PROBABILITY THEORY

MATH 154

Unit 5: Random variables

5.1. Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space. A map $X : \Omega \to \mathbb{R}$ 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 which define measurable events. Since we can add and multiply random variables, we get an **algebra** $\mathcal{L}$ of all 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**. The Borel σ algebra on the real line generated by half open intervals is denoted by $\mathcal{B}$.

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

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

5.2. A **step function** is a random variable of the form $X = \sum_{i=1}^{n} \alpha_i \cdot 1_{A_i}$ with $\alpha_i \in \mathbb{R}$ and where $A_i \in \mathcal{A}$ are disjoint sets. Denote by $\mathcal{S}$ the algebra of step functions. For $X \in \mathcal{S}$, define the **integral**

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

5.3. 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**

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

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 $\mathcal{L}^p = \{X \in \mathcal{L}, \mathrm{E}[|X|^p] < \infty\}$.

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

5.5. 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 $\mathrm{E}[X] = 0$. For such X, we can think of σX as a **length**.

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

5.7. If X is a random variable, then $\mathrm{E}[X^m]$ is called the m'th **moment** of X. The m'th **central moment** of X is defined as $\mathrm{E}[(X - \mathrm{E}[X])^m]$. The **moment generating function** (MGF) of X is defined as $M_X(t) = \mathrm{E}[e^{tX}]$. 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**.

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

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

5.9. Example. Let $\Omega = \mathbb{R}$. For given constants $m \in \mathbb{R}, \sigma > 0$, define the probability measure $\mathrm{P}[[a,b)] = \int_a^b f(x) \, dx$ 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 $(\Omega, \mathcal{A}, \mathrm{P})$ 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**. Lets justify the constants m and σ: the expectation of X is $\mathrm{E}[X] = \int X \, dP = \int_{-\infty}^{\infty} x f(x) \, dx = m$. The variance is $\mathrm{E}[(X-m)^2] = \int_{-\infty}^{\infty} x^2 f(x) \, dx = \sigma^2$ 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 $\kappa_X(t) = mt + \sigma^2 t^2/2$.

5.10. 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 $\mathrm{E}[X^n] = \frac{d^n}{dt^n} M_X(t)|_{t=0}$.*

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

Oliver Knill, knill@math.harvard.edu, Math 154, Spring, 2025