Local Regression and Likelihood
Additions and Corrections

`http://cm.bell-labs.com/stat/project/locfit/errors.pdf`
Last Updated: August 3, 2001.

This page lists mistakes, additions and clarifications to my book *Local Regression and Likelihood* (Springer, 1999). Please email any further issues to `catherine@research.bell-labs.com`.

- Page 39, line 7: ... based on the fit actually used, and not on ...
- Section 2.6 is, I believe, the simplest and most general presentation of asymptotic results for local regression that there is, although I took too much detail out in response to reviewers. Results such as (2.38), (2.39) and (2.41) are valid for any degree, and include boundary and near-boundary cases with appropriate restriction of the domains of integration.

  The bias formulae as stated are only for one dimension. Similar expressions in multiple dimensions are routine but messy to derive, and are the sum of terms for all partial derivatives of order $p+1$ and $p+2$. Ruppert and Wand (1994) give the local linear and quadratic results, in their (expletives deleted) notation. The general result was given by Loader (1996b) for density estimation; I'm unaware of a published reference for local regression.

  The leading $O(h^{p+1})$ term of the bias expansion is

$$\begin{aligned}&\mathrm{E}(\hat{\mu}(x)-\mu(x))\\ &= \frac{h^{p+1}}{(p+1)!}\mathbf{M}_1^{-1}\sum_{i_1=1}^{d}\cdots\sum_{i_{p+1}=1}^{d}\mu^{i_1,\ldots,i_{p+1}}(x)\int\prod_{j=1}^{p+1}v_{i_j}A(v)W(v)dv+o(h^{p+1}).\end{aligned}$$

  where $\mathbf{M}_1=\int A(v)A(v)^TW(v)dv$.

  For p even, this lead term is 0. Assume

$$\frac{S_n}{nh^d}=f(x)\int K(v)dv+h\sum_{j=1}^{d}f^{(j)}(x)\int v_jK(v)dv+o(h)$$

  for all bounded continuous K (for random designs, this requires $h\to 0$ and $nh^{d+2}\to\infty$), and the design density $f(\,.\,)$ is continuous, differentiable and non-zero at the fitting point x. Then

$$\begin{aligned}&\mathrm{E}(\hat{\mu}(x))-\mu(x)=h^{p+2}e_1^T\mathbf{M}_1^{-1}\\ &\quad\times\sum_{i_1=1}^{d}\cdots\sum_{i_{p+2}=1}^{d}\left(\frac{\mu^{(i_1,\ldots,i_{p+2})}(x)}{(p+2)!}+\frac{f^{(i_{p+2})}(x)\mu^{(i_1,\ldots,i_{p+1})}(x)}{f(x)(p+1)!}\right)\int\prod_{j=1}^{p+2}v_{i_j}A(v)W(v)dv\\ &\quad+o(h^{p+2}).\end{aligned}$$

- Chapter 5 (and elsewhere): In many of the density estimation figures (e.g. Figure 5.1) the data should be displayed by tick marks along the bottom – this hasn't reproduced very well. This is especially problematic in Figure 7.1, where censored observations should be shown by shorter ticks.

- Page 115: The formula in the middle of the page should be

$$\sum_{i=1}^{n} w_i(x) \left(\rho \left(\frac{Y_i - \langle a, A(x_i - x) \rangle}{s} \right) + \log(s) \right);$$

  note the $+\log(s)$.

- Chapter 7: I'm told Locfit's coding of the censoring argument (1=censored, 0=uncensored) violates the standard for survival analysis software. I won't change the code (to keep the book accurate), but please be careful!

- Page 128, example 7.4: The first `locfit.censor()` call is overkill - either call `locfit(...,lfproc=locfit.censor)` or `locfit.censor(...)` with no lfproc argument. This doesn't affect the results or Figure 7.3.

- Page 130, example 7.5: The second line should read

$$P(Y_i \geq y) = q^y = \frac{e^{\theta_i y}}{(1 + e^{\theta_i})^y}.$$

- Page 137, Ex 7.4: Second line should be

$$E(YY^*) = \Phi(c/\sigma) = P(Y < c).$$

- Page 142, middle: Code should be

```
> table(fitted(fit)>0.5, cl.train$y)
```

- Page 171: The third formula should read

$$\kappa_0 \approx \sqrt{\frac{\int W'(v)^2 dv}{W(0)}} \nu_1.$$

- Chapter 11: Some of the figures in this chapter will be slightly different with the current code, since the minimization algorithm has been improved.

- Page 204, example 11.3: Due to over-zealous editing, the `mcyc.n` dataset is no longer found in example 9.1. For distributions of Locfit after Jan 4, 2001, the `mcyc` dataset has the `vp` variable added; use `mcyc` in place of `mcyc.n`.

- Page 229: The $6/\sqrt{125}$ factor is correct only when the integrals are evaluated over the half-line. When integrated over the full line, the factor should be $3/\sqrt{125}$. This change does not affect Table 13.1.

  The formula for $\text{eff}_4(W)$ correctly assumes integration over the full line.

- Appendix A: The installation instructions for S-Plus have changed. Follow the instructions on the web page,
  `http://cm.bell-labs.com/stat/project/locfit`.