Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 115 (2003) 479 – 490
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Credible and confidence sets for restricted parameter spaces☆

Tonglin Zhang *, Michael Woodroofe

Department of Statistics, University of Michigan, 4062 Frieze Building, 105 S. State street, 48109-1285 Ann Arbor, MI, USA

Received 9 February 2001; accepted 6 March 2002

Abstract

Recent experiments in high-energy physics have sparked interest in problems where a parameter is restricted to a portion of its natural range—for example, a positive normal mean or a Poisson mean that is known to exceed some background level. A recent article has shown that the Bayesian credible intervals that arise when the parameter is given a uniform distribution over the restricted range have good frequentist coverage probabilities. Here, the latter result is examined for robustness when nuisance parameters are included in the model. It is shown that the frequentist coverage probabilities are still high, but there is a mild surprise in the nature of the intervals.

Keywords: Frequentist coverage probability; Prior and posterior distributions; Signals and noise; Total positivity; Unimodality

1. Introduction

Particle physicists often encounter problems in which measurement error produces (unbiased) estimates that are not in the physically allowable region. A particular instance was reported by Zeitnitz et al. (1998). In their model, an observed count was composed of a signal plus a background, say $X = B + S$, where $B \sim \text{Poisson}(b)$ and $S \sim \text{Poisson}(\theta)$ were independent, $b=2.88\pm0.1$ was (approximately) known, but θ was not. The unbiased estimator $X-b$ for this model can be negative and was Zeitnitz et al. who reported $X = 0$. To address problems of this nature, Feldman and Cousins (1998) proposed constructing confidence sets by taking regions of high likelihood in the space

☆ Research supported by the National Security Agency under MDA 904-99-1-0004 and U.S. Army under DAAG 55-98-1-0482.

* Corresponding author. Tel.: +1-734-764-3519; fax: +1-734-763-4676.
E-mail address: tlzhang@umich.edu (T. Zhang).

0378-3758/02/$ - see front matter
PII: S0378-3758(02)00170-2

of physically allowable parameter values. Their method is called the unified method because it automatically transforms from an upper confidence bound to a two-sided confidence interval. Roe and Woodroofe (1999) criticized the unified method in the Poisson case because the resulting intervals can depend on b when $X = 0$. In particular, they questioned the low upper confidence bound in Zeitnitz et al. that was derived from the unified method. Recently, Roe and Woodroofe (2000) have shown that if θ is given a prior uniform distribution over $[0, \infty)$, then the resulting Bayesian credible regions also have high frequentist coverage probabilities in the Poisson problem and a related normal problem.

Roe and Woodroofe considered two models, the Poisson case described above and a normal case in which one observes $X \sim \text{Normal}[\theta, \sigma^2]$, where $\theta \geqslant 0$ is unknown, but σ^2 is known. The latter model might be appropriate if θ represented a mass which is of the same order of magnitude as the measurement error. The derivations in Roe and Woodroofe (2000) require the nuisance parameter to be known, and this is seldom literally true. The purpose of this article is to provide derivations for the Poisson and normal problems when the nuisance parameters b and σ^2 are unknown and must be estimated. We find that Roe and Woodroofe's conclusions are quite robust with respect to the nuisance parameters; that is, if θ is given a uniform prior and the nuisance parameter a non-informative one, then the Bayesian credible intervals have high frequentist coverage probability. The normal case is considered in Section 2, and the Poisson case in Section 3. The normal model is more tractable, and the statistical issues less obscured by complicated formulas.

Two related recent papers have considered admissibility questions in the two models. Shao and Strawderman (1996) obtained improvements over the MLE in the normal case, and Woodroofe and Wang (2000) showed that uniformly most powerful tests may lead to inadmissible p-values when the parameter space is restricted in the Poisson case.

2. The normal case

Bayesian credible intervals. In the normal case, suppose that independent random variables $X \sim \text{Normal}[\theta, \sigma^2]$ and $W \sim \sigma^2 \chi_r^2$ are observed, where $\theta \geqslant 0$ and $\sigma > 0$ are unknown parameters and r is a positive integer. Thus, the joint density of W and X is

$$f_{\sigma,\theta}(w,x) = \frac{1}{2^{(r+1)/2}\sigma^{r+1}\sqrt{\pi}\Gamma(r/2)}\, w^{r/2-1} \exp\left[-\frac{(x-\theta)^2 + w}{2\sigma^2}\right].$$

