A note on the use of the product of spacings in Bayesian inference

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 90-35

A note on the use of the product of spacings in Bayesian inference

F.P.A. Coolen
M.J. Newby

Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Eindhoven, September 1990
The Netherlands

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, they are (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.

Consider the problem of estimating a parameter θ in the univariate distribution $F(t|\theta)$ with density function $f(t|\theta)$. 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

---

[1] Address for correspondence: Dr M J Newby, Faculty of Industrial Engineering, Department of Operational Research and Statistics, PO Box 513, 5600 MB Eindhoven, The Netherlands.

This note has been offered to *Statistics and Probability Letters*.

heavy-tailed maximum-likelihood is bound to fail (Cheng and Amin, 1983; 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.

The approach is most easily illustrated by considering a univariate distribution $F(t|\theta)$ with density $f(t|\theta)$ 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 $F(t|\theta)=0$ and $f(t|\theta)=0$ for $t<\alpha$, $F(t|\theta)=1$ and $f(t|\theta)=0$ for $t>\beta$. Let $t_1 < t_2 < t_3 < \dots < t_n$ 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.

$$u_i = F(t_i|\theta) \ ,$$

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

and the estimation follows by maximising the geometric mean of the spacings

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

As with likelihood the approach is usually to maximize $S=ln(G)$. 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

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

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. More interestingly, Ranneby showed that: (i) the estimator of θ is invariant under monotone (and therefore order preserving) transformations of the data; (ii) that $\sqrt{n}\mathcal{S}+\gamma$, where $\mathcal{S}=ln(\mathcal{G})$ and γ is 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 $\mathcal{G}(\text{data}|\theta)$ and $p(\theta)$ for the prior density of θ. Then Bayes theorem gives the posterior $\tilde{p}$ for θ in the light of the data as

$$\tilde{p}(\theta|\text{data}) = \frac{\mathcal{G}(\text{data}|\theta)p(\theta)}{\int \mathcal{G}(\text{data}|\theta)p(\theta)d\theta}$$

where the integral in the denominator is over all possible values of θ.

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 $\tilde{p}$ is asymptotically equivalent to the posterior obtained in the standard way. Further, if the prior is continuous and bounded so is $\tilde{p}$ 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

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

so that $\mathcal{L}(1|\text{data})=\exp(-1)=0.3679$ and $\mathcal{L}(4|\text{data})=64\exp(-4)=1.1722$, the posterior density is

$$\tilde{p}_L(1|\text{data}) = 0.239 \text{ and } \tilde{p}_L(4|\text{data}) = 0.761, \text{ with expected } E(\lambda) = 3.28$$

With the product of spacings the function $\mathcal{G}$ is given by

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

so that $\mathcal{G}(1|\text{data}) = 0.0016$ and $\mathcal{G}(4|\text{data}) = 0.0023$, the posterior density is

$$\tilde{p}_G(1|\text{data}) = 0.414 \text{ and } \tilde{p}_G(4|\text{data}) = 0.586, \text{ with expected value } 2.76.$$

Thus the effect of the one larger observation t_3 is seen to be smaller, this is consistent with regarding that observation as an outlier rather than an influential observation.

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, $\tilde{p}_L(\lambda=1)=0.56$, and $\tilde{p}_L(\lambda=4)=0.44$. On the other hand the product of spacings function associated with each sample is different:

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

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 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 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 $\beta(6,3)$, $E(\alpha)=0.67$, spread over the interval (0,2). The likelihood, product of spacings, and the posterior are plotted as functions of α in Figures 1–5. The likelihood shows the singularity $\alpha=t_1$, $\mathcal{G}$ has a clear

