Schneeweiss:

Estimating the endpoint of a uniform distribution under normal measurement errors with known error variance

Sonderforschungsbereich 386, Paper 339 (2003)

Online unter: http://epub.ub.uni-muenchen.de/

Estimating the endpoint of a uniform distribution under normal measurement errors with known error variance

H. Schneeweiss
Department of Statistics,
Ludwig-Maximilians-Universität München,
Akademiestr. 1/I, D-80799 München, Germany
email: Schneeweiss@stat.uni-muenchen.de

Discussion Paper 339, 2003

Abstract

The paper studies the problem of estimating the upper end point of a finite interval when the data come from a uniform distribution on this interval and are disturbed by normally distributed measurement errors with known variance. Maximum likelihood and method of moments estimators are introduced and compared to each other.

1 Introduction

The problem of estimating the end point α of a uniform distribution on the interval $[0, \alpha]$ has a long-standing history in statistics, mainly because it served as an important counterexample in the theory of maximum likelihood (ML) estimation, showing that certain regularity conditions - in particular that the support of the distribution should not depend on the unknown parameter - are necessary in order to guarantee the well-known optimality properties of ML, see, e.g., Lehmann (1983), in particular Example 6.9, p.452. The ML estimator of α, which is the maximum of the sample values, is not asymptotically normal. The likelihood function is not differentiable at the point of its maximum.

This picture changes drastically when the data are measured with normally distributed errors. In some sense, this complicates the estimation procedure. On the other hand, the likelihood function now satisfies the regularity conditions (the support does not depend on the unknown parameter) and the ML estimator has the usual optimality properties. It is asymptotically normal and efficient and its asymptotic variance can be computed from the likelihood function.

The model of a uniform distribution has not only its merits as an example in the theory of statistics. It is also useful in a number of applications, in particular in the field of image processing, e.g., Vosselman and Haralick (1996), Werman and Keren (2001). Recently Davidov and Goldenshluger (2003) studied estimators of the end-points of a line segment in a plane, where the observations are error-ridden measurements of a stochastic variate varying on the line with a support that is identical to the line segment. Because of the measurement errors the observed points will not be exactly on the line. But if the measurement error variance is known, the end-points can nevertheless be estimated. Davidov and Goldenshluger accomplished the estimation by a method of moments (MM). They compare its accuracy to that of the ML estimator, however without actually introducing the ML estimator. Werman and Keren (2001) compute ML estimators, but they skip all the details. Chan (1982) introduced a pseudo ML estimator for a line segment with a uniform distribution. His approach is a mixture of ML and MM.

The ML estimator is certainly more difficult to compute than the MM estimator. In the simpler case of a line segment on the real line with only one endpoint to be estimated, ML is easier to handle and yields reasonable estimates which compare favorably to the MM estimates. This case will be the main focus of the present paper.

In Section 2 we introduce the model, in Section 3 we study its likelihood function, and we investigate the ML estimator in Section 4. In Section 5 we compare ML to MM. Section 6 has some simulation results. The conclusion is in Section 7. Details are relegated to an extensive appendix.

2 The model

Let ξ be uniformly distributed on the interval $[0, \alpha]$ with unknown end point $\alpha > 0$. The variable ξ cannot be observed directly. Instead we observe a variable x, which is related to ξ through

$$x = \xi + \delta,$$

where δ is the measurement error. The error variable δ is supposed to be normally distributed with expectation 0 and variance $\sigma^2 > 0$. It is independent of ξ. We assume σ^2 to be known.

The density function of x, given the parameter α, can be computed as

$$f(x; \alpha) = \int k(x|\xi)h(\xi; \alpha)d\xi,$$

where $k(x|\xi)$ is the density of $N(\xi, \sigma^2)$ and

$$h(\xi; \alpha) = \begin{cases} \frac{1}{\alpha} & \text{for} \quad 0 < \xi < \alpha \\ 0 & \quad\quad \text{otherwise.} \end{cases} \tag{1}$$

Thus

$$\begin{aligned} f(x; \alpha) &= \frac{1}{\alpha} \frac{1}{\sqrt{2\pi}\sigma} \int_0^\alpha \exp[-\frac{(x-\xi)^2}{2\sigma^2}]d\xi \\ &= \frac{1}{\alpha}\Big[\Phi(\frac{x}{\sigma}) - \Phi(\frac{x-\alpha}{\sigma})\Big], \end{aligned} \tag{2}$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution.

It can be shown that for $\sigma \to 0$ the density function (2) leads to (1), see Appendix A1.

