Errata for *Statistical Models*

A. C. Davison

December 17, 2010

Paperback reprint, 2008, ISBN 978-0-521-73449-3, and earlier hardback edition

- Page 5, first line of second paragraph: replace "the kth loading" with "the lth loading"
- Page 111, display above (4.18): should be

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{n}{\sigma^2}, \quad \frac{\partial^2 \ell}{\partial \mu \partial \sigma^2} = -\frac{n}{\sigma^4}(\overline{y} - \mu), \quad \frac{\partial^2 \ell}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum (y_j - \mu)^2.$$

- Page 148, two lines before Example 4.45: X_1^2 should be χ_1^2.
- Page 169, second display: should have $b(\pi) = -m \log(1 - \pi)$
- Page 169, fourth display: integral in second line should be

$$\int \exp\left\{(\theta + t)s(y) - \kappa(\theta + t) + c(y)\right\}\, dy$$

- Page 208, (5.35): should read

$$\widehat{\mu}_0 = \frac{\sum n_j \widehat{\mu}_j}{\sum n_j}.$$

- Page 209, line 5 of text: should read $\widehat{\lambda} = 0.076$
- Page 209, line 8 of text: should read 'their meta-analysis estimate'
- Page 209, line 11 of text: insert 'based on the ten smaller trials' after 'meta-analysis'
- Page 210, line -10: should be $f(u \mid y; \theta)$
- Page 236, Table 6.5, four lines from the foot: replace 160 with 161.
- Page 288, line -4: should be Examples 10.30 and 10.31.
- Page 313, last displayed equation: A should be $\mathcal{A}$.
- Page 337, second line of second display should be changed. Display becomes

$$\begin{aligned} \frac{f_1(Y)}{f_0(Y)} &= \frac{(2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2}\sum_{j=1}^n (Y_j - \mu_1)^2\}}{(2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2}\sum_{j=1}^n (Y_j - \mu_0)^2\}} \\ &= \exp\left[\frac{n}{2\sigma^2}\left\{2\overline{Y}(\mu_1 - \mu_0) - \mu_1^2 + \mu_0^2\right\}\right]. \end{aligned}$$

- Page 396, Exercise 8.6.6: doesn't work. Will be replaced in any future printing.
- Page 483, line 11 from top of page should read "... variance matrix $(X^{\mathrm{T}}WX)^{-1}$."
- Page 490, Example 10.18: The deviances in Table 10.9 correspond not to fits of models with binary response, as stated in the text on page 491, but to fits of the corresponding binomial variables, given in Table 10.8. The correct values for the binary fit are given below.

age	stage	grade	xray	acid	df	Deviance	age	stage	grade	xray	acid	df	Deviance
					52	70.25	+	+	+			49	59.30
+					51	68.86	+	+		+		49	53.21
	+				51	62.55	+	+			+	49	55.08
		+			51	64.67	+		+	+		49	57.04
			+		51	60.93	+		+		+	49	56.24
				+	51	62.71	+			+	+	49	54.46
+	+				50	60.44		+	+	+		49	53.52
+		+			50	64.08		+	+		+	49	53.16
+			+		50	60.02		+		+	+	49	49.18
+				+	50	62.21			+	+	+	49	50.82
	+	+			50	60.54	+	+	+	+		48	52.66
	+		+		50	54.46	+	+	+		+	48	52.92
	+			+	50	55.91	+	+		+	+	48	48.76
		+	+		50	57.45	+		+	+	+	48	50.81
		+		+	50	56.26		+	+	+	+	48	47.76
			+	+	50	54.79	+	+	+	+	+	47	47.61

- Page 491, line −6: replace sentence with 'The residual scaled deviance of 49.18 on 49 degrees of freedom suggests that the model fits well, but the binomial denominators here all equal $m = 1$, so χ^2 asymptotics do not apply.'
- Page 492, replace the first two lines of text with: '$23 - 4 = 19$ rather than 49, and the deviance would be 19.64. In view of the small denominators in Table 10.8, the χ^2 asymptotics for the deviance will be unreliable even with the binomial data.'
- Page 599, the first displayed equation should be:

$$\log \Pr(y) \doteq \log f(y \mid \tilde{\theta}) + \log \pi(\tilde{\theta}) - \tfrac{1}{2} p \log n + \tfrac{1}{2} p \log(2\pi) - \tfrac{1}{2} \log \left| -\frac{1}{n} \frac{\partial^2 \ell_m(\tilde{\theta})}{\partial \theta \partial \theta^{\mathrm{T}}} \right|,$$

- Page 599, three lines after the first displayed equation: change to '... terms that are $O(1)$, such as the last term in the expression above.'
- Page 656, line before last display: should be '... where $p = m$ and'
- Page 657–9: Here is a much simpler derivation of restricted likelihood, which does not involve projections of the data y.

