M. MOTAMED, Univ. of New Mexico

Probability Theory

① probability space

Consider a random experiment.

Let Ω denote the set of all possible outcomes of the experiment

The set Ω is called the sample space, which may have a finite or infinite number of elements.

Let $\omega \in \Omega$ be an element of Ω (ω is an outcome of the experiment)

A subset of Ω (i.e. a set of outcomes) to which a probability measure is assigned is called an event (a random event).

Example.

- random experiment: flipping 3 fair coins
- sample space: $\Omega = \{(H,H,H), (H,H,T), (H,T,H), (H,T,T), (T,H,H), (T,H,T), (T,T,H), (T,T,T)\}$
- an event: the last two are heads

$$A = \{(H,H,H), (T,H,H)\} \subset \Omega$$

Example.

- random experiment: rolling 2 dice
- sample space: $\Omega = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$
- an event: the sum of the two dice is 5

$$A = \{(1,4), (2,3), (3,2), (4,1)\} \subset \Omega$$

Exercise. What happens to the last example if we play backgamon?

Let Σ be a σ-algebra on Ω, that is:

Σ is a collection of subsets of Ω with the following properties:

1. $\emptyset \in \Sigma$ (Σ contains the empty set)
2. If $A \in \Sigma$, then $A^c = \Omega \setminus A \in \Sigma$ (Σ is closed under complementation)
3. If $A_1, A_2, A_3, \dots \in \Sigma$, then $\bigcup_{i=1}^{\infty} A_i \in \Sigma$ (Σ is closed under countable unions of its members)

Note
- An element $A \in \Sigma$ is an event (Σ-measurable subset of Ω).
- Σ is also called an event space.
- Σ provides a set of events which are all measurable.

Remark

① 1 & 2 $\Rightarrow \Omega \in \Sigma$

② 1 & 2 & 3 — De Morgan's law $(A \cup B)^c = A^c \cap B^c$ → Σ is closed under countable intersections

③ $\Sigma \subset 2^{\Omega}$, where 2^{Ω} is the power set, i.e. the set of all subsets of Ω

Example. The power set of $\Omega = \{1,2,3\}$ is

$$2^{\Omega} = \{ \{1,2,3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1\}, \{2\}, \{3\}, \emptyset \}$$

It has $2^3 = 8$ members. We write $|2^{\Omega}| = 8$ and say "the cardinality of 2^{Ω} is 8".

The pair (Ω, Σ) is called a measurable space.

Let P be a probability measure on Σ satisfying Kolmogorov axioms:

- $\forall A \in \Sigma$: $P(A) \in \mathbb{R}$, $P(A) \geq 0$
- $P(\Omega) = 1$
- $P\left(\bigcup A_i\right) = \sum_i P(A_i)$ with $\begin{cases} A_i \in \Sigma \\ A_i \cap A_j = \emptyset, \ i \neq j \text{ disjoint events} \end{cases}$ (σ-additivity or countable additivity)

It follows that: $P: \Sigma \to [0,1]$, $P(\emptyset) = 0$

P returns an event probability

The triplet (Ω, Σ, P) is called a probability space (p-space).

For every event $A, B \in \Sigma$, we have the following properties:

- if $A \subseteq B \implies P(A) \leq P(B)$ (monotonicity property)
- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Remark For technical purposes (i.e. rigorous mathematical proofs) it is often useful to assume that the p-space we have is complete, in the following sense:

The p-space (Ω, Σ, P) is complete if

$B \in \Sigma$, $A \subset B$, $P(B) = 0$ implies $A \in \Sigma$.

[any subset of an event with prob. zero is an event]

Any p-space can be complete by adding to its σ-algebra all subsets of sets of prob. zero.

From now on we shall assume complete p-spaces.

Ex. The singleton points in $\mathbb{R}^n$ have Lebesgue measure zero; (length zero).
By countable additivity, any countable set of singletons has measure zero.

(2) Random variables (r.v.)

An rv is a convenient way to express the elements of Ω as numbers rather than abstract elements of sets.

Example: probability on the interval $[0,1] \subset \mathbb{R}$

Consider a p-space $([0,1], \mathcal{B}([0,1]), \mu)$, where

- The sample space is the real interval $[0,1] \subset \mathbb{R}$.
- The σ-algebra on $[0,1]$ is called the "Borel σ-algebra" on $[0,1]$ denoted by $\mathcal{B}([0,1])$, which is the σ-algebra generated by all open intervals on $[0,1]$. It is the event space.

