Dissertation Notation System

Paul Rademacher

February 10, 2023

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.

For consistency, various symbols will be reserved for a single purpose throughout this work. Table 1 lists select symbols; more detail on each can be found throughout this document.

1 Set and Function Conventions

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

Table 1: Select Reserved Symbols

$\times$ or $\prod_i \mathcal{S}_i$	Cartesian product
$\otimes$ or $\bigotimes_i f_i$	Outer product
diag	Diagonal operator
dim	Dimensionality operator
χ	Indicator function
$\mathcal{P}(\mathcal{X})$	Space of distributions over $\mathcal{X}$
supp	Support operator
$\delta[\cdot,\cdot]$	Kronecker delta function
$\delta(\cdot)$	Dirac delta function
ν	Multinomial coefficient function
β	Generalized beta function
P_{x}	Probability mass function of x
p_{x}	Probability density function of x
$\mathrm{Pr}(\cdot)$	Probability operator
$\perp\!\!\!\perp$ or $\perp\!\!\!\perp_i \mathrm{x}_i$	Probabilistic independence
$\xrightarrow{p}$	Convergence in probability
E_{x}	Expectation w.r.t. x
C_{x}	Covariance w.r.t. x
μ_{x}	Expected value of x
Σ_{x}	Covariance of x

$\mathcal{Y}^N$ 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 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 $\aleph_1$.

2 Special Operators and Functions

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 $\prod_i \mathcal{S}_i = \mathcal{S}_1 \times \mathcal{S}_2 \dots$ is employed.

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 $\bigotimes_i f_i = f_1 \otimes f_2 \dots$ for $i = 1, \dots$, where $\Big(\bigotimes_i f_i\Big)(x_1, x_2, \dots) = f_1(x_1)f_x(x_2)\dots$, is used as well. Also, the diagonal operator operates on a single real-valued function such that $\operatorname{diag}(f) \in \mathbb{R}^{\mathcal{X} \times \mathcal{X}}$. For countable sets $\mathcal{X}$, the operator values are $\operatorname{diag}(f)(x, x') = f(x)\delta[x, x']$; for Euclidean sets, the operator values are $\operatorname{diag}(f)(x, x') = f(x)\delta(x - x')$.

Numerous set functions are used throughout. The dim operator returns the dimensionality of a space (e.g. $\dim(\mathbb{R}^2) = 2$). For a given subset $\mathcal{S} \subset \mathcal{X}$, the indicator function $\chi(\mathcal{S}) : \mathcal{X} \mapsto \{0, 1\}$ is defined as

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

Numerous probability distribution functions are defined over different domains. As such, for a given set $\mathcal{X}$, a set function $\mathcal{P}$ is defined such that $\mathcal{P}(\mathcal{X})$ is the set of distributions over $\mathcal{X}$. If $\mathcal{X}$ is countable, the set is defined as $\mathcal{P}(\mathcal{X}) = \big\{p \in {\mathbb{R}_{\geq 0}}^{\mathcal{X}} : \sum_{x \in \mathcal{X}} p(x) = 1\big\}$; if $\mathcal{X}$ is a Euclidean space, the set is defined as $\mathcal{P}(\mathcal{X}) = \big\{p \in {\mathbb{R}_{\geq 0}}^{\mathcal{X}} : \int_{\mathcal{X}} p(x) \mathrm{d}x = 1\big\}$. Probability functions will be frequently used with the support operator to represent the random values with nonzero probability; for a given function $g \in \mathbb{R}^{\mathcal{X}}$, define

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

returning the subset of the function domain for which the mapped values are nonzero.

Both the Kronecker and Dirac delta functions are frequently required. As both functions are denoted by the symbol δ, they are distinguished by the use of square brackets or parentheses. The Kronecker delta function is defined as

$$\delta[x, x'] = \begin{cases} 1 & \text{if } x = x' \ , \\ 0 & \text{if } x \neq x' \ . \end{cases} \tag{3}$$

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 $\int_{\mathcal{X}} \delta(x)\mathrm{d}x = 1$. Consequently, it also satisfies $\int_{\mathcal{X}} g(x)\delta(x)\mathrm{d}x = g(0)$.