In the sequel we will use, most of the time, a different parameterisation of the model, which will render the subsequent formulas much simpler than otherwise. To this end, we introduce the new parameter

$$\beta = \frac{\alpha}{\sigma}$$

and at the same time rescale the variable x by letting

$$y = \frac{x}{\sigma}.$$

The density of y, then, becomes

$$g(y;\beta) = \frac{1}{\beta}\Big[\Phi(y) - \Phi(y-\beta)\Big]. \tag{3}$$

3 The likelihood

Suppose we have a sample $(x_1, \cdots, x_n)$ of i.i.d. realisations of the random variable x. If we transform it to a sample $(y_1, \cdots, y_n)$ via $y_i = \frac{x_i}{\sigma}$, then, due to (3), the log-likelihood function of β is given by

$$l_n(\beta) = -n\log\beta + \sum_{i=1}^{n} \log\Big[\Phi(y_i) - \Phi(y_i-\beta)\Big]. \tag{4}$$

Its first and second derivatives are

$$l_n^{'}(\beta) = -\frac{n}{\beta} + \sum_{i=1}^{n} \frac{\varphi(y_i-\beta)}{\Phi(y_i) - \Phi(y_i-\beta)}. \tag{5}$$

and

$$l_n^{''}(\beta) \;=\; \frac{n}{\beta^2} + \sum_{i=1}^{n} \frac{\{[\Phi(y_i) - \Phi(y_i-\beta)](y_i-\beta) - \varphi(y_i-\beta)\}\varphi(y_i-\beta)}{[\Phi(y_i) - \Phi(y_i-\beta)]^2} \tag{6}$$

where $\varphi(\cdot)$ is the density function of the standard normal distribution.

It is interesting to see how the log-likelihood behaves at the extreme values $\beta \to 0$ and $\beta \to \infty$. By using L'Hospital's rule (see appendix A2) one can verify that

$$l_n(0) = -n\log(\sqrt{2\pi}) - \frac{1}{2}\sum_{i=1}^{n} y_i^2 \tag{7}$$

$$l_n^{'}(0) = \frac{1}{2}\sum_{i=1}^{n} y_i^2 \tag{8}$$

$$l_n^{''}(0) = \frac{1}{12}\Big(\sum_{i=1}^{n} y_i^2 - 4n\Big) \tag{9}$$

and

$$l_n(\infty) = -\infty$$

$$l_n^{'}(\infty) = l^{''}(\infty) = 0.$$

It is possible, especially when β is small, that numerical problems will arise in computing $l(\beta)$ and its derivatives, in particular for large y. In such a case it might be helpful to approximate the individual terms of these functions by Taylor expansions at $\beta = 0$, see appendix A3. Denoting the contribution of an individual observation to the log-likelihood (4) by $l(\beta)$ and dropping the index i, we have by an expansion up to the order of β^3, β^2, and β, respectively, for $\beta \to 0$;

$$l(\beta) = \frac{1}{2}\left[\beta y + \frac{\beta^2}{12}(y^2 - 4) - \frac{\beta^3}{12}y\right] - \frac{1}{2}y^2 - \frac{1}{2}\log(2\pi) + O(\beta^4) \tag{10}$$

$$l^{'}(\beta) = \frac{1}{2}\left[y + \frac{\beta}{6}(y^2 - 4) - \frac{\beta^2}{4}y\right] + O(\beta^3) \tag{11}$$

$$l^{''}(\beta) = \frac{1}{4}\left[\frac{1}{3}(y^2 - 4) - \beta y\right] + O(\beta^2). \tag{12}$$

4 Maximum Likelihood

Both derivatives of $l_n(\beta)$ may be used to construct the iterative Newton algorithm for finding the ML estimate of β:

$$\beta_{\kappa+1} = \beta_\kappa - \frac{l_n^{'}(\beta_\kappa)}{l_n^{''}(\beta_\kappa)} \qquad\qquad \kappa = 0, 1, \cdots$$

with a starting value β_0, which may be taken to be $\max(y_1, \cdots, y_n)$. Another possible starting value is $\beta_0 = 2\overline{y}$, as this is a consistent estimate of β, see Section 5.

As $l_n(\beta)$ is bounded from above (because $l_n(0)$ is finite and $l_n(\infty) = -\infty$), $l_n(\beta)$ has a finite maximum which will typically be unique. In principle, the maximum may be attained at $\beta = 0$, but this will hardly ever happen if n is large. It *can* happen if n is small and β is small (i.e., σ is large relative to α). However, in most cases the maximum is a stationary point of $l_n(\beta)$, when n is sufficiently large.

For $n = 1$, it can be shown that $l_1(\beta)$ has an unique local (and global) maximum,which is 0 for $y_1 \leq 0$ and positive for $y_1 > 0$, see appendix A4. For $n > 1$, however, the likelihood function may have several local maxima. An extensive simulation study, not documented in this paper, has shown that this will occur only with extremely low frequency, even at small sample size (e.g., $n = 10$).

