Lecture 1: Probability

Zhentao Shi

August 9, 2017

Probability

Probability Space

A *sample space* Ω is a collection of all possible outcomes. It is a set of things.

An *event* A is a subset of Ω. It is something of interest on the sample space.

A *σ-field*, denoted by $\mathcal{F}$, is a collection of $(A_i \subseteq \Omega)_{i \in \mathbb{N}}$ events such that

(i) $\emptyset \in \mathcal{F}$;

(ii) if an event $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$;

(iii) if $A_i \in \mathcal{F}$ for $i \in \mathbb{N}$, then $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$.

It is easy to show that $\Omega \in \mathcal{F}$ and $\bigcap_{i \in \mathbb{N}} A_i \in \mathcal{F}$. The σ-field can be viewed as a well-organized structure built on the ground of the sample space. The pair $(\Omega, \mathcal{F})$ is called a *measure space*.

Let $\mathcal{G} = \{B_1, B_2, \ldots\}$ be an arbitrary collection of sets, not necessarily a σ-field. We say $\mathcal{F}$ is the smallest σ-field generated by $\mathcal{G}$ if $\mathcal{G} \subseteq \mathcal{F}$, and $\mathcal{F} \subseteq \tilde{\mathcal{F}}$ for any $\tilde{\mathcal{F}}$ such that $\mathcal{G} \subseteq \tilde{\mathcal{F}}$. A *Borel σ-field* $\mathcal{R}$ is the smallest σ-field generated by the open sets on the real line $\mathbb{R}$.

A function $\mu : (\Omega, \mathcal{F}) \mapsto [0, \infty]$ is called a *measure* if it satisfies

(i) $\mu(A) \geq 0$ for all $A \in \mathcal{F}$;

(ii) if $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, are mutually disjoint, then

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu\left(A_i\right).$$

Measure can be understand as weight or length in our daily life experience. In particular, we call μ a *probability measure* if $\mu(\Omega) = 1$. A probability measure is often denoted as P. The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space.*

For a fixed measure space $(\Omega, \mathcal{F})$, we can have different probability measure P on it.

Example Let $\Omega = \{a, b\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ where $\mathcal{P}(\Omega)$ denote the power set of Ω, i.e. the collection of all subsets of Ω. Define P_1, $P_2 : (\Omega, \mathcal{F}) \to [0, \infty)$ by

(i) $P_1(\Omega) = 1$, $P_1(\emptyset) = 0$, $P_1(\{a\}) = \frac{1}{2}$, $P_1(\{b\}) = \frac{1}{2}$

(ii) $P_2(\Omega) = 1$, $P_2(\emptyset) = 0$, $P_2(\{a\}) = 1$, $P_2(\{b\}) = 0$

Both P_1 and P_2 are probability measures on $(\Omega, \mathcal{F})$. If we simply take this abstract example as tossing a coin where a and b represent two possible symbols, say a number and a flower. P_1 means we are tossing a fair coin with a number and a flower on each side respectively, and P_2 means we are tossing a coin with a flower on both sides.

A sequence of events $\{A_i\}_{i\in\mathbb{N}}$ is increasing (resp. decreasing) if $A_1 \subseteq A_2 \subseteq \cdots$ (resp. $A_1 \supseteq A_2 \supseteq \cdots$).

Proposition If $\{A_i\}_{i\in\mathbb{N}}$ is increasing (resp. decreasing), then $P\left(\cup_{i=1}^{\infty}\right) = \lim_{i\to\infty} P(A_i)$ (resp. $P\left(\cap_{i=1}^{\infty}\right) = \lim_{i\to\infty} P(A_i)$).

Example Consider the probability space $(\Omega, \mathcal{F}, P)$ where $\Omega = (0, 1)$, $\mathcal{F}$ is the σ-field generated by $\mathcal{F}_0 = \{(a, b) : 0 < a < b < 1\}$, and P is a probability measure on $(\Omega, \mathcal{F})$ satisfying $P_{|\mathcal{F}_0}((a, b)) = b - a$.

Note that $\left\{\frac{1}{2}\right\} \in \mathcal{F}$ since $\left\{\frac{1}{2}\right\} = \cap_{n=3}^{\infty}\left(\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}\right) \in \mathcal{F}$. By the above proposition, we can easily get

$$P\left(\left\{\frac{1}{2}\right\}\right) = P\left(\cap_{n=3}^{\infty}\left(\frac{1}{2}-\frac{1}{n},\frac{1}{2}+\frac{1}{n}\right)\right) = \lim_{n\to\infty} P\left(\left(\frac{1}{2}-\frac{1}{n},\frac{1}{2}+\frac{1}{n}\right)\right) = \lim_{n\to\infty}\frac{1}{2n} = 0$$

since $\left\{\left(\frac{1}{2}-\frac{1}{n},\frac{1}{2}+\frac{1}{n}\right)\right\}_{n=1}^{\infty}$ is decreasing. This example shows that some sets in the collection of events can be 0-measure.

Random Variable

The terminology *random variable* somewhat belies its formal definition of a deterministic mapping. It is a link between two measure spaces such that any event in the σ-field installed on the range can be tracked back to an event in the σ-field installed on the domain.

Formally, a function $X : \Omega \mapsto \mathbb{R}$ is $(\Omega, \mathcal{F}) \setminus (\mathbb{R}, \mathcal{R})$ *measurable* if

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

for any $B \in \mathcal{R}$. *Random variable* is an alternative common name for a measurable function. We say a measurable is a *discrete random variable* if the set $\{X(\omega) : \omega \in \Omega\}$ is finite or countable. We say it is a *continuous random variable* if the set $\{X(\omega) : \omega \in \Omega\}$ is uncountable.

