Statistica Sinica **15**(2005), 59-72

ADMISSIBLE MINIMAX ESTIMATION OF THE SIGNAL WITH KNOWN BACKGROUND

Tonglin Zhang and Michael Woodroofe

Purdue University and The University of Michigan

Abstract: Suppose that an observed count X is of the form $X = B + S$, where the background B and the signal S are independent Poisson random variables with parameters b and θ, b is known, but θ is not. The model arises in astronomy and high-energy physics, and some recent articles have suggested conditioning on the observed bound for B; that is, if $X = n$ is observed, then the suggestion is to base the inference on the conditional distribution of X given $B \leq n$. This suggestion is used here to derive an estimator of the signal, and the estimator is shown to be admissible and minimax.

Key words and phrases: Admissible, Confidence and credible interval, coverage probability, mean square error, minimax, signal plus background.

1. Introduction

In some problems, a signal S may be combined with a background B, leaving an observed count $X = B + S$. Here we suppose that B and S are independent Poisson random variables with means b and θ respectively, so that X has a Poisson distribution with mean $b + \theta$. Further b is assumed known and θ is unknown, as might be appropriate if there were historical data on the background. Models of this nature arise in astronomy and high energy physics. For example, the KARMEN 2 Group has been searching for a neutrino oscillation reported from an earlier experiment at the Los Alamos Neutrino Detector. They had expected to see about 9.3 background events, Eitel (2000), and had observed seven events total. This example and others have sparked interest in statistical inference when maximum likelihood estimators are on or near a physical boundary of the parameter space. Recent work along these lines is reviewed by Mandelkern (2002) and discussants. Here we investigate the filtering problem, estimating (or filtering) S from X. This formulation, estimating S instead of θ, is not the usual one. Its motivation is that S exists in the physical world, while θ only exists within a mathematical model (although θ is used in the estimation of S). This is in the spirit of predictive inference (Geisser (1993)). An example in which interest centers on signals, as opposed to their expectations, is presented in the next section.

Thus, suppose that $X = n$ is observed and consider the problem of estimating S. If θ were known, then the optimal estimator of S (for squared error loss) is the conditional expectation $\hat{S}_\theta := E_\theta(S|n) = n\theta/(b+\theta)$. Substituting the maximum likelihood estimator $\hat{\theta} = \max(0, n-b)$ into this expression, then leads to

$$\hat{S}_{\hat{\theta}} = \hat{\theta} = \max(0, n-b).$$

That is, the maximum likelihood estimator subtracts the expected background from the observed count and then takes the positive part. Let us call B an ancillary variable, since its distribution is known. It is not observed, but a bound is. For if $X = n$, then $B \leq n$. In this spirit, let $\hat{B}(n) = E(B|B \leq n)$, $\hat{S}(n) = n - \hat{B}(n)$, and observe that $\hat{B}(n)$ is computable, since b is known. Thus, $\hat{S}(n)$ subtracts the conditional expected background, given the observed bound $B \leq n$; and $0 < \hat{S}(n) < n$ for all $n \geq 1$. In the example, $b = 9.3$, $X = 7$, $E(B|B \leq 7) = 5.80$, $\hat{S} = 1.20$, and $\hat{S}_{\hat{\theta}} = 0$. In this example, $\hat{S}_{\hat{\theta}}$ effectively estimates B by $n = 7$, leaving $\hat{S}_{\hat{\theta}} = 0$. This seems excessive for B can be at most 7, and $P[B = 7|B \leq 7]$ is only 0.376.

After the example in Section 2, the main results of the paper appear in Sections 3 and 4: It is shown that $\hat{S}$ is an admissible, minimax predictor of S for squared error loss and that $\hat{S}_{\hat{\theta}}$ is inadmissible. In the process, it is shown that $\hat{S}$ is generalized Bayes with respect to the uniform prior over $0 \leq \theta < \infty$. Confidence intervals for S are discussed in Section 5, and some implications of our results for estimating parameters is described briefly in Section 6. Section 6 also contains some additional references.

2. Fornax

Irwin and Hatzidimitriou (1995) report densities of star counts in elliptical annuli for several dwarf spheroidal galaxies near to the Milky Way, along with an overall background density for each galaxy. Selected values for the galaxy Fornax are listed in Table 1 and used to illustrate the difference between the MLE and $\hat{S}$. Each line of the table corresponds to the annulus in which the distance r from the center is between the values listed on the line and the previous one. Each X is regarded as the sum of a Poisson signal with an unknown mean and an independent background with mean listed under b. The counts are complete out a given magnitude (which depends on color) and, so, constitute a census of a well-defined portion of the star population. Regarding the latter stars as a finite population, there is interest in estimating the population totals within annuli for Fornax. For over half of the annuli there is no real difference between the MLE and $\hat{S}$, and most of these lines have been omitted. There are some substantial differences in the remaining lines, however. Especially, the MLE has many null values, even in cases where there are significant estimated counts at larger distances. It can be seen that the MLE and $\hat{S}$ are almost identical when the estimated values are large, since $P(B \leq n)$ is very close to 1 when the MLE is large.