Figure 1: Prior density

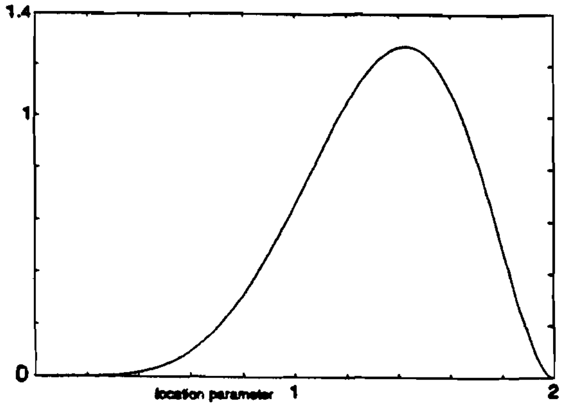


Figure 2: Likelihood

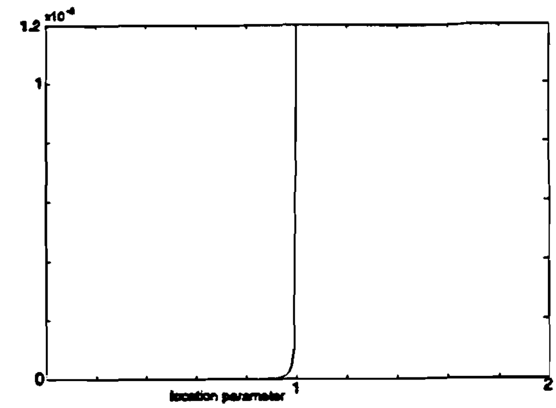


Figure 3: Product of spacings

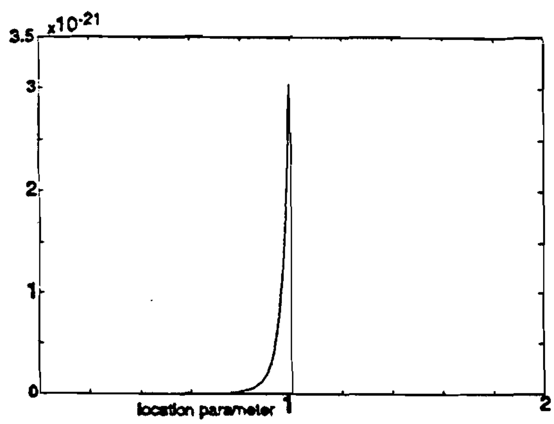


Figure 4: Un-normalized posterior from likelihood

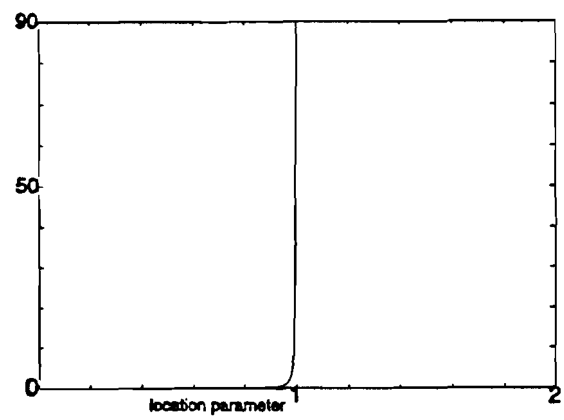
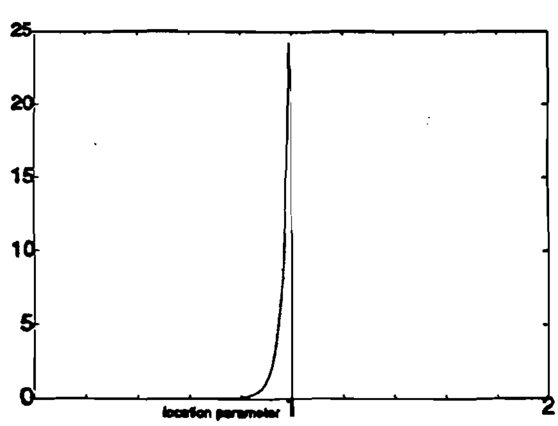


Figure 5: Posterior from product of spacings



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 $\breve{\alpha} = 0.97$.

In this example 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.

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.