Let $S^2 = W/r$, the unbiased estimator of σ^2; and let H_r be the t distribution function with r degrees of freedom. If σ and θ are given the improper prior $\rho(\sigma, \theta) = 1/\sigma$ for $0 < \sigma < \infty$ and $0 \leqslant \theta < \infty$, then the marginal density of W and X is

$$f(w,x) = \int_0^\infty \int_0^\infty f_{\sigma,\theta}(w,x)\, \frac{\mathrm{d}\sigma\, \mathrm{d}\theta}{\sigma} = \frac{1}{2w} H_r(t),$$

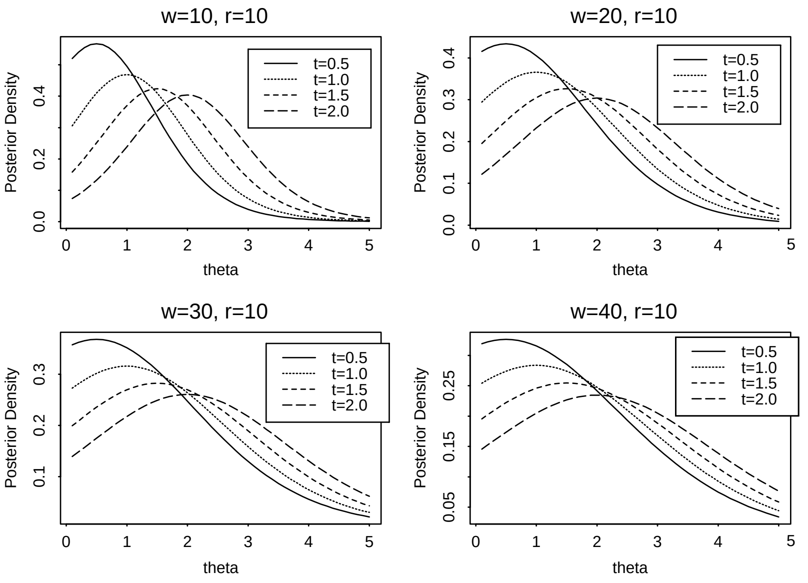

Fig. 1. Graphs of $g(\theta|w,x)$.

where $t=x/s$ and $s^2=w/r$, after some routine calculations. The posterior density of σ and θ given $W=w$ and $X=x$ is $f_{\sigma,\theta}(w,x)/\sigma f(w,x)$, and the (marginal) posterior density of θ is

$$g(\theta|w,x)=\int_0^\infty \frac{f_{\sigma,\theta}(w,x)}{f(w,x)}\frac{\mathrm{d}\sigma}{\sigma}=\frac{1}{sH_r(t)}h_r\left(\frac{\theta-x}{s}\right),$$

where h_r is the density of H_r. Equivalently, the posterior distribution of $(\theta-x)/s$ is that of a t_r random variable, truncated below at $-t$. Fig. 1 contains graphs of the posterior density functions of θ for selected x, r and w.

A level $1-\alpha$ Bayesian credible interval for θ requires two non-negative functions $\ell=\ell(w,x)$ and $u=u(w,x)$ for which

$$\int_\ell^u g(\theta|w,x)\,\mathrm{d}\theta=1-\alpha \tag{1}$$

and to minimize the length of the interval ℓ and u must satisfy

$$[\ell,u]=\{\theta\colon g(\theta|w,x)\geqslant c\} \tag{2}$$

for some constant c (see Berger, 1980, p. 266)

Theorem 1. *Conditions* (1) *and* (2) *hold with*

$$\ell=\max[0,x-bs] \quad \textit{and} \quad u=x+bs, \tag{3}$$

where

$$b = \max\{H_r^{-1}[1 - \alpha H_r(t)], H_r^{-1}[\tfrac{1}{2} + \tfrac{1}{2}(1-\alpha)H_r(t)]\} \tag{4}$$

and $t = x/s$.

Proof. First, it is clear that $g(\theta|w,x) \geqslant c$ iff $|\theta - x| \leqslant bs$, where $b = b(w,x,c)$ is strictly decreasing in c, for $0 < c < g(x|w,x)$. Thus, $\ell = \max(0, x - bs)$ and $u = x + bs$, where b is chosen to satisfy (1). There are two cases to be considered. If $t \leqslant b$, then $\ell = 0$ and (1) requires

$$\alpha = \int_{x+bs}^{\infty} g(\theta|w,x)\,\mathrm{d}\theta = \frac{1 - H_r(b)}{H_r(t)}$$

