Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: Data Scientist.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

Probabilities and Statistics Review

Definitions

σ-Algebra:

We say a collection ζ of subsets of Ω is a σ-Algebra if:

1. $\emptyset \in \zeta$
2. If $E \in \zeta \Rightarrow E^C \in \zeta$
3. If $\{E_i\}_{i\in\mathbb{N}} \subseteq \zeta$ then $\bigcup_{i\in\mathbb{N}} E_i \in \zeta$

Example: 2^Ω, $\{\emptyset, \Omega\}$.

Note: All $\sigma-$algebras are algebras.

Observation: If $\{F_i\}_{i\in I}$ are σ-algebras on Ω:

$$\bigcap_{i\in I} F_i \quad \text{are a } \sigma\text{-algebra too}$$

If C is a arbitrary collection on Ω we define:

$$\sigma(C) := \bigcap_{C\subseteq\zeta} \zeta \quad \text{where } \zeta \text{ are } \sigma\text{-algebras}$$

$\sigma(C)$ is the smallest σ-algebra which contains C.

The **borelians** are σ(open sets).

Algebra:

We say a collection $\mathcal{A}$ of subsets of Ω is an Algebra if:

1. $\emptyset \in \mathcal{A}$
2. If $E \in \mathcal{A} \Rightarrow E^C \in \mathcal{A}$
3. If $E_1, E_2, ..., E_n \in \mathcal{A}$ then $\bigcup_{i=1}^{n} E_i \in \mathcal{A}$

Example: The set of all finite disjoint unions of intervarls.

Set Function:

A set function is a function whose domain is a family of subsets of some given set (Ω for example) and that takes its values in the extended real number line $\mathbb{R} \cup \{\pm\infty\}$.

Pre-Measure:

Let's consider $\Omega \neq \emptyset$ and $\mathcal{A}$ an algebra on it. We say $\lambda : \mathcal{A} \to [0, +\infty]$ is a pre-measure if:

1. $\lambda(\emptyset) = 0$
2. If $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{A}$ is a collection of disjoint sets and if their union is contained in $\mathcal{A}$ then:

$$\lambda\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \lambda(A_i)$$

Measure:

Let's consider $\Omega \neq \emptyset$ and s ζ a σ-algebra on it. We say $\mu : \zeta \to [0, +\infty]$ is a measure if:

1. $\mu(\emptyset) = 0$
2. If $\{E_i\}_{i\in\mathbb{N}} \subseteq \zeta$ is a collection of disjoint set then:

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$$

Examples :

- Counting Measure:

$$\mu(A) = \begin{cases} |A| & if \quad A \text{ is countable} \\ +\infty & else \end{cases}$$

- If we consider any function $f : X \to [0, \infty)$ we can defines a measure μ on (Ω, ζ) vía:

$$\mu(A) := \sum_{a \in A} f(a)$$

- Let B be a set on ζ and m other measure on (Ω, ζ) such that $0 < m(B) < \infty$:

$$\mu(A) = \frac{m(A \cap B)}{m(B)}$$

Measurable Space:

Consider a set Ω and a σ-algebra ζ on Ω. Then the tuple (Ω, ζ) is called a measurable space.

Note that in contrast to a **measure space**, no measure is needed for a measurable space.

Note: In probability theory we call (Ω, ζ) **event space**.

Measure Space:

A measure space is a triple (Ω, ζ, μ), where:

1. Ω is a set.
2. ζ is a σ-algebra on the set Ω.
3. μ is a measure on (Ω, ζ)

In other words, a measure space consists of a **measurable space** (Ω, ζ) together with a measure on it.

Probability Measure:

Let's consider (Ω, ζ) an event space. We say $\mathbb{P} : \zeta \to [0,1]$ is a probability measure if:

1. $\mathbb{P}(\Omega) = 1$
2. If $\{E_i\}_{i\in\mathbb{N}} \subseteq \zeta$ is a collection of disjoint set then:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$$

We call $(\Omega, \zeta, \mathbb{P})$ **probability space**.

Notes:

a From this definition we can deduce $\mathbb{P}(\emptyset) = 0$.