As the likelihood function satisfies the usual regularity conditions (see Appendix A6,7), we can claim consistency and normality for the ML estimator $\hat{\beta}$:

$$\sqrt{n}(\hat{\beta} - \beta) \to N(0, v)$$

with $v = -[El''(\beta)]^{-1}$, for which we find, see appendix A7,

$$El''(\beta) = \frac{1}{\beta^2} - \frac{1}{\beta} \int_{-\infty}^{\infty} \frac{\varphi^2(y - \beta)}{\Phi(y) - \Phi(y - \beta)} \ dy. \tag{13}$$

For large n, the variance of $\hat{\beta}$ can be estimated by

$$\hat{\sigma}^2_{\hat{\beta}} = -\frac{1}{l''_n(\hat{\beta})}$$

or by

$$\hat{\hat{\sigma}}^2_{\hat{\beta}} = -\frac{1}{nEl''(\hat{\beta})},$$

where the integral in (13) has to be computed numerically.

For small β the asymptotic variance v can be approximated by a simple quadratic function in β (see appendix A8):

$$v = 4 + \frac{\beta^2}{9} + O(\beta^4). \tag{14}$$

Note that the remainder term in (4) is of order β^4, which renders the approximation very precize for small β.

5 Method of moments estimators

As an alternative to ML, estimation by the method of moments (MM) suggests itself. MM estimators are much easier to compute. Davidov and Goldenshluger (2003) based their investigation exclusively on MM estimation.

In our simple problem of estimating just one endpoint of an interval, it suffices to use the first order moment $\overline{x}$ to construct an MM estimator of α, but as the error variance σ^2 is supposed to be known, a second order MM estimator is also feasible. The first order MM estimator of α is given by

$$\hat{\alpha}_1 = 2\overline{x}.$$

It is obviously unbiased as $E\overline{x} = \frac{\alpha}{2}$, and due to the fact that

$$\sigma_x^2 = \frac{\alpha^2}{12} + \sigma^2, \tag{15}$$

its variance is

$$\sigma_{\hat{\alpha}_1}^2 = \frac{1}{n}\left(\frac{\alpha^2}{3} + 4\sigma^2\right). \tag{16}$$

This estimator does not depend on knowledge of the error variance, and it is also valid if ξ follows any symmetric distribution on $[0, \alpha]$. However, if σ^2 is known, we can also construct a second order MM estimator using (15):

$$\hat{\alpha}_1 = \sqrt{12(s_x^2 - \sigma^2)},$$

where s_x^2 is the empirical variance of the sample. This estimator is asymptotically normal:

$$\sqrt{n}(\hat{\alpha}_2 - \alpha) \to N(0, V_2)$$

with asymptotic variance (see appendix A9)

$$V_2 = \frac{\alpha^2}{5} + 12\sigma^2 + 72\frac{\sigma^4}{\alpha^2}. \tag{17}$$

A comparison with the corresponding asymptotic variance of $\hat{\alpha}_1$, i.e.,

$$V_1 = n\sigma_{\hat{\alpha}_1}^2 = \frac{\alpha^2}{3} + 4\sigma^2$$

shows that $V_1 > V_2$ as long as

$$\sigma^2 < (30 + 12\sqrt{10})^{-1}\alpha^2 = 0.0147\alpha^2.$$

For σ^2 larger than this bound, $\hat{\alpha}_1$ is the more efficient estimator. It is only for very small $\frac{\sigma^2}{\alpha^2}$ that the second order MM estimator is more efficient.

The corresponding estimators for the transformed parameter β are

$$\hat{\beta}_1 = 2\overline{y}$$

$$\hat{\beta}_2 = \sqrt{12(s_y^2 - 1)} \tag{18}$$

with asymptotic variances

$$v_1 = \frac{\beta^2}{3} + 4$$

$$v_2 = \frac{\beta^2}{5} + 12 + \frac{72}{\beta^2}, \tag{19}$$

and $v_1 > v_2$ if, and only if, $\beta^2 > 30 + 12\sqrt{10} = 67,95$.

According to ML theory both variances should be larger than the asymptotic variance, v, of the ML estimator $\hat{\beta}$. To see how much they differ, a simulation study was run. However, for small β we can compare the variances analytically. According to (14) the difference of v_1 and v is

$$v_1 - v = \tfrac{2}{9}\beta^2 + O(\beta^4),$$

and, of course, $v_2 - v$ is much larger for small β.

6 Simulation

In a small Monte Carlo simulation study estimates for the variance of $\hat{\alpha}$ were computed and were compared to the variance of $\hat{\alpha}_1$ see (16).(The variance of $\hat{\alpha}_2$ is quite a bit larger than the variance of $\hat{\alpha}_1$ for the values of σ chosen, see (17), and was therefore not included in the comparison). The value of the parameter α was kept fixed at $\alpha = 1$, and σ varied from 0.05 to 4. A rather small sample size, $n = 25$, and a medium sample size, $n = 100$, were chosen, so that also small sample properties could be studied.The number of replications was N=1000. The results of the Monte Carlo study are shown in the following two tables:

Table 1: Expectation and variance of estimators for $n = 25$

σ	$\widehat{E\hat{\alpha}}$	$\widehat{V\hat{\alpha}}$	$V\hat{\alpha}$	$V\hat{\alpha}$(appr)	$V\hat{\alpha}_1$
0.05	0.98	0.002	0.003	0.005	0.014
0.1	0.99	0.004	0.005	0.006	0.053
0.2	0.97	0.01	0.011	0.011	0.020
0.5	0.96	0.04	0.044	0.044	0.053
1	1.01	0.15	0.164	0.164	0.173
2	0.96	0.44	0.644	0.644	0.653
4	1.26	1.45	2.564	2.564	2.573