Commonly, a rectangular function definition of the Dirac delta function will be employed, such that $\delta(x) = \lim_{\Delta\to 0} \Delta^{-1}\chi\big(x; [0, \Delta)\big)$. Note that the Kronecker delta can similarly be represented as $\delta[x, 0] = \lim_{\Delta\to 0} \chi\big(x; [0, \Delta)\big)$. As a result, the relation $\delta[\cdot, x] = \delta(0)^{-1}\delta(\cdot - x)$ can be employed, where the concept of $\delta(0)$ is used for convenience.

To formalize the use of $\delta(0)$, define $\delta(0)^{-1}f(x) \equiv \lim_{\Delta\to 0} \int_x^{x+\Delta} f(x')\mathrm{d}x'$ for notational brevity. To motivate this representation, adopt the rectangular function form of the Dirac delta function and note that

$$\delta(0)^{-1}f(x) = \lim_{\Delta\to 0} \frac{\Delta^{-1} \int_x^{x+\Delta} f(x')\mathrm{d}x'}{\Delta^{-1}} \ . \tag{4}$$

In the limit $\Delta \to 0$, the denominator tends to $\delta(0)$ by definition. If f is bounded, the numerator tends to $f(x)$ and the entire expression tends to zero. However, if f includes Dirac functions itself, the expression may be nonzero. For example, if $f(x) = \sum_n c_n\delta(x - X_n)$, use of the rectangular delta function definition and substitution into the formula yields $\delta(0)^{-1}f(x) = \sum_n c_n\delta[x, X_n]$. This further motivates the relation between the Dirac and Kronecker deltas suggested above. A valuable use case for this notation is when f is a probability density function – if a random variable is distributed as $\mathrm{x} \sim f$, observe that $\delta(0)^{-1} f(x) \equiv \Pr(\mathrm{x} = x)$.

The multinomial coefficient and multivariate beta function, which typically operate on sequences, are defined more generally for function over countable domains. The multinomial operator ν 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

$$\nu(g) = \frac{\left(\sum_{x \in \mathcal{X}} g(x)\right)!}{\prod_{x \in \mathcal{X}} g(x)!} . \tag{5}$$

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

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

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 $\mathrm{P}_{\mathrm{x}} \in \mathcal{P}(\mathcal{X})$. 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 $\mathrm{p}_{\mathrm{x}} \in \mathcal{P}(\mathcal{X})$. The probability operator Pr relates the probability of events to these functions, such that $\Pr(\mathrm{x} \in \mathcal{S}) = \sum_{x \in \mathcal{S}} \mathrm{P}_{\mathrm{x}}(x)$ for countable domains and $\Pr(\mathrm{x} \in \mathcal{S}) = \int_{\mathcal{S}} \mathrm{p}_{\mathrm{x}}(x) \mathrm{d}x$ for Euclidean domains. Commonly, random variables are described using notation $\mathrm{x} \sim f$, where f is a valid distribution function; this indicates that $\mathrm{P}_{\mathrm{x}} = f$.

Next, consider x conditioned on another random element $\mathrm{z} \in \mathcal{Z}$. The conditional distribution is represented as $\mathrm{P}_{\mathrm{x}|\mathrm{z}} : \mathcal{Z} \mapsto \mathcal{P}(\mathcal{X})$ 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. Commonly, the dependency on the conditional variable z will typically not be expressed in terms of a specific value $z \in \mathcal{Z}$, but will be left in terms of the random element itself. In these cases, the more compact notation $\mathrm{P}_{\mathrm{x}|\mathrm{z}}$ is used to imply $\mathrm{P}_{\mathrm{x}|\mathrm{z}}(\mathrm{z})$. This is especially useful when using expectations. For example, the forms $f(\mathrm{z}) = \mathrm{E}_{\mathrm{x}|\mathrm{z}}[g(\mathrm{x})]$ and $\mathrm{P}_{\mathrm{x}} = \mathrm{E}_{\mathrm{z}}[\mathrm{P}_{\mathrm{x}|\mathrm{z}}]$ exclude the z function arguments without loss of clarity. Also, this convention enables conditional distributions to be defined using notation such as $\mathrm{x} \mid \mathrm{z} \sim f(\mathrm{z})$, which infers $\mathrm{P}_{\mathrm{x}|\mathrm{z}} = f(\mathrm{z})$.

