NOTATION

UGMTP = *User's Guide to Measure-Theoretic Probability*

Let $\mathcal{X}$ be a set equipped with a sigma-field $\mathcal{A}$. Identify sets with their indicator functions [UGMTP §1.4]. Define

$$\begin{aligned}
\mathcal{M}(\mathcal{X},\mathcal{A}) &:= \text{all } \mathcal{A}\text{-measurable functions from } \mathcal{X} \text{ into } [-\infty,\infty, \\
\mathcal{M}^+(\mathcal{X},\mathcal{A}) &:= \text{all nonnegative functions in } \mathcal{M}(\mathcal{X},\mathcal{A}) \\
\mathbb{M}(\mathcal{X},\mathcal{A}) &:= \text{all bounded functions in } \mathcal{M}(\mathcal{X},\mathcal{A}) \\
\mathbb{M}^+(\mathcal{X},\mathcal{A}) &:= \text{all bounded, nonnegative functions in } \mathcal{M}(\mathcal{X},\mathcal{A}).
\end{aligned}$$

Unless stated otherwise, measures will always be understood to be nonnegative and countably additive. Finitely additive measures on $\mathcal{A}$ correspond to maps λ from $\mathbb{M}^+(\mathcal{X},\mathcal{A})$ to $\mathbb{R}^+$ that are linear, in the sense that

$$\lambda(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1\lambda(f_1) + \alpha_2\lambda(f_2) \qquad \text{for all } \alpha_i \in \mathbb{R}^+ \text{ and } f_i \in \mathbb{M}^+(\mathcal{X},\mathcal{A}).$$

Write $\mathbb{L}^+(\mathcal{X},\mathcal{A})$ for the set of all such functionals. For functionals in $\mathbb{L}^+(\mathcal{X},\mathcal{A})$, write $\lambda_1 \ge \lambda_2$ to mean that $\lambda_1 f \ge \lambda_2 f$ for all f in $\mathbb{M}^+(\mathcal{X},\mathcal{A})$. Define a distance between functionals λ_1 and λ_2 in $\mathbb{L}^+(\mathcal{X},\mathcal{A})$ via the norm

<1> $$\|\lambda_1 - \lambda_2\| := \sup\{|\lambda_1 f - \lambda_2 f| : f \in \mathbb{M}(\mathcal{X},\mathcal{A}) \text{ and } \sup_x |f(x)| \le 1\}$$

If $\lambda_1(1) = \lambda_2(1)$ then $\|\lambda_1 - \lambda_2\| := 2\sup\{|\lambda_1 g - \lambda_2 g| : 0 \le g \le 1\}$.

Countably additive measures on $\mathcal{A}$ can be identified [UGMTP §2.3] with those increasing linear functionals μ on $\mathcal{M}^+(\mathcal{X},\mathcal{A})$ that have the ***Monotone Convergence*** property,

$$\text{if } 0 \le f_n \uparrow f \text{ then } \mu f_n \uparrow \mu f.$$

For finite countably additive measures, it is sometimes more convenient to make an identification with the functionals in $\mathbb{L}^+(\mathcal{X},\mathcal{A})$ that are ***σ-smooth at*** 0,

$$\text{if } \{f_n : n \in \mathbb{N}\} \subseteq \mathbb{M}^+(\mathcal{X},\mathcal{A}) \text{ and } f_n \downarrow 0 \text{ pointwise then } \lambda f_n \downarrow 0.$$

Write $\mathbb{L}^+_\sigma(\mathcal{X},\mathcal{A})$ for the set of all such σ-smooth functionals. If λ is a countably additive, sigma-finite measure, write then $\mathbb{L}^+_\lambda(\mathcal{X},\mathcal{A})$ for the subset of $\mathbb{L}^+_\sigma(\mathcal{X},\mathcal{A})$ consisting of the measures dominated by λ. That is, $\mu \in \mathbb{L}^+_\lambda(\mathcal{X},\mathcal{A})$ if and only if there exists a density $g \in \mathcal{M}^+(\mathcal{X},\mathcal{A})$ for which $\lambda g < \infty$ and $\mu f = \lambda(gf)$ for all f in $\mathcal{M}^+(\mathcal{X},\mathcal{A})$.

For countably additive measures, the norm defined by <1> is sometimes called the $\mathcal{L}^1$ norm, and is denoted by $\|\cdot\|_1$. If both μ and ν are probability measures, then

$$\tfrac{1}{2}\|\mu - \nu\|_1 = \sup_{A\in\mathcal{A}} |\mu A - \nu A| = \sup_{0\le f\le 1} |\mu f - \nu f|,$$

a quantity that is often called the total variation distance between the measures [UGMTP §3.3].