Table 2: Expectation and variance of estimators for $n = 100$

σ	$\widehat{E\hat{\alpha}}$	$\widehat{V\hat{\alpha}}$	$V\hat{\alpha}$	$V\hat{\alpha}$(appr)	$V\hat{\alpha}_1$
0.05	1.00	0.001	0.0006	0.0012	0.0034
0.1	0.99	0.001	0.0012	0.0015	0.0037
0.2	1.00	0.003	0.0028	0.0027	0.005
0.5	1.00	0.012	0.011	0.011	0.013
1	1.00	0.042	0.041	0.041	0.043
2	1.01	0.157	0.161	0.161	0.163
4	1.03	0.510	0.641	0.641	0.643

The second and third columns give the Monte Carlo estimates of the mean and variance of $\hat{\alpha}$, respectively. The latter can be compared to the asymptotic variance of $\hat{\alpha}$, $V\hat{\alpha}$, presented in the fourth column and evaluated with the help of (13). The approximation of $V\hat{\alpha}$ according to (14) is presented in the fifth column, and the last column has the variance of $\hat{\alpha}_1$ according (16)

It turns out that there is hardly any bias in $\hat{\alpha}$ even for large values of σ, and even if the sample size is rather small. The variance of $\hat{\alpha}$ (estimated from the Monte Carlo study) is very close to its (theoretical) asymptotic value $V\hat{\alpha}$ when $n = 100$, except when σ is extremely large ($\sigma \geq 4$). But for the smaller sample size, $n = 25$, $\hat{V\hat{\alpha}}$ is considerable smaller than $V\hat{\alpha}$ when $\sigma \geq 2$. For these values of σ the sample size $n = 25$ seems to be much too small for any asymptotic properties to hold. The fact that, for $n = 25$, asymptotics is not applicable when $\sigma \geq 2$ is reflected by the observation that in many of the Monte Carlo runs the maximum of the likelihood function was found to be at $\hat{\alpha} = 0$. For large enough n this should happen only very rarely, and only then can we expect asymptotic theory to apply.

It is interesting to note that the approximate formula (14) for $V\hat{\alpha}$, which is valid only for large values of σ, seems to give good results even for values of σ as small as 0.2, see columns four and five.

The variance of $\hat{\alpha}$ is always smaller than the variance of $\hat{\alpha}_1$. The difference is particularly noticeable for small σ, whereas for large σ the relative difference tends to decrease.

7 Conclusion

Estimating the endpoint α of a uniform distribution on the real line is a simple task: just take the maximum of a sample as the estimate. This is, in fact, the

ML estimator. However, conventional ML-theory does not apply to this model. The usual regularity conditions are not satisfied. The maximum of the likelihood function is located at a discontinuity point.

If a normally distributed measurement error with known variance σ^2 is added to the uniformly distributed random variable, the resulting model meets all the regularity conditions of ML theory. But the ML estimator is now more complicated to compute.

In this paper the ML procedure is fully developed. The maximum of the likelhood function is typically a stationary point unless it is situated at the border of the parameter space, $\alpha = 0$, a rare case in practice. Special attention is paid to the case when α/σ becomes small and approaches zero, as in this case the likelihood function tends to the form 0/0 and its computation meets numerical obstacles. The same is true for the asymptotic variance of the ML estimator $\hat{\alpha}$ when α/σ is small. However, in this case a particularly simple approximation formula, (14), for the asymptotic variance can be derived, which comes very close to the true variance formula, even if α/σ is not so small.

One can also estimate α by the method of moments. Here we studied two such estimators: a simple first order MM estimator and a second order MM estimator, which used the measurement error variance σ^2 supposed to be known. Interesting enough, the first order estimator is almost always more efficient, although it does not use the extra information σ^2. Only for very small σ/α do we find the second order MM estimator to be more efficient.

Finally we can compare the relative efficiencies of the ML and the MM estimators. It turns out that - at least for small σ/α - the asymptotic variance of the ML estimator $\hat{\alpha}$ is a good deal smaller than the asymptotic variance of the (first order) MM estimator. Only when σ/α becomes larger than 0.5 does the difference between these two variances become negligible.

These analytic results were corroborated by a simulation study, which also shed some light on the small sample properties of the estimators.

Acknowledgement: I want to thank Thomas Augustin and Ori Davidov for helpful discussions.

