

PROBABILITY THEORY

MATH 154

Unit 5: Random variables

5.1. We fix a probability space $(\Omega, \mathcal{A}, P)$. A map $X : \Omega \rightarrow \mathbb{R}$ is called a **random variable** if all sets $\{X \in [a, b]\}$ are in $\mathcal{A}$ for all $a \leq b$. Random variables are functions that can define measurable events. Most of the time we do not have to worry about measurability. On a finite probability space with the σ algebra of all subsets, all functions are random variables. On the real line with the Borel σ algebra $\mathcal{B}$ generated by half open intervals, all piecewise continuous functions are measurable.

5.2. As we can add and multiply random variables, we get an **algebra** $\mathcal{L}$ of random variables. It is a vector space over the real numbers $\mathbb{R}$. One could also look at complex-valued random variables or vector-valued random variables like **random matrices**.

Theorem 1 (Law). *Every random variable $X \in \mathcal{L}$ defines a unique probability measure $\mu([a, b]) = P[\{X \in [a, b]\}]$ on the real line $(\mathbb{R}, \mathcal{B})$. It is called the law of X .*

Proof. This is a direct consequence of the Carathéodory extension theorem: the function $\mu([a, b]) = P[\{X \in [a, b]\}]$ is a probability measure on the π -system of half open intervals on $\mathbb{R}$. We also have $\mu(\emptyset) = 0$ and $\mu(\mathbb{R}) = 1$. We can extend this uniquely to get a probability space $(\mathbb{R}, \mathcal{B}, \mu)$ $\square$

5.3. A **step function** is a random variables that are a finite sum $X = \sum_{i=1}^n \alpha_i \cdot 1_{A_i}$ with $\alpha_i \in \mathbb{R}$ and where $A_i \in \mathcal{A}$ are disjoint sets and where $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ else. Denote by $\mathcal{S}$ the algebra of step functions. It is a sub-algebra of $\mathcal{L}$. For $X \in \mathcal{S}$, define the **integral**

$$E[X] := \int_{\Omega} X \, dP = \sum_{i=1}^n \alpha_i P[A_i] .$$

5.4. Define $\mathcal{L}^1 \subset \mathcal{L}$ as the set of random variables X , for which $\sup_{Y \in \mathcal{S}, Y \leq |X|} \int_{\Omega} Y \, dP$ is finite. For $X \in \mathcal{L}^1$, define the **integral** or **expectation**

$$E[X] := \int_{\Omega} X \, dP = \sup_{Y \in \mathcal{S}, Y \leq X^+} \int_{\Omega} Y \, dP - \sup_{Y \in \mathcal{S}, Y \leq X^-} \int_{\Omega} Y \, dP ,$$

where $X^+ = X \vee 0 = \max(X, 0)$ and $X^- = -X \vee 0 = \max(-X, 0)$. The vector space $\mathcal{L}^1$ is the space of **integrable random variables**. Similarly, for $p \geq 1$, write $\mathcal{L}^p = \{X \in \mathcal{L}, E[|X|^p] < \infty\}$.

5.5. For $X \in \mathcal{L}^2$, the **variance** is defined as $\text{Var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2$. The non-negative number $\sigma[X] = \text{Var}[X]^{1/2}$ is called the **standard deviation** of X .

5.6. The expectation is also called "average" or "mean". The standard deviation measures how much we can expect the variable to deviate from the mean. Random variables are **centered**, if $E[X] = 0$. For such X , we can think of $\sigma[X]$ as a **length**.

5.7. Example: Let $\mathcal{B}$ denote the Lebesgue σ -algebra on $\Omega = [0, 1]$. The m 'th power random variable $X(x) = x^m$ on $(\Omega, \mathcal{B}, P)$ has expectation $E[X] = \int_0^1 x^m dx = \frac{1}{m+1}$, variance $\text{Var}[X] = E[X^2] - E[X]^2 = \frac{1}{2m+1} - \frac{1}{(m+1)^2} = \frac{m^2}{(1+m)^2(1+2m)}$ and so the standard deviation $\sigma[X] = \frac{m}{(1+m)\sqrt{1+2m}}$.

5.8. If X is a bounded random variable, then $E[X^m]$ is called the m 'th **moment** of X . The m 'th **central moment** of X is defined as $E[(X - E[X])^m]$. The **moment generating function** (MGF) of X is defined as $M_X(t) = E[e^{tX}]$. It is a tool for a fast simultaneous computation of all the moments. The function

$$\kappa_X(t) = \log(M_X(t))$$

is called the **cumulant generating function**.

5.9. Example: For $X(x) = x$ on $\Omega = [0, 1]$, we compute

$$M_X(t) = E[e^{tX}] = E\left[\sum_{m=0}^{\infty} \frac{t^m X^m}{m!}\right] = \sum_{m=0}^{\infty} t^m \frac{E[X^m]}{m!} = \sum_{m=0}^{\infty} \frac{t^m}{m!(m+1)} = (e^t - 1)/t.$$

5.10. Example. Let $\Omega = \mathbb{R}$. For given constants $m \in \mathbb{R}, \sigma > 0$, define the probability measure $P[[a, b]] = \int_a^b f(x) dx$ with

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Caratheodory assures that this extends to the Lebesgue σ -algebra. This is a probability measure because after a change of variables $y = (x - m)/(\sqrt{2}\sigma)$, the integral $\int_{-\infty}^{\infty} f(x) dx$ becomes $\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2} dy = 1$. The random variable $X(x) = x$ on $(\Omega, \mathcal{A}, P)$ is an example of a random variable with **Gaussian distribution**. We also say it is a **Gaussian random variable** or a random variable with **normal distribution**. Let us justify the constants m and σ : the expectation of X is $E[X] = \int X dP = \int_{-\infty}^{\infty} x f(x) dx = m$. The variance is $E[(X - m)^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \sigma^2$ so that the constant σ is indeed the standard deviation. The moment generating function of X is $M_X(t) = e^{mt + \sigma^2 t^2/2}$. The cumulant generating function is therefore $\kappa_X(t) = mt + \sigma^2 t^2/2$.

5.11. Here comes in some calculus

Theorem 2. If X has the MGF $M_X(t)$ and $M_X(t)$ has a convergent Taylor series at $t = 0$, then $E[X^n] = \frac{d^n}{dt^n} M_X(t)|_{t=0}$.

Proof. Apply the Taylor series to $e^{tX} = \sum_{n=0}^{\infty} t^n X^n/n!$. Taking the n 'th derivative on both sides $X^n e^{tX}|_{t=0} = X^n$ and taking the expectation shows that $E[X^n]$ is the n 'th derivative of $M_X(t)$ evaluated at $t = 0$. $\square$