So

$$\mathbb{P}(A|C) = \frac{1/9}{1/2} = \frac{2}{9}$$

■

Finally let us return to the stock price example.

Example: (Two day stock price – part 4)

■

We continue the example above Suppose the event B is that the value of the stock on the second day is 5/3 ie $B = \{S_2 = 5/3\}$. Notice $\mathbb{P}(B) = 0$ can we condition on it?!

That is, let A be the event that $X_1 < 3/2$. Show

$$\mathbb{P}(A|B) = 2/5$$

Of course, if we think in terms of our usual notions from calculus, we can define probabilities which are positive and then take limits.

That is let $B_\epsilon = \{|X_2 - 5/3| < \epsilon\}$, and define for an event A, $\mathbb{P}(A|B) = \lim_{\epsilon \to 0} \mathbb{P}(A|B_\epsilon)$.

■

6.2 Conditional Random Variable

We condition the random variable Y on X with joint PDF $f_{X,Y}$ with the function,

$$f_Y(y|X = x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \tag{6.2}$$

where f_X is the marginal of X defined in Section 3.1.

In the discrete setting the conditional distribution is

$$P_Y(y|X=x) = \frac{P_{X,Y}(x,y)}{P_X(x)}$$

where P_X is again the marginal of X.

Example: (Two day stock price – part 5)

■

Again let B be the value of the stock on the second day is 5/3 ie $B = \{S_2 = 5/3\}$.
If we condition on event B, what is the conditional density of S_1, ie find $f_{S_1}(s|S_2 = 5/3)$ from formula (6.2).

$$f_{S_1}(s|S_2 = 5/3) = \frac{\chi_{s\in(7/6,2)}}{5/2 - 5/3} = \frac{1}{5/6}\chi_{s\in(7/6,2)}$$

■

6.3 Conditional expectation

As the name indicates this is expectation conditioned on some function on the probability space. Formally, for jointly distributed random variables X and Y,

$$\mathbf{E}(Y|X) = \sum_y \mathbb{P}(Y = y|X).$$

The intuitive understanding is that $\mathbf{E}(\bullet|X)$ 'integrates out' all randomness *independent* of X. Notice, what remains is a function of X. Moreover,

$$\mathbb{E}(Y) = \mathbb{E}_X[\mathbf{E}(Y|X)]$$

where $\mathbb{E}_X$ indicates taking expectation with respect to X.

Example: (unfair coin flips)

■

Suppose 3 unfair coins are flipped. Let Ω be the set of outcomes. Let

$$X_i = \begin{cases} 1 & i^{th} \text{ flip is head} \\ 0 & i^{th} \text{ flip is tail} \end{cases}.$$

The total number of heads is H,

$$H = X_1 + X_2 + X_3.$$

Find $\mathbf{E}(H|X_1)$ and $\mathbf{E}(X_1|H)$.

First we find, $\mathbf{E}(H|X_1)$. We will calculate in 2 ways, first let us use the formula

$$\mathbf{E}(H|X_1) = 0\mathbb{P}(H = 0|X_1) + 1\mathbb{P}(H = 1|X_1) + 2\mathbb{P}(H = 2|X_1) + 3\mathbb{P}(H = 3|X_1)$$

consider fixing X_1,

$$\mathbb{P}(H = 0|X_1 = 0) = q^2; \quad \mathbb{P}(H = 1|X_1 = 0) = 2pq; \quad \mathbb{P}(H = 2|X_1 = 0) = p^2; \quad \mathbb{P}(H = 3|X_1 = 0) = 0$$

and

$$\mathbb{P}(H = 0|X_1 = 1) = 0; \quad \mathbb{P}(H = 1|X_1 = 1) = q^2; \quad \mathbb{P}(H = 2|X_1 = 1) = 2pq; \quad \mathbb{P}(H = 3|X_1 = 1) = p^2$$

Thus

$$\mathbb{E}(H|X_1 = 0) = 2pq + 2p^2 = 2p$$

and

$$\mathbb{E}(H|X_1 = 1) = 1 \cdot q^2 + 2 \cdot 2pq + 3 \cdot p^2 = 1 + 2pq + 2p^2 = 1 + 2p$$

consider making this a function of X_1

$$\mathbf{E}(H|X_1) \quad = \quad X_1 + 2p$$

notice this is a function of X_1 only - everything else has been integrated.