Appendix

A1 Density function for $\sigma \to 0$

Let $x > \alpha$, then

$$\lim_{\sigma \to 0} \Phi(\frac{x}{\sigma}) = \lim_{\sigma \to 0} \Phi(\frac{x-\alpha}{\sigma}) = 1,$$

and so by (2)

$$\lim_{\sigma \to 0} f(x;\alpha,\sigma) = 0.$$

Now let $0 < x < \alpha$, then

$$\lim_{\sigma \to 0} \Phi(\frac{x}{\sigma}) = 1, \qquad \lim_{\sigma \to 0} \Phi(\frac{x-\alpha}{\sigma}) = 0,$$

and so by (2)

$$\lim_{\sigma \to 0} f(x;\alpha,\sigma) = \frac{1}{\alpha}.$$

Finally let $x < 0$, then

$$\lim_{\sigma \to 0} \Phi(\frac{x}{\sigma}) = \lim_{\sigma \to 0} \Phi(\frac{x-\alpha}{\sigma}) = 0,$$

and so by (2)

$$\lim_{\sigma \to 0} f(x;\alpha,\sigma) = 0.$$

These three parts of the limit density function correspond to the density (1) of the error-free model.

A2 l(0) and l(∞)

Select an individual contribution to the log-likelihood (4) and denote it by $l(\beta)$, dropping the index i. Then

$$l(\beta) \quad = \quad -\log\beta + \log\left[\Phi(\beta - y) - \Phi(-y)\right]. \tag{20}$$

Introduce the abbreviations

$$l = l(\beta), \quad d = \Phi(\beta - y) - \Phi(-y), \quad \varphi = \varphi(\beta - y)$$

and let the prime denote the derivative with respect to β. Then

$$\begin{aligned} d^{'} &= \varphi \\ \varphi^{'} &= (y-\beta)\varphi \\ \varphi^{''} &= [(y-\beta)^2 - 1]\varphi. \end{aligned}$$

We can now write

$$l = \log \frac{d}{\beta} \tag{21}$$

$$l^{'} = -\frac{1}{\beta} + \frac{\varphi}{d} = \frac{-d + \beta\varphi}{\beta d} \tag{22}$$

$$l^{''} = \frac{1}{\beta^2} + \frac{d\varphi^{'} - \varphi^2}{d^2} \tag{23}$$

$$= \frac{d^2 + \beta^2(d\varphi^{'} - \varphi^2)}{\beta^2 d^2}. \tag{24}$$

Using L'Hospital's rule, we have for $\beta \to 0$

$$\frac{d}{\beta} \to \frac{d^{'}}{1} = \varphi \to \varphi(y)$$

and therefore

$$\lim_{\beta \to 0} l(\beta) = \log \varphi(y) = -\frac{1}{2}\log(2\pi) - \frac{y^2}{2}.$$

Similarly,

$$l^{'} \quad \to \quad \frac{\beta\varphi^{'}}{d + \beta\varphi} \quad \to \quad \frac{\varphi^{'} + \beta\varphi^{''}}{2\varphi + \beta\varphi^{'}},$$

and so

$$\lim_{\beta \to 0} l^{'}(\beta) = \frac{\varphi^{'}(-y)}{2\varphi(-y)} = \frac{y}{2}.$$

Finally,

$$l^{''} \to \frac{1}{2}\frac{2d\varphi + 2\beta(d\varphi^{'} - \varphi^2) + \beta^2(d\varphi^{''} - \varphi\varphi^{'})}{\beta d^2 + \beta^2 d\varphi}$$

$$\to \frac{1}{2}\frac{4d\varphi' + 4\beta(d\varphi'' - \varphi\varphi') + \beta^2(d\varphi''' - \varphi'^2)}{d^2 + 4\beta d\varphi + \beta^2(\varphi^2 + d\varphi')}$$

$$\to \frac{1}{2}\frac{8d\varphi'' + 6\beta(d\varphi''' - \varphi'^2) + \beta^2(\varphi\varphi''' + d\varphi^{(4)} - 2\varphi'\varphi'')}{6d\varphi + 6\beta(\varphi^2 + d\varphi') + \beta^2(3\varphi\varphi' + d\varphi'')}$$

$$\to \frac{1}{2}\frac{8\varphi\varphi'' - 6\varphi'^2 + 14d\varphi''' + \beta(\cdots)}{12(\varphi^2 + d\varphi') + \beta(\cdots)}$$

and so

$$\lim_{\beta\to 0} l''(\beta) = \frac{4\varphi\varphi'' - 3\varphi'^2}{12\varphi^2} = \frac{y^2 - 4}{12}.$$

The limits for $\beta \to \infty$ are found directly from (21) to (24) by noting that $\varphi \to 0$, $\varphi' \to 0$, and $d \to \Phi(y)$ for $\beta \to \infty$.

