Biometrika (2002), **89,** 1, *pp.* 211–224

Printed in Great Britain

Empty confidence sets for epidemics, branching processes and Brownian motion

BY FRANK G. BALL

School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, U.K.

frank.ball@nottingham.ac.uk

TOM BRITTON

Department of Mathematics, Uppsala University, P.O. Box 480, SE-751 06 Uppsala, Sweden

tomb@math.uu.se

AND PHILIP D. O'NEILL

School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, U.K.

philip.o'neill@nottingham.ac.uk

SUMMARY

This paper treats some examples where likelihood-based inference for certain model parameters may produce empty confidence sets. The first example concerns epidemics, and the parameter of interest is the basic reproduction number R_0, which is to be estimated from the final size of an epidemic in a finite population. The second example treats estimation of the mean of the offspring distribution in a branching process, based on observing the total progeny, i.e. the total number of individuals ever born in the branching process. The final example considers estimation of the linear drift in a Brownian motion, based on observing the first hitting time of some horizontal barrier.

Some key words: Branching process; Brownian motion; Confidence set; Epidemic model; Likelihood-based inference.

1. INTRODUCTION

This paper is concerned with construction of confidence sets for some examples of probability models which display a threshold phenomenon. It considers models in which a one-dimensional parameter, $\theta \in \Theta \subset \mathbb{R}$, is of interest and inference is based on an observation x of a one-dimensional random variable X that is stochastically monotone in θ.

For definiteness, suppose that X is stochastically increasing in θ. Then, prima facie, a natural method for constructing a $100(1-\alpha)$ percent confidence set for θ is by inverting the family of tests of $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$ having acceptance regions $A_\alpha(\theta_0) = [L_\alpha(\theta_0), U_\alpha(\theta_0)]$, where

$$L_\alpha(\theta_0) = \sup\{x : \mathbb{P}_{\theta_0}(X \leqslant x) \leqslant \alpha/2\}, \quad U_\alpha(\theta_0) = \inf\{x : \mathbb{P}_{\theta_0}(X \geqslant x) \leqslant \alpha/2\},$$

where $\mathbb{P}_{\theta_0}$ denotes probability conditional on $\theta = \theta_0$. Note that this gives a confidence

interval, which is conservative if X is discrete. However, for reasons outlined in § 2, this method leads to unsatisfactory confidence intervals for the examples in this paper. Thus an alternative method of constructing confidence sets is considered.

The $100(1-\alpha)$ percent confidence set is constructed directly from the probability density or mass function $p(x;\theta)$ for the random variable X as follows. First, for each $\theta_0 \in \Theta$, the acceptance region $A_\alpha(\theta_0)$ for testing $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$ is obtained by setting

$$A_\alpha(\theta_0) = \{x : p(x;\theta_0) \geqslant c_\alpha(\theta_0)\},$$

where

$$c_\alpha(\theta_0) = \sup\{c : \mathbb{P}_{\theta_0}\{p(X;\theta_0) \geqslant c\} \geqslant 1-\alpha\}.$$

Thus $A_\alpha(\theta_0)$ is constructed by including in it the most likely values of X under θ_0 until its probability is at least $1-\alpha$. The confidence set $R_\alpha(x)$ for θ given the datum x is then obtained by inverting the above family of tests, so that

$$R_\alpha(x) = \{\theta : A_\alpha(\theta) \ni x\}. \tag{1·1}$$

Again, the confidence set is conservative if X is discrete.

For ease of exposition, the above two types of confidence set are referred to as equal-tailed and inverted-likelihood-based, respectively. Note that the inverted-likelihood-based confidence set (1·1) is not generally a likelihood-based confidence region, in the terminology of Cox & Hinkley (1974, p. 218), unless $c_\alpha(\theta)$ is independent of θ.

In this paper, inverted-likelihood-based confidence sets are considered for three different examples, which are outlined below and described in §§ 2, 3 and 4.

Example 1. The first example, which largely motivates our investigations, concerns estimation of the basic reproduction number $R_0 \in [0, \infty)$ in an epidemic model. The basic reproduction number, defined in § 2, is the average number of potentially infectious contacts an individual makes during its infectious period. The community initially contains n susceptible and a infectious individuals. Estimation of R_0 is based on observing T, the total number of initially susceptible individuals that are ultimately infected, so $T \in \{0, \ldots, n\}$.

Example 2. The parameter of interest is the expected number of offspring $\lambda \in [0, \infty)$ of a given individual in a Galton–Watson branching process. This is to be estimated by observing the total progeny T, that is the number of individuals ever born in the branching process. Note that T takes the value $+\infty$ with nonzero probability if $\lambda > 1$. Hence, $T \in \{0, 1, \ldots\} \cup \{+\infty\}$.

Example 3. The last example considers a Brownian motion with unknown drift parameter $\mu \in (-\infty, +\infty)$. Estimation of μ is based on observing T_z, the time at which the Brownian motion first reaches the known level $z > 0$. If $\mu < 0$ it is well known that this may never happen, i.e. that $T_z = +\infty$ has nonzero probability. The state space for T_z is hence $(0, \infty) \cup \{+\infty\}$.

The examples share the property that the confidence set for the parameter of interest is empty when the observed value is extremely unlikely under the model, regardless of the choice of model parameter. In practical terms it seems sensible to interpret an empty confidence set as an indication that the model is inappropriate for the data. However, for each of the examples, it is possible to observe values that are close to those giving empty confidence sets, but where now the confidence set is a very narrow interval. Such an interval suggests that knowledge of the parameter of interest is very certain, a rather

different conclusion from that drawn by obtaining an empty confidence set. Further discussion regarding empty confidence sets and their interpretation can be found in Cox & Hinkley (1974, pp. 224ff).

