KM 37(1991)
pag 19 - 32

A NOTE ON THE USE OF THE PRODUCT OF SPACINGS IN BAYESIAN INFERENCE

F. P. A. Coolen & M. J. Newby[1],
Eindhoven University of Technology,
The Netherlands.

Abstract: The product of spacings is suggested as an alternative to the likelihood in Bayesian inference. It is shown the product of spacings can be used in place of the likelihood in Bayesian inference without losing the structure and properties of the Bayesian method. The method is also shown to have computational advantages.

KEYWORDS: likelihood; Bayes theorem; Bayesian inference; product of spacings; estimation; posterior densities.

1. Introduction

This note arose from the consideration of two problems that occur in classical likelihood estimation and are inherited from it by Bayesian methods. The problems arise from some of the shortcomings of the likelihood function: (i) in some circumstances the likelihood function is unbounded; (ii) the sensitivity of the likelihood function to outliers. The importance of sensitivity is to some extent context dependent, but the unboundedness of the likelihood function can be a serious impediment in both classical and Bayesian analysis. Here we want to give a brief outline of the use of the product of spacings and show its potential in Bayesian analyses.

Consider the problem of estimating a parameter ϑ in the univariate distribution $F(t|\vartheta)$ with density function $f(t|\vartheta)$. The problem of an unbounded likelihood most commonly arises when the parameter ϑ is in the boundary of the support of f, for example the maximum likelihood estimator of the left hand end-point of a domain is almost always the first order statistic (Cohen and Whitton-Jones, 1989). Indeed, for any densities which are J-shaped or heavy-tailed maximum-likelihood is bound to fail (Cheng and Amin, 1983;

---

[1] Faculty of Industrial Engineering
Technical University Eindhoven
PO Box 513
5600 MB Eindhoven tel: (040) - 472729

Ranneby, 1984). In these cases the derivation of a posterior density function for the parameter ϑ may also be problematic.

Our objective here is to summarize the properties of the maximum product of spacings method as given, with rather different perspectives, by Cheng and Amin (1983), Ranneby (1984), and Titterington (1985), and then to illustrate its use in some simple Bayesian analyses.

2. Product of Spacings

The maximum product of spacings method has been known implicitly (Titterington, 1985) for a long time, but was first formally defined and analysed by Cheng and Amin (1983) and Ranneby (1984). Cheng and Amin (1983) began by attempting to replace the likelihood function by an alternative which retained as many of the useful properties of the method of maximum likelihood as possible. Ranneby (1984) began from an information theoretic problem: he noted that the likelihood is an approximation for the Kullback-Liebler information and sought other satisfactory approximations for this measure of distance between a fitted distribution and the true distribution. The approach of Cheng and Amin (1983) is more intuitively attractive and can, to some extent, be regarded as a pragmatic solution to the problems associated with likelihood (Titterington, 1985), but that of Ranneby is more powerful theoretically and allows the derivation of the properties of maximum product of spacings estimators. Many related results and the required proof techniques can be found in the review paper by Pyke (1965).

The approach is most easily illustrated by considering a univariate distribution $F(t|\vartheta)$ with density $f(t|\vartheta)$ where it is required to estimate ϑ. The density is assumed to be strictly positive in an interval (α, β) and zero elsewhere, α and β may also be elements of ϑ, $\alpha=-\infty$ and $\beta=\infty$ are included. That is $F(t|\vartheta)=0$ and $f(t|\vartheta)=0$ for $t<\alpha$, $F(t|\vartheta)=1$ and $f(t|\vartheta)=0$ for $t>\beta$. Let $t_1 < t_2 < t_3 < \ldots < t_n$ be a complete ordered sample, further define $t_0=\alpha$, $t_{n+1}=\beta$. The spacings are defined through the probability integral transform as follows.