Table 1. Star counts from fornax.

r	d	n	b	mle	$\hat{S}$	" "	r	d	n	b	mle	$\hat{S}$
1.05	29.33	71	3.2	67.8	67.80		100.72	1.43	662	620.0	42.0	44.50
2.10	30.57	222	9.7	212.3	212.30		101.77	1.32	617	626.5	0	16.37
3.15	31.31	380	16.2	363.8	363.80		102.82	1.31	619	633.0	0	15.20
4.20	28.98	492	22.7	469.3	469.30		103.87	1.32	630	639.5	0	16.57
	...			...			104.92	1.37	660	646.0	14.0	26.03
71.34	1.44	471	438.2	32.8	35.38		105.97	1.29	628	652.5	0	12.98
72.39	1.30	431	444.7	0	12.13		107.01	1.35	658	652.7	5.3	22.08
73.44	1.54	519	451.2	67.8	67.86		108.06	1.38	685	665.5	19.5	29.28
74.49	1.45	495	457.7	37.3	39.25		109.11	1.34	672	672.0	0	20.27
75.54	1.51	523	464.2	58.8	59.03		110.16	1.29	653	678.5	0	13.14
76.59	1.49	523	470.7	52.3	52.80		111.21	1.42	726	685.0	41.0	44.23
77.64	1.40	499	477.2	21.8	27.97		112.26	1.32	681	691.5	0	17.07
78.69	1.43	516	483.7	32.3	35.48		113.31	1.34	698	697.9	0.1	20.70
79.74	1.37	501	490.2	10.8	21.94		114.36	1.33	699	704.4	0	18.86
80.79	1.51	560	496.7	63.3	63.47		115.41	1.39	737	710.9	26.1	33.84
81.83	1.55	576	498.4	77.6	77.63		116.46	1.37	734	717.4	16.6	28.40
82.88	1.34	510	509.6	0.4	17.75		117.51	1.28	692	723.9	0	12.50
83.93	1.48	570	516.1	53.9	54.47		118.56	1.32	720	730.4	0	17.67
84.98	1.46	569	522.6	46.4	47.61		119.61	1.41	775	736.9	38.1	42.45
86.03	1.29	509	529.1	0	12.05		120.65	1.33	731	736.3	0	19.36
87.08	1.50	600	535.6	64.4	64.61		121.70	1.35	755	749.9	5.1	23.44
88.13	1.47	595	542.1	52.9	53.64		122.75	1.44	813	756.4	56.6	57.97
89.18	1.35	553	548.6	4.4	20.02		123.80	1.35	769	762.9	6.1	24.05
90.23	1.57	650	555.1	94.9	94.90		124.85	1.30	746	769.4	0	14.85
91.28	1.44	604	561.6	42.4	44.38		125.90	1.36	787	775.9	11.1	26.47
92.33	1.44	611	568.1	42.9	44.86		126.95	1.28	747	782.4	0	12.62
93.38	1.46	626	574.6	51.4	52.40		128.00	1.31	771	788.9	0	16.43
94.43	1.30	564	581.1	0	13.59		129.05	1.36	807	795.4	11.6	26.97
95.47	1.35	586	582.0	4.0	20.41		130.10	1.34	802	801.9	0.1	22.21
96.52	1.43	634	594.0	40.0	42.65		131.15	1.31	790	808.3	0	16.60
97.57	1.28	574	600.5	0	11.82		132.20	1.41	857	814.8	42.2	46.27
98.62	1.33	603	607.0	0	17.81		133.24	1.26	765	813.5	0	10.95
99.67	1.29	591	613.5	0	12.84		134.29	1.36	840	827.8	12.2	27.70

3. (In) Admissibility

It is convenient to let f_μ and F_μ be the probability mass function and distribution function of the Poisson distribution with mean μ. Then

$$\hat{B}(n) = E(B|B \leq n) = b\frac{F_b(n-1)}{F_b(n)} = b\left[1 - \frac{f_b(n)}{F_b(n)}\right]. \tag{1}$$

It is also convenient to abbreviate $\hat{B}(X)$ and $\hat{S}(X)$ by $\hat{B}$ and $\hat{S}$. Now let $\tilde{S} = \tilde{S}(X)$ be any estimator for which $0 \leq \tilde{S} \leq X$. Then the risk of $\tilde{S}$ is

$$R(\tilde{S}, \theta) = E_\theta[(\tilde{S} - S)^2] = E_\theta[(\tilde{B} - B)^2], \tag{2}$$

where $\tilde{B} = X - \tilde{S}$. Recall that $\tilde{S}(X)$ is *inadmissible* if there is an $\tilde{S}'$ for which $R(\tilde{S}', \theta) \leq R(\tilde{S}, \theta)$ for all $\theta \geq 0$ with strict inequality for some θ and that $\tilde{S}$ is *admissible* otherwise.

Inadmissibility of the MLE. If π is a finite measure over $[0, \infty)$, write E^π for integration with respect to the joint distribution of θ and X when $\theta \sim \pi$; and write $E^\pi(\cdot|n)$ for conditional expectation given $X = n$. Further, let

$$\bar{R}(\tilde{S}, \pi) = \int_0^\infty R(\tilde{S}, \theta)\pi\{d\theta\} = E^\pi[(\tilde{B} - B)^2]$$

for estimators $\tilde{S}$, where $\tilde{B} = X - \tilde{S}$; and let $\bar{R}(\pi) = \inf_{\tilde{S}} \bar{R}(\tilde{S}, \pi)$.

The inadmissibility of $\hat{S}_{\hat{\theta}}$ will be deduced from Stein (1955), the applicability of which is discussed in the Section 7.

Theorem 1. *If* $b > 0$*, then* $\hat{S}_{\hat{\theta}} = \max[0, X - b]$ *is inadmissible.*

Proof. From Stein (1955), a necessary and sufficient condition for $\hat{S}_{\hat{\theta}}$ to be admissible is that, for every $\theta_0 \in [0, \infty)$ and $\epsilon > 0$, there exist a finite prior π for which $\pi\{\theta_0\} \geq 1$ and $\bar{R}(\hat{S}_{\hat{\theta}}, \pi) - \bar{R}(\pi) \leq \epsilon$. In particular, if $\hat{S}_{\hat{\theta}}$ is to be admissible, then there must be a sequence π_k of finite priors for which $\pi_k\{1\} \geq 1$ for every $k = 1, 2, \ldots$ and

$$\lim_{k\to\infty} \bar{R}(\hat{S}_{\hat{\theta}}, \pi_k) - \bar{R}(\pi_k) = 0. \tag{3}$$

