Notation

Last modified: 2026-11-18 21:58:23 (PST)

Mathematical notation is not standardized. This section states the conventions these notes use, and the alternatives you may encounter in other sources.

Table 1: Notation used in this book

symbol	meaning	LaTeX
$\neg$	not	`\neg`
$\forall$	for all	`\forall`
$\exists$	there exists	`\exists`
$\cup$	union, "or"	`\cup`
$\cap$	intersection, "and"	`\cap`
$\mid$	given, conditional on	`\mid`, `|`
$\sum$	sum	`\sum`
$\prod$	product	`\prod`
μ	population mean, $\mathbb{E}[X]$, of a random variable X	`\mu`
$\mathbb{E}$	expectation	`\mathbb{E}`
$x^\top$	transpose of x	`x\'`
$'$	transpose or derivative[1]	`'`
$\perp\!\!\!\perp$	independent	`\perp\!\!\!\perp`
$\therefore$	therefore, thus	`\therefore`
η	linear predictor of a generalized linear model[2]	`\eta`
$\lfloor x \rfloor$	floor of x: largest integer less than or equal to x	`\lfloor x \rfloor`
$\lceil x \rceil$	ceiling of x: smallest integer greater than or equal to x	`\lceil x \rceil`
$\mathbb{1}_A(x)$, $\mathbb{1}(P)$	indicator function (definition): 1 if condition holds, 0 otherwise	`\indic{A}(x)`, `\indicp{P}`

[1] depending on whether it is applied to a matrix or a function

[2] A generalized linear model (GLM) is a regression model in which a chosen function g, the link function, of the population mean μ of the outcome equals the linear predictor $\eta = x^\top\beta$, a linear combination of the predictor values x with coefficients β: $g(\mu) = \eta$.

The links for conditioning, random variables, expectation, and independence go to the Morrison Lab's probability notes, which define those terms.

Writing math in Quarto

The third column of Table 1 gives the LaTeX command for each symbol. Quarto and R Markdown documents write math in LaTeX syntax: put inline math between single dollar signs, as in `$x^2$`, and displayed equations between double dollar signs, as in `$$x^2$$`. The equations section of Quarto's Markdown guide shows the details. These notes also use shorthand macros, such as `\floor{x}` for $\lfloor x \rfloor$, defined in the `latex-macros` submodule's `macros.qmd`.

Subscripts and parentheses

> *Remark* 1 (Our convention for subscripts and parentheses). Subscripts and parentheses both attach extra information to a symbol, and sources divide the work between them differently. These notes follow two rules of thumb:
>
> - A subscript names which member of an indexed family we mean. It holds an index, such as the i in x_i, or a parameter that picks one function or distribution out of the family, such as the θ in f_θ.
> - Parentheses hold the input a function is evaluated at, as in $f(x)$.
>
> Under these rules, $f_\theta(x)$ reads "the member of the family labeled θ, evaluated at x". The notation choices page gives the reason for this split and explains how to read a source that divides the work differently.

The Notation Across Sources page of the macros site tabulates our form of each object beside the other forms you may encounter.

Example 1 (Reading $X_t(\omega)$)**.** Let X_t be the position of a particle at time $t \geq 0$. Then X_3 is the random position at time 3, and $X_3(\omega)$ is its value when the experiment's outcome is ω. Holding ω fixed and letting t vary gives the particle's whole path for that outcome. A source that writes $X(t, \omega)$ means the same thing, with both arguments in parentheses.

Why is notation in probability and statistics so inconsistent and disorganized?

In grad school, we are asked to learn from increasingly disorganized materials and lectures. Not coincidentally, as the amount of organization decreases, the amount of complexity increases, the amount of difficulty increases, the number of reliable references decreases, and the amount of inconsistency in notation and content increases (both between multiple references and within single references!). In other words, as you approach the cutting-edge of most fields, you start to run into content that hasn't been fully thought through or standardized. This lack of clarity is unfortunate and undesirable, but it is understandable and inevitable.

It's worth noting that calculus was formalized in the 1600s, elementary algebra was formalized around 820, and arithmetic even earlier. And calculus still has several competing notation systems. In contrast, the field of statistics only emerged in the late 1800s and early 1900s, so it's not surprising that the notation and terminology is still developing. Generalized linear models were only formalized in 1972 (Nelder and Wedderburn 1972), which is very recent in terms of the pace of scientific development.

Further reading

For the logical symbols and proof conventions used in these notes (whose definitions are on the Proof Writing page):

- Velleman (2019) explains how mathematical statements are built from logical connectives (definition) and quantifiers (definition), and how to read and write them.
- Barker-Plummer et al. (2011) covers the same symbols as a formal language, with propositional logic (the logic of connectives) and first-order logic (which adds quantifiers).

For set and function notation, see the Sets, Functions, and Expressions and Algorithms pages.

References

Barker-Plummer, Dave, Jon Barwise, and John Etchemendy. 2011. *Language, Proof and Logic*. 2nd ed. CSLI Publications. https://www.amazon.com/dp/1575866323.

Nelder, John Ashworth, and Robert WM Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society Series A: Statistics in Society* 135 (3): 370–84. https://doi.org/10.2307/2344614.

Velleman, Daniel J. 2019. *How to Prove It: A Structured Approach*. 3rd ed. Cambridge University Press. https://doi.org/10.1017/9781108539890.