$E = \Omega \cup \emptyset$. It's clear $\Omega \cap \emptyset = \emptyset$. By (2): $\mathbb{P}(E) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset)$. It's clear $E = \Omega$, then: $\mathbb{P}(\Omega) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset)$. How $\mathbb{P}(\Omega)$ is finite (in particular it's value is 1) then: $\mathbb{P}(\emptyset) = 0$.

b If instead of the σ-algebra ζ we have the algebra $\mathcal{A}$ and $\mathbb{P}$ satisfies (2) as long as $\bigsqcup_{i=1}^{\infty} E_i \in \mathcal{A}$, we say $\mathbb{P}$ is a probability measure in $\mathcal{A}$.

c If instead of the σ-algebra ζ we have the algebra $\mathcal{A}$ and $\mathbb{P}$ satisfies $(2')$ instead (2):

$$2'.\ \ \mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) \ \text{ if } \ E_1, E_2 \in \mathcal{A} \ \text{ and } \ E_1 \cap E_2 = \emptyset$$

We say $\mathbb{P}$ is a probability measure **finitely additive**.

Example:

a Let be (Ω, ζ, μ) a measure space where Ω its composed by a finite number of elements and where μ is the **counting measure** then we can define the next **probabily measure**:

$$\mathbb{P}(A) = \frac{\mu(A \cap \Omega)}{\mu(\Omega)} = \frac{\mu(A)}{\mu(\Omega)}$$

We usually use this probability measure when we work with **discrete random variables**.

Random Variable:

A random variable is a mathematical formalization of a quantity or object which depends on random events. It is a mapping or a function from possible outcomes in a sample space to a measurable space, often the real numbers.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a **probability space** and $(E, \mathcal{E})$ a **measurable space**. Then an $(E, \mathcal{E})$ - **value random variable** is a measurable function $X : \Omega \to E$, which means that, for every subset $B \in \mathcal{E}$, its preimage is $\mathcal{F}$-measurable i.e.

$$\forall B \in \mathcal{E},\ X^{-1}(B) \in \mathcal{F} \text{ where } X^{-1}(B) = \{w \in \Omega : X(w) \in B\}$$

The probability that X takes on a value in a measurable set $S \in E$ is written as:

$$\mathbb{P}(X \in S) = \mathbb{P}(\{w \in \Omega | X(w) \in S\})$$

If E is countable, then X is called a **discrete random variable**.

Example:

1. $\Omega = \{head, tail\}$

   $$X(w) = \begin{cases} 1 & if \quad w = head \\ 0 & if \quad w = tail \end{cases}$$

   Note $E = \{0, 1\}$.

   $$\mathbb{P}(X \in \{1\}) = \mathbb{P}(\{w \in \Omega : X(w) \in \{1\}\}) = \mathbb{P}(\{w \in \Omega : X(w) = 1\}) = \mathbb{P}(\{head\})$$

   If we use the probability measure we **introduced previously** then we have:

   $$\mathbb{P}(\{\text{head}\}) = \frac{\mu(\{head\})}{\mu(\{head, tail\})} = \frac{1}{2}$$

2. $\Omega = \{Alberto, Gustavo, Franco\}$

   $$H(w) = \begin{cases} 1{,}77 & if \quad w = Alberto, Franco \\ 1{,}79 & if \quad w = Gustavo \end{cases}$$

   Note $E = \{1{,}77, 1{,}79\}$.

$\mathbb{P}(H \in \{1{,}77\}) = \mathbb{P}(\{w \in \Omega : H(w) \in \{1{,}77\}\}) = \mathbb{P}(\{w \in \Omega : H(w) = 1{,}77\}) = \mathbb{P}(\{Alberto, Franco\})$

If we use the probability measure we **introduced previously** then we have:

$$\mathbb{P}(\{\text{Alberto, Franco}\}) = \frac{\mu(\{Alberto, Franco\})}{\mu(\{Alberto, Franco, Gustavo\})} = \frac{2}{3}$$

Distribution Functions:

If a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given, we can ask questions like "How likely is it that value of X is equal to 2?". This is the as the probability of the event $\{w \in \Omega : X(w) = 2\}$ which is often written as $\mathbb{P}(X = 2)$ or $p_X(2)$ for short.

If X is real-valued, we can always captured its cumulative distribution function:

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{w \in \Omega : X(w) \leq x\})$$

Definition: A distribution function is a function $F : \mathbb{R} \to \mathbb{R}$ such that:

(i) F is monotonous increasing (If $x \leq y \;\Rightarrow\; F(x) \leq F(y)$).

(ii) F is continuous by the right ($lim_{x \to a^+} F(x) = F(a)$).

(iii) $lim_{x \to \infty} F(x) = 1$ and $lim_{x \to -\infty} F(x) = 0$.

