

# Notation

## Contents

<b>1</b>	<b>Information matrices</b>	<b>2</b>
<b>2</b>	<b>Percent sign (“%”)</b>	<b>2</b>
<b>3</b>	<b>Proofs</b>	<b>2</b>
<b>4</b>	<b>Stochastic vs. probabilistic vs. random</b>	<b>2</b>
4.1	Key distinction: modeling approach vs. phenomena . . . . .	3
4.2	Definitions . . . . .	3
4.3	Summary of usage . . . . .	3
4.4	Additional resources . . . . .	3
<b>5</b>	<b>Why is notation in probability and statistics so inconsistent and disorganized?</b>	<b>4</b>
	<b>References</b>	<b>4</b>

Table 1: Notation used in this book

symbol	meaning	LaTeX
$\neg$	not	<code>\neg</code>
$\forall$	all	<code>\forall</code>
$\exists$	some	<code>\exists</code>
$\cup$	union, “or”	<code>\cup</code>
$\cap$	intersection, “and”	<code>\cap</code>
$ $	given, conditional on	<code>\mid,  </code>
$\sum$	sum	<code>\sum</code>
$\prod$	product	<code>\prod</code>
$\mu$	mean	<code>\mu</code>
$E$	expectation <sup>1</sup>	<code>\mathbb{E}</code>
$x^\top$	transpose of $x$	<code>x^{\top}</code>
$'$	transpose or derivative <sup>2</sup>	<code>'</code>
$\perp$	independent <sup>3</sup>	
$\therefore$	therefore, thus	<code>\therefore</code>
$\eta$	linear component of a GLM <sup>4</sup>	<code>\eta</code>
$\lfloor x \rfloor$	floor of $x$ : largest integer smaller than $x$	<code>\lfloor x</code> <code>\rfloor</code>
$\lceil x \rceil$	ceiling of $x$ : smallest integer larger than $x$	<code>\lceil x</code> <code>\rceil</code>

<sup>1</sup>[probability.qmd#def-expectation](#)

<sup>2</sup>depending on whether it is applied to a matrix or a function

<sup>3</sup>[probability.qmd#def-indpt](#)

<sup>4</sup>[https://en.wikipedia.org/wiki/Generalized\\_linear\\_model#:~:text=The%20linear%20predictor%20is%20the,data%20through%20the%20link%20function](https://en.wikipedia.org/wiki/Generalized_linear_model#:~:text=The%20linear%20predictor%20is%20the,data%20through%20the%20link%20function)

# 1 Information matrices

There is no consistency in the notation for observed and expected information matrices (see Table 2).

Table 2: notation for information matrices

book	observed information	expected information
Dobson and Barnett (2018)	$U'$	$\mathcal{J}$
Dunn and Smyth (2018)	$\mathcal{J}$	$\mathcal{J}$
McLachlan and Krishnan (2007)	$I$	$\mathcal{J}$
Wood (2017)	$\hat{I}$	$\mathcal{J}$

These notes currently have a mixture of notations, depending on my whims and what reference I had last looked at. Eventually, I will try to standardize my notation to  $I$  for observed information and  $\mathcal{J}$  for expected information.

## 2 Percent sign (“%”)

The percent sign “%” is just a shorthand for “/100”. The word “percent” comes from the Latin “per centum”; “centum” is Latin for 100, so “percent” means “per hundred” (c.f., <https://en.wikipedia.org/wiki/Percentage>)

So, contrary to what you may have learned previously,  $10\% = 0.1$  is a true and correct equality, just as  $10\text{kg} = 10,000\text{g}$  is true and correct.

**i** Proof

*Proof.*

$$\begin{aligned} 10\% &= 10/100 \\ &= \frac{10}{100} \\ &= 0.1 \end{aligned}$$

□

You are welcome to switch between decimal and percent notation freely; just make sure you execute it correctly.

## 3 Proofs

We can use any of:

- $\therefore$  (`\therefore` in LaTeX),
- $\Rightarrow$  (`\Rightarrow`),
- $\vDash$  (`\models`)

to denote logical entailments (deductive consequences).

Let's save  $\rightarrow$  (`\rightarrow`) for convergence results.

## 4 Stochastic vs. probabilistic vs. random

The terms “stochastic”, “probabilistic”, and “random” are frequently used in statistics and probability theory, often interchangeably in everyday conversation, but they carry nuanced technical distinctions.

## 4.1 Key distinction: modeling approach vs. phenomena

As noted in Wikipedia<sup>5</sup>:

*Stochasticity* and *randomness* are technically distinct concepts: the former refers to a modeling approach, while the latter describes phenomena; in everyday conversation these terms are often used interchangeably.

## 4.2 Definitions

**Random** describes something that occurs by chance, without a deterministic pattern. It is the most general term, used to describe variables or occurrences whose outcome cannot be predicted precisely, only probabilistically. For example, we speak of “random variables” and “random events”.