or

$$b = H_r^{-1}[1 - \alpha H_r(t)]. \tag{5}$$

Further, the right side of (5) is at least t iff $1 - \alpha H_r(t) \geqslant H_r(t)$ or, equivalently,

$$t \leqslant t_0 = H_r^{-1}\left(\frac{1}{1+\alpha}\right), \tag{6}$$

say. So, $\ell = 0$ and $u = x + bs$ where b is as in (5) when $t \leqslant t_0$. Similarly, if $t > t_0$, then (1) becomes

$$1 - \alpha = \int_{x-b}^{x+b} g(\theta|w,x)\,\mathrm{d}\theta = \frac{2H_r(b) - 1}{H_r(t)}$$

or

$$b = H_r^{-1}[\tfrac{1}{2} + \tfrac{1}{2}(1-\alpha)H_r(t)]. \tag{7}$$

Since the right side of (5) is decreasing in t and the right side of (7) is increasing, b is (4) in both cases $t \leqslant t_0$ and $t \geqslant t_0$. □

Remark 1. Observe that b depends only on $t = x/s$; it is denoted by $b(t)$ when this dependence is important. For later reference, observe also that

$$b(t) \geqslant b(t_0) = t_0 \tag{8}$$

for all t.

To picture the dependence of the boundaries on x and s, it is convenient to write the credible interval as

$$[\ell, u] = \left\{\theta\colon\ \max[0, t-b] \leqslant \frac{\theta}{s} \leqslant t + b\right\}.$$

Fig. 2 displays the boundaries $\max[0, t-b]$ and $t + b$ as functions of t.

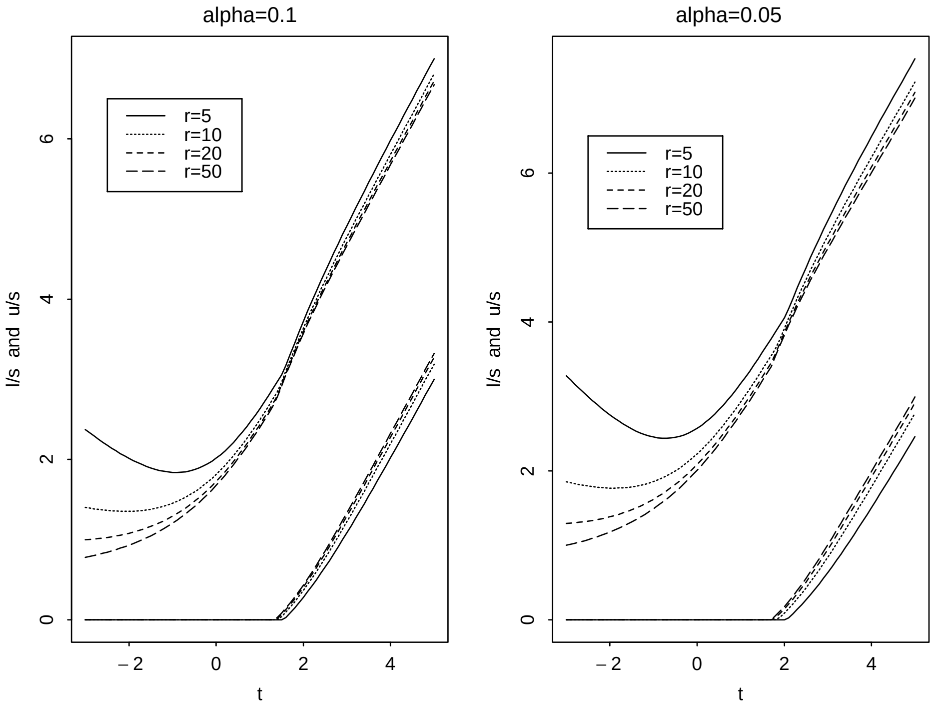

Fig. 2. Credible limits for $\theta/\hat{\sigma}$.

Remark 2. As $r \to \infty$, the boundaries converge to those of Roe and Woodroofe (2000). For finite r, however, the boundaries are decreasing in t for very small t. This may seem mildly surprising but is easily explained by long tails of the t distribution, which imply that

$$\lim_{x\to-\infty} G(-cx|w,x) = 1-(c+1)^{-r}$$

for any positive constant c. It follows that

$$\lim_{x\to-\infty} \frac{u}{|x|} = \alpha^{-1/r} - 1$$

so that $u(w,x)$ is decreasing for small x and $u(w,x) \to \infty$ as $x \to -\infty$ for any fixed s. The behavior of ℓ and u when $x \to \infty$ is easily understood because

$$\lim_{t\to\infty} H_r^{-1}\left[\frac{1}{2}+\frac{1}{2}(1-\alpha)H_r(t)\right] = H_r^{-1}\left(1-\frac{\alpha}{2}\right) < \infty.$$

Thus,

$$\frac{\ell(w,x)}{s} = t - H_r^{-1}\left(1-\frac{\alpha}{2}\right) + \mathrm{o}(1) \quad \text{and} \quad \frac{u(w,x)}{s} = t + H_r^{-1}\left(1-\frac{\alpha}{2}\right) + \mathrm{o}(1)$$

as $x \to \infty$ for fixed s.