Remark Since $\mathcal{B}([0,1])$ is a σ-algebra, it contains:

all open sub-intervals on $[0,1]$
all closed sub-intervals on $[0,1]$
all unions of closed or open sub-intervals on $[0,1]$
all intersections of closed or open sub-intervals on $[0,1]$
all semi-open sub-intervals on $[0,1]$

For example, it contains things like $\{\frac{1}{2}\}$, $[0, \frac{1}{2}] \cup (\frac{2}{3}, 1]$, $[\frac{1}{2}, \frac{2}{3}]$, ...

- $\forall A \in \mathcal{B}([0,1])$, $\mu(\cdot)$ is a probability measure, defined in terms of the length of the intervals contained in A:

$$\mu\left([\tfrac{1}{2}, \tfrac{2}{3}]\right) = \tfrac{2}{3} - \tfrac{1}{2} = \tfrac{1}{6}$$

$$\mu\left([\tfrac{1}{2}]\right) = 0$$

$$\mu\left([0, \tfrac{1}{2}] \cup (\tfrac{2}{3}, 1]\right) = \tfrac{1}{2} + (1 - \tfrac{2}{3}) = \tfrac{5}{6}$$

$$\mu\left([0, \tfrac{1}{2}] \cup [\tfrac{1}{3}, 1]\right) = \mu\left([0, \tfrac{1}{2}]\right) + \mu\left([\tfrac{1}{3}, 1]\right) - \mu\left([0, \tfrac{1}{2}] \cap [\tfrac{1}{3}, 1]\right)$$

$$= \tfrac{1}{2} + \tfrac{2}{3} - (\tfrac{1}{2} - \tfrac{1}{3}) = 1$$

Question : How to define probability on the real line $\mathbb{R}$?

Let us consider a p-space (Ω, Σ, P).

We want to define a p-space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$.

$\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$, which is the σ-algebra generated by all open intervals of $\mathbb{R}$ (or equivalently all closed or all semi-open intervals of $\mathbb{R}$).

For instance the sets $\{(-\infty, y_0], \; y_0 \in \mathbb{R}\}$ generate $\mathcal{B}(\mathbb{R})$.

We now consider a function $y : \Omega \to \mathbb{R}$ (domain, range) from the measurable space (Ω, Σ) to the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

The function y is said to be measurable from (Ω, Σ) to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ if

$$\forall \; A \in \mathcal{B}(\mathbb{R}) : \quad y^{-1}(A) \equiv \{\omega \mid y(\omega) \in A\} \in \Sigma$$

It is also called a Borel function.

We can now define a probability measure μ, denoted by P_y as:

$$\forall \; A \in \mathcal{B}(\mathbb{R}) : \quad P_y(A) = P\left(y^{-1}(A)\right) \equiv P(\omega \in \Omega : y(\omega) \in A)$$

The probability measure P_y is called the probability induced by y, or the distribution of y. P_y is also called a distribution measure.

The function y is called a real-valued random variable.

Remark The main assumption is that $\forall \; A \in \mathcal{B}(\mathbb{R}) : \; y^{-1}(A) \in \Sigma$. Because, otherwise, $P(y^{-1}(A))$ may not be well-defined.

In other words, y is a random variable iff $y^{-1}((-\infty, y_0]) \in \Sigma, \; \forall \; y_0 \in \mathbb{R}$.

Remark We can replace (Ω, Σ, P) with the p-space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_y)$.

An example of a discrete random variable:

Ex. Two-coin toss with $\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$

Let $t_i = \begin{cases} 1 & \text{if H in i-th toss} \\ 2 & \text{if T in i-th toss} \end{cases}$ for $i = 1, 2$

one random variable is for instance $y = t_1 + t_2$

ω	$y(\omega)$
(H, H)	1 + 1 = 2
(H, T)	1 + 2 = 3
(T, H)	2 + 1 = 3
(T, T)	2 + 2 = 4

Another random variable is for instance y = the no. of heads

	ω	$y(\omega)$	
$P = \frac{1}{4}$ ←	(H, H)	2	→ $P_y = \frac{1}{4}$
$P = \frac{1}{4}$ ←	(H, T)	1	→ $P_y = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
$P = \frac{1}{4}$ ←	(T, H)	1	
$P = \frac{1}{4}$ ←	(T, T)	0	→ $P_y = \frac{1}{2}$

In the latter case, let (Ω, Σ, P) be the original p-space.

we can define the p-space for the random variable to be:

$(\{0, 1, 2\}, \text{All subsets of } \{0,1,2\}, P_y)$

where P_y can be defined in terms of P, as shown in green in the second table above.