Propositions:

a) Let be $\mathbb{P}(\cdot)$ a probability measure on $(\mathbb{R}, B(\mathbb{R}))$ then:

$$F(x) = \mathbb{P}((-\infty, x])$$

b) Let be F a distribution function then exists an unique $\mathbb{P}(\cdot)$ on $(\mathbb{R}, B(\mathbb{R}))$ such that:

$$F(x) = \mathbb{P}((-\infty, x])$$

Density function:

Let be X a random variable. If exists a measurable function $f_X : \mathbb{R} \to \mathbb{R}_+$ such that:

$$\forall B \in B(\mathbb{R}) \qquad \int_B f_X(x)dx = \mathbb{P}(X \in B) = P_X(B)$$

Then we say that the random variable and its distribution have the density function f_X.

Note:

i) If F is differentiable then:

$$f_X(x) = \frac{dF(x)}{dx}$$

ii) If X is a discrete random variable $\Leftrightarrow \ \exists\, S \subseteq \mathbb{R}$ countable such that:

$$\sum_{x \in S} \mathbb{P}(X = x) = 1$$

Expectation:

Let be X a random variable over $(\Omega, \zeta, \mathbb{P})$.

a) The expectation of X is given by:

$$\boxed{\mathbb{E}(X) = \int X d\mathbb{P} = \int X dF}$$

if $X_+ \in L^1(\Omega, \zeta, \mathbb{P})$ or $X_- \in L^1(\Omega, \zeta, \mathbb{P})$.

b) For all $k \geq 1, k \in \mathbb{N}$ the moment of k order is given by:

$$\boxed{\mathbb{E}(X^k)}$$

if $X \in L^k(\Omega, \zeta, \mathbb{P})$.

c) If $X \in L^2(\Omega, \zeta, \mathbb{P})$ then we define its variance:

$$\boxed{\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2}$$

Note: $\mathbb{V}[X] = \mathbb{E}\left([X - \mathbb{E}(X)]^2\right)$

Jensen inequality:

Let be $(\Omega, \zeta, \mathbb{P})$ a probability space, X integrable random variable and $X \in L^1(\Omega, \zeta, \mathbb{P})$.

Then for every convex function $\phi(\cdot)$ defined in the rangeof X we have that:

$$\boxed{\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]}$$

Chebyshev inequality:

Let be X a random variable on a probability space $(\Omega, \zeta, \mathbb{P})$. Let be f an increasing and positive function, then:

$$\boxed{\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[f(X)]}{f(a)}}$$

Moment-generating function:

$$\boxed{M_X(t) = \mathbb{E}[e^{t \cdot X}]}$$

You can see:

$$M'_X(t) = \mathbb{E}[X \cdot e^{t \cdot X}] \quad \Rightarrow \quad M'_X(0) = \mathbb{E}[X]$$

$$M''_X(t) = \mathbb{E}[X^2 \cdot e^{t \cdot X}] \quad \Rightarrow \quad M''_X(0) = \mathbb{E}[X^2]$$

$$M^{(n)}_X(t) = \mathbb{E}[X^n \cdot e^{t \cdot X}] \quad \Rightarrow \quad M^{(n)}_X(0) = \mathbb{E}[X^n]$$

Independent random variable:

$$X_1, ..., X_n \text{ are independents} \Leftrightarrow \forall A_1, ..., A_n \in B(\mathbb{R})$$

$$\mathbb{P}(X_1 \in A_1 \wedge ... \wedge X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdot ... \cdot \mathbb{P}(X_n \in A_n)$$

Theorem: Let be $(\Omega, \zeta, \mathbb{P})$ a probability space. Let be X_1, X_2 independent random variables, then:

a) If $X_1, X_2 \in L^1(\Omega, \zeta, \mathbb{P})$, then

$$\boxed{\mathbb{E}[X_1 \cdot X_2] = \mathbb{E}[X_1] \cdot \mathbb{E}[X_2]}$$

b) If $f, g : \mathbb{R} \to \mathbb{R}$ measurable and $f(X_1), g(X_2) \in L^1(\Omega, \zeta, \mathbb{P})$ then:

$$\boxed{\mathbb{E}[f(X_1) \cdot g(X_2)] = \mathbb{E}[f(X_1)] \cdot \mathbb{E}[g(X_2)]}$$

Definition: Let be $X, Y, X \cdot Y \in L^1(\Omega, \zeta, \mathbb{P})$. The covariance between X and Y is defined by:

$$\boxed{Cov(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\right] = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]}$$

We can also define the **correlation** between two random variables:

$$\boxed{corr(X, Y) = \frac{\mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]}{\sqrt{\mathbb{V}[X]} \cdot \sqrt{\mathbb{V}[Y]}}}$$

<u>Observation or realization:</u>

Observation, realization or observed value of a random variable is the value that is actually observed (what actually happened). The random variable itself is the process dictating how the observation comes about.

$$x = X(w)$$

When we have an observation X_i from a random variable $X \sim D(u, \sigma)$ then $X_i \overset{\text{iid}}{\sim} D(\mu, \sigma)$ that is **every observation have the same distribution of the random variable.**

We can also say every observation is also a random variable.