A3 The likelihood near $\beta = 0$

We have the following Taylor series expansion of $\Phi(y - \beta)$ for $\beta \to 0$:

$$\Phi(y - \beta) = \Phi(y) - \beta\varphi(y) + \frac{\beta^2}{2}\varphi'(y) - \frac{\beta^3}{6}\varphi''(y) + \frac{\beta^4}{24}\varphi'''(y) + O(\beta^5),$$

and therefore

$$\begin{aligned}\frac{\Phi(y) - \Phi(y-\beta)}{\beta} &= \varphi(y) - \frac{\beta}{2}\varphi'(y) + \frac{\beta^2}{6}\varphi''(y) - \frac{\beta^3}{24}\varphi'''(y) + O(\beta^4) \\ &= \left[1 + \frac{\beta}{2}y + \frac{\beta^2}{6}(y^2 - 1) + \frac{\beta^3}{24}y(y^2 - 3)\right]\varphi(y) + O(\beta^4), \qquad (25)\end{aligned}$$

where we used the following formulas for the derivatives of $\varphi(x)$:

$$\varphi'(x) = -x\varphi(x), \quad \varphi''(x) = (x^2 - 1)\varphi(x),$$

$$\varphi'''(x) = x(3 - x^2)\varphi(x), \quad \varphi^{(4)}(x) = (3 - 6x^2 + x^4)\varphi(x).$$

Using the Taylor series

$$\log(1 + z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 + O(z^4),$$

we get

$$l(\beta) = \frac{1}{2}\left(\beta y + \frac{\beta^2}{12}(y^2-4) - \frac{\beta^3}{12}y\right) + \log\varphi(y) + O(\beta^4),$$

which is (10). Differentiating (10) with respect to β results in (11) and (12).

Letting $\beta \to 0$ we receive, once again, (7) to (9).

A4 Maximum of the log-likelihood

Consider first a sample with $n = 1$ and let $l_1(\beta) = l(\beta)$. We consider two cases.

Case 1: $y \leq 0$. In this case, $l^{'}(\beta) < 0$ for all $\beta > 0$. Indeed, $l^{'}(\beta) < 0$ is equivalent to

$$\frac{\Phi(y) - \Phi(y-\beta)}{\beta} > \varphi(y-\beta),$$

and this is true if $y \leq 0$ and $\beta > 0$, because $\Phi(y)$ is concave for $y < 0$. Note that the left hand side is the slope of the secant of the graph of Φ at the points y and $y - \beta$ and the right hand side is the slope of the tangent at $y - \beta$. Thus in this case, the maximum of $l(\beta)$ is at $\beta = 0$ and there are no other local maxima.

Case 2: $y > 0$. In this case, $l^{'}(\beta) > 0$ for all $0 < \beta \leq y$. Indeed, $l^{'}(\beta) > 0$ is equivalent to

$$\frac{\Phi(y) - \Phi(y-\beta)}{\beta} < \varphi(y-\beta),$$

and this is true as long as $y > 0$ and $y - \beta \geq 0$, because $\Phi(y)$ is convex for $y > 0$. On the other hand, $l(\infty) = -\infty$. Therefore $l(\beta)$ has a maximum at some $\beta_m > y$.

Let β_0 be a stationary point of $l(\beta)$. As $l^{'}(\beta) > 0$ for $\beta \leq y$ including $\beta = 0$ (for which $l^{'}(0) = \frac{y^2}{2} > 0$), we must have $\beta_0 > y$. At β_0, $l^{'}(\beta_0) = 0$, which implies

$$d := \Phi(y) - \Phi(y-\beta_0) = \beta_0\varphi(y-\beta_0). \tag{26}$$

Consider the second derivative of l at β_0. According to (24) and substituting d from (26), we find

$$\begin{aligned} l^{''}(\beta_0) &= \frac{1}{\beta_0^2} + \frac{\beta_0(y-\beta_0)\varphi(y-\beta_0) - \varphi^2(y-\beta_0)}{\beta_0^2\varphi^2(y-\beta_0)} \\ &= \frac{y-\beta_0}{\beta_0\varphi(y-\beta_0)} < 0. \end{aligned}$$

Thus any stationary point of $l(\beta)$ is a local maximum. Suppose there were two distinct local maxima, then there would be a local minimum in-between, which would be a stationary point, contradicting the previous statement. It follows that there is only one local maximum β_m, which is the absolute maximum of $l(\beta)$, and $\beta_m > y$.

For samples with $n > 1$, no such statement can be made. Indeed, $l_n(\beta)$ may have several local maxima. An example is n=165 with $y_i = 0.2$ for $i = 1, \ldots, 155$ and $y_i = 8$ for $i = 156, \ldots, 165$. In this (artificial) example, $l(\beta)$ has two local maxima at around $\beta_1 = 2.1$ and $\beta_2 = 4.8$, β_2 being the absolute maximum.

A5 Moments of y

As $y = \frac{\xi}{\sigma} + \frac{\delta}{\sigma}$ and ξ and δ are independent, the moment generating function of y is just the product of the moment generating functions of $\frac{\delta}{\sigma}$ and $\frac{\xi}{\sigma}$, i.e. :

$$E\exp(ty) = \exp(\frac{1}{2}t^2)\frac{\exp(t\beta) - 1}{t\beta}, \qquad -\infty < t < \infty.$$

It follows that all the moments of y exist and therefore, by the Schwarz inequality, that

$$E(|y|^k \exp(ty)) < \infty, \qquad k = 0, 1, \ldots, -\infty < t < \infty.$$

The first four moments of y are

$$Ey = \frac{\beta}{2}, \qquad \sigma_y^2 = \frac{\beta^2}{12} + 1$$
$$m_3(y) = 0, \;\; m_4(y) = \frac{\beta^4}{80} + \frac{\beta^2}{2} + 3,$$

where m_3 and m_4 are third and fourth central moments.