  Consider data y whose distribution $f(y; \psi, \lambda)$ depends on two parameters, suppose that interest is focused on ψ, and that for fixed ψ there is a sufficient statistic for λ, s_ψ, say, which may depend on ψ. Then we have

$$f(y; \psi, \lambda) = f(y \mid s_\psi; \psi) f(s_\psi; \psi, \lambda),$$

  and since the first density on the right is a proper conditional density not depending on λ, we can base inference for ψ on it.

In the case of the model (12.10), we take $\beta \equiv \lambda$, and note that if all the variance parameters are fixed, then $\widehat{\beta}_\psi = (X^{\mathrm{T}}\Upsilon X)^{-1}X^{\mathrm{T}}\Upsilon y$ is sufficient for β; its distribution is $N_p(\beta, \sigma^2(X^{\mathrm{T}}\Upsilon X)^{-1})$. Apart from constants, the logarithm of the required conditional density is therefore

$$\begin{aligned}\log f(y;\psi,\beta,\sigma^2) - \log f(\widehat{\beta}_\psi;\psi,\beta,\sigma^2) \quad &\equiv \quad -\frac{n}{2}\log\sigma^2 + \tfrac{1}{2}\log|\Upsilon| - \frac{1}{2\sigma^2}(y-X\beta)^{\mathrm{T}}\Upsilon(y-X\beta) \\ &\quad +\frac{p}{2}\log\sigma^2 - \tfrac{1}{2}\log|X^{\mathrm{T}}\Upsilon X| + \frac{1}{2\sigma^2}(\widehat{\beta}_\psi-\beta)^{\mathrm{T}}X^{\mathrm{T}}\Upsilon X(\widehat{\beta}_\psi-\beta),\end{aligned}$$

which reduces to (12.13) on writing $y - X\beta = (y - X\widehat{\beta}_\psi) + X(\widehat{\beta}_\psi - \beta)$ in the first quadratic term and expanding out, noting that $(y - X\widehat{\beta}_\psi)^{\mathrm{T}}\Upsilon X(\widehat{\beta}_\psi - \beta) = 0$.

- Page 725, first column: "Rényi representation" is misplaced

Original hardback edition, 2003, ISBN 0-521-77339-3

- Page 40, just before (2.26): Iterating this argument yields ... (Insert 'iterating' before 'this').
- Page 44, fourth display: should have $M''(0)$ and not $M''(O)$.
- Page 58, line −5: Replace $(-0.008, 0.034)$ with $(-0.007, 0.047)$.
- Page 63, third line of display leading up to (3.8) should be:

$$= \frac{1}{(2\pi\sigma^2)^{1/2}}\int_{-\infty}^{\infty}\exp\left\{\mu t + \sigma^2\frac{t^2}{2} - \frac{1}{2\sigma^2}(y-\mu-t\sigma^2)^2\right\}dy$$

- Page 77: Delete side note: Some authors call them *quasi-random*.
- Page 83, line −17: ... several women arrive virtually together. (Replace 'almost' with 'virtually'.)
- Page 124, two lines after (4.33): summation should be $-\sum d^2\log f(Y_j;\theta)/d\theta^2$
- Page 138, line 13: and and $\mapsto$ and (delete second 'and')
- Page 148, just after second display: Should read

> and the normal distribution (4.52) of $\widehat{\theta}$ implies that the likelihood ratio statistic has distribution $\sum_{r=1}^{p}\zeta_r V_r$, where the ζ_r are the eigenvalues of $I_g(\theta_g)^{-1}K(\theta_g)$ and the V_r are independent χ^2_1 variables; thus its expected value is $\mathrm{tr}\{I_g(\theta_g)^{-1}K(\theta_g)\}$. Analogous limiting results can be shown to hold for tests on subsets of θ.

