Signals and Background
Statistics 710
September 5 - 14, 2006

Prologue:*An Identity.* Let q_λ and Q_λ denote the Poisson mass function and distribution function with parameter λ,

$$q_\lambda(n) = \frac{\lambda^n e^{-\lambda}}{n!} \qquad \text{and} \qquad Q_\lambda(n) = \sum_{k=0}^{n} q_\lambda(k) \tag{1}$$

and let h_n and H_n denote the $\Gamma(n,1)$ density and distribution function

$$h_n(y) = \frac{y^{n-1}e^{-y}}{(n-1)!} \qquad \text{and} \qquad H_n(y) = \int_0^y \mathtt{h}_n(x)dx.$$

Then

$$Q_\lambda(n) = 1 - H_{n+1}(\lambda) = P[\chi^2_{2n+2} \le 2\lambda];$$

that is

$$\sum_{k=0}^{n} \frac{\lambda^k e^{-\lambda}}{k!} = \int_\lambda^\infty \frac{s^n e^{-s}}{n!} ds. \tag{2}$$

Equation (2) can be verified by noting that both sides approach 0 as $\lambda \to \infty$ and then showing that they have the same derivatives (There is some cancellation of the left). In particular, it follows from (2) that $Q_\lambda(n)$ is decreasing in λ for all n.

The Problem: Let N_b and N_s be independent Poisson variables with means b and s, where b is known but s is not and suppose that (only)

$$N = N_b + N_s \sim \text{Poisson}(b+s)$$

is observed. The goal is to set confidence intervals, or upper confidence bounds for s.

Frequentist Solutions: *UMA Bounds*: Uniformly most accurate confidence bounds may be found by inverting tests of $H_0 : s \ge t$. Ignoring randomization, the UMP test of this hypothesis rejects when $N \le n_t$, where n_t is the greatest integer n for which

$$Q_{b+t}(n) = P_t\left[N \le n\right] \le \alpha$$

and $1-\alpha$ is the desired confidence coefficient. So, if $N = n$, then the confidence set is $\mathcal{C}_n = \{t : n_t > n\}$. To understand the nature of the n_t, let λ_n solve the equation

$$Q_\lambda(n) = \alpha.$$

Then $0 < \lambda_0 < \lambda_1 < \cdots < \lambda_n \to \infty$. Then $s \in \mathcal{C}_n$ iff $Q_{b+s}(n) > \alpha$ iff $s \leq t_n := \lambda_n - b$. Observe that t_n can be negative if n is small and b is large.

The Unified Method: Now consider testing $H_0 : s = t$, using a likeihood ratio test. The likelihood function, MLE, and likelihood ratio test statistics

$$L(s|n) = \frac{1}{n!}(b+s)^n e^{-(b+s)},$$

$$\hat{s} = \max[0, n-b],$$

and

$$\Lambda_t(n) = \frac{L(t|n)}{L(\hat{s}|n)}.$$

Let c_t be the largest value of c for which

$$P_t\left[\Lambda_t(N) \geq c\right] \geq 1-\alpha.$$

The the unified confidence sets are $\mathcal{C} = \{s : \Lambda_s(n) \geq c_s\}$. These is just the usual likelihood ratio intervals, but without the chi-squared approximation. The c_t have to computed numerically. This was done by Feldman and Cousins, [3].

Example 1 : *The KARMEN Data:* In a study of nuetrino oscillations, a group at the Las Alamos claimed to have detected a poitive signal (which showed that the nuetrino has mass). In a followup study the KARMEN Group found $N = 0$ with $b = 3$ and a unified confidence interval of $0 \leq s \leq 1.??$. The Las Alamos data suggested a signal in the range $1 \leq s \leq 1.5$, and the KARMEN Group came perilously close to saying that they had disproved the Las Alamos results. ♢