A6 Local uniform boundedness of the derivatives of the log-likelihood.

There are several alternative sets of sufficient "regularity" condition that guarantee the asymptotic properties of the ML estimator $\hat{\beta}$. Among these, the requirement of local uniform boundedness of the first and second derivatives of $l(\beta)$ is quite common. By way of example, we here consider the first derivatives $l'(\beta)$. Fix an interval $(\beta_1, \beta_2), \beta_1 > 0$, which contains the true parameter point β_0, say, and let $\beta \in (\beta_1, \beta_2)$. Then $l'(\beta) = -\frac{1}{\beta} + m(y; \beta)$ with

$$m(y; \beta) = \frac{\varphi(y - \beta)}{\Phi(y) - \Phi(y - \beta)},$$

and $m(y;\beta)$ can be bounded by

$$0 \leq m(y;\beta) \leq \frac{q(y)}{\Phi(y) - \Phi(y-\beta_1)} =: Q(y),$$

where

$$q(y) = \begin{cases} \varphi(y-\beta_1) & \text{for} \quad y < \beta_1 \\ \varphi(0) & \text{for} \quad \beta_1 \leq y \leq \beta_2 \\ \varphi(y-\beta_2) & \text{for} \quad y > \beta_2. \end{cases}$$

We want to show that

$$EQ(y) < \infty,$$

where here and in the sequel the expectation is always understood to be taken at the true parameter value β_0.

To prepare for the proof, we state the following properties of $m(y;\beta)$ and of the ratio $\varphi(y-\beta)/\varphi(y)$:

1. $m(y;\beta) = y - \beta + o(y)$ as $y \to \infty$
2. $m(y;\beta) \to 0$ as $y \to -\infty$
3. $\frac{\varphi(y-\beta)}{\varphi(y)} = \exp(-\frac{1}{2}\beta^2)\exp(\beta y)$.

Statement 1 and 2 can be proved with the help of L'Hospital's rule; statement 3 follows directly from computing the ratio.

Now,

$$EQ(y) = E\left[Q(y)I(y<\beta_1)\right] + E\left[Q(y)I(\beta_1 \leq y \leq \beta_2)\right] + E\left[Q(y)I(y>\beta_2)\right].$$

The middle term is finite due to the continuity of $Q(y)$. The first term on the right hand side equals (with g from (3))

$$\int_{-\infty}^{\beta_1} \frac{\varphi(y-\beta_1)}{\Phi(y)-\Phi(y-\beta_1)} g(y;\beta_0)dy = \int_{-\infty}^{\beta_1} m(y;\beta_1)g(y;\beta_0)dy,$$

which is finite because of property 2 above. The last term equals

$$\begin{aligned} &\int_{\beta_2}^{\infty} \frac{\varphi(y-\beta_2)}{\Phi(y)-\Phi(y-\beta_1)} g(y;\beta_0)dy \\ =& \int_{\beta_2}^{\infty} m(y,\beta_1)\frac{\varphi(y-\beta_2)}{\varphi(y-\beta_1)} g(y;\beta_0)dy \end{aligned}$$

and is finite because of properties 1 and 3 above and because of the existence of $E(|y|\exp(ty))$, see A5.

A7 $\quad$ El$'(\beta)$ and El$''(\beta)$.

Although it follows, under regularity conditions, from general likelihood theory that $El'(\beta) = 0$ and $El''(\beta) = -E[l'(\beta)]^2$, these equations can also be derived directly from the model.
Indeed,

$$El'(\beta) = -\frac{1}{\beta} + \frac{1}{\beta}\int \frac{\varphi(y-\beta)}{\Phi(y)-\Phi(y-\beta)}\left[\Phi(y)-\Phi(y-\beta)\right]dy = 0$$

and

$$\begin{aligned} E[l'(\beta)]^2 &= \frac{1}{\beta^2} - \frac{2}{\beta^2}\int \varphi(y-\beta)dy + \frac{1}{\beta}\int \frac{\varphi^2(y-\beta)}{\Phi(y)-\Phi(y-\beta)}dy \\ &= -\frac{1}{\beta^2} + \frac{1}{\beta}\int \frac{\varphi^2(y-\beta)}{\Phi(y)-\Phi(y-\beta)}dy. \end{aligned}$$

In the same way $El''(\beta)$ can be computed and is seen to be equal to $-E[l'(\beta)]^2$.

A8 $\quad$ Variance of $\hat{\beta}$ for small β.

