Dissertation Notation Conventions

Paul Rademacher

September 1, 2022

This article details the mathematical notation and typesetting conventions used throughout my dissertation and related publications. Note that many variable scalars and functions including x, y, g, etc. are repeatedly redefined and reused to avoid introducing an excess of symbols. Unless explicitly stated, none of these variable definitions will hold in subsequent sections.

1 Sets and Function Arguments

Sets are typically typeset with a calligraphic font (e.g. $\mathcal{X}$), with the exception of some common number sets which are typeset using blackboard bold (e.g. real numbers $\mathbb{R}$). Function spaces, such as the set of functions $\mathcal{X} \mapsto \mathcal{Y}$, are compactly represented as $\mathcal{Y}^{\mathcal{X}}$.

Various mappings are defined for which the domain and/or the range [5] are function spaces. For a mapping $g : \mathcal{Z} \mapsto \mathcal{Y}^{\mathcal{X}}$, the argument notation $g(z) \in \mathcal{Y}^{\mathcal{X}}$ denotes a function, while $g(x; z) \in \mathcal{Y}$ is a specific value of that function. Semicolons are used to distinguish between the arguments referring to the domain and arguments that access the resulting function. The mapping $\{1, \dots, N\} \mapsto \mathcal{Y}$ is represented as $\mathcal{Y}^N$ for brevity. Items of an indexed tuple $g \in \mathcal{Y}^N$ are accessed with subscripts rather than parentheses, such that $g_i \in \mathcal{Y}$.

The Cartesian product of sets is frequently used, such that for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, the pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$. For a general product of sets $\mathcal{S}_i$, the notation $\prod_i \mathcal{S}_i = \mathcal{S}_1 \times \mathcal{S}_2 \ldots$ is employed.

The convention adopted for natural numbers is $\mathbb{N} = \{1, 2, \ldots\}$; the set of non-negative integers is denoted $\mathbb{Z}_{\geq 0} = \mathbb{N} \cup \{0\}$. The set of positive real numbers $\mathbb{R}^+$ excludes zero, while nonnegative real numbers are represented as $\mathbb{R}_{\geq 0} = \mathbb{R}^+ \cup \{0\}$. The cardinality of countably infinite sets, including the set of natural numbers, is denoted $\aleph_0 = |\mathbb{N}|$, where $\aleph$ is the aleph number [1]; the cardinality of uncountable sets, such as $\mathbb{R}$, is at least $\aleph_1$.

Numerous probability distribution functions are defined over different domains. As such, for a given set $\mathcal{X}$, a set function $\mho$ is defined such that $\mho(\mathcal{X})$ is the set of distributions over $\mathcal{X}$. If $\mathcal{X}$ is countable, the set is defined as $\mho(\mathcal{X}) = \big\{p \in {\mathbb{R}_{\geq 0}}^{\mathcal{X}} : \sum_{x \in \mathcal{X}} p(x) = 1\big\}$; if $\mathcal{X}$ is a Euclidean space, the set is defined as $\mho(\mathcal{X}) = \big\{p \in {\mathbb{R}_{\geq 0}}^{\mathcal{X}} : \int_{\mathcal{X}} p(x) \mathrm{d}x = 1\big\}$.

2 Special Operators and Functions

Various operators commonly used in linear algebra are generalized for functions. Specifically, the outer product operator $\otimes$ is used on two real-valued functions $f \in \mathbb{R}^{\mathcal{X}}$ and $g \in \mathbb{R}^{\mathcal{Y}}$ such that $\big(f \otimes g\big) \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$ is defined as $\big(f \otimes g\big)(x, y) = f(x)g(y)$. A general outer product of functions is denoted $\bigotimes_i f_i = f_1 \otimes f_2 \ldots$ for $i = 1, \ldots$, where $\Big(\bigotimes_i f_i\Big)(x_1, x_2, \ldots) = f_1(x_1) f_x(x_2) \ldots$, is used as well. Also, the diagonal operator operates on a single real-valued function such that $\mathrm{diag}(f) \in \mathbb{R}^{\mathcal{X} \times \mathcal{X}}$. For countable sets X, the operator values are $\mathrm{diag}(f)(x, x') = f(x)\delta[x, x']$; for Euclidean sets, the operator values are $\mathrm{diag}(f)(x, x') = f(x)\delta(x - x')$.