Frequentist coverage. The frequentist coverage probability $P_{\sigma,\theta}[\ell(W,X) \leqslant \theta \leqslant u(W,X)]$ of the intervals may be $< 1 - \alpha$, but a non-trivial lower bound can be obtained.

Theorem 2.

$$P_{\sigma,\theta}[\ell(W,X) \leqslant \theta \leqslant u(W,X)] \geqslant \frac{1-\alpha}{1+\alpha} \tag{9}$$

for all $\theta \geqslant 0$ *and* $\sigma > 0$.

Proof. From (3), (4), and (6), it is clear the $\ell(c^2w, cx) = c\ell(w,x)$ and $u(c^2w,x) = cu(w,x)$ for $c > 0$, and it follows easily that the coverage probability depends on σ and θ only through θ/σ. So, there is no loss of generality in supposing that $\sigma = 1$.

It is clear that $\ell(w,x)$ is strictly increasing in x where $\ell(w,x) > 0$. So, for each $\theta > 0$ and $w > 0$, the equation $\theta = \ell(w,x)$ has a unique solution $x = c_2(w,\theta)$. For the upper boundary, let $v(w) = \min_x(w,x)$. Then for each $\theta \geqslant v(w)$ and $w > 0$, the equation $\theta = u(w,x)$ has a largest solution $x = c_1(w,\theta)$. Let $c_1(w,\theta) = -\infty$ for $\theta < v(w)$. With this notation

$$\begin{aligned} \theta &= u(w,c_1) = c_1 + sb\left(\frac{c_1}{s}\right) \quad \text{when } \theta \geqslant v(w), \\ \theta &= \ell(w,c_2) = c_2 - sb\left(\frac{c_2}{s}\right) \end{aligned} \tag{10}$$

and $\ell(w,x) \leqslant \theta \leqslant u(w,x)$ whenever $c_1(w,\theta) \leqslant x \leqslant c_2(w,\theta)$. So the conditional coverage probability given W is at least

$$P_{1,\theta}[c_1 \leqslant X \leqslant c_2 | W = w] \geqslant \Phi(c_2 - \theta) - \Phi(c_1 - \theta). \tag{11}$$

Using (8) and (10), and considering the case $\theta < v(w)$, the right side of (11) is

$$\Phi\left[sb\left(\frac{c_2}{s}\right)\right] - \Phi\left[-sb\left(\frac{c_1}{s}\right)\right] = \Phi\left[sb\left(\frac{c_2}{s}\right)\right] + \Phi\left[sb\left(\frac{c_1}{s}\right)\right] - 1 \geqslant 2\Phi(st_0) - 1$$

with the convention $b(-\infty) = \infty$. Integrating the latter result over w then yields

$$P_{1,\theta}[c_1 \leqslant X \leqslant c_2] \geqslant 2H_r(t_0) - 1 = \frac{1-\alpha}{1+\alpha}.$$

This establishes (9) for $\theta > 0$. For $\theta = 0$, the result follows from continuity. □

Fig. 3 displays the actual coverage probability as a function of θ/σ for selected values of r and α. From these graphs, the minimum coverage probability appears to decrease as r increases. Thus, we can conclude that the minimum coverage probability with unknown σ is always greater than that with known σ which is given in Roe and Woodroofe (2000). From Fig. 3, the actual coverage probability can be substantially higher than the right side of (9). For example, when $\alpha = 0.1$, the right side of (9) is 0.8181, and the minimum coverage probability is 0.8657; when $\alpha = 0.05$, the right side of (9) is 0.9048, and the minimum coverage probability is 0.9308.