(3) Cumulative Distribution Function (CDF)

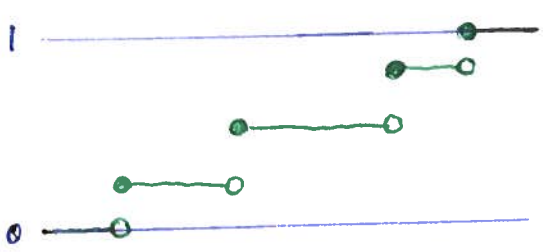
CDF of a random variable $y \in \mathbb{R}$ defined on (Ω, Σ, P), or equivalently on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_y)$, is defined by:

$$\forall y_0 \in \mathbb{R}: \quad F_y(y_0) := P\left(y^{-1}((-\infty, y_0])\right) = P(\{\omega \in \Omega : y(\omega) \le y_0\}) = P_y(y \le y_0)$$

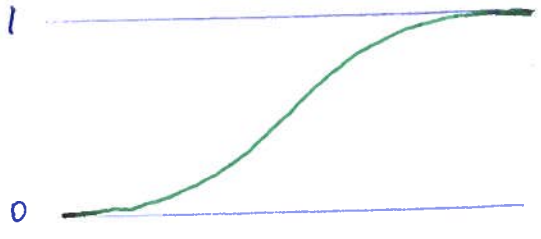
Properties of CDF:

- CDF is non-decreasing with range $[0,1]$
- $\lim_{y_0 \to -\infty} F(y_0) = 0$
- $\lim_{y_0 \to \infty} F(y_0) = 1$

Remark If CDF of a real-valued random variable y is continuous, then y is a continuous random variable.



CDF of a discrete r.v.



CDF of a continuous r.v.

④ Probability Density Function (PDF)

- If the CDF F_y of a real-valued random variable y is absolutely continuous, then there exists a Lebesgue-integrable function $\pi = \pi(y)$ called PDF of y, such that

$$\underbrace{F_y(b) - F_y(a)}_{P_y(a < y \leq b)} = \int_a^b \pi(y_0)\, dy_0 \quad , \qquad a \leq b$$

- The function π is equal to the derivative of F_y almost everywhere.

$$\pi(y_0) = F_y'(y_0) \qquad \text{for almost every } y_0 \in \mathbb{R}$$

- we also have: $\pi(y) \geq 0$

$$\int_{-\infty}^{\infty} \pi(y)\, dy = 1$$

Remark A function $f: [a,b] \to \mathbb{R}$ is <u>absolutely continuous</u> on $[a,b]$ if

$$\begin{cases} f \text{ has a derivative } f' \text{ almost everywhere} \\ \text{and} \\ f(x) = f(a) + \int_a^x f'(\tau)\, d\tau \qquad \forall x \in [a,b] \end{cases}$$

- Absolute continuity is stronger than continuity and weaker than Lipschitz continuity:

$$f \text{ Lipschitz continuous} \Rightarrow f \text{ absolutely continuous} \Rightarrow f \text{ continuous}$$

<u>Exercise</u>. Verify that $f(x) = \begin{cases} 0 & x = 0 \\ x \sin\left(\frac{1}{x}\right) & x \neq 0 \end{cases}$ is continuous, but not abs. continuous.

(5) Expection and higher moments of a random variable

- Expected value of a random variable $y : \Omega \to \mathbb{R}$ is defined as an integral with respect to the underlying probability measure.

Let (Ω, Σ, P) be the original P-space.

Let y be a real-valued random variable, which is a measurable function from (Ω, Σ) into a measurable space $(\Gamma, \mathcal{B}(\Gamma))$ where $\Gamma \subseteq \mathbb{R}$. The ex

the expected value of y is defined as the integral of y w.r.t. the P-measure P (a Lebesgue integral):

$$\boxed{\mathbb{E}[y] = \int_{\Omega} y(\omega)\, dP(\omega)}$$

Note that the integrand is the random variable, which is viewed as a function on the sample space Ω taking values in $(\Gamma, \mathcal{B}(\Gamma))$

Let P_y be the probability measure on $(\Gamma, \mathcal{B}(\Gamma))$ given by

$P_y(A) = P(\omega \in \Omega : y(\omega) \in A)$, $\forall\ A \in \mathcal{B}(\Gamma)$. In other words, we consider the P-space $(\Gamma, \mathcal{B}(\Gamma), P_y)$.

