A crash course in probability Theory

Once developed all the theory concerning measure theory and the lebesgue integral, we can go back to "redefine" the first principles in probability theory:

Def/ A measure space $(X, \mathcal{X}, \mu)$ where $\mu(X) = 1$ is called probability space.

This last condition tells us that $\mu(E) \leq 1$ for all $E \in \mathcal{X}$. In this context, $\mathcal{X}$ is called the set of events, and each $E \in \mathcal{X}$ is called an event. The measure is called probability function, it is usually written with P instead of μ. We usually write Ω instead of X, and $\mathcal{A}$ instead of $\mathcal{X}$.

Ex/ Take Ω finite set, $\mathcal{A} = \mathcal{P}(X)$, and for each $E \in \mathcal{X}$, define $P(E) = \frac{|E|}{|X|}$. This defines a measure over $(\Omega, \mathcal{A})$, which is called the uniform probability.

Def/ A random variable is a measurable function $X : \Omega \to Y$, where $(Y, \mathcal{Y})$ is a measurable space.

In application we will consider random variables with $Y = \mathbb{R}$, $\mathcal{Y} = \mathcal{B}$, but as we did in measure theory, the previous definition works as well.

Density function

Given a probability space $(\Omega, \mathcal{A}, P)$, another measurable space $(Y, \mathcal{Y})$ and a random variable $X : \Omega \to Y$, we can define a probability measure over $(Y, \mathcal{Y})$ in the following way: for each $E \in \mathcal{Y}$, define

$$P_X(E) = P(\overbrace{X^{-1}(E)}^{\in \mathcal{A}})$$

it is an easy exercise to see that $P_X : \mathcal{Y} \to \mathbb{R}$ satisfies all the properties of a probability function. Usually, we will take $(Y, \mathcal{Y})$ to be $(\mathbb{R}, \mathcal{B})$.

Def/ The law of X is the probability measure P_X.

One interprets this in the following way: when $(Y, \mathcal{Y})$ is $(\mathbb{R}, \mathcal{B})$ the law of X "models" X not in the abstract probability space $(\Omega, \mathcal{A}, P)$, but in the prototype of $(\mathbb{R}, \mathcal{B})$.

Particular cases:

Very particular!

① Discrete random variables: $(\Omega, \mathcal{A}, P)$, and $X : \Omega \to Y$, Y is finite. Then the law of X is

$$P_X = \sum_{y \in Y} p_y \mu_y$$

where $p_y = P(\{\omega \in \Omega : X(\omega) = y\}) = P(X = y)$, and μ_y is the measure $\mu_y(E) = \begin{cases} 1, & y \in E \\ 0, & y \notin E \end{cases}$.

② Continuous random variables: assume that now $(Y, \mathcal{Y}) = (\mathbb{R}, \mathcal{B})$, and consider the lebesgue measure over $(\mathbb{R}, \mathcal{B})$ (λ). Then, use Radon-Nikodym theorem to build a "density" for P_X:

i) λ is σ-finite: $\mathbb{R}$ is the countable union of intervals $[i, i+1)$, and each has finite measure.

ii) Assuming that P_X is absolutely continuous with respect to λ, by Radon-Nikodym there is a function $f_X \in \mathcal{M}^+(\mathbb{R}, \mathcal{B})$ such that

$$P(B) = \int_B f \, d\lambda = \int_B f$$

Def / We say that f_X is the density probability function of (the law) P_X (also called mass function)

In particular, $1 = P_X(\mathbb{R}) = \int_{\mathbb{R}} f_X \, d\lambda = 1$.

Def / The distribution function of the law P_X is $F_X(x) = P_X((-\infty, x]) = \int_{(-\infty,x]} f_X \, d\lambda$.

Obs / $F_X(x)$ is an increasing function that satisfies $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to +\infty} F_X(x) = 1$. It is right continuous: for all $a \in \mathbb{R}$, $\{x_n\}_{n \geq 1}$ sequence such that $x_n \downarrow a$, we have that $F(x_n) \downarrow F(a)$.

Expectation and related operators.