2. Epidemic example

2·1. *Model and distribution of size of epidemic*

Consider the following model for the spread of an epidemic among a closed population comprising initially a infectives and n susceptibles. The infectious periods of different infectives are independently and identically distributed according to a random variable T_I, having an arbitrary but specified distribution. Throughout its infectious period, a given infective makes contacts at the points of a Poisson process with rate $n\beta$. Successive contacts are with individuals chosen independently and uniformly from the n initial susceptibles. If a contacted individual is still susceptible then it becomes infected and it is immediately able to infect other susceptibles, otherwise nothing happens. The contact processes of different infectives are mutually independent. The epidemic ceases as soon as there is no infective present in the population.

When the initial number of susceptibles n is large, during the early stages of an epidemic all infectious contacts are very likely to be with susceptible individuals and consequently the process of infectives can be approximated by a branching process, in which a typical individual lives for a time T_I, during which it has offspring at the points of a Poisson process with rate $n\beta$; see for example Ball & Donnelly (1995). The mean number of contacts made by a given infective is $R_0 = n\beta E(T_I)$. Thus, by standard branching process theory, for large n, the epidemic has a nonzero probability of taking off only if $R_0 > 1$ (Whittle, 1955; Williams, 1971). The threshold parameter R_0 is usually called the basic reproduction number or ratio of the epidemic (Heesterbeek & Dietz, 1996).

Let T denote the total size of the epidemic, and let $P_k = \mathrm{pr}(T = k)$ $(k = 0, 1, \ldots, n)$. A triangular system of linear equations for P_k $(k = 0, 1, \ldots, n)$ is given in Ball (1986). However, this system of equations is numerically unstable, because of rounding errors, even for relatively small values of n, for example for $n = 50$ or $n = 100$. When the infectious period T_I follows a negative exponential distribution with mean γ^{-1}, the epidemic model reduces to the general stochastic epidemic (Bailey, 1975, Ch. 6), whose total size distribution can be determined by the following, numerically stable two-dimensional system of equations, derived from Bailey (1975, eqn (6.49)). Let $\rho = \gamma/\beta$,

$$E = \{(i, j) : i = 0, 1, \ldots, n,\ j = 1, 2, \ldots, n + a,\ 1 \leqslant i + j \leqslant n + a\}$$

and let $h_{i,j}$, $(i, j) \in E$, be determined by

$$h_{n,a} = (n + \rho)^{-1}, \tag{2·1a}$$

$$(i + 1)h_{i+1,j-1} - (i + \rho)h_{i,j} + \rho h_{i,j+1} = 0, \quad (i, j) \in E \backslash (n, a), \tag{2·1b}$$

where $h_{i,j} = 0$ if $(i, j) \notin E$. Then

$$P_k = \rho h_{n-k,1} \quad (k = 0, 1, \ldots, n). \tag{2·2}$$

Note that, for the general stochastic epidemic, $R_0 = n/\rho$ and $P_k = P_k(R_0)$, for $k = 0, 1, \ldots, n$.

2·2. *Equal-tailed confidence intervals*

In practice, the total size of an epidemic is often used as a basis for inference since, unlike the temporal course of an epidemic, it can usually be obtained accurately from

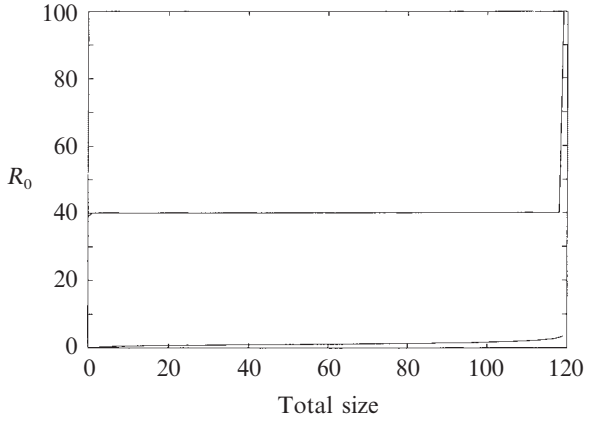

Fig. 1. Smallpox epidemic example: 95 percent equal-tailed confidence intervals for reproduction number, R_0.

data. Consider the problem of constructing a confidence set for R_0 from an observation, t_{obs} say, of the total size of a general stochastic epidemic. For this model, T is stochastically increasing in R_0, so equal-tailed confidence intervals can be constructed as described in § 1.

As an illustration, consider the case $(n, a) = (119, 1)$, motivated by a dataset obtained from a smallpox epidemic in Abakaliki, Nigeria (Bailey, 1975, p. 125). The 95 percent equal-tailed confidence interval for R_0 for different values of t_{obs} is shown in Fig. 1; for the smallpox dataset $t_{\text{obs}} = 29$. The intervals are clearly unsatisfactory. First, they are very wide, because the distribution of T is bimodal when $R_0 > 1$. Secondly, the upper end of the confidence interval is the same for most values of t_{obs}, which is a consequence of the threshold behaviour of the epidemic. The distribution of T for various values of R_0 is shown in Fig. 2. When $R_0 \leqslant 1$, the distribution of T is unimodal with the mode at zero. When $R_0 > 1$, the distribution of T still has a mode at zero but there is a second mode, corresponding to the epidemic taking off, at a value of T that tends to n as $R_0 \to \infty$; in fact, the second mode first appears when R_0 is slightly larger than one (Ball & Nåsell, 1994). It is clear from Fig. 2 that when $R_0' \geqslant 2$ the equal-tailed test of $H_0 : R_0 = R_0'$ versus $H_1 : R_0 \neq R_0'$ has an acceptance region which contains several extremely unlikely values of T, thus leading to very wide confidence intervals. For large n, the probability of a 'minor' epidemic, for the general stochastic epidemic, is approximately $\min\{1, R_0^{-a}\}$ (Whittle, 1955). Also, it is clear from Fig. 2 that, for $R_0 \geqslant 5$, the distribution of T is concentrated around 0 and n, so for such R_0 and $a = 1$, $\text{pr}(T = 0) \doteq R_0^{-1} \doteq 1 - \text{pr}(T = n)$. Thus the upper end of the conservative confidence interval is at $R_0 = (\alpha/2)^{-1}$, unless t_{obs} is close to 0 or n.