We can now use the change of variables and write:

$$\boxed{\mathbb{E}[y] = \int_{\Omega} y(\omega)\, dP(\omega) = \int_{\Gamma} y_0\, dP_y(y_0)}$$

Suppose now that P_y is absolutely continuous and hence y has a PDF $\pi = \pi(y)$ $\left(\pi(y_0) = F_y'(y_0) \text{ for almost every } y_0 \in \Gamma\right)$. Then

$$\boxed{\mathbb{E}[y] = \int_{\Omega} y(\omega)\, dP(\omega) = \int_{\Gamma} y_0\, dP_y(y_0) = \int_{\Gamma} y_0\, \pi(y_0)\, dy_0}$$

where the last integral is an ordinary Riemann integral of calculus.

- Some properties of expectation:

1) $y \geq 0$ a.s. $\Rightarrow$ $\mathbb{E}[y] \geq 0$

2) $y_1 \leq y_2$ a.s. $\Rightarrow$ $\mathbb{E}[y_1] \leq \mathbb{E}[y_2]$

3) $|\mathbb{E}[y]| \leq \mathbb{E}[|y|]$

4) linearity: $\mathbb{E}[a y_1 + b y_2] = a\,\mathbb{E}[y_1] + b\,\mathbb{E}[y_2]$

Notation. almost surely (a.s.)

Ex. $y \geq 0$ a.s. $\equiv$ $P(\{\omega \in \Omega : y(\omega) < 0\}) = 0$

- Higher moments

Expection is the 1st moment.

Moments of order $r \geq 1$ are defined as:

$$m_r(y) = \mathbb{E}[y^r] = \int_\Gamma y_o^r \, \pi(y_o)\, dy_o$$

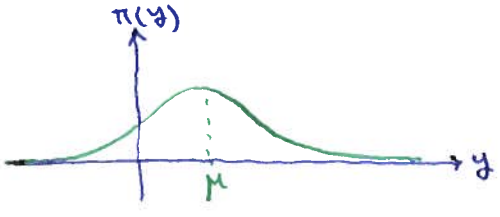
where $y : \Omega \to \Gamma$

For the r-th moment to exist, we need the integral in the right hand side to exist.

Ex. A Gaussian (or normal) random variable

$y \sim N(\mu, \sigma^2)$ where $\begin{cases} \mu = \mathbb{E}[y] & \text{mean} \\ \sigma^2 = \mathbb{E}[(y-\mu)^2] & \text{variance = 2nd central moment} \end{cases}$

$$\pi(y) = \pi(y; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$



(6) Random functions

Let $y \in \Gamma \subset \mathbb{R}$ be a real-valued random variable

The random function $g = g(y)$ where $g : \Gamma \to \mathbb{R}$ is defined by the mapping

$$(\Omega, \Sigma) \xrightarrow{y} (\Gamma, \mathcal{B}(\Gamma)) \xrightarrow{g} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

If y is measurable from (Ω, Σ) to $(\Gamma, \mathcal{B}(\Gamma))$ and if g is measurable from $(\Gamma, \mathcal{B}(\Gamma))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ $\Rightarrow g(y(\omega)) : (\Omega, \Sigma) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable

Hence, g is a real-valued random variable.

- Expectation of a random function $g = g(y) : \Gamma \to \mathbb{R}$

$$\mathbb{E}[g(y)] = \int_\Gamma g(y) \cdot \pi(y)\, dy$$

- Moments of a random function:

$$m_r(g(y)) = \mathbb{E}[g(y)^r] = \int_\Gamma g(y)^r \pi(y)\, dy$$

For the moments to exist, we need the right hand side integral to exist.

To make this more precise, we define L^r_π-norm as follows:

$$\|g\|_{L^r_\pi(\Gamma)} := \left(\int_\Gamma |g(y)|^r \pi(y)\, dy \right)^{1/r} = \left(\mathbb{E}[|g(y)|^r] \right)^{1/r}$$

We then define the L^r_π-space of random functions:

$$L^r_\pi(\Gamma) := \{ g : \Gamma \to \mathbb{R} \mid \|g\|_{L^r_\pi(\Gamma)} < \infty \}$$

Therefore, for the r-th moment $m_r(g(y))$ to exist, we need $g \in L^r_\pi(\Gamma)$.