A measurable function connects two measurable spaces. No probability is involved in its definition. However, if a probability measure P is installed on $(\Omega, \mathcal{F})$, the measurable function X will induce a probability measure on $(\mathbb{R}, \mathcal{R})$. It is easy to verify that $P_X : (\mathbb{R}, \mathcal{R}) \mapsto [0, 1]$ is also a probability measure if defined as $P_X(B) = P(X^{-1}(B))$ for any $B \in \mathcal{R}$.
(If $B_1, B_2 \in \mathcal{R}$ are disjoint, then $X^{-1}(B_1), X^{-1}(B_2) \in \mathcal{F}$ are also disjoint.) This P_X is called the probability measure *induced* by the measurable function X. The induced probability measure P_X is an offspring of the parent probability measure P though the channel of X.

Distribution Function

We go back to some terms that we have learned in the undergraduate probability course. A *(cumulative) distribution function* $F : \mathbb{R} \mapsto [0, 1]$ is defined as

$$F(x) = P(X \leq x) = P(\{X \leq x\}) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

It is often abbreviated as CDF, and it has the following properties.

(i) $\lim_{x \to -\infty} F(x) = 0$,

(ii) $\lim_{x \to \infty} F(x) = 1$,

(iii) non-decreasing,

(iv) right-continuous $\lim_{y \to x^+} F(y) = F(x)$.

For continuous distribution, if there exists a function f such that for all x,

$$F(x) = \int_{-\infty}^{x} f(y)\, dy,$$

then f is called the *probability density function* of X, often abbreviated as PDF. It is easy to show that $f(x) \geq 0$ and $\int_a^b f(x)\, dx = F(b) - F(a)$.

Example We have learned many parametric distributions like the binary distribution, the Poisson distribution, the uniform distribution, the normal distribution, χ^2, t, F and so on. They are parametric distributions, meaning that the CDF or PDF can be summarized in a few parameters.

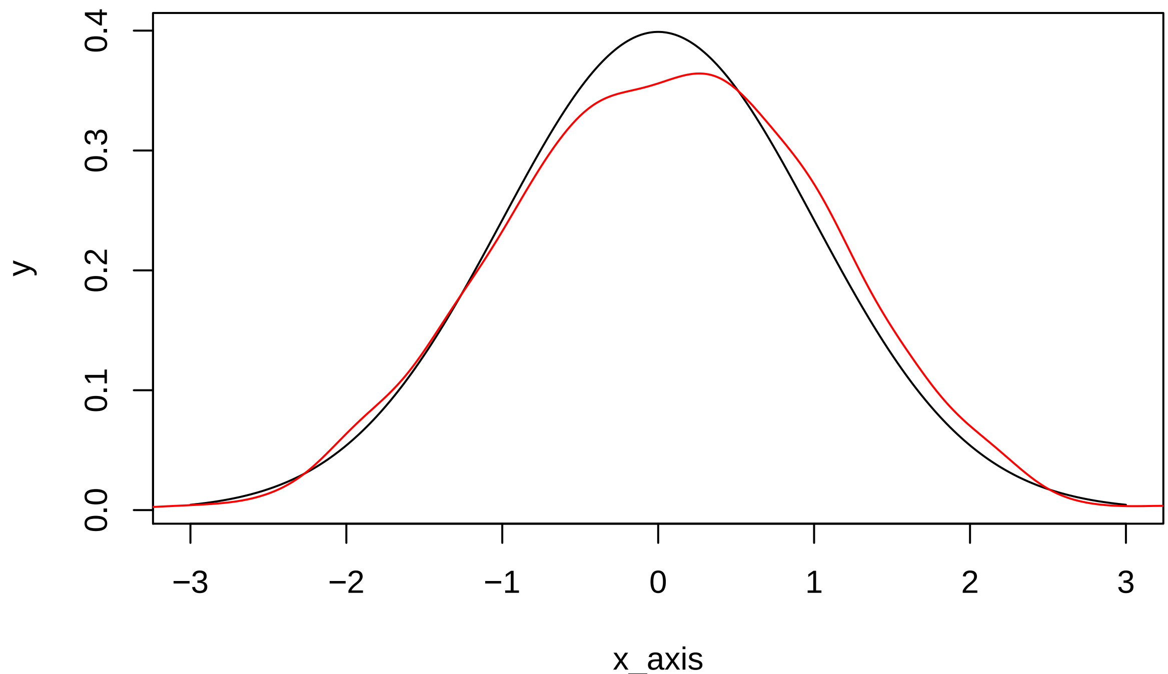
Example `R` has a rich collection of distributions implemented in a unified rule: `d` for density, `p` for probability, `q` for quantile, and `r` for random variable generation. For instance, `dnorm`, `pnorm`, `qnorm`, and `rnorm` are the corresponding functions of the normal distribution, and the parameters μ and σ can be specified in the arguments of the functions.

Below is a piece of `R` code for demonstration.

1. Plot the density of standard normal distribution over an equally spaced grid system

`x_axis = seq(-3, 3, by = 0.01)` (black line).

2. Generate 1000 observations for $N(0, 1)$. Plot the kernel density, a nonparametric estimation of the density (red line).
3. Calculate the 95-th quantile and the empirical probability of observing a value greater than the 95-th quantile. In population, this value should be 5%. What is the number in this experiment?




```
## [1] 0.053
```