2·3. *Inverted-likelihood-based confidence sets*

In view of the above, it seems sensible to consider inverted-likelihood-based confidence sets for R_0. Write P_k as $P_k(R_0)$, for $k = 0, 1, \ldots, n$, and for fixed R_0 let $I(j, R_0)$ denote the suffix of the jth smallest $P_k(R_0)$, for $j = 1, 2, \ldots, n+1$. For $\alpha \in (0, 1)$, let $k^*(\alpha, R_0) = \max\{k : \sum_{j=1}^{k} P_{I(j,R_0)}(R_0) < \alpha\}$. The conservative $100(1-\alpha)$ percent confidence set for R_0 is obtained by inverting the family of tests of $H_0 : R_0 = R_0'$ versus $H_1 : R_0 \neq R_0'$ having acceptance regions

$$A_\alpha(R_0') = \{I(j, R_0) : j = k^*(\alpha, R_0') + 1, k^*(\alpha, R_0') + 2, \ldots, n+1\}.$$

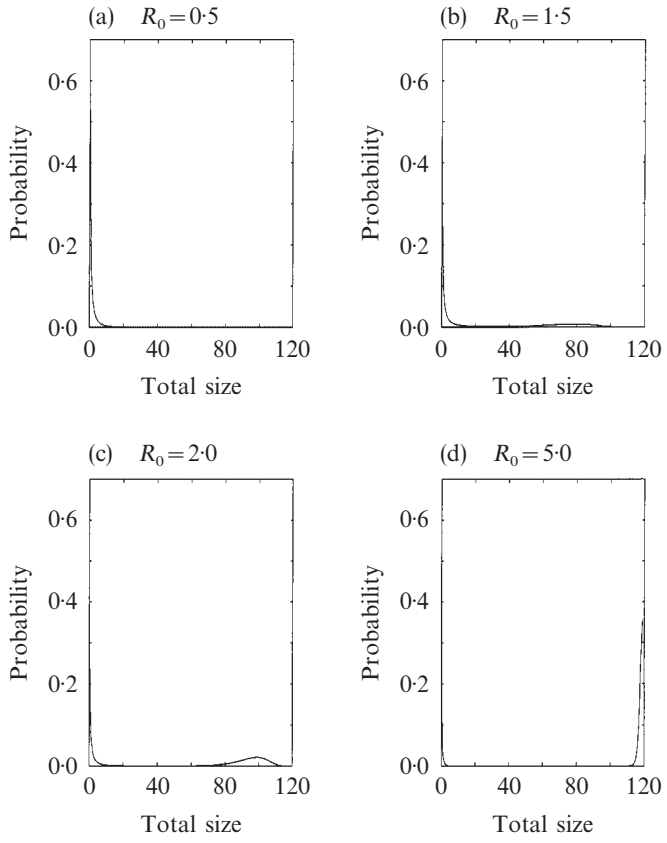

Fig. 2. Smallpox epidemic example: total size distribution for general stochastic epidemic with $(n, a) = (119, 1)$ and various R_0.

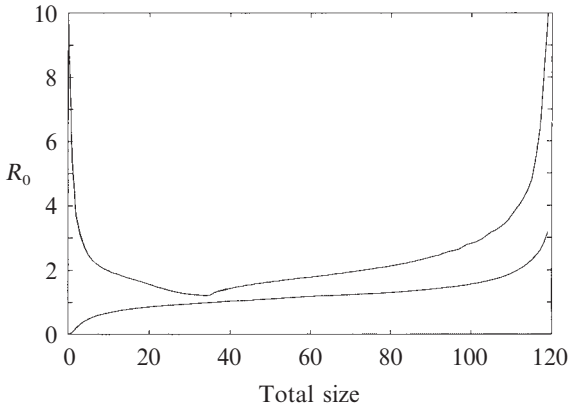

Fig. 3. Smallpox epidemic example: 95 percent inverted-likelihood-based confidence sets for reproduction number, R_0.

The 95 percent confidence sets for the case $(n, a) = (119, 1)$ are shown in Fig. 3. These were based on tests of $R_0 = R_0'$ using the grid of values $R_0' = 0{\cdot}0001j$, for $j = 1, 2, \ldots, 100\,000$. For most values of t_{obs}, the confidence set is a single interval, though for a few values it is the union of at most four intervals, with either the lowest or the

highest interval being appreciably longer than the others. Figure 3 shows the bottom end of the lowest interval, the top end of the highest interval and the longest interval, if the confidence set comprises two or more intervals. A finer grid of values for R_0' might result in the confidence set comprising more, but not fewer, intervals for some values of t_{obs}, but the overall picture is unlikely to be very different from that shown in Fig. 3. Also, the top ends of the confidence intervals when $t_{\text{obs}} = 0$ and 119 should be 20·0188 and ∞, respectively.

The confidence sets depicted in Fig. 3 are clearly more sensible than the intervals shown in Fig. 1. The sets are still wide for very small and very large values of t_{obs} but this reflects inherent features of the model. Small epidemics are likely to occur for a wide range of values for R_0, and very large epidemics occur with high probability for all sufficiently large R_0.