Note that if $\int |g| \pi \, dy < \infty$, then $\int g \pi \, dy < \infty$.

Ex. $L^2_\pi(\Gamma)$ is the space of random functions with finite second moment, or equivalently the space of square integrable functions w.r.t. $\pi(y)\,dy$, or equivalently the space of π-square integrable functions.

(7) Random vectors

$Y = [y_1, y_2, \cdots, y_N] \in \Gamma \subseteq \mathbb{R}^N$ is called a random vector if $\{y_n\}_{n=1}^N$ are random variables (measurable from (Ω, Σ) to $(\Gamma, \mathcal{B}(\Gamma))$).

- Distribution measure P_Y of a random vector Y can be defined similar to that of a random variable.

  Consider two p-spaces (Ω, Σ, P) and $(\Gamma, \mathcal{B}(\Gamma), P_Y)$.

  Let $Y: (\Omega, \Sigma) \longrightarrow (\Gamma, \mathcal{B}(\Gamma))$ be a random N-vector. Its distribution measure (or induced probability measure) is defined as:

$$P_Y(A) = P(Y \in A) = P(Y^{-1}(A)) \quad \text{for } A \in \mathcal{B}(\Gamma)$$

- **Joint distribution function (Joint CDF)**

$$F_Y(Y_0) = P\left(\bigcap_{n=1}^{N} \{ y_n \le y_{0n} \} \right) \qquad \begin{cases} Y_0 = [y_{01}, \cdots, y_{0N}] \subset \mathbb{R}^N \\ Y = [y_1, \cdots, y_N] \subset \mathbb{R}^N \end{cases}$$

or equivalently, $F_Y(Y_0) = P(y_1 \le y_{01}, \cdots, y_N \le y_{0N})$

- $y_n \longrightarrow F_Y(Y)$ is non-decreasing, $n = 1, \cdots, N$

- $\lim_{y_{0n} \to -\infty} F_Y(Y_0) = 0$, $n = 1, 2, \cdots, N$

- $\lim_{y_{0n} \to \infty} F_Y(Y_0) = F_{\hat{Y}_n}(\hat{Y}_{0n})$ where $\begin{cases} \hat{Y}_{0n} = [y_{01}, \cdots, y_{0,n-1}, y_{0,n+1}, \cdots, y_{0N}] \subset \mathbb{R}^{N-1} \\ \hat{Y}_n = [y_1, \cdots, y_{n-1}, y_{n+1}, \cdots, y_N] \subset \mathbb{R}^{N-1} \end{cases}$

- **Joint density function (Joint PDF)**

$$\pi(Y_0) = \frac{\partial^N F_Y(Y_0)}{\partial y_1 \cdots \partial y_N}$$

- $\int_\Gamma \pi(Y)\, dY = 1$ Note: $dY = dy_1\, dy_2 \cdots dy_N$

This integral is a multiple integral. If $\Gamma = \Gamma_1 \times \Gamma_2 \times \cdots \times \Gamma_N$, then $\int_\Gamma \pi(Y)\, dY = \int_{\Gamma_1} \int_{\Gamma_2} \cdots \int_{\Gamma_N} \pi(y_1, \cdots, y_N)\, dy_1 \cdots dy_N$.

Remark We can define a random vector $Y = [y_1, \cdots, y_N] \in \Gamma \subset \mathbb{R}^N$ as a collection of $N \ge 2$ random variables $\{y_n\}_{n=1}^N$ defined on a prob. space $(\Gamma, \mathcal{B}(\Gamma), \pi(Y))$ where π is the joint PDF of Y.

- **Marginalization**

It is an important technique to find the joint density of some specific coordinates of $Y \subset \mathbb{R}^N$.

Ex. To find joint PDF of $\hat{Y}_n = [y_1, \cdots, y_{n-1}, y_{n+1}, \cdots, y_N] \subset \mathbb{R}^{N-1}$, we marginalize $\pi(Y)$ over y_n :

$$\pi(\hat{Y}_n) = \int_{\Gamma_n} \pi(Y)\, dy_n$$

This is also called y_n - marginalized density.

Ex. To find the marginal density of $y_n \in \Gamma_n \subset \mathbb{R}$:

$$\pi(y_n) = \int_{\Gamma_1 \times \cdots \times \Gamma_{n-1} \times \Gamma_{n+1} \times \cdots \times \Gamma_N} \pi(Y)\, d\hat{Y}_n$$

- **Independence of random variables**

N random variables are independent if and only if (iff)

$$\pi(Y) = \prod_{n=1}^{N} \pi(y_n)$$

i.e., Joint density is the product of marginal densities.