Ever need the full functional?

The notion of probabilistic independence is used extensively. The independence of two random variables x and y is notated as $\mathrm{x} \perp\!\!\!\perp \mathrm{y}$. Formally, this indicates that the joint distribution can be factored as $\mathrm{P}_{\mathrm{x},\mathrm{y}} = \mathrm{P}_{\mathrm{x}} \otimes \mathrm{P}_{\mathrm{y}}$, such that $\mathrm{P}_{\mathrm{x},\mathrm{y}}(x, y) = \mathrm{P}_{\mathrm{x}}(x)\mathrm{P}_{\mathrm{y}}(y)$. Similarly, conditional independence on another random variable z is denoted $(\mathrm{x} \perp\!\!\!\perp \mathrm{y}) \mid \mathrm{z}$, implying $\mathrm{P}_{\mathrm{x},\mathrm{y}|\mathrm{z}}(z) = \mathrm{P}_{\mathrm{x}|\mathrm{z}}(z) \otimes \mathrm{P}_{\mathrm{y}|\mathrm{z}}(z)$. To express the independence of a sequence of variables $(\ldots, \mathrm{x}_i, \ldots)$ from one another, the notation $\perp\!\!\!\perp_i \mathrm{x}_i$ is used.

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 $\lim_{n\to\infty} \Pr\big(|\mathrm{x}_n - \mathrm{y}\ | < \epsilon\big) = 0$ 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 PDFs and PMFs 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 $\rho \in \mathcal{P}(\mathcal{X})$; the PMF will be notated as $\mathrm{Emp} : \mathbb{Z}_{\geq 0} \times \mathcal{P}(\mathcal{X}) \mapsto \mathcal{P}(\mathcal{X})$, where the range is the set of valid PMFs. 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, respectively.

3.1 Expectations

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

$$\mathrm{E}_{\mathrm{x}}\big[g(\mathrm{x})\big] = \sum_{x} \mathrm{P}_{\mathrm{x}}(x) g(x) \ , \tag{7}$$

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

$$\mathrm{C}_{\mathrm{x}}\big[g(\mathrm{x})\big] = \mathrm{E}_{\mathrm{x}}\left[\Big(g(\mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(\mathrm{x})\big]\Big)^2\right] \ . \tag{8}$$

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

$$\mathrm{E}_{\mathrm{x|z}}\big[g(\mathrm{x})\big](z) = \sum_{x} \mathrm{P}_{\mathrm{x|z}}(x \mid z) g(x) \ . \tag{9}$$

Similarly, the conditional variance is notated $\mathrm{C}_{\mathrm{x|z}}\big[g(\mathrm{x})\big](z)$. When g is the identity operator, the conditional mean and variance as represented by $\mu_{\mathrm{x|z}}(z)$ and $\Sigma_{\mathrm{x|z}}(z)$, 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 $\mathrm{E}_{\mathrm{x|z}}\big[g(\mathrm{x})\big]$ implies the dependency on z. This convention also holds for the conditional variance operator $\mathrm{C}_{\mathrm{x|z}}$, as well as for the $\mu_{\mathrm{x|z}}$ and $\Sigma_{\mathrm{x|z}}$ 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

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

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

$$\mathrm{C}_{\mathrm{x}}\big[g(\mathrm{x})\big](y,y') = \mathrm{E}_{\mathrm{x}}\left[\Big(g(y;\mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(y;\mathrm{x})\big]\Big)\Big(g(y';\mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(y';\mathrm{x})\big]\Big)\right] . \tag{11}$$

$$\mathrm{C}_{\mathrm{x}}\big[g(\mathrm{x})\big] = \mathrm{E}_{\mathrm{x}}\left[\Big(g(\mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(\mathrm{x})\big]\Big) \otimes \Big(g(\mathrm{x}) - \mathrm{E}_{\mathrm{x}}\big[g(\mathrm{x})\big]\Big)\right] . \tag{12}$$

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.