The results of the next section suggest that the inverted-likelihood-based confidence set is empty for some values of t_{obs} if the initial susceptible population size n is sufficiently large. The smallest value of n for which this phenomenon arises clearly decreases in α. For $\alpha = 0{\cdot}05$, numerical calculation shows that empty confidence sets do not occur when $(n, a) = (1000, 1)$ so, given the computational effort required to compute the total size distribution for large n, the possible existence of empty confidence sets was investigated by keeping $(n, a) = (119, 1)$ and increasing α. If we use the grid of values $R_0' = 0{\cdot}0001j$, for $j = 1, 2, \ldots, 50\,000$, when $\alpha = 0{\cdot}1$, $t_{\text{obs}} = 42$ gives rise to an empty confidence set whilst all other values for t_{obs} do not; $t_{\text{obs}} = 43$ yields a confidence set comprising 13 intervals! It is possible that the empty confidence set when $t_{\text{obs}} = 42$ would disappear if a finer grid for R_0' was used. However, if we use the same grid for R_0', when $\alpha = 0{\cdot}2$, an observed total size of $22 \leqslant t_{\text{obs}} \leqslant 57$ gives rise to an empty confidence set. Note from Fig. 2 that values of t_{obs} in the interval $[22, 57]$ have low probabilities for all R_0. It is difficult to explain analytically the fact that the confidence set can comprise so many disjoint intervals, though it seems likely to be because $P_k(R_0)$, for $k = 0, 1, \ldots, n$, are complicated rational functions of R_0.

It is interesting to note that $t_{\text{obs}} = 29$, the observed value for the smallpox dataset, lies in the interval for which the confidence set is empty when $\alpha = 0{\cdot}2$, casting doubt on the correctness of the general stochastic epidemic for these data. Indeed, closer scrutiny of the smallpox epidemic reveals that to be the case. As indicated in Bailey & Thomas (1971), the population was partitioned into compounds, making the assumption of homogeneous mixing questionable, and some of the susceptibles appeared actually to have been vaccinated.

3. Branching process example

3·1. *Assumptions and confidence sets*

Consider a Galton–Watson branching process with a initial ancestors, in which the number of offspring of a single individual is distributed according to a random variable Z_λ, where λ is a parameter that governs the distribution of Z. We shall take λ to be the mean number of offspring, so that $\lambda = E(Z_\lambda)$. For $k = 0, 1, \ldots$, define $p_k(\lambda) = \text{pr}(Z_\lambda = k)$. Let T_λ denote the total progeny of the branching process, not including the initial ancestors, and for $k = 0, 1, \ldots$ define $\tilde{p}_k(\lambda) = \text{pr}(T_\lambda = k)$. Then

$$\tilde{p}_k(\lambda) = \frac{a}{a+k} \text{pr}\{Z(1) + Z(2) + \ldots + Z(a+k) = k\} \quad (k = 0, 1, \ldots), \tag{3·1}$$

where $Z(1), Z(2), \ldots$ are independent and identically distributed copies of Z_λ (Jagers, 1975, Theorem 2.11.2). Note that (3·1) provides an explicit expression for the total progeny mass function, provided that the probability on the right-hand side can be evaluated.

Suppose now that the following three conditions hold:

(i) for fixed $\lambda > 0$, $\tilde{p}_k(\lambda) \geqslant \tilde{p}_{k+1}(\lambda)$ $(k = 0, 1, \ldots)$;

(ii) $T_\lambda \geqslant_{\mathrm{st}} T_{\lambda'}$ for any $\lambda > \lambda'$, where $\geqslant_{\mathrm{st}}$ denotes stochastic ordering;

(iii) for $k = 0, 1, \ldots, \tilde{p}_k(\lambda) < \tilde{p}_k(\lambda')$ for $\lambda > \lambda' > 1$.

A sufficient condition for (ii) is that $Z_\lambda \geqslant_{\mathrm{st}} Z_{\lambda'}$ for $\lambda > \lambda'$, and this condition is easily checked in practice. Consider a size-α test of $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda \neq \lambda_0$ constructed using the most likely values of T_λ in the manner described in § 1. Then conditions (i) and (iii) imply that the acceptance region $A_\alpha(\lambda_0)$ of the test takes the form

$$A_\alpha(\lambda_0) = \begin{cases} [0, b(\lambda_0, \alpha)], & \text{if } \lambda_0 < \lambda^*(\alpha), \\ [0, b(\lambda_0, \alpha)] \cup \{\infty\}, & \text{if } \lambda_0 \geqslant \lambda^*(\alpha), \end{cases}$$

for some $\lambda^*(\alpha)$. Note that $\lambda^*(\alpha) > 1$, since $T_\lambda < \infty$ almost surely if $\lambda \leqslant 1$.

Condition (ii) ensures that $b(\lambda, \alpha) \leqslant b(\lambda', \alpha)$ if $\lambda < \lambda' < \lambda^*(\alpha)$. Similarly, condition (iii) ensures that $b(\lambda, \alpha) \leqslant b(\lambda', \alpha)$ if $\lambda > \lambda' > \lambda^*(\alpha)$. Consequently, $b(\lambda, \alpha) \leqslant b(\lambda^*(\alpha), \alpha)$ for all $\lambda \geqslant 0$ and, if the value $T_\lambda = t$ is observed, the inverted-likelihood-based confidence set for λ is given by

$$R_\alpha(t) = \begin{cases} [\min\{\lambda : b(\lambda, \alpha) \geqslant t\}, \max\{\lambda : b(\lambda, \alpha) \leqslant t\}], & \text{if } t \leqslant b(\lambda^*(\alpha), \alpha), \\ \varnothing, & \text{if } t \in (b(\lambda^*(\alpha), \alpha), \infty), \\ [\lambda^*(\alpha), \infty), & \text{if } t = \infty. \end{cases}$$

Note that the probability of observing an empty confidence set is

$$\mathrm{pr}\{b(\lambda^*(\alpha), \alpha) < T_\lambda < \infty\},$$

which by conditions (ii) and (iii) is maximised when $\lambda = 1$. Since $\lambda^*(\alpha) > 1$, this maximised probability is slightly less than α.

3·2. *Examples*

We now consider two examples of different offspring distributions. Both examples satisfy conditions (i)–(iii) implying that confidence sets for λ may be empty. The first example, namely negative binomial, corresponds to the limiting branching process for an epidemic model where T_I follows a gamma distribution with an integral-valued shape parameter, of which the general stochastic epidemic is a special case. The second example, with a Poisson offspring distribution, corresponds similarly to a Reed–Frost epidemic model, in which T_I is constant (Ball, 1986).