Bayes' Theorem:

In probability theory Bayes' theorem (alternatively Bayes' law or Bayes' rule) describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

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

Where A and B are events ($A, B \subseteq \Omega$) and $\mathbb{P}(B) \neq 0$.

Applications:

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

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

Law of total probability:

Discrete case:

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$$

Continuous case:

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A|X = x) \cdot f_X(x)dx$$

Time series:

In time series when we talk about $\{X_t\}_{t=1,\ldots,T}$ these **aren't observations**, but **random variables** and for each random variable generally **we only have** 1 **observation**.

For example let's consider the autoregressive model:

$$X_t = \phi \cdot X_{t-1} + \epsilon_t$$

where $\epsilon_t \sim N(0, \sigma^2)$.

We can see clearly that $\{\epsilon_t\}_{t=1,\ldots,T}$ are different random variables too.

Notes:

- Despite $\epsilon_1, , ..., \epsilon_T$ are different random variables and not observations, similarly we use it to estimate σ^2 because all these random variables have the **same distribution** and also are **independent**.
- Analogously, when we have a stationary time series we use for example the observations $X_1^1, X_2^1, X_3^1, X_4^1, X_5^1$ from the random variables X_1, X_2, X_3, X_4, X_5 to estimate $corr(X_{t+1}, X_t)$ because:

  $$corr(X_5, X_4) = corr(X_4, X_3) = corr(X_3, X_2) = corr(X_2, X_1)$$

  By the way, we estimate $corr(X_{t+1}, X_t)$ in the following way:

  $$corr\left(\begin{bmatrix} X_2^1 \\ X_3^1 \\ X_4^1 \\ X_5^1 \end{bmatrix} \begin{bmatrix} X_1^1 \\ X_2^1 \\ X_3^1 \\ X_4^1 \end{bmatrix}\right)$$

Properties of expection and covariance

Consider $X_1, X_2, ..., X_n,\ Y_1, Y_2, ..., Y_n$ random variables and $\alpha_1, ..., \alpha_n, \beta_1, ..., \beta_n, c, c_1, c_2 \in \mathbb{R}$ then:

$$\boxed{\mathbb{E}\left[\sum_{i=1}^{n} \alpha_i \cdot X_i + c\right] = \sum_{i=1}^{n} \alpha_i \cdot \mathbb{E}\left[X_i\right] + c} \quad (1)$$

$$\boxed{Cov\left(\sum_{i=1}^{n} \alpha_i \cdot X_i + c_1, \sum_{j=1}^{n} \beta_j \cdot Y_j + c_2\right) = Cov\left(\sum_{i=1}^{n} \alpha_i \cdot X_i, \sum_{j=1}^{n} \beta_j \cdot Y_j\right)} \quad (2)$$

$$\boxed{Cov\left(\sum_{i=1}^{n} \alpha_i \cdot X_i, \sum_{j=1}^{n} \beta_j \cdot Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \cdot \beta_j \cdot Cov\left(X_i, Y_j\right)} \quad (3)$$

$$\boxed{\mathbb{V}\left(\sum_{i=1}^{n}\alpha_i \cdot X_i\right) = Cov\left(\sum_{i=1}^{n}\alpha_i \cdot X_i, \sum_{j=1}^{n}\alpha_j \cdot X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i \cdot \alpha_j \cdot Cov(X_i, X_j)} \quad (4)$$

If $X_1, X_2, ..., X_n$ are independents:

$$\mathbb{V}\left(\sum_{i=1}^{n}\alpha_i \cdot X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i \cdot \alpha_j \cdot Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1:j\neq i}^{n}\alpha_i \cdot \alpha_j \cdot Cov(X_i, X_j) + \sum_{i=1}^{n}(\alpha_i)^2 \cdot Cov(X_i, X_i)$$

$$= \sum_{i=1}^{n}(\alpha_i)^2 \cdot Cov(X_i, X_i) = \sum_{i=1}^{n}(\alpha_i)^2 \cdot \mathbb{V}(X_i)$$

$$\boxed{\mathbb{V}\left(\sum_{i=1}^{n}\alpha_i \cdot X_i\right) = \sum_{i=1}^{n}(\alpha_i)^2 \cdot \mathbb{V}(X_i)} \quad (5)$$