A variety of special functions are used throughout. For a given subset $\mathcal{S} \subset \mathcal{X}$, the indicator function $\chi(\mathcal{S}) : \mathcal{X} \mapsto \{0, 1\}$, defined as

$$\chi(x; \mathcal{S}) = \begin{cases} 1 & \text{if } x \in \mathcal{S} \ , \\ 0 & \text{if } x \notin \mathcal{S} \ , \end{cases} \tag{1}$$

will be used repeatedly. Also, for a given function $g \in \mathbb{R}^{\mathcal{X}}$, define the support operator,

$$\mathrm{supp}(f) = \{x \in \mathcal{X} : f(x) \neq 0\} \ , \tag{2}$$

returning the subset of the function domain for which the mapped values are nonzero. This operator will be frequently used on probability functions to represent the random values with nonzero probability.

Both the Dirac and Kronecker delta functions are frequently required. The Dirac delta function [2] over a Euclidean domain $\mathcal{X}$ is represented as $\delta(\cdot)$; it has support only at the point $x = 0$ and satisfies

$$\int_{\mathcal{X}} \delta(x) \mathrm{d}x = 1 \ . \tag{3}$$

Consequently, it also satisfies

$$\int_{\mathcal{X}} g(x)\delta(x) \mathrm{d}x = g(0) \ . \tag{4}$$

Consider a set $\mathcal{X}$; the Kronecker delta function has domain $\mathcal{X} \times \mathcal{X}$ and is defined as

$$\delta[x, x'] = \begin{cases} 1 & \text{if } x = x' \ , \\ 0 & \text{if } x \neq x' \ . \end{cases} \tag{5}$$

As both functions are denoted by the symbol δ, they are distinguished by the use of parentheses or square brackets. The relation $\delta(\cdot - x) = \delta(0)\delta[\cdot, x]$ may be used to relate the two functions.

The multinomial coefficient and multivariate beta function, which typically operate on sequences, are defined more generally for function inputs. The multinomial operator $\mathcal{M}$ is used for functions $g : \mathcal{X} \mapsto \mathbb{Z}_{\geq 0}$ that map to nonnegative integers from an arbitrary countable domain $\mathcal{X}$. The output of the operator is

$$\mathcal{M}(g) = \frac{\left(\sum_{x \in \mathcal{X}} g(x)\right)!}{\prod_{x \in \mathcal{X}} g(x)!} \ . \tag{6}$$

Similarly, the beta function β operates on functions $g : \mathcal{X} \mapsto \mathbb{R}^+$ that map to positive real numbers from an arbitrary countable domain $\mathcal{X}$, such that

$$\beta(g) = \frac{\prod_{x \in \mathcal{X}} \Gamma\big(g(x)\big)}{\Gamma\left(\sum_{x \in \mathcal{X}} g(x)\right)} \ . \tag{7}$$

Note that the countable domains of the input functions may have an infinite number of elements.

3 Random elements, variables, and processes

Random elements are denoted with roman font (e.g., x), while specific values are denoted with italics (e.g., x). Random elements that assume numerical scalars or functions are referred to as random variables or processes, respectively.

Consider a random element $\mathrm{x} \in \mathcal{X}$. If $\mathcal{X}$ is countable, either finite with $|\mathcal{X}| \in \mathbb{N}$ or countably infinite with $|\mathcal{X}| = \aleph_0$, then x is a discrete random element and is characterized by a probability mass function (PMF) [4] denoted $\mathrm{P}_{\mathrm{x}} \in \mho(\mathcal{X})$. If $\mathcal{X}$ is a Euclidean space and is thus uncountable with $|\mathcal{X}| \geq \aleph_1$, then x is a continuous random variable or process characterized by a probability density function (PDF) denoted $\mathrm{p}_{\mathrm{x}} \in \mho(\mathcal{X})$.

