

LETTER TO THE EDITOR: FITTING A FOLDED NORMAL DISTRIBUTION WITHOUT EM

BY IAIN L. MACDONALD

Centre for Actuarial Research, University of Cape Town, iaain.macdonald@uct.ac.za

The problem of fitting a folded normal distribution by maximum likelihood has been described as ‘not straightforward’, and alternatives such as EM proposed. We suggest here that it is in fact straightforward to fit such a distribution by direct numerical maximization of the likelihood. We demonstrate this in an example. The relevant R code is included.

1. Introduction. As a very small part of a much larger (and to me fascinating) piece of work, [Jung, Foskey and Marron \(2011\)](#) needed to find maximum-likelihood estimates (MLEs) of μ and σ from (independent) observations r_i assumed to arise as absolute values $R_i = |Y_i|$, where $Y_i \sim N(\mu, \sigma^2)$. They describe the problem as ‘not straightforward’ and early attempts to solve it, those of [Leone, Nelson and Nottingham \(1961\)](#), [Elandt \(1961\)](#) and [Johnson \(1962\)](#), as ‘very complicated’.

They therefore consider the use of EM. They derive the necessary E and M steps, which are indeed straightforward to implement once derived; the M step is available in closed form. They describe how to carry out EM, and use the resulting MLEs of μ and σ to estimate the ratio μ/σ , the quantity they actually need. After a simulation comparing the MLEs as found by EM with another (robust) estimator, they recommend the use of the robust estimator, partly because it is computationally light.

Even that small part of their paper has proved influential—see [Tsagris and Stewart \(2020\)](#), p. 257—so it seems worthwhile to point out to readers that this problem conveniently falls into a class of problems in which the MLEs can be found more efficiently via a readily available general-purpose unconstrained optimizer, for example, `nlm` in R. We illustrate this alternative, more direct, approach in the example of [Leone, Nelson and Nottingham \(1961\)](#), who did not have available to them the computational tools and power that we are fortunate to have now. The approach described here seems to be applicable, without alteration, to the problem of [Jung, Foskey and Marron](#).

2. Example of Leone, Nelson and Nottingham (1961). [Leone, Nelson and Nottingham \(1961\)](#) presented 497 grouped observations of the camber (in milliradians) of leads that needed to be placed straight in a certain manufacturing process. Ideally the camber is zero, and in general it is nonnegative. (The observations range in approximate magnitude from 1 to 43, with sample mean 14.01 and sample standard deviation 7.79.) For these observations, [Leone, Nelson and Nottingham](#) proposed and fitted a normal distribution folded at zero.

The relevant density is, for $r > 0$,

$$f(r) = n(r; \mu, \sigma^2) + n(-r; \mu, \sigma^2),$$

where n denotes the density of a normal with the mean and variance indicated. The (observed-data) log likelihood of independent observations r_i is then $l = \sum_i \log f(r_i)$, which can be evaluated and numerically maximized with respect to μ and σ , with the constraint that $\sigma > 0$. To constrain σ to be positive, the usual technique is to optimize over $\log \sigma$ rather than over

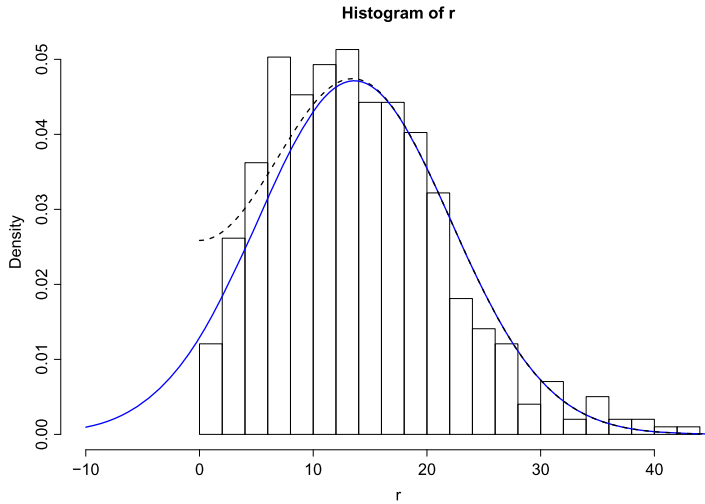


FIG. 1. Camber data of Leone, Nelson and Nottingham (1961); compare with their Figure 1. Also shown here are the $N(\hat{\mu}, \hat{\sigma}^2)$ density and the corresponding folded normal density (dashed line).