$$u_i = F(t_i|\vartheta) \;,\; i=0, \;\dots\; ,n+1$$

$$D_i = u_i - u_{i-1} \quad , \; i=1, \;\dots\; ,n+1.$$

If the true distribution, F, with a true parameter value, ϑ_0, is chosen then the u_i are order statistics from the standard uniform density. In this case the sum of the D_i's is 1 and the expected value $\boldsymbol{E}[D_i] = \frac{1}{n+1}$. The product of spacings method utilises the geometric mean of the spacings

$$G(\vartheta) = \left\{ \prod_{i=1}^{n+1} D_i \right\}^{1/(n+1)} .$$

Which in view of the preceding remarks is a bounded function of the parameters. Furthermore the maximum value of G will only be obtained if the D_i are all equal which corresponds to choosing the true value ϑ_0 in the parameter space. Thus as Ranneby points out the function $G(\vartheta)$ is a measure of deviation from the true model. The maximum–product–of–spacings method obtains estimators by maximising G as a function of ϑ. As with likelihood the approach is usually to maximize $\mathcal{S}=ln(G)$. It is clear that estimation can also proceed directly from the product of spacings itself and that the same estimators will be obtained. Since Bayes theorem requires probabilities we use the product of spacings

$$\mathcal{G}(\vartheta) = \prod_{i=1}^{n+1} D_i \quad .$$

in the rest of this note. The above observation is also a natural consequence of what in essence has been a pragmatic version of the product of spacings obtained by grouping data to give a grouped–likelihood without singularities (Titterington, 1985).

The function $\mathcal{G}$ has many of the properties of a likelihood, the simpler forms of censoring and truncation can also be handled exactly as in the usual likelihood approach, with each censored observation, t^*, contributing a term $1-F(t^*)$ to the product, and truncation at t_a and t_b dividing each contribution by $F(t_b)-F(t_a)$. It follows that the likelihood principle can be maintained (Press, 1989). The product can readily be updated to take account of new observations, but without the simplicity of the likelihood. For discrete distributions there is no problem with the likelihood, and in some senses the use of $\mathcal{G}$ in place of a standard likelihood can be seen as replacing some unpleasant qualities of a continuous density function with the more attractive properties of a discrete probability mass function. The product of spacings is itself a probability function on the sample space. It is also clear that the invariance properties of maximum product of spacings estimators are the same as those of maximum likelihood estimators. That is if $\phi=\psi(\vartheta)$ is a 1–1 transformation then the estimator of ϕ is $\hat{\phi}=\psi(\hat{\vartheta})$ where $\hat{\vartheta}$ is the estimator of ϑ. More interestingly, Ranneby showed that: (i) the estimator of ϑ is invariant under monotone (and therefore order preserving) transformations of the data; (ii) that $\sqrt{n}\mathcal{S}(\vartheta_0)+\gamma$, with $\mathcal{S}(\vartheta_0)=ln[G(\vartheta_0)]$, ϑ_0 the true value of ϑ, and γ Euler's constant, is asymptotically normally distributed with zero mean and variance $\frac{\pi^2}{6} - 1$, thus providing an immediate classical test of fit along with the estimates; (iii) that the estimators themselves are asymptotically normally distributed around the true values.

3. Bayesian Inference

Now that $\mathcal{G}$ has been described in the context of an approximation to a likelihood, or as an estimating function in its own right, its rôle in Bayesian inference can be examined. Firstly, as an approximation to a likelihood $\mathcal{G}$ can be used directly in the Bayes equation, and secondly, it is a probability function in its own right as the product of the probability masses associated with the spacings t_i-t_{i-1}. Parameter free estimates, for example the Kaplan–Meier, of the distribution F yield parameter free versions of $\mathcal{G}$. More importantly, as noted above, $\mathcal{G}$ maintains the likelihood principle so that the handling of new observations and censoring will still fall within the usual Bayesian framework.

The idea of a conjugate prior may no longer be of use, the definition of a conjugate depends on the likelihood, and it is not clear whether there are classes of distributions which would be conjugate with respect to the product of spacings. Although the loss of the idea of a conjugate prior may make it harder to see the separate contributions of the prior and the data to an estimator, it is no loss from the technical point of view since there are now sufficiently many effective numerical methods available to handle the integrations required in the Bayesian context (Smith *et al.*, 1985).