Example 1: *Negative binomial offspring distribution.* Suppose that Z_λ has a negative binomial distribution with probability mass function

$$p_k(\lambda) = \left(\frac{\lambda}{r+\lambda}\right)^k \left(\frac{r}{r+\lambda}\right)^r \binom{k+r-1}{r-1} \quad (k = 0, 1, \ldots).$$

By (3·1),

$$\tilde{p}_k(\lambda) = \frac{a}{a+k} \left(\frac{\lambda}{r+\lambda}\right)^k \left(\frac{r}{r+\lambda}\right)^{r(a+k)} \binom{k+r(a+k)-1}{r(a+k)-1} \quad (k = 0, 1, \ldots).$$

It is straightforward to verify condition (ii), while, for fixed k,

$$dp_k(\lambda)/d\lambda < 0 \tag{3·2a}$$

if and only if

$$\lambda > k/(a+k), \tag{3·2b}$$

so that condition (iii) is satisfied. For condition (i), it can be shown that

$$r_k(\lambda) = \frac{\tilde{p}_{k+1}(\lambda)}{\tilde{p}_k(\lambda)}$$
$$= \left\{\frac{r(a+k+1)+k}{k+1}\right\}\left(\frac{a+k}{a+k+1}\right)\left\{\frac{\lambda r^r}{(r+\lambda)^{r+1}}\right\}\prod_{j=1}^{r}\frac{r(a+k)-1+k+j}{r(a+k)-1+j}.$$

In the case $a=1$, we find that

$$\prod_{j=1}^{r}\frac{r(1+k)-1+k+j}{r(1+k)-1+j} \leqslant \left\{\frac{r(k+1)+k}{r(k+1)}\right\}^r,$$

while it is straightforward to show that, for all $\lambda > 0$, $g(\lambda) = \lambda/(r+\lambda)^{r+1}$ is maximised at $\lambda = 1$, so that

$$\frac{\lambda}{(r+\lambda)^{r+1}} \leqslant g(1) = \frac{1}{(1+r)^{r+1}}.$$

We thus obtain that

$$r_k(\lambda) \leqslant \frac{r^r}{(1+r)^{r+1}}\left\{\frac{r(k+2)+k}{k+2}\right\}\left\{\frac{r(k+1)+k}{r(k+1)}\right\}^r$$
$$= \left\{\frac{r(k+2)+k}{(k+2)(r+1)}\right\}\left\{\frac{r(k+1)+k}{(r+1)(k+1)}\right\} \leqslant 1,$$

so that condition (i) is satisfied in this case.

For numerical illustration, suppose first that $r=1$, so that Z_λ is geometric, $\alpha = 0{\cdot}05$ and $a=1$. We find that an observed final progeny $T_\lambda \in [128, \infty)$ give rise to an empty confidence set. For $\alpha = 0{\cdot}01$ this becomes $T_\lambda \in [3183, \infty)$. The probability of observing an empty confidence set when $\alpha = 0{\cdot}05$ is maximised when $\lambda = 1$ by $\mathrm{pr}\{T_1 \in [128, \infty)\} = 0{\cdot}0498$. In Fig. 4 the upper and lower limits of the 95 percent inverted-likelihood-based confidence interval for λ are plotted as a function of the observed total progeny. As seen from Fig. 4 the confidence set is empty for $T_\lambda \in [128, \infty)$. For $T_\lambda = \infty$, not shown in Fig. 4, the confidence set is not empty; rather, the confidence set is $\lambda \in (1{\cdot}001, \infty)$.

As a final illustration, if $r=4$, $\alpha = 0{\cdot}05$ and $a=1$, then any observation $T_\lambda \in [204, \infty)$ will give an empty confidence set.

Example 2: *Poisson offspring distribution.* If $Z_\lambda \sim \mathrm{Po}(\lambda)$ then

$$p_k(\lambda) = \frac{\lambda^k e^{-k}}{k!} \quad (k = 0, 1, \ldots),$$

and, by (3·1),

$$\tilde{p}_k(\lambda) = \frac{a}{a+k}\,\frac{\{\lambda(a+k)\}^k}{k!}\,e^{-\lambda(a+k)} \quad (k = 0, 1, \ldots).$$

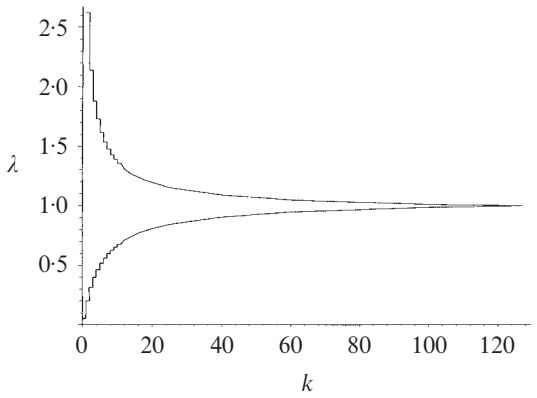

Fig. 4. Example 1: 95 percent inverted-likelihood-based confidence sets for λ. For an observed total progeny $k < 128$, the confidence set for λ is the interval between the two curves. For finite $k \geqslant 128$ the confidence set is empty; for $k = 0$ the interval goes from 0 up to $\lambda = 20$, this upper limit being omitted.

It is straightforward to verify condition (ii), and condition (iii) can be verified in the same way as in the previous example; once again we obtain the relationship given at (3·2). For condition (i), proceeding as before we obtain that

$$r_k(\lambda) = \frac{\tilde{p}_{k+1}(\lambda)}{\tilde{p}_k(\lambda)} = \frac{(a+k+1)^k}{(a+k)^{k-1}(k+1)}\,\lambda e^{-\lambda},$$

so condition (i) will not hold if a is sufficiently large. For the case $a = 1$ we have

$$r_k(\lambda) = \left(\frac{k+2}{k+1}\right)^k \lambda e^{-\lambda} = \frac{k+1}{k+2}\left(1 + \frac{1}{k+1}\right)^{k+1} \lambda e^{-\lambda} < \frac{k+1}{k+2}\, e\, \lambda e^{-\lambda} < 1,$$

since $\lambda e^{-\lambda}$ is maximised at $\lambda = 1$. Since all three conditions are satisfied for $a = 1$, we conclude that inverted-likelihood-based confidence sets may be empty. For example, if $\alpha = 0{\cdot}05$ we find that $T_\lambda \in [255, \infty)$ yields an empty confidence set for λ.