σ , and to use an unconstrained optimizer. It is convenient that `nlm`, a modified Newton optimizer, does not require that derivatives be supplied. If one follows this route there is of course no need to derive or code the E and M steps of EM, nor to implement any EM-based method of finding standard errors or confidence intervals.

The obvious starting values to use for μ and σ are the sample mean and standard deviation of the observations. But when applied to this problem with deliberately poor starting values, `nlm` (with defaults) terminated successfully in 12 iterations. The (excellent) estimates given by Leone, Nelson and Nottingham for μ and σ were 13.6424 and 8.4254, which imply that $-l = 1703.994$. The MLEs found here are $\hat{\mu} = 13.6120$ and $\hat{\sigma} = 8.4626$, with $-l = 1703.986$. Corresponding approximate 95% confidence intervals of Wald type are found from the numerical Hessian supplied by `nlm`: $\mu \in (12.823, 14.401)$ and $\log \sigma \in (2.056, 2.215)$. The latter interval implies the interval $\sigma \in (7.813, 9.166)$, which is (slightly) asymmetric about the MLE and probably preferable to an interval of Wald type for σ . Figure 1 shows a plot of the data, the $N(\hat{\mu}, \hat{\sigma}^2)$ density, and the corresponding density for the absolute values R_i . The R code needed to find the MLEs appears in the Appendix.

3. Discussion. This note serves as a reminder of the broad array of powerful numerical methods that can often be applied to solve problems that currently arise in computational statistics. Although EM conveniently solves some of these problems, other more powerful methods can be considerably more efficient. This approach is illustrated here by an example from Leone, Nelson and Nottingham (1961), but appears to be quite generally applicable to finding the MLEs of a folded normal. It is hoped that this note will provide a useful (if minor) addition to the impressive work of Jung, Foskey and Marron.

APPENDIX: R CODE

The R code in this Appendix finds the MLEs for the data of Leone, Nelson and Nottingham.

```
midpoints=2*(1:22)-1
counts=c(12,26,36,50,45,49,51,44,44,40,32,18,14,12,4,7,2,5,2,2,1,1)
# Data of Leone et al. (1961)
(r=rep(midpoints,counts))
f=function(rr,m,s) dnorm(rr,mean=m,sd=s)+dnorm(-rr,mean=m,sd=s)
```

```
minusl=function(params) {  
mu=params[1]  
sigma=exp(params[2])  
-sum(log(f(r,mu,sigma)))  
}  
minusl(c(13.6424,log(8.4234))) #These are Leone's estimates.  
start=c(10,log(10))  
(model=nlm(minusl,start))  
muest=model$estimate[1]; sigmaest=exp(model$estimate[2])  
c(muest,sigmaest)
```

REFERENCES

- ELANDT, R. C. (1961). The folded normal distribution: Two methods of estimating parameters from moments. *Technometrics* **3** 551–562. [MR0130738 https://doi.org/10.2307/1266561](https://doi.org/10.2307/1266561)
- JOHNSON, N. L. (1962). The folded normal distribution: Accuracy of estimation by maximum likelihood. *Technometrics* **4** 249–256. [MR0137221 https://doi.org/10.2307/1266622](https://doi.org/10.2307/1266622)
- JUNG, S., FOSKEY, M. and MARRON, J. S. (2011). Principal arc analysis on direct product manifolds. *Ann. Appl. Stat.* **5** 578–603. [MR2810410 https://doi.org/10.1214/10-AOAS370](https://doi.org/10.1214/10-AOAS370)
- LEONE, F. C., NELSON, L. S. and NOTTINGHAM, R. B. (1961). The folded normal distribution. *Technometrics* **3** 543–550. [MR0130737 https://doi.org/10.2307/1266560](https://doi.org/10.2307/1266560)
- TSAGRIS, M. and STEWART, C. (2020). A folded model for compositional data analysis. *Aust. N. Z. J. Stat.* **62** 249–277.