This implies a factorization of joint density.

mean vector:

$$\bar{m} = \mathbb{E}[Y] = \left[\mathbb{E}[y_1], \cdots, \mathbb{E}[y_N]\right]^T$$

Covariance

- Covariance of two random variables is a measure of how much two random variables change together.
- Covariance function of two random variables y_n and y_m is:

$$\text{COV}(y_n, y_m) = \mathbb{E}\left[(y_n - \mathbb{E}[y_n])(y_m - \mathbb{E}[y_m])\right] = \mathbb{E}[y_n y_m] - \mathbb{E}[y_n]\mathbb{E}[y_m].$$

- If $\text{COV}(y_n, y_m) > 0 \Rightarrow$ y_n & y_m have similar behavior
  large (small) values of y_n correspond to large (small) values of y_m
- If $\text{COV}(y_n, y_m) < 0 \Rightarrow$ y_n & y_m have opposite behavior
  large (small) values of y_n correspond to small (large) values of y_m
- Covariance matrix for a random vector $Y = [y_1, \cdots, y_N]^T \in \mathbb{R}^N$:

$$\text{COV}(Y) = \left[\text{cov}(y_n, y_m)\right] = \mathbb{E}\left[(Y - \bar{m})(Y - \bar{m})^T\right] \in \mathbb{R}^{N \times N}$$

- properties of COV:
  1) symmetric: $\text{cov}(y_n, y_m) = \text{cov}(y_m, y_n)$
  2) semi-positive definite: $\forall\, \bar{\alpha} \in \mathbb{R}^N$: $\bar{\alpha}^T \text{COV}\, \bar{\alpha} \geq 0$

  covariance matrix invertible $\Leftrightarrow$ COV is positive definite

Variance { a measure of the dispersion of y_n around its mean }

$$\text{Var}(y_n) = \text{cov}(y_n, y_n) = \sigma_n^2 \geq 0$$

- $\text{Cov}(y_1, y_2) = 0 \iff y_1$ and y_2 are uncorrelated.
- If all N random variables in $Y = [y_1, \dots, y_N]$ are uncorrelated, then the covariance matrix is diagonal: $\text{COV} = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$.
- y_1 and y_2 independent $\Rightarrow \text{Cov}(y_1, y_2) = 0 \iff y_1$ and y_2 uncorrelated.
  $(\mathbb{E}[y_1 y_2] = \mathbb{E}[y_1]\,\mathbb{E}[y_2])$
- The converse is usually not true. See the example below.
- In general: independent $\Rightarrow$ uncorrelated
- For Gaussian random vectors: independence $\iff$ uncorrelation

---

Correlation

- correlation function: $\text{CORR}(y_n, y_m) = \dfrac{\text{COV}(y_n, y_m)}{\sigma_n \cdot \sigma_m}$ (a normalized version of covariance)
- Correlation matrix: $\text{CORR}(Y) = [\text{CORR}(y_n, y_m)]$

Ex. $X \sim \mathcal{U}[-1, 1]$

$$Y \sim \begin{cases} -X & \text{if } X \le 0 \\ X & \text{if } X \ge 0 \end{cases} \Rightarrow Y \sim \mathcal{U}[0, 1]$$

$$\mathbb{E}[XY] = \mathbb{E}[XY \mid X \le 0] + \mathbb{E}[XY \mid X > 0] = \int_{-1}^{0} -x^2 dx + \int_{0}^{1} x^2 dx = -\frac{1}{3} + \frac{1}{3} = 0$$

$$\mathbb{E}[X] = 0 \Rightarrow \mathbb{E}[X]\,\mathbb{E}[Y] = 0$$

$\Rightarrow \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ — X & Y uncorrelated

but we know that X & Y are dependent.

Ex $\begin{cases} y_1 \sim N(0,1) \\ y_2 = y_1^2 \end{cases} \Rightarrow \text{cov}(y_1, y_2) = \mathbb{E}[y_1 y_2] - \cancel{\mathbb{E}[y_1]}^{\,0}\,\mathbb{E}[y_2] = \mathbb{E}[y_1^3] = 0$

uncorrelated but dependent

Gaussian Vector

- A Gaussian vector $Y \in \mathbb{R}^N$ with mean $\bar{m} = \mathbb{E}[Y]$ and covariance matrix $C = \mathbb{E}[(Y-\bar{m})(Y-\bar{m})^T]$ has the joint density function:

$$\pi(Y) = \frac{1}{(2\pi)^{N/2}\sqrt{|C|}} \exp\left(-\frac{1}{2}(Y-\bar{m})^T C^{-1}(Y-\bar{m})\right)$$

where $|C|$ is the determinant of the covariance matrix. (C is assumed to be invertible).