- Page 153, caption to Figure 4.10: Should read

> Model selection using likelihood criteria. Upper left: $21n$ observations (blobs) with true mean (solid) and polynomial fits $r = 1$, 2, 3 (dots, small dashes, large dashes); $n = 3$. Upper right: empirical versions of AIC, BIC and NIC for data on left. All are minimized with $r = 3$. Lower left: minus twice expected log likelihood $-2\mathrm{E}_g(\ell(\theta_g))$ (blobs) and theoretical versions of AIC, BIC and NIC for the panel above. The crosses show how $-2\mathrm{E}_g\{\ell(\widehat{\theta})\}$ decreases with the dimension of the fitted model. Lower right: as lower left panel, but with $n = 8$ observations at each value of x.

- Page 194, line 2 of Example 5.29: ... conditions has led ... (replace 'have' with 'has')

- Page 209, line 10 of text: insert space between 1.22 and 1.86 (should be $(1.22, 1.86)$)
- Page 210, line 5 of §5.5.2: replace *estimation-maximization* with *expectation-maximization*
- Page 268, line 4: $\rho_t = \alpha^{|t|}$ (no minus sign in exponent)
- Page 313, line 2 of Section 7.1.4: ... when the estimator ... (insert 'the' before 'estimator')
- Page 355, last line: replace 'dependences' with 'dependencies'
- Page 367, first line of display (8.8): $\widehat{\beta}$ and $\widehat{\beta}_1$ should be interchanged.
- Page 375, fourth display should be

$$\rho'(u) = \begin{cases} u\left\{1 - (u/c')^2\right\}^2, & \text{if } |u| < c', \\ 0, & \text{otherwise;} \end{cases}$$

- Page 375, line -6: replace $\sigma^{-2}\sum x_j x_j^{\mathrm{T}} g'(y_j - x_j^{\mathrm{T}}\beta)$ with $\sigma^{-2}\sum x_j x_j^{\mathrm{T}} \rho''(y_j - x_j^{\mathrm{T}}\beta)$
- Page 375, line -5: replace $\sigma^{-2} X^{\mathrm{T}} X \times \mathrm{E}\{g'(\varepsilon)\}$ with $\sigma^{-2} X^{\mathrm{T}} X \times \mathrm{E}\{\rho''(\varepsilon)\}$
- Page 375, last display should be

$$\sigma^2 (X^{\mathrm{T}} X)^{-1} \times \mathrm{E}\left\{\rho'(\varepsilon)^2\right\} / \mathrm{E}\left\{\rho''(\varepsilon)\right\}^2$$

- Page 376, first line of text: replace $\mathrm{E}\{g'(\varepsilon)\}/\mathrm{E}\left\{g(\varepsilon)^2\right\}$ by $\mathrm{E}\{\rho''(\varepsilon)\}^2/\mathrm{E}\left\{\rho(\varepsilon)^2\right\}$
- Page 377, last line should be

$$f(u;\sigma) = (2^{1/2}\sigma)^{-1} \exp\{-2^{1/2}|u|/\sigma\}, \quad -\infty < u < \infty, \sigma > 0.$$

- Page 423, line 9: ... pollution increase levels ... (not increases)
- Page 431, line 1 of Example 9.3: Thirty-two pigs ... (not twelve!)
- Page 472, second display should be

$$\log f(y_j; \eta_j, \phi) = -\tfrac{1}{2}\left\{\log(2\pi\phi) + (y_j - \eta_j)^2/\phi\right\}$$

- Page 472, line -8: replace 'numerator' with 'denominator'
- Page 475, line -12: replace 'estimates' with 'estimate'
- Page 481, first display: should be

$$M(t) = \exp\left\{\frac{b(\theta + t\phi) - b(\theta)}{\phi}\right\},$$

- Page 483, (10.21): should be

$$P = \frac{1}{\phi}\sum_{j=1}^{n} \frac{(y_j - \widehat{\mu}_j)^2}{a_j V(\widehat{\mu}_j)}, \qquad (10.21)$$

- Page 488, Table 10.7: logit link function should be $\eta = \log\{\pi/(1-\pi)\}$

- Page 488, line after displayed equation: replace 'If F' with 'If the density of ε'
- Page 505, caption to Table 10.16: should read