3·3. *Additional remarks*

Remark 1: *Censored observations.* In practice, $T_\lambda = \infty$ will never be observed. Suppose instead that observations are censored at some finite positive integer m, so that possible observations lie in the set $\{0, 1, \ldots, m\}$, with m corresponding to $T_\lambda \geqslant m$. Then empty confidence sets are still possible, as we now illustrate.

In similar notation to that used in § 3·1, the acceptance region $A_\alpha^m(\lambda_0)$ takes the form

$$A_\alpha^m(\lambda_0) = \begin{cases} [0, b_m(\lambda_0, \alpha)], & \text{if } \lambda_0 \leqslant \lambda_m^*(\alpha), \\ [0, b_m(\lambda_0, \alpha)] \cup \{m\}, & \text{if } \lambda_0 > \lambda_m^*(\alpha). \end{cases}$$

Let

$$b_m^*(\alpha) = \max\{b_m(\lambda, \alpha) : \lambda > 0\}.$$

Then, if $m > b_m^*(\alpha) + 1$, any observation in $[b_m^*(\alpha) + 1, m - 1]$ yields an empty confidence set. Note that, since $\mathrm{pr}(T_\lambda \geqslant m) > \mathrm{pr}(T_\lambda = \infty)$, $\lambda_m^*(\alpha) \leqslant \lambda^*(\alpha)$, $b_m(\lambda, \alpha) = b(\lambda, \alpha)$ if $\lambda \leqslant \lambda_m^*(\alpha)$,

and $b_m(\lambda, \alpha) \leqslant b(\lambda, \alpha)$ if $\lambda > \lambda_m^*(\alpha)$. Thus $b_m^*(\alpha) \leqslant b(\lambda^*(\alpha), \alpha)$. Finally, if $\lambda_m^*(\alpha) \geqslant 1$, which does not necessarily hold, unlike in the uncensored case, then condition (iii) ensures that $b_m^*(\alpha) = b_m(\lambda_m^*(\alpha), \alpha)$.

Remark 2: *Initial number of ancestors.* It is clear that empty confidence sets are still possible for $a > 1$, although it is less straightforward to work with the distribution of T_λ in this case; for example, condition (i) will not be true in general.

4. BROWNIAN MOTION EXAMPLE

4·1. *The distribution of the hitting time for a Brownian motion*

Let $\{B_t: t \geqslant 0\}$ be a Brownian motion with $B_0 = 0$ almost surely, linear drift μ and infinitesimal variance σ^2. Define the hitting time $T_z = \inf\{t \geqslant 0; B_t = z\}$, that is the first time the process reaches z, where $z > 0$ is assumed without loss of generality. Below, μ is treated as a parameter whereas σ and z are assumed to be fixed and known.

The hitting time T_z has distribution function given by

$$F(t; \mu) = e^{2z\mu}\Phi\left(\frac{-\mu t - z}{\sigma\sqrt{t}}\right) + \Phi\left(\frac{\mu t - z}{\sigma\sqrt{t}}\right) \quad (t \geqslant 0), \tag{4·1}$$

where $\Phi(.)$ is the standard normal distribution function (Asmussen, 1987, p. 263).

If $\mu \geqslant 0$ then $\mathbb{P}_\mu(T_z < \infty) = 1$, where $\mathbb{P}_\mu$ denotes probability given that the drift equals μ, and T_z has an inverse Gaussian distribution with parameters z/μ and $(z/\sigma)^2$. Thus, the density function of T_z is

$$f(t; \mu) = \frac{z}{\sigma t^{3/2}\sqrt{(2\pi)}}\, e^{-(\mu t - z)^2/2t\sigma^2} \quad (t > 0). \tag{4·2}$$

If $\mu < 0$, the distribution of T_z also has point mass at $T_z = \infty$ corresponding to the process never hitting the barrier z: $\mathbb{P}_\mu(T_z = \infty) = 1 - e^{2\mu z/\sigma^2}$. The density for a negative drift $-|\mu|$ is related to the corresponding density for positive drift $|\mu|$ by

$$f(t; -|\mu|) = e^{-2|\mu|z/\sigma^2} f(t; |\mu|) \quad (t > 0).$$

It follows that, conditional on T_z being finite, the two distributions are identical (Asmussen, 1987, p. 265).

4·2. *Constructing confidence sets for* μ

We now derive confidence sets for μ assuming σ and z known, where inference is based on one observation of T_z. The distribution of T_z is stochastically decreasing in μ, as is easily proved by a coupling argument, using the well-known fact that if $\{B_t\}$ is a Brownian motion without drift then $\{B_t + \mu t\}$ is a Brownian motion with drift μ. Thus it is straightforward to construct an equal-tailed confidence interval for μ, using the method described in § 1. However, the inverted-likelihood-based confidence set is rather more complicated and it is necessary to examine the density $f(t; \mu)$ more closely.