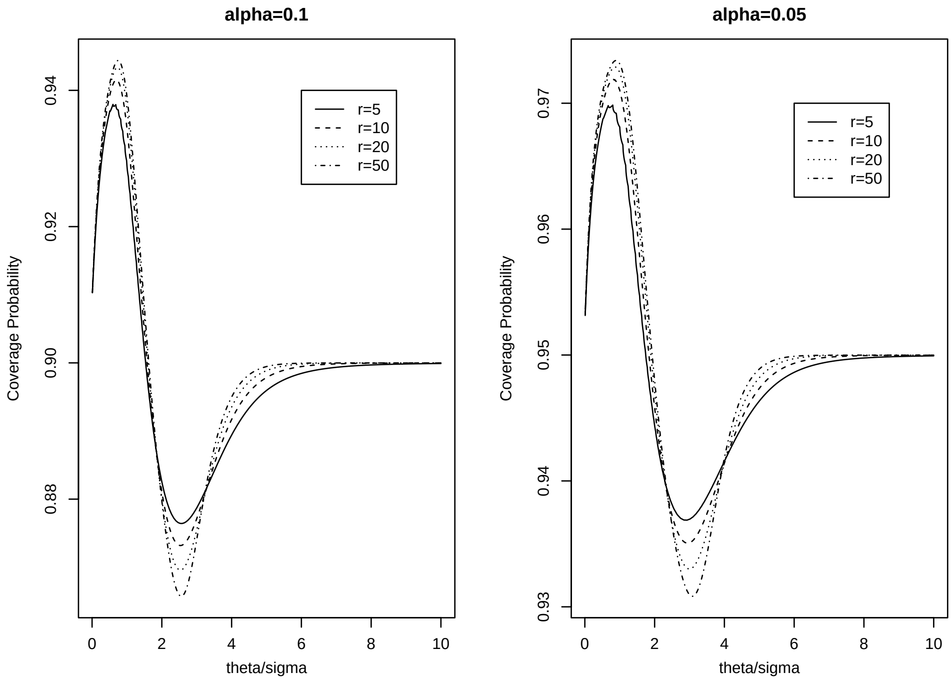

Fig. 3. Graphs of $P_{\sigma,\theta}[\ell(w,x) \leqslant \theta \leqslant u(w,x)]$.

Similar numerical calculations are possible when a point mass at zero is added to the prior for θ. In these calculations, the minimum coverage probability is decreased as the size of the point mass increased.

3. The Poisson case

Credible intervals. In the Poisson case, suppose that we observe $X = B + S$ and W, where B, S, and W are independent random variables for which $B \sim \text{Poisson}(\xi)$, $S \sim \text{Poisson}(\theta)$, and $W \sim \text{Poisson}(m\xi)$, where $\theta, \xi \geqslant 0$ are unknown parameters and $m > 0$ is known. We will find credible intervals for θ with ξ treated as a nuisance parameter. The joint probability mass function of W and X is

$$p_{\xi,\theta}(w,x) = \frac{1}{w!x!} m^w \xi^w (\xi + \theta)^x \, \mathrm{e}^{-\theta-(m+1)\xi}$$

$$= \sum_{k=0}^{x} \frac{1}{w!k!(x-k)!} m^w \xi^{w+k} \theta^{x-k} \, \mathrm{e}^{-\theta-(m+1)\xi}.$$

Suppose now θ and ξ are given an improper prior density, say $\xi^{a-1}\mathrm{e}^{-b\xi}$, for $0 \leqslant \xi, \theta < \infty$ where $a > 0$, $b \geqslant 0$. Then the marginal probability mass function