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a (real-valued) random variable. Assume that $X \in L_1(\Omega, \mathcal{A}, P)$.

Def / The expected value of X is defined as

$$\mathbb{E}[X] = \int_\Omega X \, dP < +\infty.$$

If P_X is absolutely continuous w.r.t λ, then this is in fact $\mathbb{E}[X] = \int_{\mathbb{R}} x \, f_X(x) \, dx$.

In particular, for every measurable function $g : \mathbb{R} \to \mathbb{R}$, $g(X) : \Omega \to \mathbb{R}$ is another random variable, and

$$\mathbb{E}[g(X)] = \int_\Omega g(X) \, dP = \int_{\mathbb{R}} g(x) f_X(x) \, dx$$

Def / $\mathbb{E}[X^r]$ is called the factorial moment of order r of X. (If $X \in L_r(\Omega, \mathcal{A}, P)$ this has sense).

Def / $\mathbb{E}[(X - \mathbb{E}[X])^2]$ is called the variance of X. It is written as $\mathbb{V}[X]$ ($X \in L_2(\Omega, \mathcal{A}, P)$!)

A wide variety of results can be deduced from previous facts about L_p spaces. For instance, we have the following observation:

$$\mathbb{E}[|X|^p] < +\infty \text{ iff } \int_\Omega |X|^p \, dP < +\infty \text{ iff } X \in L_p(\Omega, \mathcal{A}, P)$$

In particular, we have the following results: ($\frac{1}{p} + \frac{1}{q} = 1$)

(A) Hölder: if X, Y are random variables over $(\Omega, \mathcal{A}, P)$, $X \in L_p(\Omega, \mathcal{A}, P)$, $Y \in L_q(\Omega, \mathcal{A}, P)$, then $XY \in L_1(\Omega, \mathcal{A}, P)$ and

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \, \mathbb{E}[|X|^q]^{1/q}$$

(B) Minkowsky: if $X, Y \in L_p(\Omega, \mathcal{A}, P)$, then $X + Y \in L_p(\Omega, \mathcal{A}, P)$ and $\mathbb{E}[|X+Y|^p]^{1/p} \leq \mathbb{E}[|X|^p]^{1/p} + \mathbb{E}[|Y|^p]^{1/p}$

i) Monotone convergence: $\{X_n\}_{n \geq 0}$ sequence of r.v., $X_n \geq 0$ (X_n non-dec.), $X_n \to X$ (pointwise) then $\lim \mathbb{E}[X_n] = \mathbb{E}[X]$

ii) Fatou lemma: given a sequence of random variables $\{X_n\}_{n \geq 0}$, $X_n \geq 0$, $\mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$

iii) Dominant convergence: $\{X_n\}_{n \geq 0}$ sequence of r.v., such that $|X_n| \leq Z$ with $\mathbb{E}[Z] < +\infty$. Then if $X_n \to X$ pointwise, $\lim \mathbb{E}[X_n] = \mathbb{E}[X]$.

Two important inequalities are the following:

Prop/ (Markov's inequality) Let $a>0$ and X a random variable on (Ω, A, P). Assuming that $X \geq 0$, we have that

$$P(\{\omega \in \Omega : X(\omega) \geq a\}) = P(X \geq a) \leq \frac{1}{a} \mathbb{E}[X]$$

Prop/ (Tchebicheff's inequality) Let $X \in L_2(\Omega, A, P)$ and $a>0$. Then

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{a^2} \mathbb{V}[X].$$

Independence of random variables

The next notion is the first important fact which mates the theory of probabilities differ from the one developed in Measure Theory.