First consider the case $\mu \geqslant 0$. The density $f(t; \mu)$ is unimodal and its maximum is obtained at the point

$$t_{\max} = \sqrt{\left\{\left(\frac{3\sigma^2}{2\mu^2}\right)^2 + \frac{z^2}{\mu^2}\right\}} - \frac{3\sigma^2}{2\mu^2} \quad (\mu > 0), \quad t_{\max} = \frac{z^2}{3\sigma^2} \quad (\mu = 0).$$

Consequently, the inverted-likelihood-based acceptance region $A_\alpha(\mu_0)$ for a test of $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$, where $\mu_0 \geqslant 0$, consists of the interval $I^{(\alpha)}_{\mu_0} = [l^{(\alpha)}(\mu_0), u^{(\alpha)}(\mu_0)]$, where $l^{(\alpha)}$ and $u^{(\alpha)}$ depend on the chosen confidence level $1-\alpha$. The lower and upper limits always satisfy $l^{(\alpha)}(\mu_0) < t_{\max}$ and $u^{(\alpha)}(\mu_0) > t_{\max}$. Since the density is continuous in both t and μ, it follows that $l^{(\alpha)}(.)$ and $u^{(\alpha)}(.)$ are continuous and $u^{(\alpha)}(0)$ is some finite value. As μ becomes large the distribution of T_z becomes concentrated on small values, implying that $u^{(\alpha)}(\mu) \to 0$ as $\mu \to \infty$. Hence $u^{(\alpha)}(.)$ is bounded, by $u_{\max} = u_{\max}(\alpha, z, \sigma)$ say, on $[0, \infty)$.

Now consider the case $\mu < 0$. To emphasise μ being negative, write $\mu = -|\mu|$. As mentioned above, the density is proportional to the density having drift parameter $|\mu|$. This implies that the order of inclusion of t's in the acceptance region is the same as for the positive case, only now the point $\{\infty\}$ is the first outcome to be included in the acceptance region $A_\alpha(-|\mu_0|)$ for the hypothesis $\mu = -|\mu_0|$, since it has positive point mass, as opposed to all other values in the state space. Comparing the acceptance region for $-|\mu_0|$ with that of $|\mu_0|$ we note that

$$\begin{aligned}\mathbb{P}_{-|\mu_0|}(T_z \in I^{(\alpha)}_{|\mu_0|} \cup \{+\infty\}) &= e^{-2|\mu_0|z/\sigma^2}\mathbb{P}_{|\mu_0|}(T_z \in I^{(\alpha)}_{|\mu_0|}) + (1 - e^{-2|\mu_0|z/\sigma^2}) \\ &= e^{-2|\mu_0|z/\sigma^2}(1-\alpha) + (1 - e^{-2|\mu_0|z/\sigma^2}) \\ &> 1 - \alpha,\end{aligned}$$

which implies that $I^{(\alpha)}_{|\mu_0|}$ is too wide. Thus, the acceptance region $A_\alpha(-|\mu_0|)$ consists of an interval $I^{(\alpha)}_{-|\mu_0|} \subset I^{(\alpha)}_{|\mu_0|}$ together with $\{\infty\}$. In particular, the upper limit of the interval satisfies $u^{(\alpha)}(-|\mu_0|) \leqslant u^{(\alpha)}(|\mu_0|)$, so that the upper limit is bounded by the same $u_{\max} = u_{\max}(\alpha, z, \sigma)$ as for positive μ. Consequently, no parameter value μ, positive or negative, has t's satisfying $u_{\max} < t < \infty$ in the acceptance region $A_\alpha(\mu)$.

Now suppose that the value $T_z = t$ is observed. Then the $100(1-\alpha)$ percent inverted-likelihood-based confidence set for μ is given by

$$R_\alpha(t) = \{\mu\colon A_\alpha(\mu) \ni t\}. \tag{4·3}$$

If $t \in (u_{\max}, \infty)$ then the confidence set is empty, that is $R_\alpha(t) = \varnothing$. The probability of obtaining an empty confidence set depends on the actual parameter value. The probability is largest in the case $\mu = 0$, when there is probability α of obtaining an empty $100(1-\alpha)$ percent confidence set.

As a numerical illustration, suppose that $\sigma = 1$ and that the barrier $z = 1$. Then, for $\alpha = 0{\cdot}05$, an observed finite hitting time T_1 greater than about 250 yields an empty confidence set for μ. In Fig. 5 the 95 percent inverted-likelihood-based confidence set for μ is plotted as a function of the observed hitting time t_1. Note that Fig. 5 is plotted using the scaling $\mu \to \mu^{1/4}$ and $t_1 \to t_1^{1/4}$. This scaling was chosen to ensure that certain salient features of the confidence sets are clearly visible. For observations of t_1 less than about 0·04, note that $0{\cdot}04^{1/4} \simeq 0{\cdot}43$, there is a single interval containing large positive values of μ; the confidence intervals for very small values of t_1 are not shown in Fig. 5 as the corresponding μ values are very large. An intuitive interpretation here is that a short hitting time is only likely when there is a large positive drift. For t_1 between about 0·04 and 0·07, note that $0{\cdot}07^{1/4} \simeq 0{\cdot}51$, the confidence sets consist of two intervals, one of which includes μ values around zero, and the other of which contains large positive μ values. It is harder to explain these disjoint intervals intuitively. They arise because of the way that the density of T_1 changes as μ increases; the right tail of the density becomes thinner as μ becomes positive and increases, with the effect that the width of the acceptance region becomes shorter.

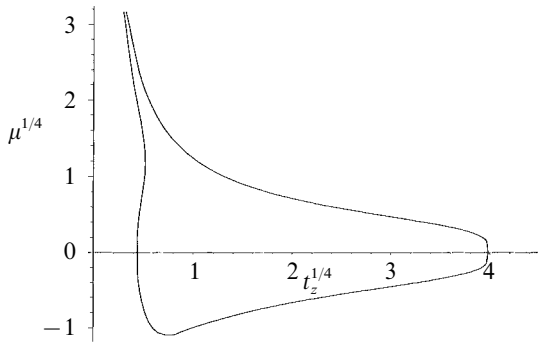

Fig. 5. Brownian motion example: 95 percent inverted-likelihood-based confidence sets for the drift parameter μ of a Brownian motion with $\sigma = 1$, based on observing the hitting time $T_z = t_z$ for the barrier $z = 1$. Note that both axes are scaled to ensure that certain interesting features of the confidence sets are clearly visible.