To introduce $\mathcal{G}$ into the Bayesian framework consider the ordered sample $\{t_i\}$ used above and the calculation of a posterior density for ϑ. Write $\mathcal{G}$ as $\mathcal{G}(\text{data}|\vartheta)$ and $p(\vartheta)$ for the prior density of ϑ. In standard Bayes we can derive the posterior density from a prior p and likelihood $\mathcal{L}$ as follows.

$$p_L(\vartheta|\{t_i\}_{i=1}^{n})d\vartheta = \frac{\left\{\prod_{i=1}^{n} f(t_i|\vartheta)dt_i\right\} \times p(\vartheta)d\vartheta}{\int\left\{\prod_{i=1}^{n} f(t_i|\vartheta)dt_i\right\} \times p(\vartheta)d\vartheta}$$

$$p_L(\vartheta|\{t_i\}_{i=1}^{n}) = \frac{\mathcal{L}(\text{data}|\vartheta)p(\vartheta)}{\int\mathcal{L}(\text{data}|\vartheta)p(\vartheta)d\vartheta}$$

where the integral in the denominator is over all possible values of ϑ. Now in a rough and ready way we can write

$$f(t_i|\vartheta)dt_i \cong \Delta F(t_i|\vartheta) = D_i$$

$$p_L(\vartheta|\{t_i\}_{i=1}^{n})d\vartheta \cong p_G(\vartheta|\{t_i\}_{i=1}^{n})d\vartheta = \frac{\prod_{i=1}^{n}\Delta F(t_i|\vartheta)\times p(\vartheta)d\vartheta}{\int\prod_{i=1}^{n}\Delta F(t_i|\vartheta)\times p(\vartheta)d\vartheta}$$

$$p_L(\vartheta|\{t_i\}_{i=1}^{n}) \cong p_G(\vartheta|\{t_i\}_{i=1}^{n}) = \frac{\mathcal{G}(\text{data}|\vartheta)\times p(\vartheta)}{\int\mathcal{G}(\text{data}|\vartheta)\times p(\vartheta)d\vartheta}$$

In view of the remarks above and in section **2** there are at this point no theoretical problems associated with using the product of spacings in place of a likelihood. The posterior is certainly not the posterior obtained from the likelihood, but following Cheng and Amin (1983), Ranneby (1984), and Titterington (1985) the asymptotic equivalence of $\mathcal{G}$ and the likelihood show that p_G is asymptotically equivalent to the posterior obtained in the standard way. Further, if the prior is continuous and bounded so is p_G as the product of two continuous bounded functions. This removes some of the problems associated with distributions defined on finite intervals with unknown endpoints.

4. Examples

Now that the validity of the product of spacings as an alternative to the likelihood has been demonstrated it is useful to compare the performance of a standard Bayesian approach to one where the product of spacings is used. We give three examples to illustrate the differences in the case where there is a simple parameter estimation problem, and one in which the endpoint of the support is also a parameter.

Example 1: *sensitivity*

Here the problem is to estimate the parameter λ of an exponential distribution

$$f(t|\lambda) = \lambda \cdot \exp(-\lambda t)$$

$$F(t|\lambda) = 1 - \exp(-\lambda t)$$

with a simple discrete prior $p(\lambda=1)=0.5$, $p(\lambda=4)=0.5$, and with three observations, $t_1=0.1$, $t_2=0.3$, $t_3=0.6$, with $t_0=0$ and $t_4=\infty$.

The likelihood is

$$\mathcal{L}(\lambda|\text{data}) = \lambda^3 \exp(-\lambda[t_1+t_2+t_3])$$

and the product of spacings is

$$\mathcal{G}(\lambda|\text{data}) = \prod_{i=1}^{4} [\exp(-\lambda t_{i-1}) - \exp(-\lambda t_i)] \ .$$

The estimators are:

maximum likelihood $\hat{\lambda} = 3.00$

maximum product of spacings $\tilde{\lambda} = 2.36$.

For the posteriors calculate:

$\mathcal{L}(1\,|\,\text{data}) = \exp(-1) = 0.3679;$ $\quad \mathcal{L}(4|\text{data}) = 64\exp(-4) = 1.1722;$

$\mathcal{G}(1|\text{data}) = 0.0016;$ $\quad \mathcal{G}(4|\text{data}) = 0.0023.$

The posterior densities are

$$p_L(1|\text{data}) = 0.239,\ p_L(4|\text{data}) = 0.761, \text{ with } \boldsymbol{E}(\lambda) = 3.28$$

and

$$p_G(1|\text{data}) = 0.414,\ p_G(4|\text{data}) = 0.586, \text{ with } \boldsymbol{E}(\lambda) = 2.76.$$

Thus the effect of the one larger observation t_3 is seen to be smaller both in estimation directly from the product of spacings and in the Bayesian estimates. Thus the product of spacings appears to give an outlier less weight than the likelihood.

