Examples of bad notation

Examples of bad notation in "G. Casella & R.L. Berger (1990). Statistical Inference. Duxbury Press, Belmont, CA.", the Stat 461 text used in the year 2000.

1. page 137, bottom:

$$f(y|x) = \Pr(Y = y|X = x) = f(x, y)/f_X(x)$$

$$f(x|y) = \Pr(X = x|Y = y) = f(x, y)/f_Y(y)$$

Why bad? Although these equations can be understood from the context, this is bad notation. For example, what is the meaning of $f(1|2)$? We don't know without x, y whether this means that first or the second conditional probability. The notation is bad because the function symbol 'f' is used for three different probability mass functions: two conditional and one joint.

Better notation would be $f_{Y|X}$ and $f_{X|Y}$ for the two conditional pmf's. Note that functions are themselves objects in some (function) space, so symbols for them should be understood without arguments.

2. page 221, bottom displayed equation.

$$f(y_1, ...y_n) = \cdots \exp\Big\{-(1/2)(y_1 - \sum_{i=2}^{n} y_i)^2\Big\} \cdots -\infty < y_i < \infty$$

Why bad? y_i is useful inside a sum in the function definition (i is a dummy variable for the summation), and it is used as one of the arguments of the function. The domain in the definition of the function is incomplete. Better would be $-\infty < y_j < \infty$, $j = 1, \ldots, n$; or $(y_1, ..., y_n) \in \Re^n$, where $\Re$ is the real line.

3. page 298, Example 7.2.9: The joint distribution of Y and p is $f(y, p) = \cdots$.
Why bad? p is being used as a random variable and as an argument of a function [note that this is different for random variable Y and argument y].

4. page 386, Problem 8.5, and page 388.

   1. $X_i/\min_i X_i$
   2. $\pi(\theta|x) = f(x-\theta)\pi(\theta) \,/ \int f(x-\theta)\pi(\theta)d\theta$

   Why bad? i is used as both a real index (in numerator) and as a dummy variable (in denominator); θ is used as an argument of a function and as a dummy variable (of the integral). Better notation is

   1. $X_i/\min_{i'} X_{i'}$
   2. $\pi^*(\theta|x) = f(x-\theta)\pi(\theta) \,/ \int f(x-\theta')\pi(\theta')d\theta'$

5. page 468:

   $R = \{\mathbf{x} : \delta(\mathbf{x}) = a_1\}$

   $R(\theta, \delta) = \cdots$

   Why bad? R is denoting two different objects, a set and a function, in the same paragraph. Better notation would be to change the set R to a script letter or another symbol.

Examples 1 and 3 are common abuses of notation in discussion of Bayesian statistics; this is only acceptable if the writer says that notation is being abused and the meaning of any function or variable is clear from the written context.

Examples of bad typesetting and notation

- In math writing, math variables are in italics, and special functions are in roman font (not slanted).
  The following is not good: $E(y_{ij}) = \mu_{ij}$, $Var(y_{ij}) = v_{ij}$, and $Cov(y_{ij}, y_{ik}) = v_{ijk}$
  Special functions like $\sin, \exp, \log, \ln$ are not slanted. Define $\mathrm{E}\,, \mathrm{Var}\,, \mathrm{Cov}$ in a similar way to get: $\mathrm{E}\,(y_{ij}) = \mu_{ij}$, $\mathrm{Var}\,(y_{ij}) = v_{ij}$, and $\mathrm{Cov}\,(y_{ij}, y_{ik}) = v_{ijk}$.

- Inline fraction may be lead to a small font.
  The function $f^*(y_i) = \frac{f_R(r_i|z_i,b_i,v_i;\psi^{(t)})}{f_Y(y_{obs,i},r_i|z_i,b_i,v_i;\psi^{(t)})}$ is a constant with respect to ..
  Better is:
  The function $f^*(y_i) = f_R(r_i|z_i, b_i, v_i; \psi^{(t)})/f_Y(y_{obs,i}, r_i|z_i, b_i, v_i; \psi^{(t)})$ is a constant with respect to ..

- \cdots, \ldots, \ddots, \vdots (for center, lower, diagonal, vertical dots)
  $A_1 \times \ldots \times A_k$, $k = 1, \cdots, n$: should be $A_1 \times \cdots \times A_k$, $k = 1, \ldots, n$

$$\begin{pmatrix} 1 & \rho_1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_2 \\ \rho_1 & \rho_2 & 1 & \cdots & \rho_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_1 & \rho_2 & \rho_2 & \cdots & 1 \end{pmatrix} \quad \text{should be} \quad \begin{pmatrix} 1 & \rho_1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_2 \\ \rho_1 & \rho_2 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_1 & \rho_2 & \rho_2 & \cdots & 1 \end{pmatrix}$$

- Size of parentheses, brackets or braces:

$$\begin{aligned} \Omega_2 &= C_2 - \frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\Sigma_i|) \\ &\quad -\frac{1}{2\sigma^2}\int \frac{(y_i - X_i\beta - T_ib_i - T_i\Sigma_i^{1/2}k_i)^T(y_i - X_i\beta - T_ib_i - T_i\Sigma_i^{1/2}k_i)}{(\sqrt{2\pi})^s|\Sigma_i^{1/2}|} \\ &\quad \times\exp(-\frac{1}{2}k_i^Tk_i)d(b_i + \Sigma_i^{1/2}k_i) \end{aligned}$$