For t_1 between about 0·07 and 250, note that $250^{1/4} \doteq 3{\cdot}98$, a single interval is obtained, including both positive and negative values. The confidence set is empty for observations of t_1 larger than about 250. Finally, for $t_1 = \infty$, not shown in the figure, the confidence set is not empty; on the contrary it is $\mu \in (-\infty, 0)$.

4·3. *Additional remarks*

Remark 3: *Censored observations.* In practice it is of course never possible to observe $T_z = +\infty$. A more realistic scenario would be to observe $T_z = t$ if $t \leqslant \tau$, or else $T_z > \tau$ is observed, where τ is some pre-defined time horizon to which the observer is willing to wait. This is similar to the case of censored observations in the branching process example, studied in § 3·3, and, provided $\tau > u_{\max}$, there exists $u_{\max}(\tau) < u_{\max}$ such that observations $t \in (u_{\max}(\tau), \tau)$ give rise to an empty confidence set.

Remark 4: *Related models.* Brownian motion can be thought of as the limit of a sequence of random walks, so inverted-likelihood-based confidence sets derived from observing the first time that a random walk reaches a horizontal barrier may also be empty. Also, certain birth-and-death processes can give empty confidence sets if the process is started above the horizontal barrier.

5. DISCUSSION

We have given several examples where it is possible to obtain empty confidence sets for a parameter. The examples share the probabilistic property that the random variable, on which the analysis is based, is stochastically monotone in the parameter. Furthermore, in all of the examples the parameter has a threshold value at which a phase transition occurs in that the probability distribution becomes bimodal as the parameter passes the threshold value. As a consequence, the confidence sets are empty for observed values that have low probability regardless of the true underlying parameter.

The examples all have a 'symmetry' property concerned with conditioning upon a certain event. In the Brownian motion example, if $\mu < 0$, then the distribution of T_z, conditioned upon the barrier being hit in finite time, coincides with that of Brownian

motion with positive drift $|\mu|$. For the branching process example, a supercritical branching process conditioned on extinction is probabilistically indistinguishable from a subcritical branching process with a suitable offspring distribution (Athreya & Ney, 1972, Theorem I.12.3). As has already been stated, for large population sizes the initial stages of an epidemic can be approximated by a branching process; in particular, a supercritical epidemic that dies out quickly behaves similarly to a subcritical epidemic.

The examples share several other features. For observations resulting in empty confidence sets the maximum likelihood estimate lies close to the threshold value in all examples. Furthermore, the estimated value of Fisher's information, obtained by using the maximum likelihood estimate, is very large for observations with empty confidence sets. In spite of this, it seems sensible to interpret an empty confidence set by saying that the model is inappropriate for the data for any choice of parameter value. Also, as mentioned in the introduction, there are observations, close to those having empty confidence sets, that give a narrow confidence interval and have high observed information. Such narrow confidence intervals are usually interpreted by saying that the parameter can be estimated accurately. This conclusion is drastically different from nearby observations, with empty confidence sets, where the result suggests that the model is inappropriate. An alternative conclusion, given an observation close to one which would yield an empty confidence set, could therefore be to question the model. This raises an important general question: how should one relate the possibility for the model to be incorrect with the possibility that the model is correct and the observation is extreme, for all parameter choices?

Finally, a Bayesian analysis of the three examples would proceed as follows. For a given prior density $\pi(\theta)$ of the parameter of interest, the posterior density given a datum $X = x$ is proportional to $\pi(\theta)L_x(\theta)$, where $L_x(\theta)$ denotes the likelihood. Suppose now that x is such that an empty confidence set is obtained. If $\pi(\theta)$ is sufficiently diffuse, then the posterior density will be sharply peaked around its modal value, which in turn is close to the threshold value. Consequently, an equal-tailed credible interval would typically consist of a short interval containing the modal value, suggesting that there is little uncertainty about the value of θ. This is clearly a very different conclusion from that inferred from an empty confidence set.

Acknowledgement

The confidence sets discussed in § 2·3 arose out of discussions of Frank Ball with Peter Donnelly and Denis Mollison. This research was supported by the U.K. Engineering and Physical Sciences Research Council, and it was carried out in part at the Research Kitchen on Statistical Inference using Stochastic Epidemic Models, held in July 2000 at Cowan Bridge, England, and funded by the European Science Foundation initiative on Highly Structured Stochastic Systems. Tom Britton also acknowledges financial support from The Swedish Natural Science Research Council.

References

Asmussen, S. (1987). *Applied Probability and Queues*. Chichester: Wiley.

Athreya, K. B. & Ney, P. E. (1972). *Branching Processes*. New York: Springer Verlag.

Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd ed. London: Griffin.

Bailey, N. T. J. & Thomas, A. S. (1971). The estimation of parameters from population data on the general stochastic epidemic. *Theor. Pop. Biol.* **2**, 253–70.

BALL, F. G. (1986). A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. *Adv. Appl. Prob.* **18**, 289–310.

BALL, F. G. & DONNELLY, P. J. (1995). Strong approximations for epidemic models. *Stoch. Proces. Appl.* **55**, 1–21.

BALL, F. G. & NÅSELL, I. (1994). The shape of the size distribution of an epidemic in a finite population. *Math. Biosci.* **123**, 167–81.

COX D. R. & HINKLEY D. V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

HEESTERBEEK, J. A. P. & DIETZ, K. (1996). The concept of R_0 in epidemic theory. *Statist. Neer.* **50**, 89–110.

JAGERS, P. (1975). *Branching Processes with Biological Applications.* London: Wiley.

WHITTLE, P. (1955). The outcome of a stochastic epidemic — a note on Bailey's paper. *Biometrika* **42**, 116–22.

WILLIAMS, T. (1971). An algebraic proof of the threshold theorem for the general stochastic epidemic (abstract). *Adv. Appl. Prob.* **3**, 223.

[*Received April* 2001. *Revised August* 2001]