Consider x conditioned on another random element $\mathrm{z} \in \mathcal{Z}$. The conditional distribution is represented as $\mathrm{P}_{\mathrm{x}|\mathrm{z}} : \mathcal{Z} \mapsto \mho(\mathcal{X})$ such that $\mathrm{P}_{\mathrm{x}|\mathrm{z}}(z)$ is a PMF over $\mathcal{X}$ and $\mathrm{P}_{\mathrm{x}|\mathrm{z}}(x|z)$ is a specific value of that PMF. Often, the dependency on the conditional variable z will not be expressed in terms of a specific value z, but will be left in terms of the random element itself; in this case, the more compact notation $\mathrm{P}_{\mathrm{x}|\mathrm{z}}$ is used to imply $\mathrm{P}_{\mathrm{x}|\mathrm{z}}(\mathrm{z})$, a function of z.

A frequent consideration will be how the statistics of a sequence of random elements tend. The notion of convergence in probability [3] will be used throughout; the notation $\mathrm{x}_n \xrightarrow{p} \mathrm{y}$ indicates that $\lim_{n\to\infty} \Pr\left(|\mathrm{x}_n - \mathrm{y}| < \epsilon\right) = 0$ for all values $\epsilon > 0$. In this work, dependency on the index will typically be excluded. Also, the sequence will often converge to a deterministic value; this implies convergence of the probability distribution to a delta function (Kronecker or Dirac for discrete or continuous random elements, respectively). For example, as $N \to \infty$, the notation $\mathrm{x} \xrightarrow{p} y$ indicates that $\mathrm{p}_{\mathrm{x}} \to \delta(\cdot - y)$.

Many distributions will be repeatedly used and thus special functions will be defined for the PDF's and PMF's of interest. For example, consider a random process $\mathrm{x} \in \mathcal{X}$ characterized by an Empirical distribution with parameters $N \in \mathbb{Z}_{\geq 0}$ and $\rho \in \mathcal{P}$; the PMF will be notated as $\mathrm{Emp} : \mathbb{Z}_{\geq 0} \times \mathcal{P} \mapsto \mho(\mathcal{X})$, where the range is the set of valid PMF's. More compactly, the notation $\mathrm{x} \sim \mathrm{Emp}(N, \rho)$ implies that $\mathrm{P}_{\mathrm{x}} = \mathrm{Emp}(N, \rho)$. Other distribution functions repeatedly used include Dir, DE, DP, and DEP, representing the Dirichlet distribution, the Dirichlet-Empirical distribution, the Dirichlet process, and the Dirichlet-Empirical process.

4 Expectation Operators

For a discrete random element x, the expectation operator E_{x} is defined as

$$\mathrm{E}_{\mathrm{x}}\big[g(\mathrm{x})\big] = \sum_{x} \mathrm{P}_{\mathrm{x}}(x) g(x) \;, \tag{8}$$

where the argument g is an arbitrary scalar function of x with range $\mathbb{R}$. Additionally, define the variance operator C_{x} as

$$\mathrm{C}_{\mathrm{x}}\big[g(\mathrm{x})\big] = \mathrm{E}_{\mathrm{x}}\left[\Big(g(\mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(\mathrm{x})\big]\Big)^2\right] \;. \tag{9}$$

When x is a random variable and the function g is the identity operator, such that $g(\mathrm{x}) = \mathrm{x}$, the mean and variance are compactly represented as μ_{x} and Σ_{x}, respectively.

These operations can be performed with respect to a conditional distribution as well. In this case, the expectation operator is a function of the observed value of z, such that

$$\mathrm{E}_{\mathrm{x}|\mathrm{z}}\big[g(\mathrm{x})\big](z) = \sum_{x} \mathrm{P}_{\mathrm{x}|\mathrm{z}}(x|z) g(x) \;. \tag{10}$$

Similarly, the conditional variance is notated $\mathrm{C}_{\mathrm{x}|\mathrm{z}}\big[g(\mathrm{x})\big](z)$. When g is the identity operator, the conditional mean and variance as represented by $\mu_{\mathrm{x}|\mathrm{z}}(z)$ and $\Sigma_{\mathrm{x}|\mathrm{z}}(z)$, respectively.