example 2: *sufficiency*

Cheng and Amin (1983) considered how far the idea of sufficiency could be retained in the product of spacings method. Continuing with the above example on the exponential distribution, $F(t|\lambda) = 1-\exp(-\lambda t)$, shows that the product of spacings distinguishes between samples with the same total time on test, whereas likelihood sees all samples with the same value of the total time on test as the same because the total time on test is a sufficient statistic for λ. Consider the situation of example 1 but now with three samples, $\mathcal{X}_1=\{0.01,\ 0.99\}$, $\mathcal{X}_2=\{0.2,\ 0.8\}$, and $\mathcal{X}_3=\{0.4,\ 0.6\}$. The likelihood does not distinguish between these samples because the total time on test is 1 for all three and so all three give the same posterior density, $p_L(\lambda=1)=0.56$, and $p_L(\lambda=4)=0.44$. On the other hand the product of spacings function associated with each sample is different:

sample 1 – $\mathcal{G}(\mathcal{X}_1|\lambda=1)=0.002$, $\mathcal{G}(\mathcal{X}_1|\lambda=4)=0.0007$; $p_G(\lambda=1)=0.76$, $p_G(\lambda=4)=0.24$;
sample 2 – $\mathcal{G}(\mathcal{X}_2|\lambda=1)=0.030$, $\mathcal{G}(\mathcal{X}_2|\lambda=1)=0.009$; $p_G(\lambda=1)=0.77$, $p_G(\lambda=4)=0.23$;
sample 3 – $\mathcal{G}(\mathcal{X}_3|\lambda=1)=0.022$, $\mathcal{G}(\mathcal{X}_3|\lambda=4)=0.008$; $p_G(\lambda=1)=0.73$, $p_G(\lambda=4)=0.27$.

Thus there is a different posterior associated with each sample. Since the product of spacings is asymptotically equivalent to the likelihood this should be a small sample phenomenon.

Example 3: *singularities in the likelihood*

Cheng and Amin give an example of a truncated exponential density, $f(t|\alpha)=\exp[-(t-\alpha)]$, $f(t|\alpha)=0$ for $t<\alpha$, to demonstrate how the product of spacings method handles the estimation of the location parameter α. To make the point more forcibly we consider both their example and the estimation of the location parameter in a Weibull distribution $F(t|\alpha) = 1-\exp(-[t-\alpha]^{\frac{1}{2}})$, with density $f(t|\alpha) = \frac{1}{2}(t-\alpha)^{-\frac{1}{2}}\exp(-[t-\alpha]^{\frac{1}{2}})$. In this case both the likelihood and the product of spacings exist, but the likelihood has a singularity at the smallest sample value. An un-normalised posterior can be obtained from the likelihood, but we have not investigated whether the singularity prevents the calculation of a normalised posterior. Certainly such a singularity causes numerical problems requiring careful handling when writing computer programs to carry out Bayesian analyses. Because the product of spacings is a bounded continuous function taking the value zero for $t-\alpha<0$, the minimum observation may be an interior point of the support of the prior without causing problems. Indeed, the product of spacings results in a posterior which assigns zero probability to values of the location parameter greater than the smallest observation.

We simulated a sample of 15 observations from the distribution $F(t|1)$ and compared the product of spacings method and likelihood. The ordered data, $\{t_i\}_{i=1}^{15}$, are

1.0006	1.0087	1.0682	1.1084	1.1823
1.2256	1.3357	1.4616	1.9437	2.2487
3.0994	3.9001	4.0802	7.8657	9.9195

The prior was $\boldsymbol{\beta}(6,3)$, $\boldsymbol{E}(\alpha)=\frac{4}{3}$, spread over the interval (0,2). The likelihood, product of spacings, and the posterior are plotted as functions of α in Figure 1. The likelihood shows the singularity $\alpha=t_1$, $\mathcal{G}$ has a clear maximum, and the posterior density obtained from the product of spacings has a well defined maximum.

