# A Short Guide to Typesetting Math in NLP Papers

**Chris Dyer (DeepMind), Kevin Gimpel (TTIC), and Noah A. Smith (UW)**

## Abstract

Sloppy typesetting is a good way to turn off your reviewers (they'll think you're an idiot or lazy, or a lazy idiot), get less useful attention from your coauthors (they don't want to fix your mess), and be ignored by your readers. This document describes an easy convention to follow that works well in the authors' experience.

## 1 Introduction

The advice given in this document outlines one good typesetting convention among many. It is meant to be a short list of examples of good and consistent style, not to provide exhaustive instruction for every typesetting challenge. The style is not necessarily appropriate for every paper (e.g., vectors and matrices are in bold face, and this "weightiness" could be distracting if every variable you use is a vector or matrix)—at the end of the day, you must rely on your own aesthetic judgment. But if you deviate, you should be consistent.

If one can speak of such things as being principled, the style we recommend here is based on the principle that typefaces should be used to communicate information about the type of the symbol to the reader, a bit like the way Hungarian notation[1] for variable naming is used to communicate type information to people reading source code.

## 2 Basics

- Use either `\begin{align}` ...`\end{align}` or `align*` for equations; do not use `equation` or `eqnarray`. Only number equations that you refer to in the text. Use `$...$` for inline math.
- If you subscript or superscript symbols with text, use a text environment like `text` or `textit`. It is bad to write $p_{base}(x)$ instead of $p_{base}(x)$. Yes, people notice this. It is especially noticeable with particular characters (like $f$) due to LaTeX's expectations of their use: compare $p_{fact}(x)$ and $p_{fact}(x)$.
- Use ... (`\ldots`) instead of three periods (...) for ellipsis. A good list is $1, 2, \ldots, n$. A bad list is $1, 2 \ldots n$. In sequences of operations, $\cdots$ (`\cdots`) is often appropriate.

If an equation or sequence of equations is part of a sentence, even if it is not inline, e.g.,

$$\mu_t = 1 + f(t)$$
$$p(x_t \mid \mu_t, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x_t - \mu_t)^2}{2\sigma^2}},$$

then the last line should probably also end with a comma or period. There is some room for disagreement on this; is the comma after the exponent a bit distracting? Use your judgment, and be consistent.

[1] `https://en.wikipedia.org/wiki/Hungarian_notation`

## 3 Typefaces

### 3.1 Variables

- Use lowercase Latin letters for scalar values; $a, b, c$ preferentially for symbols, $i, j, k, \ell, m, n$ preferentially for whole numbers/integers (prefer $\ell$ to $l$ since the latter is confusable for 1), and $e, f, w, x, y, z$ for the values taken by random variables/predicted by models (be they symbols, whole numbers, or reals).

- To indicate a structured value, like a sequence or tree, use lowercase `boldsymbol` Latin letters. For example, the input sentence $\boldsymbol{x}$, and the predicted tree $\boldsymbol{y}$.

- To indicate a vector (column usually, but row if you are consistent), use lowercase `mathbf` Latin letters $\mathbf{u}$, $\mathbf{v}$. If you want to use Greek letters to indicate a vector (often used for parameters/hyperparameters), use lowercase `boldsymbol` Greek letters, e.g. $\boldsymbol{\theta}_0$, $\boldsymbol{\varphi}$, or $\boldsymbol{\pi}$.

- To indicate a matrix or higher-order tensor, use capital `mathbf` Latin letters, e.g., $\mathbf{U}$, or capital `boldsymbol` Greek letters when convention dictates, e.g., $\boldsymbol{\Sigma}^{-1}$.

- An element of a vector, a sequence, or matrix is (usually) a scalar, so its typeface should reflect this. So we write that $v_i$ is the $i$th element of $\mathbf{v}$, $m_{ij}$ (or perhaps $m_{i,j}$) is an element of $\mathbf{M}$, and $w_1$ is the first word in sentence $\boldsymbol{w}$. Similarly, slices of sequences are themselves sequences, so we write $\boldsymbol{w}_{\leq t}$ and $\boldsymbol{z}_{-i}$ (this convention is used to designate all of the values in $\boldsymbol{z}$ but the $i$th).

- Sets of structured values are often $\mathcal{X}$ or $\mathcal{Y}$. If you have sets of symbols (or lots of sets, e.g., in a formal language paper), using capital Latin letters is the standard convention, e.g., $A, B, C$. Standard sets like $\mathbb{R}$, $\mathbb{Z}$, and $\Sigma$ have their own conventions that should be followed. Note that $\mathbb{R}$ is strongly preferable to $\Re$ or $R$ when denoting real values.

- Scalar-valued random variables are sometimes capital letters, e.g., $X, Y, Z$. Random variables that take a vector, matrix or sequence value are either capital letters or `boldsymbol` capital letters $\boldsymbol{X}$. But a specific value taken by a random variable follows the standard conventions. Thus we can write $p(X = x_0)$, although we would just write $p(x_0)$ if the random variable in question is clear from context.

- When defining ordered pairs or tuples, use parentheses, e.g., $(a, b)$, or `langle`/`rangle`, e.g., $\langle a, b \rangle$. Either is OK if consistent. Do not use angle brackets: $< a, b >$.

### 3.2 Probability

- $P$ or $p$ or $\Pr$ or $\mathbb{P}$? If you are talking about a particular probability mass/density function (e.g., a language model or something), definitely $p$. Actually, almost always $p$ if you are talking about models. If you are writing about probabilities and event sets, get advice elsewhere.

- Use `\mid` instead of |, e.g. $p(x_i \mid \boldsymbol{x}_{<i})$ rather than $p(x_i|\boldsymbol{x}_{<i})$.

- Parameters can be subscripted or separated by a semicolon, e.g. $p_{\boldsymbol{\theta}}(y \mid x)$ or $p(y \mid x; \boldsymbol{\theta})$.

- Use $\mathbb{E}_{p(X)}[f(X)]$ for expectations, which is produced by `\mathbb{E}`.

- Use $\sim$ (`\sim`) to indicate "is drawn from" or "is distributed according to", e.g., $\boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\alpha})$ or $w_i \sim \text{Categorical}(\boldsymbol{\theta}_i)$. The names of distributions should be in `\mathrm` environments, except where standard conventions have precedence, like $\mathcal{N}(\mu, \sigma^2)$ for a normal distribution.

### 3.3 Operations

- Don't confuse $\Pi$ (`\Pi`) and $\prod$ (`\prod`). The latter is used for products.
- The same goes for $\Sigma$ (`\Sigma`) and $\sum$ (`\sum`). The latter is used for sums.
- If LaTeX has a macro for a multi-letter function, e.g., `\tanh` which produces $\tanh$, use that. If it doesn't use `\mathrm{softmax}`, which produces $\mathrm{softmax}$. Never write $softmax(\mathbf{x})$ or $tanh(x)$ or $log(x)$.
- A function like $\mathit{score}(x)$ should be `$\textit{score}(x)$`, never `$score(x)$`. See Section 2.