Dependence on b when $N = 0$. For the KARMEN Data, consider testing the hpothesis $H_0 : s \geq 1$, and let N^* be an independent copy of N, representing a hypothetical replication of the experiment. When $N = 0$ and $b = 3$, the p-value is

$$p = P_1[N^* = 0] = e^{-(b+1)} = e^{-4} = .01833\cdots,$$

which is significant at the usual levels. However, if $N = 0$, then necessarily $N_b = 0$ and $N_s = 0$. So, the background count was lower than expected. It seems unfair (and also not very wise) to regard a low background count as evidence against $H_0 : s \geq 1$. To go further observe that if N_b and N_s were both observed N_b would be an ancillary statistic and N_s a complete sufficient statistic, and N_b and N_s are observed when $N = 0$. So, when $N = 0$,

$$P_1[N^* = 0 | N_b^* = 0] = e^{-1} = .3679\cdots$$

seems a more reasonable p value.

Litmus Tests. The simple problem is made much more interesting and difficult by imposing:

(L1) The solution should not depend on b when $N = 0$.

(L2) The solution should be equivariant under monotone transformations of s.

As noted above, if $N = 0$, then $N_b = 0$ and $N_s = 0$. Once this is known the apriori expectation of N_b cannot be relevant. It is implicit in (L1) that there should be limited dependence on b when $N = 1$, say. The consensus among the physicists that I know is that (L2) is necessary. The UMA and Unified Intervals satisfy (L2), but not (L1). Bayesian solutions will satisfy (L1).

A Bayesian Solution. The Bayesian solution with a uniform prior has several things to recommend it. If s has a uniform prior, then the posterior density is

$$g(s|n) = \frac{1}{c_n n!}(b+s)^n e^{-(b+s)} \tag{3}$$

where

$$c_n = \int_0^\infty \frac{1}{n!}(b+s)^n e^{-(b+s)} ds = \int_b^\infty \frac{t^n e^{-t}}{n!} dt.$$

Using the identity (1),

$$c_n = \sum_{k=0}^{n} \frac{b^k e^{-b}}{k!} = P[N_b \le n]$$

(The distribution of N_b does not depend on s). Let P^n denote posterior probability when $N = n$, and let $G(\cdot|n)$ denote the posterior distribution function. Then

$$1 - G(t|n) = P^n[s > t] = \frac{\int_t^\infty (s+b)^n e^{-(s+b)} ds/n!}{\int_0^\infty (s+b)^n e^{-(s+b)} ds/n!} = \frac{\int_{b+t}^\infty s^n e^{-s} ds}{\int_b^\infty s^n e^{-s} ds},$$

which can be recognized as

$$1 - G(t|n) = \frac{P_t[N \le n]}{P[N_b \le n]} = P_t[N \le n | N_b \le n].$$

A level $1-\alpha$ upper credible limit, u_n say, is determined by the condition that $P^n[s > u_n] = \alpha$; that is,

$$\alpha = 1 - G(u_n|n) = P_{u_n}[N \le n | N_b \le n] \tag{4}$$

or equivalently, $Q_{b+u_n}(n) = \alpha Q_b(n)$.

The Frequentist Coverage Probabilities. It follows from (4) that the frequentist coverage probability of the Bayesian intervals is at least $1 - \alpha$,

$$P_s\left[s \le u_N\right] > 1 - \alpha \tag{5}$$

for all $0 \le s < \infty$. The proof of (5) is quite simple, given the following intuitive facts: $1 - G(t|n) = P^n[s > t]$ is increasing in n for fixed t and $0 < u_0 < u_1 < \cdots < u_n \to \infty$. Assuming this for the moment, $P_s[s \le u_N] = 1$ for all $s \le u_0$. For a fixed $s > u_0$, let $m = m_s$ be largest value of n for which $u_n < s$, so that $s > u_N$ iff $N \le m$. Thus

$$P_s[s > u_N] = P_s[N \le m_s] \le P_{u_m}[N \le m] < P_{u_m}[N \le m | N_b \le m] = \alpha$$