When the choice of $\mathcal{X}$ or of $\mathcal{A}$ is clear, abbreviate $\mathcal{M}^+(\mathcal{X},\mathcal{A})$ to $\mathcal{M}^+(\mathcal{X})$, or $\mathcal{M}^+(\mathcal{A})$, or even just $\mathcal{M}^+$. And so on.

Let $\mathcal{Y}$ be another set, equipped with a sigma-field $\mathcal{B}$. If T is an $\mathcal{A}\backslash\mathcal{B}$-measurable map from $\mathcal{X}$ to $\mathcal{Y}$, and μ is a measure on $\mathcal{A}$, the ***image measure*** $T(\mu)$, or $T\mu$, is defined on $\mathcal{B}$ by $(T\mu)(B) := \mu\{x : T(x) \in B\}$ for each $B \in \mathcal{B}$. Equivalently,

$$(T\mu)^y g(y) := \mu^x g(T(x)) \qquad \text{for } g \in \mathcal{M}^+(\mathcal{Y}, \mathcal{B}).$$

Markov kernels

A Markov kernel from $(\mathcal{X}, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$ is a family of probability measures $\mathbb{K} := \{\mathbb{K}_x : x \in \mathcal{X}\}$ such that $x \mapsto \mathbb{K}_x B$ is $\mathcal{A}$-measurable, for each $B \in \mathcal{B}$. As a consequence, for each f in $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{A} \otimes \mathcal{B})$ the function $x \mapsto \mathbb{K}_x^y f(x, y) := \int f(x, y)\, \mathbb{K}_x(dy)$ is $\mathcal{A}$-measurable. If μ is a measure on $\mathcal{A}$ then the measure $\mu \otimes \mathbb{K}$ is defined on $\mathcal{A} \otimes \mathcal{B}$ by

$$(\mu \otimes K)\, f := \mu^x \left(\mathbb{K}_x^y f(x, y)\right) \qquad \text{for } f \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{A} \otimes \mathcal{B}).$$

It has marginals μ and λ, with λ the measure on $\mathcal{B}$ defined by

$$\lambda^y g(y) := \mu^x \left(\mathbb{K}_x^y g(y)\right) \qquad \text{for } g \in \mathcal{M}^+(\mathcal{Y}, \mathcal{B}).$$

The measure λ is also denoted by $\mathbb{K}\mu$ or $\mu^x \mathbb{K}_x$ for λ. The map $\mu \mapsto \mathbb{K}\mu$ from $\mathbb{L}_\sigma^+(\mathcal{X}, \mathcal{A})$ to $\mathbb{L}_\sigma^+(\mathcal{Y}, \mathcal{B})$ is linear, and it takes probability measures to probability measures.

If μ is a probability measure, the pair (x, y) generated by

$$x \sim \mu \qquad \text{and} \qquad y|x \sim \mathbb{K}_x$$

has joint distribution $\mu \otimes \mathbb{K}$. The y has marginal distribution $\mathbb{K}\mu$.

Decision theory

A ***statistical model*** (or statistical experiment) is defined to be a family of probability measures $\mathcal{P} := \{\mathbb{P}_\theta : \theta \in \Theta\}$, all defined on the same sigma-field $\mathcal{A}$ on a set $\mathcal{X}$. Let $\mathbb{D}$ be some set, equipped at least with a sigma-field $\mathcal{D}$. A ***decision procedure*** for $\mathcal{P}$ is a measurable map T from $\mathcal{X}$ to $\mathbb{D}$. (If $\mathbb{D} = \Theta$, then T is usually called an estimator for the parameter θ.) A randomized procedure is defined as a Markov kernel τ from $(\mathcal{X}, \mathcal{A})$ to $(\mathbb{D}, \mathcal{D})$.

A map L from $\mathbb{D} \times \Theta$ into $[-\infty, \infty]$ is called a ***loss function***. Typically I will assume L is either nonnegative or bounded, so that there are no problems with the next definition. The risk function for a procedure T is defined as

$$R(T, \theta) := \mathbb{P}_\theta^x L\,(T(x), \theta) = (T\mathbb{P}_\theta)^t L(t, \theta) \qquad \text{for } \theta \in \Theta.$$

The risk function for a randomized procedure τ is defined as

$$R(\tau, \theta) := \mathbb{P}_\theta^x \tau_x^t L\,(t, \theta) = (\tau\mathbb{P}_\theta)^t L(t, \theta) \qquad \text{for } \theta \in \Theta.$$