  Joint distribution of visual impairment on both eyes by race and age combinations (Liang *et al.*, 1992). Combination $(0, 0)$ means neither eye is visually impaired.
- Page 507, line -10: insert 'if data are missing at random then' after semi-colon.
- Page 518, line 10 (third line of part (b) of Exercise 4): replace $s^2 = ab/\{(a+b)(a+b+1)\}$ with $\sigma^2 = ab/\{(a+b)^2(a+b+1)\}$.
- Page 527, line -4: replace 'taking $\phi > 0$' with 'taking $\phi > 1$'
- Page 548, line –8: replace $J_j(\widehat{\beta})$ by $J_j(\widehat{\beta})^{-1/2}$.
- Page 554, line -7: Replace **?)** by Davison and Snell (1991)
- Page 557, line -14: (Cox and Davison, 1989) should be on separate line.
- Page 580, equation (11.13): this display should read

$$\begin{aligned} \pi(\mu, \sigma^2 \mid \overline{y}, s^2) &\propto f(\overline{y}, s^2 \mid \mu, \sigma^2)\pi(\mu, \sigma^2) \\ &= f(\overline{y} \mid \mu, \sigma^2) f(s^2 \mid \mu, \sigma^2)\pi(\mu, \sigma^2) \\ &= f(\overline{y} \mid \mu, \sigma^2) f(s^2 \mid \sigma^2)\pi(\mu)\pi(\sigma^2) \\ &\propto \pi(\mu \mid \overline{y}, \sigma^2) f(\overline{y} \mid \sigma^2) f(s^2 \mid \sigma^2)\pi(\sigma^2), \qquad (11.13) \end{aligned}$$

  (Missing $f(\overline{y} \mid \sigma^2)$ in (11.13).)
- Page 581, lines 1–8: Sentences should read:

  ... conditional on $\overline{y}$ and σ^2. Integration of (11.13) with respect to μ shows that $\pi(\sigma^2 \mid \overline{y}, s^2) \propto f(\overline{y} \mid \sigma^2) f(s^2 \mid \sigma^2)\pi(\sigma^2)$: in general the marginal posterior density of σ^2 depends on s^2 and on $\overline{y}$. Integration of (11.13) with respect to σ^2 shows that the marginal posterior for μ also depends on both $\overline{y}$ and s^2.

  Let us use the improper priors $\pi(\mu) \propto 1$, $\pi(\sigma^2) \propto \sigma^{-2}$. Example 11.7 shows that the posterior density for μ when σ^2 is known is $N(\overline{y}, \sigma^2/n)$, and it is easy to verify that $f(\overline{y} \mid \sigma^2) = 1$. Conditional ...
- Page 596, line 4 before (11.27): insert 'is' after that (should be '... that is unavailable ...').
- Page 597, line before (11.30): insert 'the' after 'of' (should be 'of the form')
- Page 604, line -7: This example shows ...
- Page 610, line 6: $\theta_1, \ldots, \theta_T \overset{\text{iid}}{\sim} N(\mu, \sigma_\theta^2)$ (μ replaces ν)
- Page 611, line -13: $\mu_0 = 0$ (not $\mu = 0$)
- Page 606, last display should be:

$$\sum_{t=1}^{m}\sum_{s=1}^{n} \pi_2(t) p_{ts} q_{sr} = \sum_{t=1}^{m}\sum_{s=1}^{n} \pi_2(t) \frac{\pi(s,t)}{\pi_2(t)} \frac{\pi(s,r)}{\pi_1(s)} = \pi_2(r),$$

- Page 607, line −16: insert 'in a given number of iterations' after 'probability mass'. (Sentence should read ' ... is more likely to visit all regions of significant probability mass in a given number of iterations.')

- Page 610, 4th display should be:

$$\begin{aligned}\pi(\sigma_\theta^2 \mid \sigma^2, \mu, \theta, y) &= \frac{f(y \mid \theta, \sigma^2) f(\theta \mid \mu, \sigma_\theta^2) \pi(\mu) \pi(\sigma^2) \pi(\sigma_\theta^2)}{\int f(y \mid \theta, \sigma^2) f(\theta \mid \mu, \sigma_\theta^2) \pi(\mu) \pi(\sigma^2) \pi(\sigma_\theta^2)\, d\sigma_\theta^2} \\ &= \frac{f(\theta \mid \mu, \sigma_\theta^2) \pi(\sigma_\theta^2)}{\int f(\theta \mid \mu, \sigma_\theta^2) \pi(\sigma_\theta^2)\, d\sigma_\theta^2} \\ &= \pi(\sigma_\theta^2 \mid \mu, \theta).\end{aligned}$$

- Page 676, line after (12.38), replace $\log\{r(\psi)/v(\psi)\}$ with $\log\{v(\psi)/r(\psi)\}$.

- Page 681, line before Example 12.21: Replace 'Example 12.18' by 'Example 12.24'.

- Page 708, column 2, line −5: Volume number for Robinson (1991) should be **6**, not **3**.