We write $Y \sim N(\bar{m}, C)$.
We say N random variables $y_1, \ldots, y_N$ are jointly Gaussian.

- An important property: if C diagonal (uncorrelated r.v.'s) $\Rightarrow$ r.v.'s are mutually independent.
  In this case we can show $\pi(Y) = \prod_{n=1}^{N} \pi(y_n)$.

- If Y and X are two $\mathbb{R}^N$-valued Gaussian vectors, then $(\alpha Y + \beta X)$ is a Gaussian random vector.

- Y Gaussian $\equiv \{y_n\}_{n=1}^N$ jointly Gaussian $\Rightarrow \{y_n\}_{n=1}^N$ individually Gaussian

  ($\Rightarrow$ In general converse is not true.)

$\Rightarrow$ The assumption that Y is a Gaussian vector is stronger than having y_n's be Gaussian random variables.

- **Theorem** $Y \in \mathbb{R}^N$ Gaussian random vector $Y \sim N(\bar{m}, \Sigma)$
  Let $X \in \mathbb{R}^M$ and $X = \underbrace{A}_{\mathbb{R}^{M\times N}} Y \Rightarrow X \sim N(A\bar{m}, A\Sigma A^T)$

- In particular, we have: if $X \in \mathbb{R}^N$, $X \sim N(0, I)$ iid standard/unit r.v.'s, then $Y \in \mathbb{R}^N$, $Y \sim N(\bar{m}, \Sigma)$ is obtained by $Y = \bar{m} + AX$ where $AA^T = \Sigma$.

(8) Stochastic Processes

We are interested in the computation of the stochastic solutions of ODEs/PDEs.

These solutions are functions of space and/or time, in addition to ω.

The notion of random variables and random vectors needs to be extended to include dependence on space and/or time.

Let $D \subset \mathbb{R}^d$ be a compact d-dimensional spatial domain. (compact = {closed + bounded})

Let (Ω, Σ, P) be a probability space corresponding to a random experiment.

Consider a function $u = u(t, \bar{x}, \omega) : [0,T] \times D \times \Omega \longrightarrow \mathbb{R}$

- t: time
- $\bar{x}$: spatial coordinate
- ω: an outcome of a random experiment

For example u can be the solution to a time-dependent PDE with random coefficients.

If $u(t, \bar{x}, \cdot) : \Omega \to \mathbb{R}$ is a $\mathbb{R}$-valued random variable on (Ω, Σ, P) for all $t \in [0,T]$ and all $\bar{x} \in D \subset \mathbb{R}^d$, then $u : [0,T] \times D \times \Omega \to \mathbb{R}$ is called a stochastic process.

$\Rightarrow u$ is a collection of infinite random variables.

The function $u(\cdot, \cdot, \omega) : [0,T] \times D \to \mathbb{R}$ for given $\omega \in \Omega$ is called a realization of u (or a sample path).

To simplify notation, we will drop t and restrict ourself to stochastic processes $a(\bar{x}, \omega) : D \times \Omega \longrightarrow \mathbb{R}$.

Such processes are called spatial processes or random fields.

Random fields

Let $D \subset \mathbb{R}^d$ be a compact d-dimensional spatial domain.

Let (Ω, Σ, P) be a complete probability space.

Consider a random field $a = a(\bar{x}, \omega) : D \times \Omega \longrightarrow \mathbb{R}$.

This means that $\forall\ \bar{x} \in D$, $a(\bar{x}, \cdot) : \Omega \to \mathbb{R}$ is a random variable on (Ω, Σ, P).