Supposing that this is the case, let $\hat{B}_k(n) = E^{\pi_k}(B|n) = E^{\pi_k}[nb/(b + \theta)|n]$ and $\hat{B}_0(n) = b \wedge n$. Then

$$\begin{aligned}\bar{R}(\hat{S}_{\hat{\theta}}, \pi_k) - \bar{R}(\pi_k) &= \sum_{n=0}^\infty [\hat{B}_0(n) - \hat{B}_k(n)]^2 \int_0^\infty \frac{1}{n!}(b + \theta)^n e^{-(b+\theta)}\pi_k\{d\theta\} \\ &\geq e^{-(b+1)} \sum_{n=0}^\infty \frac{1}{n!}(b + 1)^n[\hat{B}_0(n) - \hat{B}_k(n)]^2.\end{aligned}$$

So, $\lim_{k\to\infty} \hat{B}_k(n) = \hat{B}_0(n) = b \wedge n$ for all $n = 0, 1, 2, \ldots$. Now

$$\hat{B}_k(n) = \left[\int_0^\infty f_{b+\theta}(n)\pi_k\{d\theta\}\right]^{-1} \int_0^\infty \frac{nb}{b + \theta} f_{b+\theta}(n)\pi_k\{d\theta\}.$$

After some simple algebra, this may be written as $\hat{B}_k(n) = nb\mu_{k,n-1}/\mu_{k,n}$, where $\mu_{k,n} = \int_0^\infty (b+\theta)^n \pi_k^*\{d\theta\}$, $\pi_k^*\{d\theta\} = (1/C_k)e^{-\theta}\pi_k\{d\theta\}$, and $C_k = \int_0^\infty e^{-\theta}\pi_k\{d\theta\}$, so that each π_k^* is a probability measure and $\mu_{k,0} = 1$. So, if π_k exist, then

$$\lim_{k\to\infty} \frac{\mu_{k,n-1}}{\mu_{k,n}} = \frac{1}{nb}(b \wedge n) = \frac{1}{b \vee n}$$

for $n = 1, 2, \ldots$. This requires

$$\lim_{k\to\infty} \mu_{k,n} = \begin{cases} b^n, & \text{if } n \le m, \\ b^m n!/m!, & \text{if } n > m, \end{cases} \tag{4}$$

where $m = \lfloor b \rfloor$ is the greatest integer that is less than or equal to b. In turn, (4) requires that the distribution functions $G_k(\omega) = \pi_k^*\{\theta : b + \theta \le \omega\}$ converge weakly to a distribution function G that is supported by $[b, \infty)$ and has moments $\mu_n = b^n$ or $b^m n!/m!$ for $n \le m$ or $n > m$. See, for example, Billingsley (1995, Chap.30). Such a G cannot exist, however, if b is positive $-$ For if $b \ge 1$, then G must be degenerate at b, since $\mu_1 = b$, and the higher moments of the degenerate distribution are not $b^m n!/m!$. On the other hand if $0 < b < 1$, then $\mu_n = n!$ are the moments of the standard exponential distribution, which is not supported by $[b, \infty)$. So, the assumed existence of π_k in (3) leads to a contradiction.

Admissbility of $\hat{S}$. The admissibility of $\hat{S}$ will be deduced from a Bayesian approximation. Consider the (unnormalized) priors $\pi_\alpha(d\theta) = e^{-\alpha\theta}$, $\theta \ge 0$ for $\alpha \ge 0$. Thus, π_α is finite if $\alpha > 0$, and π_0 is the (infinite) uniform distribution over $[0, \infty)$. Write E^α for integration with respect to π_α, $E^\alpha(\cdot|n)$ for posterior expectation (which is proper), and

$$\bar{R}(\tilde{S}, \alpha) = E^\alpha[(S - \tilde{S})^2] = \int_0^\infty R(\tilde{S}, \theta)e^{-\alpha\theta}d\theta$$

for the *integrated risk* of a predictor $\tilde{S}$. This is minimized by $\hat{S}^\alpha(n) = E^\alpha(S|n)$.

The computation of $\hat{S}^\alpha$ is straightforward. First observe that $\hat{S}^\alpha(n) = n - \hat{B}^\alpha(n)$, where $\hat{B}^\alpha(n) = E^\alpha(B|n)$. From $P_\theta[B = k, X = n] = b^k\theta^{n-k}e^{-(b+\theta)}/[k!(n-k)!]$, follows, for $n \ge 0$,

$$\begin{aligned} P^\alpha[B = k, X = n] &= \int_0^\infty P_\theta[X = n, B = k]e^{-\alpha\theta}d\theta = \frac{b^k e^{-b}}{k!(1+\alpha)^{n-k+1}}, \\ P^\alpha[X = n] &= \sum_{k=0}^n P^\alpha[B = k, X = n] = \frac{e^{\alpha b}}{(1+\alpha)^{n+1}} F_{(1+\alpha)b}(n), \end{aligned} \tag{5}$$

$$\hat{B}^\alpha(n) := E^\alpha(B|n) = (1+\alpha)b\frac{F_{(1+\alpha)b}(n-1)}{F_{(1+\alpha)b}(n)}. \tag{6}$$

Thus $\hat{B}^0(n) = \hat{B}(n)$ in (1). Let

$$\delta_{\alpha,n} = \frac{f_{(1+\alpha)b}(n)}{F_{(1+\alpha)b}(n)} = \frac{(1+\alpha)^n b^n/n!}{\sum_{k=0}^n (1+\alpha)^k b^k/k!}, \tag{7}$$

and denote $\delta_{0,n}$ by δ_n. Then, $\hat{B}^\alpha(n) = b(1+\alpha)(1-\delta_{\alpha,n})$ and $\hat{B}(n) = b(1-\delta_n)$.

Theorem 2. *If $\hat{B}^\alpha$ and $\hat{B}$ are as in* (6) *and* (1), *then* $\lim_{\alpha\to 0} E^\alpha[\hat{B}^\alpha - \hat{B}]^2 = 0$.