Use \half for 1/2 so that it appears smaller and like a single symbol (part of extended ASCII). Also use larger parentheses, brackets or braces, with \bigr, \bigl, \Bigl etc., if needed for readability, and use half-spaces between functions etc.

$$\begin{aligned} \Omega_2 &= C_2 - \tfrac{1}{2}n_i\log\sigma^2 - \tfrac{1}{2}\log(|\Sigma_i|) \\ &\quad -\frac{1}{2\sigma^2}\int \frac{\bigl(y_i - X_i\beta - T_ib_i - T_i\Sigma_i^{1/2}k_i\bigr)^T\bigl(y_i - X_i\beta - T_ib_i - T_i\Sigma_i^{1/2}k_i\bigr)}{\bigl(\sqrt{2\pi}\bigr)^s\bigl|\Sigma_i^{1/2}\bigr|} \\ &\quad \times\exp\bigl(-\tfrac{1}{2}k_i^Tk_i\bigr)\,d\bigl(b_i + \Sigma_i^{1/2}k_i\bigr) \end{aligned}$$

- Left-hand side of equation:

$$\begin{aligned}
& f(z_{mis,i}|z_{obs,i}, y_i, b_i, v_i, r_i; \psi^{(t)}) \\
= & \frac{f(z_i, y_i, b_i, r_i|v_i; \psi^{(t)})}{f(z_{obs,i}, y_i, b_i, r_i|v_i; \psi^{(t)})} \\
= & \frac{f(z_i|v_i; \psi^{(t)}) f(b_i|z_i, v_i; \psi^{(t)}) f(r_i|z_i, b_i, v_i; \psi^{(t)}) f(y_i|r_i, z_i, b_i, v_i; \psi^{(t)})}{f(z_{obs,i}, y_i, b_i, r_i|v_i; \psi^{(t)})} \\
\propto & f(b_i|\psi^{(t)}) f(r_i|b_i, z_i, v_i; \psi^{(t)}) f(y_i|b_i, r_i, z_i, v_i; \psi^{(t)})
\end{aligned}$$

The start of a multiline displayed equation should be on the left-hand side. If alignment on the first line with the equal sign doesn't fit, use \lefteqn as below. Note also the better use of spacing with \, for a half-space etc.

$$\begin{aligned}
\lefteqn{f(z_{mis,i}|z_{obs,i}, y_i, b_i, v_i, r_i; \psi^{(t)}) = \frac{f(z_i, y_i, b_i, r_i|v_i; \psi^{(t)})}{f(z_{obs,i}, y_i, b_i, r_i|v_i; \psi^{(t)})}} \\
= & \frac{f(z_i|v_i; \psi^{(t)})\, f(b_i|z_i, v_i; \psi^{(t)})\, f(r_i|z_i, b_i, v_i; \psi^{(t)})\, f(y_i|r_i, z_i, b_i, v_i; \psi^{(t)})}{f(z_{obs,i}, y_i, b_i, r_i|v_i; \psi^{(t)})} \\
\propto & f(b_i|\psi^{(t)})\, f(r_i|b_i, z_i, v_i; \psi^{(t)})\, f(y_i|b_i, r_i, z_i, v_i; \psi^{(t)})
\end{aligned}$$

Note that there is abuse of notation in this example, since f stands for many different densities.

- Bad notation

$$\begin{aligned}
& Y_{H1}, \ldots, Y_{Hk} \mid \pi \overset{iid}{\sim} \text{Bernoulli}(\pi) \\
& \pi \sim G(\,\cdot\,; \boldsymbol{\theta}),
\end{aligned}$$

Better is:

$$\begin{aligned}
& Y_{H1}, \ldots, Y_{Hk} \mid P = \pi \overset{iid}{\sim} \text{Bernoulli}(\pi) \\
& P \sim G(\,\cdot\,; \boldsymbol{\theta}),
\end{aligned}$$

Exercises: Each of the following example is a "math" sentence that can be improved; can you make the improvement.

- Suppose $X \sim F$. If for each c, $0 \le c \le 1$, we have

$$X \stackrel{d}{=} c * X + \epsilon_c = \sum_{i=1}^{X} I_i + \epsilon_c, \quad I_1, I_2, \ldots \text{i.i.d. Bernoulli}(c),$$

  where ϵ_c is independent of X, then F is said to be discrete self-decomposable.

- The proof is similar to Theorem 2.1.

- $X(t)$ can be viewed to be the sum of n independent random variables, all of which are distributed as $X(t/n)$.

- Consider K is a degenerate random variable.

- Proof: Apply the same reasoning of Theorem 6.1.

- If the distribution of $S_n + \beta_n$ tends to a probability distribution of U, then U is infinitely divisible.

- The first order term is $1/2\ n$.

My improvements are on the next page.

- Suppose $X \sim F$. If for each c, $0 \le c \le 1$, there exists ϵ_c independent of X, such that

$$X \stackrel{d}{=} c * X + \epsilon_c = \sum_{i=1}^{X} I_i + \epsilon_c, \quad I_1, I_2, \ldots \text{ i.i.d. Bernoulli}(c),$$

  then F is said to be discrete self-decomposable.

- The proof is similar to *that of* Theorem 2.1.

- $X(t)$ can be viewed to be the sum of n independent random variables, each having the distribution of $X(t/n)$.

- Let K be a degenerate random variable.

- Proof: Apply the same reasoning *as in the proof* of Theorem 6.1.

- If $S_n + \beta_n$ converges in distribution to U, then U is infinitely divisible.

- "The first order term is $n/2$"; or: "The first order term is $1/(2n)$", depending on where the multiplication is.