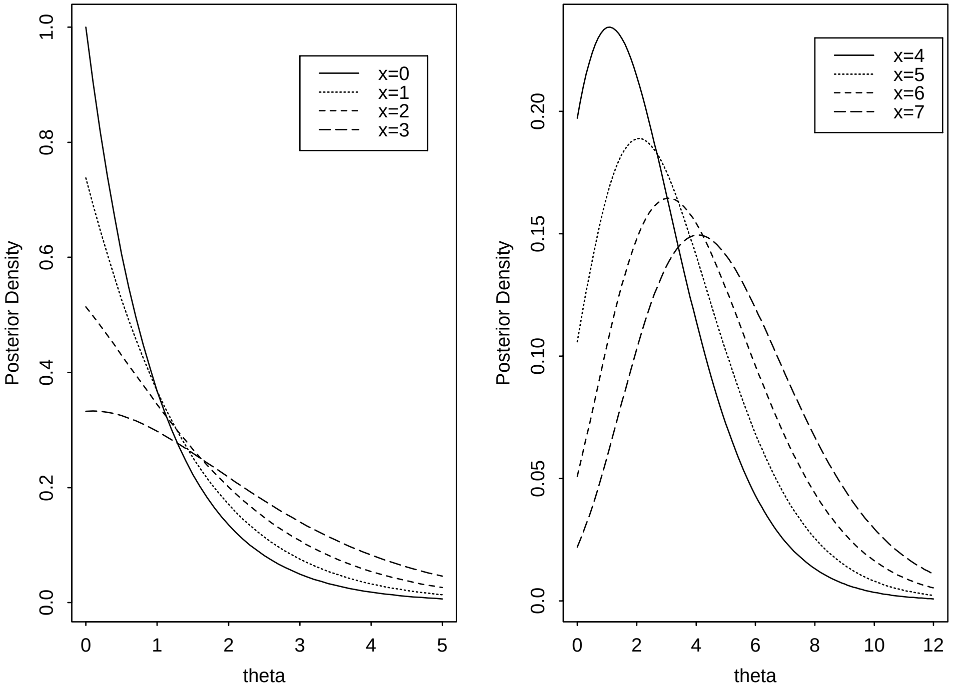

Fig. 4. Graphs of $g(\theta|w,x)$. Here, $a=b=1.0$, $w=60$, and $m=20$.

of W and X is

$$p(w,x)=\int_0^\infty\int_0^\infty p_{\xi,\theta}(w,x)\xi^{a-1}\,\mathrm{e}^{-b\xi}\,\mathrm{d}\xi\,\mathrm{d}\theta$$
$$=\sum_{k=0}^{x}\frac{\Gamma(a+w+k)}{w!k!}\,\frac{m^w}{(b+m+1)^{a+w+k}}$$

after some routine calculations. The posterior density and distribution function of θ are

$$g(\theta|w,x)=\frac{1}{p(w,x)}\sum_{k=0}^{x}\frac{\Gamma(a+w+k)}{w!k!(x-k)!}\,\frac{m^w}{(b+m+1)^{a+w+k}}\,\theta^{x-k}\,\mathrm{e}^{-\theta} \tag{12}$$

and

$$G(\theta|w,x)=\frac{1}{p(w,x)}\sum_{k=0}^{x}\frac{\Gamma(a+w+k)}{w!k!}\,\frac{m^w}{(b+m+1)^{a+w+k}}\,H_{x-k+1}(\theta), \tag{13}$$

where now H_j denotes the gamma distribution function with shape parameter j and unit scale parameter.

Fig. 4 shows the posterior density function of θ when $w=60$ and $m=20$. Based on W, $\hat{\xi}=W/m$ is an estimator of ξ. In Fig. 4, the posterior density is a decreasing function of θ when $x\leqslant\hat{\xi}=3$, and unimodal with a positive mode when $x>\hat{\xi}$. The unimodality is no accident.

Theorem 3. *If $a + w \geqslant 1$, then $g(\theta|w,x)$ is log-concave (and therefore unimodal).*

The proof is presented to the end of the section. Of course, $a+w \geqslant 1$ for all $w \geqslant 0$ if $a \geqslant 1$.

As above, a Bayesian credible interval consists of two functions $\ell = \ell(w,x)$ and $u = u(w,x)$ for which

$$1 - \alpha = G(u|w,x) - G(\ell|w,x),$$
$$[\ell, u] = \{\theta\colon g(\theta|w,x) \geqslant c\}$$

for some $c = c(w,x)$. Finding ℓ and u is more involved in the Poisson case than the normal one, but the following algorithm can be used:

(i) First solve the equation

$$G(z|w,x) = 1 - \alpha$$

for z.

(ii) If $g(0|w,x) \geqslant g(z|w,x)$, then let $u = z$ and $\ell = 0$. Otherwise let $c_0 = g(0|w,x)$ and $c_1 = g(z|w,x)$. Find the mode $\hat{\theta}$ of $g(\theta|w,x)$; and iterate step (iii) to convergence.

(iii) Let $c = (c_0 + c_1)/2$; solve the equations $g(y|w,x) = c = g(z|w,x)$ for y and z in the range $0 < y < \hat{\theta} < z$. If $G(z|w,x) - G(y|w,x) \leqslant 1 - \alpha$, let $c_1 = c$ otherwise, let $c_0 = c$ and iterate.

In view of the log-concavity of $g(\theta|w,x)$ when $a + w \geqslant 1$, solving for $\hat{\theta}$, y and z is straightforward. Fig. 5 displays the credible intervals for selected x and w when $a = b = 1$ and $m = 20$. Both the upper and lower bounds appear to decrease when w increases.