Proof. For $0 \le \alpha \le 1$,

$$\begin{aligned} E^\alpha[(\hat{B}^\alpha - \hat{B})^2] &= \int_0^\infty E_\theta(\hat{B}^\alpha - \hat{B})^2 e^{-\alpha\theta} d\theta \\ &= \int_0^\infty \sum_{n=0}^\infty \{\alpha b - [(1+\alpha)b\delta_{\alpha,n} - b\delta_n]\}^2 \frac{(b+\theta)^n}{n!} e^{-(b+\theta)} e^{-\alpha\theta} d\theta \\ &\le 2\int_0^\infty \alpha^2 b^2 e^{-\alpha\theta} d\theta + 2\sum_{n=0}^\infty \int_0^\infty [(1+\alpha)\delta_{\alpha,n} - \delta_n]^2 \frac{(b+\theta)^n}{n!} \\ &\quad \times e^{-(b+\theta)} e^{-\alpha\theta} d\theta \\ &\le 2\alpha b^2 + 2\sum_{n=0}^\infty [(1+\alpha)\delta_{\alpha,n} - \delta_n]^2. \end{aligned}$$

Clearly, $\alpha b^2 \to 0$ and $(1+\alpha)\delta_{\alpha,n} - \delta_n \to 0$ as $\alpha \to 0$ for fixed n. That the summation approaches 0 then follows from the Dominated Convergence Theorem, since

$$|(1+\alpha)\delta_{\alpha,n} - \delta_n|^2 \le 2[(1+\alpha)\delta_{\alpha,n}]^2 + 2\delta_n^2 \le 10\frac{2^{2n} b^{2n}}{n!}$$

for $0 \le \alpha \le 1$, and the right side is summable over $n \ge 0$.

Corollary 1. $\lim_{\alpha\to 0}[\bar{R}(\hat{S},\alpha) - \bar{R}(\hat{S}^\alpha,\alpha)] = 0$.

Proof. For $\alpha > 0$, $\bar{R}(\hat{S},\alpha) = E^\alpha[(S-\hat{S}_\alpha)^2 + (\hat{S}-\hat{S}_\alpha)^2] = \bar{R}(\hat{S}_\alpha,\alpha) + E^\alpha[(\hat{B} - \hat{B}_\alpha)^2]$, and $E[(\hat{B}-\hat{B}^\alpha)^2] \to 0$ by the theorem.

Since $0 \le \tilde{S}(n) \le n$ for any predictor $\tilde{S}$ (under consideration), it is clear from the Dominated Convergence Theorem that $R(\tilde{S},\theta)$ is continuous in θ for any predictor.

Corollary 2. *$\hat{S}$ is admissible.*

Proof. If $\hat{S}$ were inadmissible, then there would be an $\tilde{S}$ for which $R(\tilde{S},\theta) \le R(\hat{S},\theta)$ for all $\theta \ge 0$, and $R(\tilde{S},\theta_0) < R(\hat{S},\theta_0)$ for some $\theta_0 \ge 0$. Let $\epsilon_0 = [R(\hat{S},\theta_0) - R(\tilde{S},\theta_0)]/2$. Then, there is an $\eta > 0$, such that $R(\hat{S},\theta) \ge R(\tilde{S},\theta) + \epsilon_0$, for all non-negative θ such that $|\theta - \theta_0| < \eta$; and then

$$\bar{R}(\hat{S},\alpha) - \bar{R}(\hat{S}_\alpha,\alpha) \ge \int_{\theta_0}^{\theta_0+\eta} [R(\hat{S},\alpha) - R(\tilde{S},\alpha)] e^{-\alpha\theta} d\theta$$

$$\geq \epsilon_0 \frac{e^{-\alpha\theta_0} - e^{-\alpha(\theta_0+\eta)}}{\alpha} \to \epsilon_0 \eta > 0$$

as $\alpha \to 0$, contradicting Corollary 1.

Corollary 3. $\hat{S}(n) = E^0(S|n)$.

Proof. This is clear from the proof of Theorem 2.

Of course, it does not follow that $\hat{S}$ dominates $\hat{S}_{\hat{\theta}}$. The mean squared errors of the two predictors are shown in Figure 1 for selected b. The graphs show that the MSEs for $\hat{S}$ are less than the MSEs for $\hat{S}_{\hat{\theta}}$ for large θ but greater than the MSEs for $\hat{S}_{\hat{\theta}}$ for small θ. $\hat{S}$ does dominate $\hat{S}_{\hat{\theta}}$ in terms of conditional risk, however, for the equations

$$E_\theta[(\hat{S} - S)^2 | B \leq n] = \mathrm{Var}(B|B \leq n), \tag{8}$$

$$E_\theta[(\hat{S}_{\hat{\theta}} - S)^2 | B \leq n] = \mathrm{Var}(B|B \leq n) + [b \wedge n - \hat{B}(n)]^2 \tag{9}$$

do not depend on θ, and (8) is less than (9). Graphs of the two conditional risks are included in Figure 2 for selected b.