Let us calculate in a second way remember that for any expectation of any random variables X, Y, we have $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. Apply this to H..

$$\begin{aligned}\mathbf{E}(H|X_1) &= \mathbf{E}(X_1 + X_2 + X_3|X_1) \\ &= \mathbf{E}(X_1|X_1) + \mathbf{E}(X_2 + X_3|X_1)\end{aligned}$$

But $\mathbf{E}(X_1|X_1) = X_1$ and X_2, X_3 are independent of X_1 so they may be integrated as normal.

$$\mathbf{E}(H|X_1) = \mathbf{E}(X_1|X_1) + \mathbf{E}(X_2 + X_3|X_1) = X_1 + 2p$$

Now let us find $\mathbf{E}(X_1|H)$. First let us calculate for fixed H, of course,

$$\mathbb{E}(X_1|H = 0) = \mathbb{P}(X_1 = 1|H = 0) = 0$$

For $H = 1$ consider that each coin has equal probability of being heads,

$$\mathbb{E}(X_1|H = 1) = \mathbb{P}(X_1 = 1|H = 1) = 1/3$$

For $H = 2$ we have exactly the symmetric case that each coin has equal probability of being tails,

$$\mathbb{E}(X_1|H = 2) = \mathbb{P}(X_1 = 1|H = 2) = 2/3$$

Finally, it is clearly true that,

$$\mathbb{E}(X_1|H = 3) = \mathbb{P}(X_1 = 1|H = 3) = 1$$

Thus, we have

$$\mathbf{E}(X_1|H) = H/3.$$

■

The first part of the example illustrates, if X and Y are any random variables and $f(X)$ is a function of the random variable X and $g(X, Y)$ is a function of both random variables, that

$$\mathbf{E}(f(X) + g(X, Y)|X) = f(X) + \mathbf{E}(g(X, Y)|X).$$

Similarly:

$$\mathbf{E}(f(X)g(X, Y)|X) = f(X)\mathbf{E}(g(X, Y)|X)$$

Moreover, if Y is independent of X, and $g(Y)$ is only a function of Y, then $\mathbf{E}(g(Y)|X) = \mathbb{E}(g(Y))$ which is a number - no longer a function.

The above can be generalized to the case conditioning on several random variables.

Example: (biased coin flips)

■

Again suppose $\mathbb{P}(X_i = 1) = p$, and $\mathbb{P}(X_i = 0) = 1 - p = q$ for $i = 1, ..., n$. For $m = 1, .., n$, let us write $H_m = X_1 + \cdots + X_m$ so for $m < n$ we have

$$\mathbf{E}(H_m|H_n) = \frac{m}{n}H_n$$

on the other hand,

$$\mathbf{E}(H_n|H_m) = H_m + \mathbb{E}(X_{m+1} + \cdots + X_n) = H_n + p(n - m).$$

■

In the above example we can let $\mathcal{F}_i$ contain 'all the information obtained from the first i coins.' This is just notation, so we will write,

$$\mathbf{E}(H_n|\mathcal{F}_m) \equiv \mathbf{E}(H_n|X_1, \cdots, X_m).$$

Example: (Sum of i.i.d.)

■

Let X_i for $i = 1, 2, \ldots$ be independent and identically distributed (iid) random variables. That is, for given random variable X for all i suppose X_i has the same probability distribution as X, written $X_i \sim X$. Let $\mathbb{E}(X) = \mu$ and $Var(X) = \sigma^2$.

For $m = 1, .., n$, let $S_m = X_1 + \cdots + X_m$. For $m < n$, calculate $\mathbf{E}[S_n^2|\mathcal{F}_m]$ the expected value of S_n^2 given the first m flips.

We can write, $S_n^2 = ((S_n - S_m) + S_m)^2$, so let us write $Y = S_n - S_m$. Then we have

$$\begin{aligned}
\mathbf{E}[S_n^2|\mathcal{F}_m] &= \mathbf{E}[(Y + S_m)^2|\mathcal{F}_m] \\
&= \mathbf{E}[Y^2 + 2YS_m + S_m^2|\mathcal{F}_m] \\
&= \mathbb{E}[Y^2] + 2S_m\mathbb{E}[Y] + S_m^2
\end{aligned}$$

where we used $\mathbf{E}(YS_m|\mathcal{F}_m) = S_m\mathbb{E}(Y)$ as Y is independent of $\mathcal{F}_m$ and S_m is a function of $\mathcal{F}_m$.

Now $\mathbb{E}(Y) = \mathbb{E}(X_{m+1}) + \cdots + \mathbb{E}(X_n) = (n - m)\mu$. On the other hand, $var(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2$, but $var(Y) = \sigma^2(n - m)$, so that $\mathbb{E}(Y^2) = \mu^2(n - m)^2 + \sigma^2(n - m)$.

Finally we have

$$\mathbf{E}[S_n^2|\mathcal{F}_m] = \mu^2(n - m)^2 + \sigma^2(n - m) + 2S_m\mu(n - m) + S_m^2$$

■

7 Asymptotic behavior

Let $(T_i)_{i\geq 1}$ be i.i.d. random variables, with distribution $F(x) = \mathbb{P}(T_i \leq x)$. We assume for all $i = 0, 1, 2, ...$ that $\mathbb{E}(|T_i|) < \infty$ and define $\mu := \mathbb{E}(T_1)$.