Coverage probability. The coverage probability

$$P_{\theta,\xi}[l(W,X) \leqslant \theta \leqslant u(W,X)] = \sum_{x=0}^{\infty}\sum_{w=0}^{\infty} I_{\{l(w,x) \leqslant \theta \leqslant u(w,x)\}}\, p_{\theta,\xi}(w,x)$$

can be calculated by approximating the double infinite sum. Graphs of the coverage probability are shown in Fig. 6, and the minimum of the coverage probability again decreases when m increases. The curves fluctuate up and down as θ changes because the Poisson distribution is discrete.

Unimodality. It remains to show that $g(\theta|w,x)$ is log-concave in θ for fixed w and x. Recall that a non-negative function $K(x,y)$ defined on a subset of R^2 is said to be TP_2 in (x,y), if $K(x_1,y_1)K(x_2,y_2) \geqslant K(x_1,y_2)K(x_2,y_1)$, for any $x_1 < x_2$ and $y_1 < y_2$. It can be shown (Karlin, 1996, p. 157) that densities $f_\theta(x)$ have monotone likelihood ratio in x, iff $f_\theta(x)$ is TP_2 in (θ,x).

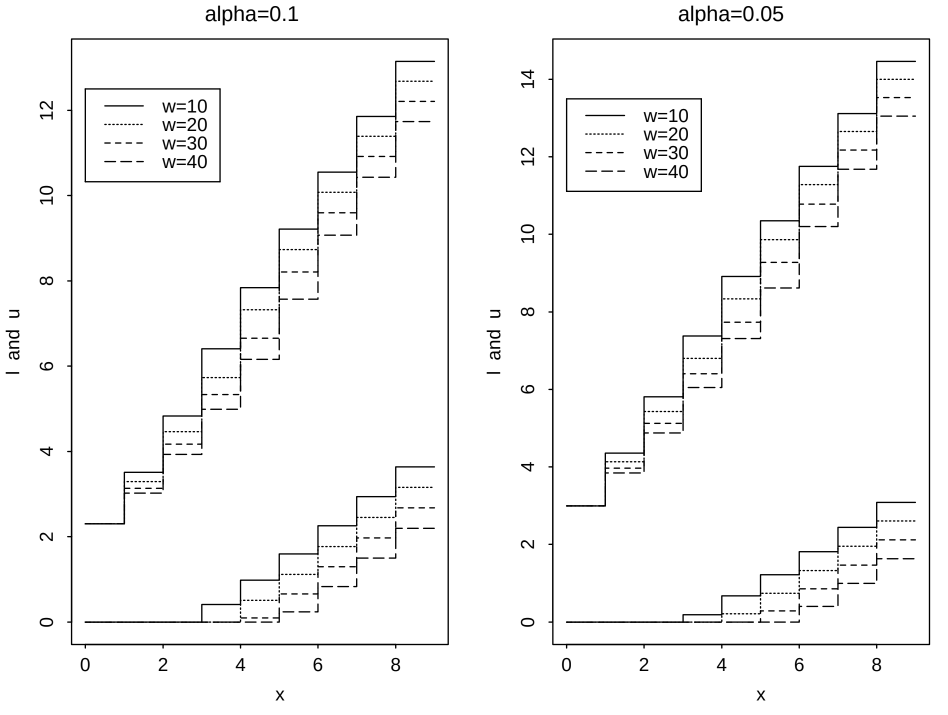

Fig. 5. Credible limits for θ. Here, $m=20$ and $a=b=1.0$.

To show log-concavity first write

$$g(\theta|w,x)=\frac{m^w}{w!\,p(w,x)c^{a+w}}\,\tilde{g}(\theta,x),$$

where

$$\tilde{g}(\theta,x)=\sum_{k=0}^{x}\frac{\Gamma(a+w+k)}{k!(x-k)!}\,\frac{1}{c^k}\,\theta^{x-k}\,\mathrm{e}^{-\theta}$$

and $c=b+m+1$. Then

$$\frac{\partial\tilde{g}(\theta,x)}{\partial\theta}=\tilde{g}(\theta,x-1)-\tilde{g}(\theta,x)$$

and, therefore,

$$\frac{\partial}{\partial\theta}\log[\tilde{g}(\theta,x)]=\frac{\tilde{g}(\theta,x-1)}{\tilde{g}(\theta,x)}-1. \tag{14}$$

If it were known that $\tilde{g}(\theta,x)$ is TP_2 in (θ,x), then it would follow that the right side of (14) is a decreasing function of θ; and from that it would follow that $g(\theta|x)$