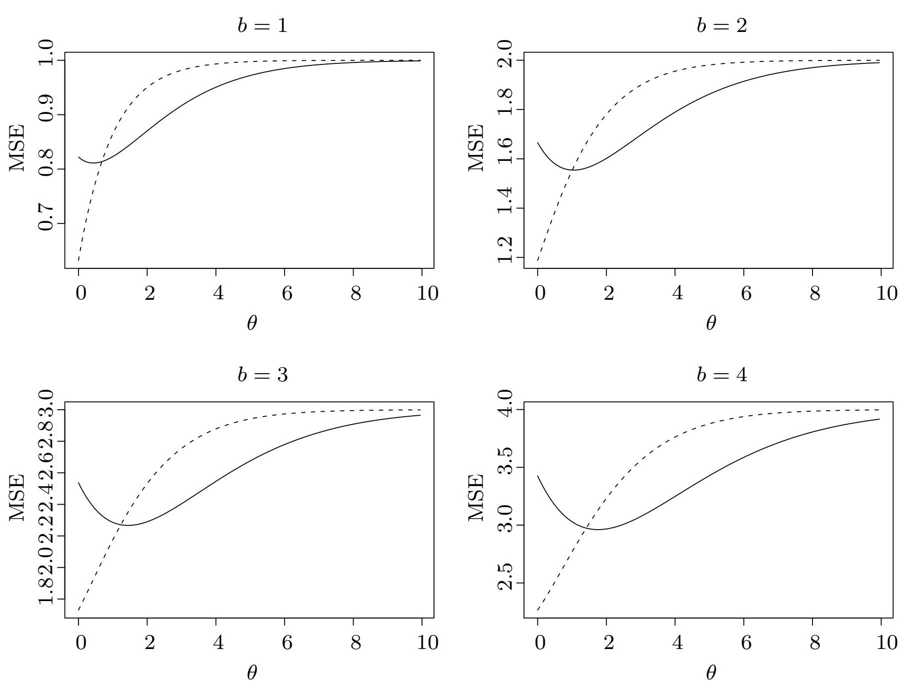

Figure 1. The MSE for $\hat{S}(n)$ (solid) and $\hat{S}_{\hat{\theta}}(n)$ (dotted) for selected b.

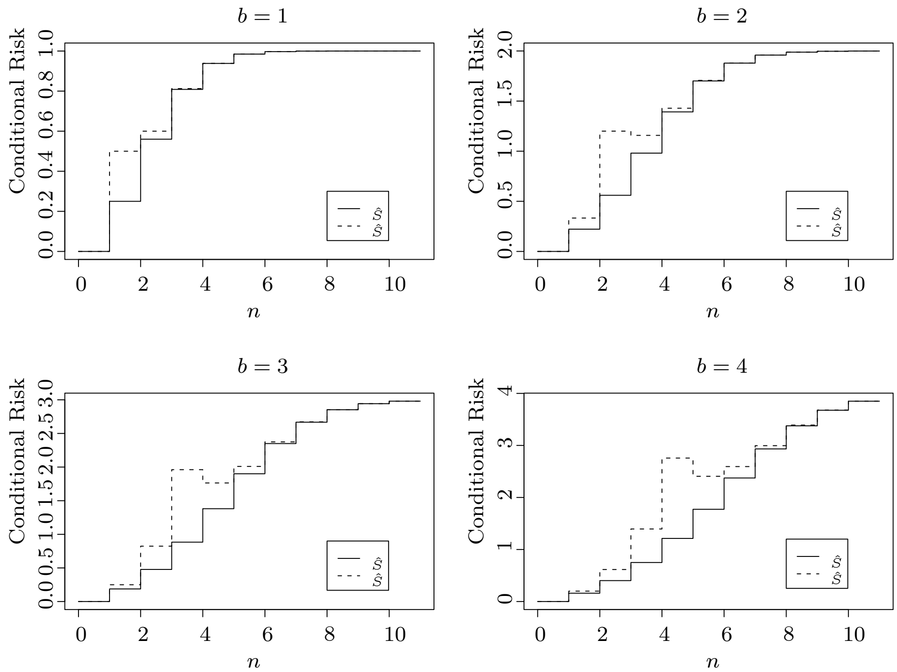

Figure 2. The conditional MSE for $\hat{S}(n)$ (solid) and $\hat{S}_{\hat{\theta}}(n)$ (dashed) for selected b.

4. Minimaxity

Recall that an estimator $\tilde{S}$ is *minimax* if it minimizes $\sup_\theta R(\tilde{S};\theta)$. In showing that $\hat{S}$ is minimax, it is necessary to consider randomized predictors, because the use of a randomized predictor could reduce the minimax risk in principle. Let $I\!N$ denote the non-negative integers and $\mathcal{B}^+$ the Borel sets of $[0,\infty)$. Then a randomized predictor is a function $\gamma : I\!N \times \mathcal{B}^+ \to [0,1]$ for which $\gamma(n,\cdot)$ is a probability measure on $\mathcal{B}^+$ for each n; and the risk of a randomized predictor is

$$\begin{aligned} r(\gamma,\theta) &= E_\theta\{\int_{[0,\infty)}[a-S]^2\gamma(X;da)\} \\ &= E_\theta\{\int_{[0,\infty)}\{[a-\frac{X\theta}{b+\theta}]^2+[S-\frac{X\theta}{b+\theta}]^2\}\gamma(X;da)\}. \end{aligned}$$

A (non-randomized) predictor $\tilde{S}$ then corresponds to $\gamma(n;A) = \mathbf{1}_A[\tilde{S}(n)]$ in which case $r(\gamma,\theta) = R(\tilde{S},\theta)$ in (2).

Recalling the optimal predictor for known θ, $\hat{S}_\theta(n) = n\theta/(b+\theta)$, it is clear that $r(\gamma,\theta) \geq E_\theta[(S-\hat{S}_\theta)^2] = b\theta/(b+\theta)$, so that $\sup_\theta r(\gamma,\theta) \geq b$ for any predictor γ. So, minimaxity of $\hat{S}$ means that $\sup_\theta R(\hat{S},\theta) = b$. Again recalling the optimal predictor for known θ,

$$E_\theta[(S-\hat{S})^2|N=n] = [\hat{B}(n) - \frac{nb}{b+\theta}]^2 + \frac{nb\theta}{(b+\theta)^2}.$$

Then, using $\hat{B}(n) = b(1-\delta_n)$ and some simple algebra (essentially, summation by parts), it follows that $R(\hat{S},\theta) = b + h_b(\theta)$, where

$$h_b(\theta) = b^2 \sum_{n=0}^{\infty} [\delta_n^2 - 2(\delta_n - \delta_{n+1})] f_{b+\theta}(n); \tag{10}$$

and minimaxity of $\hat{S}$ is equivalent to $h_b(\theta) \leq 0$ for all $\theta \geq 0$.