Conditional Frequentist Intervals. The Bayesian intervals are conditional frequentist intervals in the following sense: Suppose that $N = n$ is observed and let N^* denote an independent copy of N. If $0 < s \le u_n$, then there is a $k \le n - 1$ for which $u_k < s \le u_{k+1}$, so that

$$P_s[s > u_{N^*} | N_b^* \le n] = P_s[N^* \ge k | N_b^* \le n] \le P_{u_k}[N^* \ge k | N_b^* \le n]$$

and, therefore,

$$P_s[s > u_{N^*} | N_b^* \le n] \le P_{u_k}[N^* \ge k | N_b^* \le k] = \alpha \tag{6}$$

The uniform prior is not invariant under transformations of s, but invariance of the interval can be obtained by simply forgetting the derivation and keeping the result. As just explained, the solutions to (6) may reasonably be called *conditional frequentist* upper limits. With this definition, the conditional frequentist limits satisfy both litmus tests: the construction is equivariant under increasing transformations and does not depend on b when $N = 0$. The conditional frequentist intervals are conservative for small n when compared to the frequentist solutions. A conservative solution seems necessary in order to avoid dependence on b when $N = 0$ and the embarassing possibility of a degenerate interval.

Problem 1 *With $\alpha = .1$ and $b = 3$, write computer code to compute u_n for $n = 0, \cdots, 25$, and display the results in a graph. Then compute the left side of (5) for $s = 0\ (0.1)\ 5$, and graph the result. Comment on any peculiar or interesting aspects of the graphs.*

An Inequality: *If X is any random variable and u and v are two non-decreasing functions for which $u(X)$ and $v(X)$ have finite expections, then*

$$E[u(X)v(X)] \ge E[u(X)] \times E[v(X)], \tag{7}$$

Let $\mu = E[u(X)]$ and observe that $u(x) \geq \mu$ for all sufficiently large x. If $u(x) \geq \mu$ for all x, then $P[u(X) = \mu] = 1$, and (7) is clear. Otherwise, let $x_0 = \inf\{x \in I\!\!R : u(x) \geq \mu\}$. Then $[u(x) - \mu][v(x) - v(x_0)] \geq 0$ for all x. It follows that $E\{[u(X) - \mu][v(X) - v(x_0)]\} \geq 0$ and, therefore,

$$E[u(X)v(X)] \geq \mu E[v(X)] + v(x_0)E[u(X)] - \mu v(x_0) = E[u(X)] \times E[v(X)], \tag{8}$$

as asserted. *If X is non-degenerate and v is strictly increasing, then there is strict inequality in (7) unless $u(X) = \mu$ w.p.1.* For if it is not the case that $u(X) = \mu$ *w.p.*1, then $u(X) < \mu$ and $u(X) > \mu$ with positive probability; and then $[u(X) - \mu][v(X) - v(x_0)] > 0$ with positive probability, so that $\{E[u(X) - \mu][v(X) - v(x_0)]\} > 0$ and there is strict inequality in (8).

Total Positivity. Let $\mathcal{X} \subseteq I\!\!R$ and $\mathcal{Y} \subseteq I\!\!R$ be Borel sets and let $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ be a non-negative Borel measurable function defined on $\mathcal{X} \times \mathcal{Y}$. Then f is said to be *totally positive of order two (*TP_2*)* iff

$$f(x_1, y_1)f(x_2, y_2) \geq f(x_1, y_2)f(x_2, y_1) \tag{9}$$

whenver $x_1 < x_2$ and $y_1 < y_2$. Observe that if $f(x, y) > 0$ for all x and y, then (9) is equivalent to

$$\frac{f(x_2, y_2)}{f(x_1, y_2)} \geq \frac{f(x_2, y_1)}{f(x_1, y_1)};$$

that is, the ratio $f(x_2, y)/f(x_1, y)$ must be non-decreasing whenever $x_1 < x_2$. If f has the form $f(\theta, y) = p_\theta(y)$, where p_θ is a family of probability densities, then the latter property is called *monotone likelihood ratio.*

Here are two simple examples.