The estimates of α are:

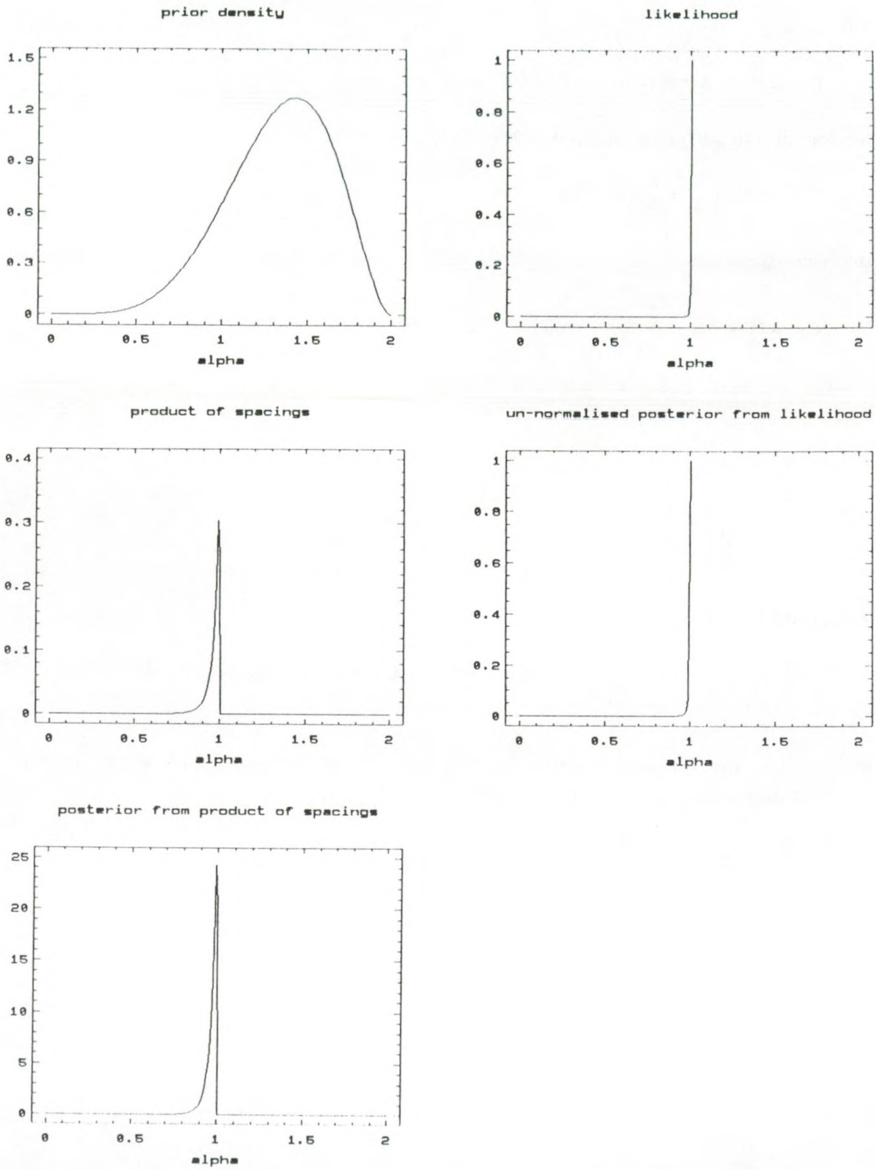
maximum likelihood estimator $\hat{\alpha} = 1.0006$;

maximum product of spacings estimator $\tilde{\alpha} = 0.99$.

and the squared error loss function estimator is:

posterior mean using spacings $\boldsymbol{E}(\alpha) = 0.97$.

FIGURE 1: Weibull example



The method also copes well with situations in which more than one parameter is to be estimated. Consider a shifted exponential distribution

$$F(t|\alpha,\lambda) = 1 - \exp\{-\lambda(t-\alpha)\} = 1 - e^{\lambda\alpha}\exp\{-\lambda t\}, \ t\geq\alpha$$

with density

$$f(t|\alpha,\lambda) = \lambda \exp\{-\lambda(t-\alpha)\} = \lambda e^{\lambda\alpha}\exp\{-\lambda t\}, \ t\geq\alpha$$

the log-likelihood for an ordered sample $\{t_i\}_{i=1}^{n}$ is

$$\mathcal{L} = n\mathit{ln}(\lambda) - \lambda\sum_{i=1}^{n}(t_i - \alpha)$$

the derivatives are

$$\mathcal{L}_\lambda = \frac{n}{\lambda} + n\alpha - \sum_{i=1}^{n} t_i \qquad \text{and} \qquad \mathcal{L}_\alpha = n\lambda$$