Lemma 1. *δ_n and δ_{n+1}/δ_n are both decreasing in $n \geq 0$.*

Proof. Since $b(1-\delta_n) = E(B|B \leq n)$, it is clear that δ_n is decreasing. So, the issue is δ_{n+1}/δ_n; and since

$$\frac{\delta_{n+1}}{\delta_n} = \frac{F_b(n)}{F_b(n+1)} \frac{b}{n+1},$$

it suffices to show that $F_b(n-1)F_n(n+1) > (n/(n+1))F_b(n)^2$. For this,

$$\begin{aligned}
&(n+1)F_b(n-1)F_b(n+1) - nF_b(n)^2 \\
&= F_b(n)^2 + (n+1)f_b(n+1)F_b(n-1) - (n+1)f_b(n)F_b(n) \\
&= [\sum_{j=0}^{n}\sum_{k=0}^{n} \frac{b^{j+k}}{j!k!} + \frac{b^{n+1}}{n!}\sum_{k=0}^{n-1}\frac{b^k}{k!} - (n+1)\frac{b^n}{n!}\sum_{k=0}^{n}\frac{b^k}{k!}]e^{-2b}.
\end{aligned}$$

Let c_ℓ denote the coefficient of b^ℓ in $[\cdots]$. Then clearly $c_\ell \geq 0$ for $0 \leq \ell < n$; and for $n \leq \ell \leq 2n$,

$$\begin{aligned}
c_\ell &= \sum_{k=\ell-n} \frac{1}{k!(\ell-k)!} + \frac{1}{n!(\ell-n-1)!} - \frac{n+1}{n!(\ell-n)!} \\
&= \sum_{k=\ell-n} \frac{1}{k!(\ell-k)!} - \frac{2n-\ell+1}{n!(\ell-n)!} \\
&\geq 0,
\end{aligned}$$

since $(k!(\ell-k)!)^{-1} \geq (n!(\ell-n)!)^{-1}$ for all $k = \ell-n\cdots,n$ (because $n\times(n-1)\times\cdots\times(k+1) \geq (\ell-k)\times\cdots\times(\ell-n+1)$).

Lemma 2. *As a function of n, $\delta_n^2 + 2\delta_{n+1} - 2\delta_n$ has at most one sign change; and any change of sign is from $+$ to $-$.*

Proof. Clearly, $\delta_n^2 + 2\delta_{n+1} - 2\delta_n < 0$ iff

$$\delta_n \leq 2(1 - \frac{\delta_{n+1}}{\delta_n}). \tag{11}$$

By Lemma 1, the left side of (11) is decreasing, and the right side is increasing. So, if $\delta_n^2 + 2\delta_{n+1} - 2\delta_n < 0$ for some n, then $\delta_n^2 + 2\delta_{n+1} - 2\delta_n < 0$ for all large values of n, and the lemma follows.

Theorem 3. *$\hat{S}$ is minimax.*

Proof. As noted above, it suffices to show that $h_b(\theta) \leq 0$ for all $\theta \geq 0$. By Lemma 2 $\delta_n^2 + 2\delta_{n+1} - 2\delta_n$ has at most one sign change in (10), and any change of sign is from $+$ to $-$. By Theorem 5.1 of Karlin (1966), the same is then true of $h_b(\theta)$, regarded as a function of θ. Clearly $\lim_{n\to\infty} \hat{B}(n) = b$. So, $B - \hat{B}$ converges in distribution to $B - b$ as $\theta \to \infty$. Since $0 \leq B - \hat{B} \leq B$, it then follows that $\lim_{\theta\to\infty} R(\hat{S}, \theta) = \lim_{\theta\to\infty} E[(B - \hat{B})^2] = b$ and, therefore, $\lim_{\theta\to\infty} h_b(\theta) = 0$. It is shown below that $h_b(0) < 0$, and it then follows from the sign regularity of $h_b(\theta)$ that $h_b(\theta) \leq 0$ for all θ. To see that $h_b(0) < 0$, simply observe that

$$\begin{aligned}
\sum_{n=0}^{\infty} \delta_n^2 f_b(n) &= \sum_{n=0}^{\infty} \delta_n^2 [F_b(n) - F_b(n-1)] \\
&= \sum_{n=0}^{\infty} [\delta_n^2 - \delta_{n+1}^2] F_b(n) \\
&= \sum_{n=0}^{\infty} [\delta_n + \delta_{n+1}][\delta_n - \delta_{n+1}] F_b(n) \\
&< \sum_{n=0}^{\infty} 2\delta_n [\delta_n - \delta_{n+1}] F_b(n) \\
&= \sum_{n=0}^{\infty} 2[\delta_n - \delta_{n+1}] f_b(n),
\end{aligned}$$

so that $h_b(0) < 0$ by (10) and Lemma 1.

5. Confidence and Credible Interval for S

The confidence or credible interval $[\ell(n), u(n)]$ for S can be directly obtained through the confidence and the credible interval $[a(n), b(n)]$ for B by subtracting $[a(n), b(n)]$ from n. The confidence interval of B can be calculated from the conditional distribution of B on the event $B \leq n$ when $X = n$ is observed. The conditional mass function of B on $B \leq n$ is

$$f_{b|n}(k) = \frac{P[B = k]}{P[B \leq n]} = \frac{f_b(k)}{F_b(n)} \tag{12}$$

if $0 \leq k \leq n$, and $f_{b|n}(k) = 0$ if $k > n$. Let $F_{b|n}$ be the corresponding conditional cumulative distribution function of B on $B \leq n$. Then, the $1 - \alpha$ confidence interval $[a(n), b(n)]$ for B on $B \leq n$ can be computed by solving

$$\begin{aligned} &[a(n), b(n)] = \{k : f_{b|n}(k) \geq c_n\}, \\ &F_{b|n}[b(n)] - F_{b|n}[a(n)-] \geq 1 - \alpha, \end{aligned} \tag{13}$$

for some $c_n \geq 0$. Therefore, the $1 - \alpha$ confidence interval for S is

$$[\ell(n), u(n)] = [n - b(n), n - a(n)]. \tag{14}$$

Then we have $\sum_{k=\ell(n)}^{u(n)} f_b(n-k)/F_b(n) \geq 1 - \alpha$.