In order to derive (14), we have to take the expansions of $l(\beta)$ and its derivatives in (10) to (12) one step further. The next term in the expansion of (25) is

$$\frac{\beta^4}{5!}(3-6y^2+y^4)\varphi(y).$$

From this and by using a further term, $-\frac{1}{4}z^4$, in the expansion of $\log(1+z)$, we get with some algebra, the next term of $l(\beta)$ in (10) as follows:

$$\frac{1}{2}\cdot\frac{\beta^4}{180}\left(-\frac{1}{8}y^4 - \frac{1}{2}y^2 + 4\right).$$

The next terms in the expansions of $l'(\beta)$ and $l''(\beta)$ in (11) and (12), then are, respectively,

$$\frac{1}{2}\cdot\frac{\beta^3}{45}\left(-\frac{1}{8}y^4 - \frac{1}{2}y^2 + 4\right). \tag{27}$$

and

$$\frac{1}{4} \cdot \frac{\beta^2}{15} \left(-\frac{1}{4} y^4 - y^2 + 8 \right). \tag{28}$$

We can now compute $El''(\beta)$ up to the order of β^2. Using (12) with the further term (28) and the expansion (25) up to the order of β^2, we have

$$\begin{aligned} El''(\beta) = \frac{1}{60} \int & \left[5y^2 - 20 - 15\beta y + \beta^2 (-\frac{1}{4} y^4 - y^2 + 8) \right] \\ & \cdot \left[1 + \frac{\beta}{2} y + \frac{\beta^2}{6} (y^2 - 1) \right] \varphi(y) dy + O(\beta^3). \end{aligned} \tag{29}$$

Multiplying both terms and computing the integral results in

$$El''(\beta) = -\frac{1}{4}(1 - \frac{\beta^2}{36}) + O(\beta^3). \tag{30}$$

From this we find (14), as $v = -\left[El''(\beta) \right]^{-1}$, but with $O(\beta^3)$ instead of $O(\beta^4)$.

If we go even one step further in the expansion of $l(\beta)$, we see that the remainder term in (14) is not $O(\beta^3)$, but actually $O(\beta^4)$. To see this, we need not expand $l(\beta)$ in any detail. It suffices to note that the next term of order β^5 in the expansion of $l(\beta)$ has a coefficient which is a polynomial in y with only uneven powers: y, y^3, y^5. The corresponding term in l'' is β^3 with the same polynomial as its coefficient. An evaluation of the integral in (29) with this further term and with the full expansion (25) shows that there is no term of order β^3 after the main term of (30) because the expected values of odd powers of y are all zero. Thus the term $O(\beta^3)$ in (30) can be replaced with $O(\beta^4)$, and consequently the remainder term in (14) is $O(\beta^4)$.

Another way to derive (14) is to start from

$$v = -\left[El''(\beta) \right]^{-1}$$

and use (13):

$$v = \beta \left[\int \frac{\varphi^2(y - \beta)}{\Phi(y) - \Phi(y - \beta)} dy - \frac{1}{\beta} \right]^{-1}$$

Owing to (5), for $n = 1$, or (22) this reduces to

$$v = \beta \left[\int l'(\beta) \varphi(y - \beta) dy \right]^{-1}. \tag{31}$$

Now we can use the expansion of $l^{'}(\beta)$ in (11) with the further term (27) and get for the integral in (31), after some algebra,

$$\int l^{'}(\beta)\varphi(y-\beta)dy = \frac{\beta}{4}(1-\frac{\beta^2}{36}) + O(\beta^4)$$

and from this again (14), albeit with the remainder term $O(\beta^3)$

A9 Asymptotic Variance of $\hat{\beta}_2$.

First note that

$$\sqrt{n}(s_y^2 - \sigma_y^2) \to N(0, V_y)$$

with

$$V_y = m_4(y) - \sigma_y^4.$$

According to A5, this is

$$\begin{aligned} V_y &= \left(\frac{\beta^4}{80} + \frac{\beta^2}{2} + 3\right) - \left(\frac{\beta^2}{12} + 1\right)^2, \\ &= \frac{\beta^4}{180} + \frac{\beta^2}{3} + 2. \end{aligned}$$

According to (18), the asymptotic variance of $\hat{\beta}_2$ is then given by

$$v_2 = 12\frac{V_y}{4(\sigma_y^2 - 1)}.$$

With $\sigma_y^2 - 1 = \frac{\beta^2}{12}$ this reduces to (19).

References

- Chan, N.N. (1982): Linear structural relationship with unknown error variances. *Biometrika* 69, 277-279.
- Davidov, O. and Goldenschluger, A. (2003): Fitting a linear segment to noisy data. To be published in *Journal of Statistical Planning and Inference.*

- E. L. Lehmann (1983): *Theory of Point Estimation.* J. Wiley; 1991 edition by Wadsworth, Belmont, Cal.
- Vosselman G. and Haralick, R. M. (1996): Performance analysis of line and circle fitting in digital images. *Proc. Workshop Performance Characteristics of Vision Algorithms.*
- Werman, M. and Keren, D. (2001): A Bayesian method for fitting parametric and nonparametric models to noisy data. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 23, 528-534.