We discretize the domain D into n grid points: $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$. Then, similar to random vectors, we can define distribution and density of random fields:

- finite dimensional distribution of order n: $F_n(\bar{z}; \bar{x}_1, \dots, \bar{x}_n) = P(a(\bar{x}_1, \omega) \le z_1, \dots, a(\bar{x}_n, \omega) \le z_n)$
- finite dimensional density of order n: $\pi_n(\bar{z}; \bar{x}_1, \dots, \bar{x}_n) = \partial_{\bar{z}}^n F_n(\bar{z}; \bar{x}_1, \dots, \bar{x}_n)$

where $\bar{z} = [z_1, \dots, z_n] \in \mathbb{R}^n$ and $\partial_{\bar{z}}^n := \dfrac{\partial^n}{\partial z_1 \partial z_2 \cdots \partial z_n}$

- Expected value: $\mu_a(\bar{x}) = \mathbb{E}[a(\bar{x}, \cdot)] : D \to \mathbb{R}$
- Covariance function: $COV_a(\bar{x}_1, \bar{x}_2) = \mathbb{E}[(a(\bar{x}_1, \cdot) - \mu_a(\bar{x}_1))(a(\bar{x}_2, \cdot) - \mu_a(\bar{x}_2))]$. $: D \times D \to \mathbb{R}$
- variance: $VAR_a(\bar{x}) = COV_a(\bar{x}, \bar{x}) : D \to \mathbb{R}$

- $a(\bar{x}, \omega)$ is said to be a second-order random field if $\forall\ \bar{x} \in D, \mathbb{E}[(a(\bar{x}, \cdot))^2] < \infty$.
- $a(\bar{x}, \omega)$ is said to be a stationary field if its F_n is invariant under arbitrary translation (or shift) $\bar{\xi} \in \mathbb{R}^d$ $(\bar{x} \longrightarrow \bar{x} + \bar{\xi})$, or equivalently if its law is invariant under translation: $a(\bar{x}, \omega) \sim a(\bar{x} + \bar{\xi}, \omega)\ \forall \bar{\xi} \in \mathbb{R}^d$.
- As a results, all moments of a stationary field are invariant under translation.
- Stationary fields are also called strictly stationy.

- A random field is said to be weakly stationary if

  1) μ_a is constant

  2) $Cov_a(\bar{X}_1, \bar{X}_2) = Cov_a(\bar{x}_1 - \bar{x}_2) \Rightarrow Var_a = \text{constant}$

- Weakly stationarity is weaker than stationarity, since it imposes conditions only on the 1st and 2nd moments of the random field.
- A weakly stationary random field is said isotropic if the covariance function depends only on $\| \bar{x}_1 - \bar{x}_2 \|$.

a stationary $\xrightarrow{Var_a < \infty}$ a weakly stationary

a stationary $\xrightarrow[\text{a Gaussian}]{}$ a weakly stationary

properties of covariance function

For a second order random field $\left(\forall \bar{x} \in D, \mathbb{E}\left[|a(\bar{x}, \cdot)|^2\right] < \infty\right)$ the covariance function is:

1) bounded: $Cov_a(\bar{x}_1, \bar{x}_2) \leq \sqrt{var_a(\bar{x}_1)} \sqrt{var_a(\bar{x}_2)}$

2) symmetric: $Cov_a(\bar{x}_1, \bar{x}_2) = Cov_a(\bar{x}_2, \bar{x}_1)$

3) semi-positive definite: $\forall \bar{\xi}_1, \dots, \bar{\xi}_n \in D$ and $\bar{\alpha} \in \mathbb{R}^N$: $\bar{\alpha}^T \left[Cov_a(\bar{\xi}_i, \bar{\xi}_j)\right] \bar{\alpha} \geq 0$

Gaussian Random Fields

$a = a(\bar{x}, \omega) : D \times \Omega \to \mathbb{R}$ is a Gaussian field if $\forall \bar{x}_1, \dots, \bar{x}_n \in D$, the random vector $Y(\omega) = \left(a(\bar{x}_1, \omega), \dots, a(\bar{x}_n, \omega)\right) \in \mathbb{R}^n$ is Gaussian, that is all finite dimensional densities Π_n are Gaussian for every choice of $\bar{x}_1, \dots, \bar{x}_n$.

$\Rightarrow$ $\boxed{\forall \bar{x} \in D : \; a(\bar{x}, \cdot) \sim N\left(\mu_a(\bar{x}), cov_a(\bar{x}, \bar{x})\right)}$

For example:

$cov_a(\bar{x}_1, \bar{x}_2) = \sigma^2 e^{-\frac{\| \bar{x}_1 - \bar{x}_2 \|}{L_c}}$ exponential cov. model

$cov_a(\bar{x}_1, \bar{x}_2) = \sigma^2 e^{-\frac{\| \bar{x}_1 - \bar{x}_2 \|^2}{2 L_c^2}}$ squared exponential cov. model

where L_c is the correlation length of the field a.