### **i** Note

The term “random” is sometimes used as shorthand for a uniform distribution (especially the discrete uniform distribution), but it can refer to any probability distribution.

**Stochastic** comes from the Greek “στόχος” (*stókhos*), meaning “aim” or “guess” (see etymology<sup>6</sup>). In mathematics, a **stochastic process** is formally defined as a collection of random variables indexed by time or space. The term is almost always used in the context of processes or systems evolving in time or space under uncertain rules. Note that in probability theory, “stochastic process” and “random process” are synonyms (Adler and Taylor 2009; Stirzaker 2005; Kallenberg 2002).

**Probabilistic** refers to any model, reasoning, or method that explicitly involves probability theory. Probabilistic models assign probabilities to events or outcomes; they focus on quantifying and reasoning about uncertainty based on known or estimated distributions. While all stochastic models are probabilistic (since they use probabilities), not all probabilistic models need to describe processes evolving in time.

## 4.3 Summary of usage

Table 3: Comparison of “random”, “stochastic”, and “probabilistic”

Term	What it describes	Typical use	Example
Random	Single variable or event	Random variable, random outcome	Coin toss, die roll
Stochastic	System or process in time/space	Stochastic process	Stock price evolution, Markov chain
Probabilistic	Approach/model using probability	Probabilistic model/reasoning	Bayesian inference, regression

While some sources treat “stochastic” and “random” as practically synonymous, the academic preference is to use “random” for variables and events, and “stochastic” for processes, especially to highlight temporal or spatial structure in the modeling.

## 4.4 Additional resources

- Wikipedia: Stochastic<sup>7</sup>
- Wikipedia: Stochastic process<sup>8</sup>
- Mathematics Stack Exchange: What’s the difference between stochastic and random?<sup>9</sup>
- Cross Validated: Probability model vs statistical model vs stochastic model<sup>10</sup>

<sup>5</sup><https://en.wikipedia.org/wiki/Stochastic>

<sup>6</sup><https://www.etymonline.com/search?q=stochastic>

<sup>7</sup><https://en.wikipedia.org/wiki/Stochastic>

<sup>8</sup>[https://en.wikipedia.org/wiki/Stochastic\\_process](https://en.wikipedia.org/wiki/Stochastic_process)

<sup>9</sup><https://math.stackexchange.com/questions/114373/whats-the-difference-between-stochastic-and-random>

<sup>10</sup><https://stats.stackexchange.com/questions/421462/probability-model-vs-statistical-model-vs-stochastic-model>

## 5 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 encounter 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<sup>11</sup>, elementary algebra was formalized around 820<sup>12</sup>, and arithmetic even earlier<sup>13</sup>. And calculus still has several competing notation systems<sup>14</sup>. In contrast, the field of statistics only emerged in the late 1800s and early 1900s<sup>15</sup>, 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<sup>16</sup>.

### References

- Adler, Robert J., and Jonathan E. Taylor. 2009. *Random Fields and Geometry*. Springer. <https://doi.org/10.1007/978-0-387-48116-6>.
- Dobson, Annette J, and Adrian G Barnett. 2018. *An Introduction to Generalized Linear Models*. 4th ed. CRC press. <https://doi.org/10.1201/9781315182780>.
- Dunn, Peter K, and Gordon K Smyth. 2018. *Generalized Linear Models with Examples in R*. Vol. 53. Springer. <https://link.springer.com/book/10.1007/978-1-4419-0118-7>.
- Kallenberg, Olav. 2002. *Foundations of Modern Probability*. 2nd ed. Springer. <https://doi.org/10.1007/978-1-4757-4015-8>.
- McLachlan, Geoffrey J, and Thriyambakam Krishnan. 2007. *The EM Algorithm and Extensions*. 2nd ed. John Wiley & Sons. <https://doi.org/10.1002/9780470191613>.
- 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.
- Stirzaker, David. 2005. *Stochastic Processes and Models*. Oxford University Press. <https://global.oup.com/academic/product/stochastic-processes-and-models-9780198568131>.
- Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with r*. Chapman; hall/CRC.

---

<sup>11</sup>[https://en.wikipedia.org/wiki/Leibniz%27s\\_notation](https://en.wikipedia.org/wiki/Leibniz%27s_notation)

<sup>12</sup><https://en.wikipedia.org/wiki/Al-Jabr>

<sup>13</sup><https://en.wikipedia.org/wiki/Arithmetic#History>

<sup>14</sup>[https://en.wikipedia.org/wiki/Notation\\_for\\_differentiation](https://en.wikipedia.org/wiki/Notation_for_differentiation)

<sup>15</sup>[https://en.wikipedia.org/wiki/History\\_of\\_statistics#Development\\_of\\_modern\\_statistics](https://en.wikipedia.org/wiki/History_of_statistics#Development_of_modern_statistics)

<sup>16</sup>[https://en.wikipedia.org/wiki/The\\_Structure\\_of\\_Scientific\\_Revolutions](https://en.wikipedia.org/wiki/The_Structure_of_Scientific_Revolutions)