Clearly $\mathcal{L}_\alpha$ is a positive increasing function of α for all λ and so the likelihood estimates are:

$$\hat{\alpha} = t_1$$

and

$$\hat{\lambda} = n/\left\{\sum_{i=2}^{n}(t_i - t_1)\right\}.$$

The product of spacings is

$$\mathcal{G} = \left\{1 - e^{-\lambda(t_1-\alpha)}\right\}\left\{\prod_{i=2}^{n}\left[e^{-\lambda(t_{i-1}-\alpha)} - e^{-\lambda(t_i-\alpha)}\right]\right\}e^{-\lambda(t_n-\alpha)}$$

and we use the function $\mathcal{H}=\ell n(\mathcal{G})$ as the basis for estimation. With some care the required derivatives can be written as

$$\mathcal{H}_\alpha = \frac{\lambda\, e^{-\lambda(t_1-\alpha)}}{1 - e^{-\lambda(t_1-\alpha)}} - n\lambda$$

$$\mathcal{H}_\lambda = \frac{(t_1-\alpha)e^{-\lambda(t_1-\alpha)}}{1 - e^{-\lambda(t_1-\alpha)}} + \sum_{i=1}^{n} \frac{(t_i-\alpha)e^{-\lambda(t_i-\alpha)} - (t_{i-1}-\alpha)e^{-\lambda(t_{i-1}-\alpha)}}{e^{-\lambda(t_{i-1}-\alpha)} - e^{-\lambda(t_i-\alpha)}} - (t_n-\alpha)$$

Solving $\mathcal{H}_\alpha=0$ yields

$$\tilde{\alpha} = t_1 - \frac{1}{\lambda}\,\ell n(1 + \frac{1}{n})$$

and substituting $\tilde{\alpha}$ in $\mathcal{H}_\lambda$ gives

$$\mathcal{H}_\lambda(\tilde{\alpha}) = \sum_{i=1}^{n} \frac{(t_i-t_1)e^{-\lambda(t_i-t_1)} - (t_{i-1}-t_1)e^{-\lambda(t_{i-1}-t_1)}}{e^{-\lambda(t_{i-1}-t_1)} - e^{-\lambda(t_i-t_1)}} - (t_n-t_1)\ .$$

Neither $\mathcal{H}_\lambda=0$ or $\mathcal{H}_\lambda(\tilde{\alpha})=0$ yields an explicit solution, but $\mathcal{H}_\lambda(\tilde{\alpha})$ is readily evaluated and is a monotone decreasing function of λ so that a simple search gives an estimator $\tilde{\lambda}$ which in turn can be used to evaluate $\tilde{\alpha}$.

To illustrate the method a simulated sample of size 10 was drawn from the distribution with $\lambda=5$ and $\alpha=1$. The ordered data were

1.0331	1.0422	1.0428	1.0549	1.0977
1.1455	1.1586	1.3109	1.4993	1.9482

From the above the estimators are:

maximum–likelihood	$\hat{\alpha} = 1.0331$	$\hat{\lambda} = 4.99$
product of spacings	$\tilde{\alpha} = 1.0079$	$\tilde{\lambda} = 3.78$