Alternatively under the uniform prior on $[0, \infty)$ for θ, the Bayesian credible interval for S can be computed from the posterior conditional Bayesian mass function of S on $X = n$, which is

$$P^0[S = k|X = n] = \frac{P^0[B = n - k, X = n]}{P^0[X = n]}, \tag{15}$$

where $P^0[B = n - k, X = n]$ and $P^0[X = n]$ are given in (5). It is clear that the Bayesian conditional mass function of B given $X = n$ under the uniform prior on $[0, \infty)$ for θ is exactly the same as the conditional mass function of B given in (12). Therefore, the Bayesian credible interval and the conditional confidence interval for S are identical if the prior is uniformly distributed on $[0, \infty)$. Figure 3 displays the numerical results of estimated value $\hat{S}(n)$, the 90% confidence bounds $\ell(n)$ and $u(n)$ as functions of n for selected b.

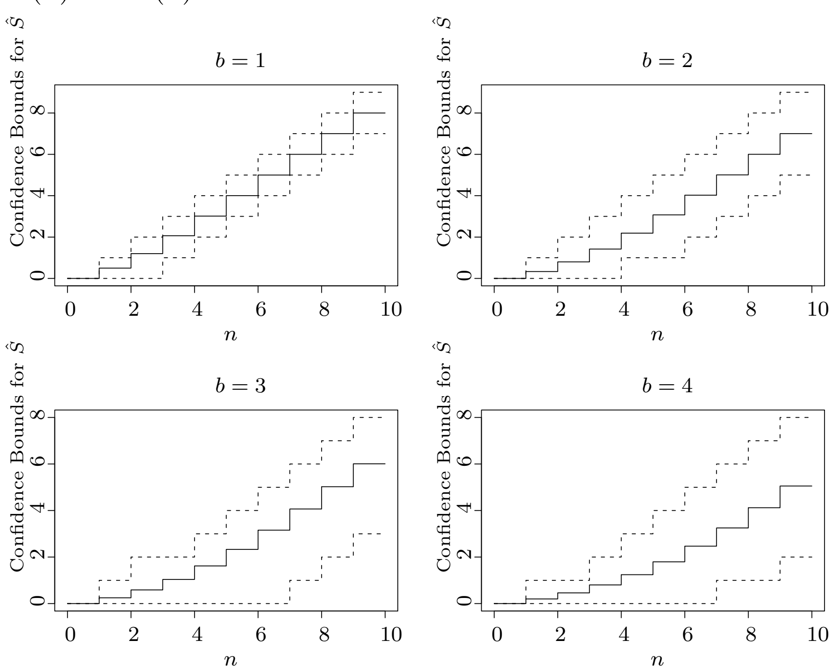

Figure 3. The estimated values (solid) and their 90% confidence intervals (dashed) for S as functions of n for selected b.

The frequentist coverage probability of the confidence interval for S is the probability of the interval $[\ell(n), u(n)]$ in (14) to contain S, i.e.,

$$P_\theta[\ell(n) \le S \le u(n)] = \sum_{n=0}^{\infty} \sum_{k=\ell(n)}^{u(n)} \frac{\theta^k e^{-\theta}}{k!} \frac{b^{n-k} e^{-b}}{(n-k)!}. \tag{16}$$

The numerical results of the frequentist coverage probabilities of the 90% confidence interval for S as functions of θ for selected b are given in Figure 4. As suggested by Figure 4, the coverage probabilities in (16) are continuous in θ but discontinuous in b. This is an effect of the discreteness of the Poisson, since $\ell(n)$ and $u(n)$ depend on b.

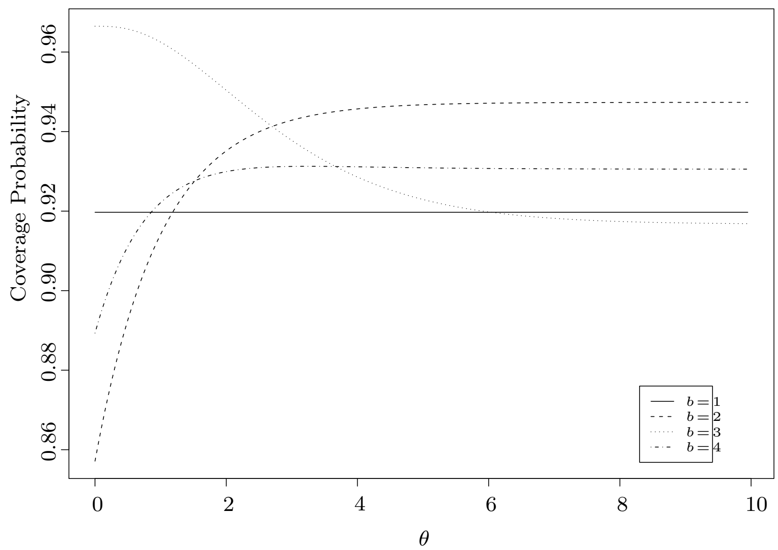

Figure 4. The coverage probabilities of the 90% confidence intervals for S as functions of θ for selected b.

6. Remarks and References

Viewed as estimators of θ, neither $\hat{S}$ nor $\hat{S}_{\hat{\theta}}$ is admissible for squared error loss for essentially the same reasons in the proof of Theorem 1. However, interpreting $0/0$ as 0, $\hat{p} = \hat{S}/X$ is an admissible estimator of $p = \theta/(b+\theta)$ for the loss function $L(p, \hat{p}, x) = (\hat{p} - p)^2 x^2$, which depends on x as well as p and $\hat{p}$, and $\hat{S}_{\hat{\theta}}$ is inadmissible. To see this simply observe that $R(\tilde{S}; \theta) = E_\theta[(\tilde{S} - S_\theta)^2] + b\theta/(b+\theta)$ for any estimator $\tilde{S}$, that the second term on the right does not depend on $\tilde{S}$, and that $E_\theta[(\tilde{S} - S_\theta)^2] = E_\theta[((\tilde{S}/X) - p)^2 X^2]$.