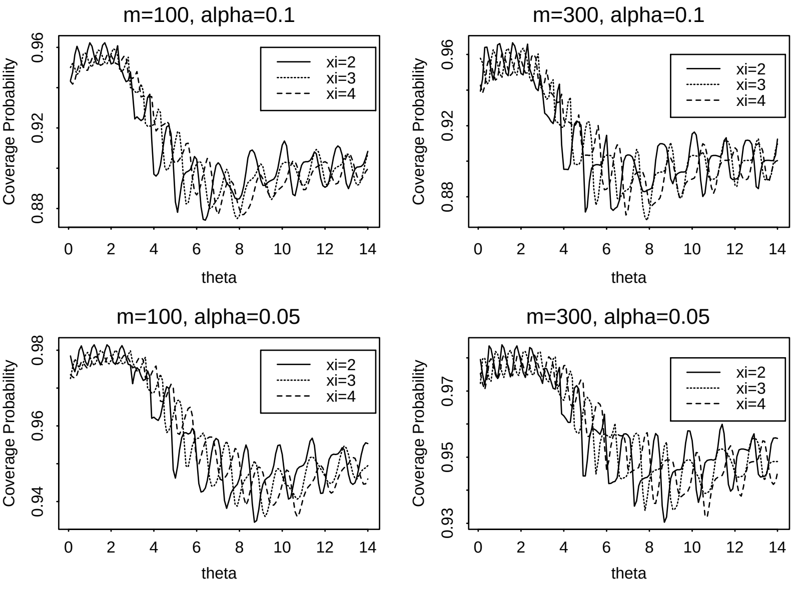

Fig. 6. Graphs of $P_{\theta,\xi}[\ell(w,x) \leqslant \theta \leqslant u(w,x)]$ in Poisson model. Here, $a = b = 1.0$.

is log-concave in θ. To see that $\tilde{g}(\theta,x)$ is TP_2 in (θ,x), write

$$\tilde{g}(\theta,x) = \sum_{j=0}^{\infty} f(x,j)h(j,\theta),$$

where $f(x,j) = 0$ for $j > x$,

$$f(x,j) = \frac{\Gamma(a+w+x-j)}{(x-j)!c^{x-j}}$$

for $j = 0,\ldots,x$, and

$$h(j,\theta) = \frac{1}{j!}\,\theta^j\,\mathrm{e}^{-\theta}$$

for $j = 0,1,2,\ldots$. It is clear that h is TP_2 in (j,θ), since it is a Poisson probability mass function (see Karlin, 1996, p. 19). Next,

$$\frac{f(x+1,j)}{f(x,j)} = \frac{a+w+x-j}{(x+1-j)c} = \frac{1}{c}\left[1 + \frac{a+w-1}{x+1-j}\right]$$

for $j \leqslant x$. Since the right side is increasing in $j \leqslant x$ if $a+w \geqslant 1$, and $f(x+1,j)/f(x,j) = \infty$ when $j = x$, it follows that f is TP_2 in (x,j) (see Karlin, 1996, p. 157). That $\tilde{g}$ is TP_2 in (θ,x) now follows from the basic composition theorem (see Karlin, 1996, p. 98).

Remark 3. We have been unable to prove that g is unimodal when $a + w < 1$. It is unimodal, however, in several examples that we considered ($w = 0$, $a = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$, $b = 0, 1, 2$, and $x = 0, 1, 2, 3$).

References

Berger, J., 1980. Statistical Decision Theory, Foundations, Concepts, and Methods. Springer, Berlin.

Feldman, G.J., Cousins, R., 1998. Unified approach to the classical statistical analysis of small signals. Phys. Rev. D 57, 3873–3889.

Karlin, S., 1996. Total Positivity. Addison-Wesley, Reading, MA.

Roe, B., Woodroofe, M., 1999. Improved probability method for estimating signal in the presence of background. Phys. Rev. D 60, 053009.

Roe, B., Woodroofe, M., 2000. Setting confidence belts. Phys. Rev. D 63, 013009.

Shao, P.Y., Strawderman, W.E., 1996. Improving of the MLE of a positive normal mean. Statis. Sinica 6, 259–274.

Woodroofe, M., Wang, H., 2000. On the problem of low counts in a signal plus noise model. Ann. Statist. 28, 1561–1569.

Zeitnitz, B., et al., 1998. Neutrino oscillation results from KARMEN. Progr. Particle Nucl. Phys. 40, 169–181.