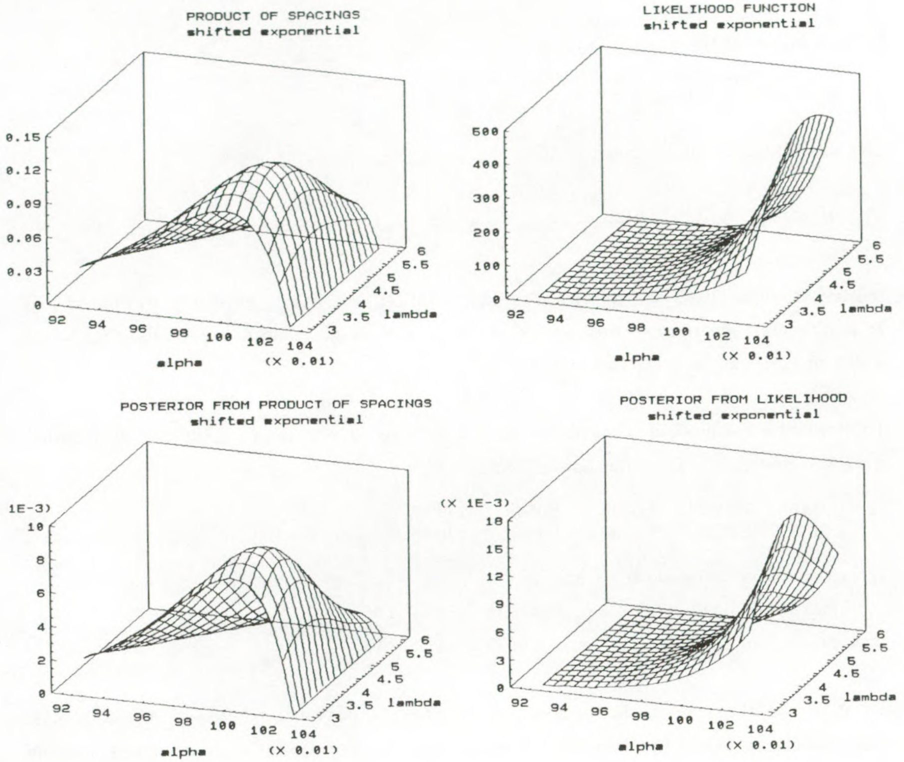
To illustrate the use of the product of spacings in Bayesian analysis a prior, $p(\alpha,\lambda)$, was chosen in which the two parameters α and λ were taken as independent random variables with marginal distributions $\beta(6,3)$ on $(0,2)$ and $\Gamma(3, \frac{4}{3})$ respectively. Thus the prior density is just the product of the two marginal densities. The prior means for α and λ are $\boldsymbol{E}(\alpha)=\frac{4}{3}$ and $\boldsymbol{E}(\lambda)=4$. The posteriors are then simply obtained from

$$p_G \propto \mathcal{G}(\alpha,\lambda\,|\,\text{data})\times p(\alpha,\lambda)$$

and

$$p_L \propto \mathcal{L}(\alpha,\lambda\,|\,\text{data})\times p(\alpha,\lambda)\ .$$

FIGURE 2: Shifted exponential



The squared loss Bayes estimators are, based on a crude numerical calculation:

posterior means using spacings $E(\alpha) = 0.99; E(\lambda) = 3.98;$

posterior means using likelihood $E(\alpha) = 1.01; E(\lambda) = 4.35$.

The product of spacings, likelihood, and the two versions of the posterior density can be seen in Figure 2.

In these examples the product of spacings shows clear advantages over the likelihood. Firstly, from a theoretical point of view there is no problem dealing with values of the location parameter interior to the support of the prior, secondly, as a result of this first remark there are no numerical problems caused by singularities.

5. Conclusion

This note shows how the maximum product of spacings can be used to replace the likelihood in a Bayesian argument, and that all the necessary properties of the likelihood are also possessed by the product of spacings. Some properties of the likelihood are lost because the ordering by magnitude of the observations is required. Further, since Bayes theorem requires only a conditional probability and a prior, it can be seen that choices other than the likelihood are available as the joint probability of a particular set of observations conditioned on a parameter. The idea of a conjugate prior may be lost, and the rôle of sufficiency is less clear.

From a number of simulations the product of spacings appears to give less weight to the data than the likelihood function. However since the product of spacings is a bounded continuous function of the parameters its numerical behaviour is better than that of the likelihood.

References

Cheng, R. C. H., and Amin, N. A. K., (1983), Estimating parameters in continuous univariate distributions with shifted origin, *Journal of the Royal Statistical Society*, **B**, **45**, No. 3, 394–403.

Cohen, A. C., and Jones Whitton, B., (1988), *Parameter Estimation in Reliability and Life Span Models*, Marcel Dekker, NY.

Press, S. J, (1989), *Bayesian Statistics: Principles, Models and Applications*, John Wiley, NY.

Pyke, R., (1965) Spacings (with discussion), *Journal of the Royal Satistical Society*, **B**, **27**, 395–449

Ranneby, Bo, (1984), The maximum spacing method. An estimation method related to the maximum likelihood method, *Scandinavian Journal of Statistics*, **11**, 93–112.

Smith, A. F. M., Skene, A. M., Shaw, J. E. H., Naylor, J. C., & Dransfield, M., (1985), The implementation of the Bayesian paradigm, *Communications in Statistics Theory and Methods*, **14**(5), 1079–1102.

Titterington, D. M., (1985), Comment on "Estimating parameters in continuous univariate distributions", *Journal of the Royal Statistical Society*, **B**, **47**, No. 1, 115–116.

Ontvangen: 25-09-1990
Geaccepteerd: 03-04-1991