For large b, the prediction problem is closely related to the problem of estimating a positive normal mean. To see how, let $Y = 2[\sqrt{X} - \sqrt{b}]$ and

$\mu = 2[\sqrt{b+\theta} - \sqrt{b}]$. If $b,\ \theta \to \infty$ in such manner that μ remains fixed, then $Y \Rightarrow$ Normal$[\mu, 1]$ and $S/\sqrt{b} \to^p \mu$, where $\Rightarrow$ and $\to^p$ denote convergence in distribution and convergence in probability. So, the asymptotic version of the problem is to estimate μ when $Y = \mu + \epsilon$, $\mu \geq 0$, and $\epsilon \sim$ Normal$[0, 1]$. Like B, ϵ is an ancillary variable in that its distribution is known; and if $Y = y$ is observed, then $\epsilon \leq y$. It is known that the maximum likelihood estimator $\hat{\mu} = \max[0, Y]$ is inadmissible and that the formal Bayes estimator with respect to a uniform prior is admissible (Katz (1961)). Moreover, the latter can be written as $y - E(\epsilon|\epsilon \leq y)$. Shao and Strawderman (1996) provide a recent contribution to the problem of estimating a positive normal mean.

The derivations in Sections 3 and 4 use several properties of the Poisson distribution: non-negativity, reproductivity, and the exponential family structure. An interesting potential generalization is to marked Poisson processes where a mark is observed with each event and the marks have different distributions for background and signal events.

Bayesian calculations of the nature used here have appeared in the physics literature in Helene (1984) and Zech (1989), for example. Conditioning on a partially observed ancillary variable $B \leq n$ was first suggested by Roe and Woodroofe (1999), who derived regions of high conditional likelihood for θ. Cousins (2000) showed that this approach leads to undesirable intervals when applied to the positive normal mean problem. Other uses of conditioning appear in Roe and Woodroofe (2000), which shows that Bayesian credible intervals from flat priors have nearly exact conditional frequentist coverage probabilities, and in Woodroofe and Wang (2000) which shows that an unconditional p-value is inadmissible but its conditional analogue is admissible.

7. More About Stein's Condition

The applicability of Stein's Theorem is outlined in this appendix. In the present context, the sets A and B of Stein (1955) are, respectively, the parameter space $[0, \infty)$ and the set of randomized predictors for which $\gamma(n, [0, n]) = 1$ for all n; the function K of Stein (1955) is $K(a, b) = r(\gamma, \theta)$. Two conditions are imposed in Stein (1955): compactness of B in the sense of Wald, and the applicability of the Minimax Theorem. Letting d denote the Levy metric for probability measures on $[0, \infty)$, as in Billingsley (1995, p.198), for example, the set B is compact in the topology generated by the metric $d^*[\beta, \gamma] = \sum_{n=0}^{\infty} 2^{-n} d[\beta(n, \cdot), \gamma(n, \cdot)]$. For if γ_k is any sequence in B, then there are subsequences along which $\gamma_k(n, \cdot)$ converge weakly for each n. By diagnolization, there is then a subsequence along which $\gamma_k(n, \cdot)$ converges weakly for all n, and this is equivalent to convergence in the metric d^*. When combined with the (easily verified) lower semi-continuity of $r(\gamma, \theta)$ in γ for fixed θ, compactness implies compactness in the sense of Wald;

when combined with the obvious continuity of $r(\gamma,\theta)$ in θ for each $\gamma \in B$, compactness also implies the Minimax Theorem. See Ferguson (1965, pp.81ff), for example.

References

Billingsley, P. (1995). *Probability and Measure*. John Wiley, New York.

Cousins, R. (2000). Comment on "Improved probability method for estimating a signal in the presence of background, by B. P. Roe and Michael Woodroofe." *Phys. Rev. D* **62**, 098301.

Eitel, K., et al. (2000). Update of the KARMEN2 $\nu_\mu \to \nu_e$ oscillation search. In *Proc. 14th Lake Louise Winter Institute Electroweak Physics*, 353-360. World Scientific, Singapore.

Ferguson, T. (1965). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.

Geisser, S. (1993). *Predictive Inference: An Introduction*. Chapman Hall, New York.

Helene, O. (1984). Errors in experiments with small numbers of events. *Nuclear Instr. and Meth.* **A228**, 120-128

Irwin, M. and Hatzidimitriou, D. (1995). Structural parameters for the Galactic dwarf spheroidals. *Monthly Notices Roy. Astronom. Soc.* **277**, 1354-1378.

Karlin, S. (1966). *Total Positivity*. Addison Wesley, Boston.

Katz, M. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Ann. Math. Statist.* **32**, 136-142.

Mandelkern, M. (2002). Setting confidence intervals for bounded parameters (with discussions). *Statist. Sci.* **17**, 149-172.

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

Roe, B. and Woodroofe, M. (2000). Setting confidence belts. *Phys. Rev. D* **63**, 013009

Shao, P. Y. and Strawderman, W. E. (1996). Improving of the MLE of a positive normal mean. *Statist. Sinica* **6**, 259-274.

Stein, C. (1955). A Necessary and Sufficient Condition for Admissibility. *Ann. Math. Statist.* **26**, 518-522.

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

Zech, G. (1989). The upper limit in experiments with background or measurement errors. *Nuc. Instr. & Meth.* **A277**, 608-610.

Department of Statistics, Purdue University, 150 N. University Street, West Lafayette, IN 47907-2067, U.S.A.

E-mail: tlzhang@stat.purdue.edu

Department of Statistics, University of Michgan, 439 West Hall, 550 East University, Ann Arbor, MI 48109-1092, U.S.A.

E-mail: michaelw@umich.edu

(Received January 2003; accepted April 2004)