Law of large numbers (LLN) The law of large numbers states, for T_i that with probability 1 the average converges to the mean i.e.

$$\lim_{N\to\infty} \sum_{i=1}^{N} \frac{T_i}{N} \to \mu.$$

Notice the statement 'with probability 1', there are conceivable other sequences not averaging to the mean but they total zero in probability - like flipping infinite heads in a row. An easier to understand, but weaker, statement is

$$\mathbb{P}\left(\left|\frac{T_1 + \cdots T_N}{N} - \mu\right| > \epsilon\right) \to 0 \text{ as } N \to \infty.$$

The moral is $T_1 + \cdots + T_N \approx \mu N$, if we set $t = \mu N$ then $N = t/\mu$ so that

$$T_1 + \cdots + T_{t/\mu} \approx t$$

Central Limit Theorem (CLT) We assume $\text{Var}\,(T_1) = \sigma^2$. The central limit theorem states, for $N(0,1)$ a normal variable with mean 0 and variance 1,

$$\frac{T_1 + \cdots + T_N - N\mu}{\sigma\sqrt{N}} \to N(0,1)$$

by this we mean that

$$\mathbb{P}\left(\frac{T_1 + \cdots + T_N - N\mu}{\sigma\sqrt{N}} \le x\right) \to \Phi(x).$$

Where Φ is the cumulative distribution of a $N(0,1)$,

$$\Phi(x) = \int_{-\infty}^{x} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}.$$

We write down the sum, (here $Z \sim N(0,1)$),

$$T_1 + \cdots + T_N = S_N \approx N\mu + Z\sigma\sqrt{N}$$

Central Limit Theorem - for multivariate random variables (CLT) Suppose X_i are iid Random variables in $\mathbb{R}^d$ we may write

$$X_i = \begin{pmatrix} X_{i(1)} \\ \vdots \\ X_{i(d)} \end{pmatrix}.$$

Suppose $\mathbb{E}(X_i) = \mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ is the covariance matrix $\Sigma_{j,k} = cov(X_{i(j)}, X_{i(k)})$.

Let us suppose that $det\Sigma \neq 0$ *(this is equivalent to saying that each $X_{i(j)}$ has some randomness not contained in the other $X_{i(k)}$ - that is one of the $X_{i(j)}$ is not a function of the other $X_{i(j)}$.)*

Then a sum of the X_i properly normalized approaches a Gaussian random variable. Let $S_n = X_1 + \cdots + X_n$ That is for $A \in \mathbb{R}^d$ we have

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}} \in A\right) = \int \cdots \int_{x\in A} e^{-(x-\mu)^T\Sigma^{-1}(x-\mu)} \frac{dx_1 \cdots dx_d}{\sqrt{(2\pi)^d \det\Sigma}}.$$

As a short hand we write $\frac{1}{\sqrt{n}}(S_n - n\mu) \xrightarrow{\mathcal{D}} Z$ where $Z \sim N(0,\Sigma)$.

8 ODE review

We will review first and second order linear ODEs.

8.1 First Order

Let $f : [0, \infty) \to \mathbb{R}$. The notation $f : U \to V$ means f is a map from the set U to the set V.

Consider the ODE,

$$\frac{d}{dt}y(t) = f(t)y(t) \tag{8.1}$$

with $y(0) = y_o$.

Let $F : [0, \infty) \to \mathbb{R}$ be the antiderivative of f, that is $\frac{d}{dt}F(t) = f(t)$ and set $F(0) = 0$. Let $y(t) = y_o e^{F(t)}$, then $y(t)$ solves (8.1).

EG Let $y' = 1.01y$ and $y_o = 100$. The solution is

$$y(t) = 100e^{1.01t}$$

EG Let $y' = 10ty$ and $y_o = 10$. The solution is

$$y(t) = 10e^{5t^2}$$

8.2 Second Order

Let us consider

$$a\frac{d^2}{dt^2}y(t) + b\frac{d}{dt}y(t) + cy(t) = 0 \tag{8.2}$$

with $y(0) = y_o$ and $y'(0) = y_1$.

To solve this equation we need to first find the roots of the characteristic polynomial,

$$ax^2 + bx + c = 0.$$

Suppose the roots are r_1, r_2.

If $r_1 \neq r_2$ then the solution of (8.2) is

$$y(t) = A_1 e^{r_1 t} + A_2 e^{r_2 t}$$

where $\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ r_1 & r_2 \end{pmatrix}^{-1} \begin{pmatrix} y_o \\ y_1 \end{pmatrix} = \frac{1}{r_2 - r_1} \begin{pmatrix} r_2 & -1 \\ -r_1 & 1 \end{pmatrix} \begin{pmatrix} y_o \\ y_1 \end{pmatrix}$.

Recall the inverse of a 2x2 matrix,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

If $r = r_1 = r_2$ then the solution of (8.2) is

$$y(t) = A_1 e^{rt} + A_2 t e^{rt}$$

where $A_1 = y_o$ and $A_2 = y_1 - r y_o$.

EG Let $y'' = -4y$ and $y_o = 1$, $y_1 = 0$. The characteristic polynomial $x^2 + 4 = 0$ has roots $r_1 = 2i$ and $r_2 = -2i$. The solution is:

$$y(t) = \frac{1}{2}e^{i2t} + \frac{1}{2}e^{-i2t} = \cos(2t)$$

We used the Euler formula $e^{ix} = \cos(x) + i\sin(x)$