Def/ Let (Ω, A, P) be a probability space. We say that a family of events $\{A_i\}_{i \in I}$ is independent if for any finite set J, $P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$. (A and B)

Def/ Let (Ω, A, P) be a probability space, $A, B \in A$, $P(B) > 0$. The conditional probability $P(A|B)$ is

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

Next step is to extend this notion to random variables:

Def/ Let (Ω, A, P) be a probability space, and $\{X_i\}_{i \in I}$ a set of random variables. We say that these random variables are independent if for every finite set of index $J = \{i_1, \dots, i_n\}$ and Borel sets $B_1, \dots, B_n$,

$$P(\{X_{i_1} \in B_1\} \cap \{X_{i_2} \in B_2\} \cap \dots \cap \{X_{i_n} \in B_n\}) = \prod_{j=1}^{n} P(X_{i_j} \in B_i)$$

This definition is related to the product of measure spaces in the following way

Theorem/ The n random variables $X_1, \dots, X_n$ are independent iff

$P_{(X_1, \dots, X_n)}$ on $\Omega_1 \times \Omega_2 \times \dots \times \Omega_n$ and with $A_1 \otimes \dots \otimes A_n$ $= P_{X_1} \otimes \dots \otimes P_{X_n}$

Proof/ By definition, $P_{(X_1, \dots, X_n)}(F_1 \times \dots \times F_n) = P(\{X_1 \in F_1\} \cap \{X_2 \in F_2\} \cap \dots \cap \{X_n \in F_n\}) = \prod_{i=1}^{n} P(X_i \in F_i)$ $= P_{X_1}(F_1) \otimes \dots \otimes P_{X_n}(F_n)$. As they have the same value over $A_1 \times \dots \times A_n$ by the construction of sets they must be the same.

With this approach one can show that the expectation works fine with respect to product of independent random variables; assume f_i a measurable function (in the convenient L_+ space):

$$\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] = \int_{\mathbb{R}^n} \prod_{i=1}^{n} f_i \, dP_{(X_1, \dots, X_n)} \overset{\text{(ind)}}{=} \int_{\mathbb{R}^n} \prod_{i=1}^{n} f_i \, dP_{X_1} \otimes \dots \otimes dP_{X_n} \overset{\text{(Fubini)}}{=}$$

$$= \prod_{i=1}^{n} \int_{\mathbb{R}} f_i \, dP_{X_i} = \prod_{i=1}^{n} \mathbb{E}[f_i(X_i)]$$

Obs/ When $f_1 = \dots = f_n = e$, then $\mathbb{E}[X_1 \dots X_n] = \prod_{i=1}^{n} \mathbb{E}[X_i]$; however, in general, the product of two r.v. in L_1 are not in L_1 and we need to be careful here.

We can now extend some definitions concerning the conditional expectation of random variables:

Def / Let $X \in L_1(\Omega, \mathcal{A}, P)$ be a random variable, $B \in \mathcal{A}$. The conditioned expectation of X w.r.t. B is

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X \mathbb{1}_B]}{P(B)}$$

We can extend now this definition to two random variables, but we will deal with this when studying Markov chains.

Some well-used random variables

Here we present a list of retated used random variables that would appear a lot

Name	Type	Distribution	Statistics
Bernoulli $Ber(p)$	discrete	$P(X=0)=p$ $P(X=1)=1-p$	$\mathbb{E}[X]=p$, $\mathbb{V}[X]=p(1-p)$
Binomial $Bin(n,p)$	discrete	- sum of n ind. bernoulli r.v. $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	$\mathbb{E}[X]=np$ $\mathbb{V}[X]=np(1-p)$
Geometric $Geom(p)$	discrete	an Bernoulli is repeated ind. until there is success $P(X=n)=(1-p)^{n-1}p$	$\mathbb{E}[X]=\frac{1}{p}$ $\mathbb{V}[X]=\frac{1-p}{p^2}$
Poisson $Po(\lambda)$	discrete	$P(X=n)=\frac{e^{-\lambda}}{n!}\lambda^n$	$\mathbb{E}[X]=\lambda$ $\mathbb{V}[X]=\lambda$
Uniform $U([a,b])$	continuous	Picks a value uniformly at random in $[a,b]$ $f_X(x)=\frac{1}{b-a}\mathbb{1}_{[a,b]}$	$\mathbb{E}[X]=\frac{b+a}{2}$ $\mathbb{V}[X]=\frac{1}{12}(b-a)^2$
Normal $N(\mu,\sigma^2)$	continuous	$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mathbb{E}[X]=\mu$ $\mathbb{V}[X]=\sigma^2$