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science
PROBABILITY THEORY, STATISTICS, OPERATIONS RESEARCH AND SYSTEMS THEORY
P.O. Box 513
5600 MB Eindhoven - The Netherlands
Secretariate: Dommelbuilding 0.03
Telephone: 040 - 47 3130

List of COSOR-memoranda - 1990

Number	Month	Author	Title
M 90-01	January	I.J.B.F. Adan J. Wessels W.H.M. Zijm	Analysis of the asymmetric shortest queue problem Part 1: Theoretical analysis
M 90-02	January	D.A. Overdijk	Meetkundige aspecten van de productie van kroonwielen
M 90-03	February	I.J.B.F. Adan J. Wessels W.H.M. Zijm	Analysis of the assymmetric shortest queue problem Part II: Numerical analysis
M 90-04	March	P. van der Laan L.R. Verdooren	Statistical selection procedures for selecting the best variety
M 90-05	March	W.H.M. Zijm E.H.L.B. Nelissen	Scheduling a flexible machining centre
M 90-06	March	G. Schuller W.H.M. Zijm	The design of mechanizations: reliability, efficiency and flexibility
M 90-07	March	W.H.M. Zijm	Capacity analysis of automatic transport systems in an assembly factory
M 90-08	March	G.J. v. Houtum W.H.M. Zijm	Computational procedures for stochastic multi-echelon production systems

Number	Month	Author	Title
M 90-09	March	P.J.M. van Laarhoven W.H.M. Zijm	Production preparation and numerical control in PCB assembly
M 90-10	March	F.A.W. Wester J. Wijngaard W.H.M. Zijm	A hierarchical planning system versus a schedule oriented planning system
M 90-11	April	A. Dekkers	Local Area Networks
M 90-12	April	P. v.d. Laan	On subset selection from Logistic populations
M 90-13	April	P. v.d. Laan	De Van Dantzig Prijs
M 90-14	June	P. v.d. Laan	Beslissen met statistische selectiemethoden
M 90-15	June	F.W. Steutel	Some recent characterizations of the exponential and geometric distributions
M 90-16	June	J. van Geldrop C. Withagen	Existence of general equilibria in infinite horizon economies with exhaustible resources. (the continuous time case)
M 90-17	June	P.C. Schuur	Simulated annealing as a tool to obtain new results in plane geometry
M 90-18	July	F.W. Steutel	Applications of probability in analysis
M 90-19	July	I.J.B.F. Adan J. Wessels W.H.M. Zijm	Analysis of the symmetric shortest queue problem
M 90-20	July	I.J.B.F. Adan J. Wessels W.H.M. Zijm	Analysis of the asymmetric shortest queue problem with threshold jockeying
M 90-21	July	K. van Ham F.W. Steutel	On a characterization of the exponential distribution
M 90-22	July	A. Dekkers J. van der Wal	Performance analysis of a volume shadowing model

Number	Month	Author	Title
M 90-23	July	A. Dekkers J. van der Wal	Mean value analysis of priority stations without preemption
M 90-24	July	D.A. Overdijk	Benadering van de kroonwielflank met behulp van regeloppervlakken in kroonwieloverbrengingen met grote overbrengverhouding
M 90-25	July	J. van Oorschot A. Dekkers	Cake, a concurrent Make CASE tool
M 90-26	July	J. van Oorschot A. Dekkers	Measuring and Simulating an 802.3 CSMA/CD LAN
M 90-27	August	D.A. Overdijk	Skew-symmetric matrices and the Euler equations of rotational motion for rigid systems
M 90-28	August	A.W.J. Kolen J.K. Lenstra	Combinatorics in Operations Research
M 90-29	August	R. Doornbos	Verdeling en onafhankelijkheid van kwadratensommen in de variantie-analyse
M 90-30	August	M.W.I. van Kraaij W.Z. Venema J. Wessels	Support for problem solving in manpower planning problems
M 90-31	August	I. Adan A. Dekkers	Mean value approximation for closed queueing networks with multi server stations
M 90-32	August	F.P.A. Coolen P.R. Mertens M.J. Newby	A Bayes-Competing Risk Model for the Use of Expert Judgment in Reliability Estimation
M 90-33	September	B. Veltman B.J. Lageweg J.K. Lenstra	Multiprocessor Scheduling with Communication Delays
M 90-34	September	I.J.B.F. Adan J. Wessels W.H.M. Zijm	Flexible assembly and shortest queue problems

Number	Month	Author	Title
M 90-35	September	F.P.A. Coolen M.J. Newby	A note on the use of the product of spacings in Bayesian inference