As with conditional distributions, it is common that an explicit value z of the conditional random element will not be used, but rather the expectation will be left as a function of the random element z. In these cases, the argument is suppressed and the notation $\mathrm{E}_{\mathrm{x}|\mathrm{z}}\big[g(\mathrm{x})\big]$ implies the dependency on z. This convention also holds for the conditional variance operator $\mathrm{C}_{\mathrm{x}|\mathrm{z}}$, as well as for the $\mu_{\mathrm{x}|\mathrm{z}}$ and $\Sigma_{\mathrm{x}|\mathrm{z}}$ functions.

If the range of g is a Hilbert space, such that $g(\mathrm{x})$ is itself a function with a domain $\mathcal{Y}$, then the notation for these operators is expanded. The output of the expectation operator is a function over $\mathcal{Y}$ represented by

$$\mathrm{E}_{\mathrm{x}}\big[g(\mathrm{x})\big](y) = \sum_{x} \mathrm{P}_{\mathrm{x}}(x) g(y; x) \ . \tag{11}$$

Similarly, the covariance function notation is modified and the output is a function over $\mathcal{Y} \times \mathcal{Y}$,

$$\mathrm{C}_{\mathrm{x}}\big[g(\mathrm{x})\big](y, y') = \mathrm{E}_{\mathrm{x}}\Big[\Big(g(y; \mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(y; \mathrm{x})\big]\Big)\Big(g(y'; \mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(y'; \mathrm{x})\big]\Big)\Big] \ . \tag{12}$$

$$\mathrm{C}_{\mathrm{x}}\big[g(\mathrm{x})\big] = \mathrm{E}_{\mathrm{x}}\Big[\Big(g(\mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(\mathrm{x})\big]\Big) \otimes \Big(g(\mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(\mathrm{x})\big]\Big)\Big] \ . \tag{13}$$

As before, the notation is simplified when the function g is the identity operator. If x is a random process over a domain $\mathcal{Y}$, then the mean and covariance functions are defined over domains $\mathcal{Y}$ and $\mathcal{Y} \times \mathcal{Y}$ with values notated such as $\mu_{\mathrm{x}}(y)$ and $\Sigma_{\mathrm{x}}(y, y')$.

If the expectations are evaluated with respect to a conditional distribution $\mathrm{P}_{\mathrm{x}|\mathrm{z}}$, the additional argument for the observed random element is added and the notation for the above operators is extended to $\mathrm{E}_{\mathrm{x}|\mathrm{z}}\big[g(\mathrm{x})\big](y; z)$ and $\mathrm{C}_{\mathrm{x}|\mathrm{z}}\big[g(\mathrm{x})\big](y, y'; z)$ for nonscalar outputs. When g is the identity operator, the notation $\mu_{\mathrm{x}|\mathrm{z}}(y; z)$ and $\Sigma_{\mathrm{x}|\mathrm{z}}(y, y'; z)$ is used. As for probability distributions, it is common for the conditional random element z to be left as a random quantity instead of being explicitly defined; in these cases, the dependency on z is implied.

References

[1] Paul Cohen. *Set theory and the continuum hypothesis*. Dover Publications, Inc., 2008. ISBN: 0486469212.

[2] P. A. M. Dirac. *The Principles of Quantum Mechanics*. International series of monographs on physics (Oxford, England). Clarendon Press, 1930.

[3] R. M. Dudley. *Real Analysis and Probability*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002. DOI: 10.1017/CBO9780511755347.

[4] Athanasios Papoulis and S. Unnikrishna Pillai. *Probability, Random Variables and Stochastic Processes*. Fourth. McGraw-Hill, 2002.

[5] Walter Rudin. *Real and Complex Analysis, 3rd Ed.* USA: McGraw-Hill, Inc., 1987. ISBN: 0070542341.