Example 2 *a)* If $\mathcal{X} = \mathcal{Y} = I\!\!R$, then $f(x, y) = e^{xy}$ is strictly TP_2; for in this case, the difference between the left side of (9) and the right is

$$e^{-(x_2 y_1 + x_1 y_2)} \left[e^{(x_2 - x_1)(y_2 - y_1)} - 1\right],$$

which is positive when $x_1 < x_2$ and $y_1 < y_2$.

b) If $\mathcal{X} = \mathcal{Y} = (0, \infty)$ then $f(x, y) = \mathbf{1}\{y \leq x\}$, is TP_2*, as may be seen by considering cases.* ♢

These examples may appear in slightly disguised form. First, if $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ is TP_2, $\mathcal{X}_0 \subseteq \mathcal{X}$, and $\mathcal{Y}_0 \subseteq \mathcal{Y}$, then the restriction of f to $\mathcal{X}_0 \times \mathcal{Y}_0$, is again TP_2. Next, if $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ is TP_2, $g : \mathcal{X} \to [0, \infty)$, and $h : \mathcal{Y} \to [0, \infty)$, then $\tilde{f}(x, y) = g(x)f(x, y)h(y)$ is again TP_2. Finally, if $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ is TP_2, $g : \tilde{\mathcal{X}} \to \mathcal{X}$, and $h : \tilde{\mathcal{Y}} \to \mathcal{Y}$, then

$\tilde{f}(x,y) := f[g(x),h(y)]$ is again TP_2. For example, the Poisson probability mass function $f_\lambda(n)$ in (1) is TP_2, as is the posterior density $g(s|n)$ of s in (3).

Non-decreasing functions. Now let $f : \mathcal{X} \times \mathcal{Y} \to (0,\infty)$ be TP_2, let ν be a sigma finite measure on $\mathcal{Y}$, and suppose that

$$\int_{\mathcal{Y}} f(x,y)d\nu(y) \equiv 1.$$

If v is a non-decreasing function for which

$$w(x) = \int_{\mathcal{Y}} v(y)f(x,y)d\nu(y)$$

is finite for all $x \in \mathcal{X}$, then w is a non-decreasing function. For, if $x_1 < x_2$, then

$$\begin{aligned} w(x_2) - w(x_1) &= \int_{\mathcal{Y}} v(y)\left[\frac{f(x_2,y)}{f(x_1,y)} - 1\right] f(x_1,y)d\nu(y) \\ &= \int_{\mathcal{Y}} [v(y) - w(x_1)]\left[\frac{f(x_2,y)}{f(x_1,y)} - 1\right] f(x_1,y)d\nu(y) \\ &\geq 0 \end{aligned}$$

by (7).

Back to Signals and Noise. Let G denote the posterior distribution function of s given $N = n$,

$$G(s|n) = \int_0^s g(t|n)dt.$$

Then clearly $G(s|n)$ is increasing s for each n. Next, I claim that $1 - G(s|n)$ is increasing in n for fixed s. For

$$1 - G(s|n) = \int_0^\infty \mathbf{1}_{(s,\infty)}g(t|n)dt,$$

$g(s|n)$ is TP_2, and $\int_0^\infty g(t|n)dt \equiv 1$. It follows easily that $0 < u_0 < u_1 < \cdots < u_n \to \infty$. For $1 - G(u_n|n) = \alpha$, so that $\alpha = 1 - G(u_n|n) < 1 - G(u_n|n+1)$ and, therefore, $u_{n+1} > u_n$.

Problem 2 *Show that if $\mathcal{X}$ and $\mathcal{Y}$ are open intervals and f is positive and continuously differentiable, then f is* TP_2 *iff*

$$\frac{\partial^2 \log f(x,y)}{\partial x \partial y} \geq 0.$$

Marked Poisson Variables. Now let $N \sim \mathrm{Poisson}(b+s)$, as above, and let $(J_1, X_1), (J_2, X_2), \cdots$ be i.i.d. random vectors for which

$$P[J_i = 1] = \frac{s}{b+s} = 1 - P[J_i = 0]$$

$$X_i|J_i = 0 \sim f_b \qquad \text{and} \qquad X_i|J_i = 1 \sim f_s,$$

where b, f_b, and f_s are assumed known, but s is unknown. Then

$$X_i \sim f = \frac{bf_b + sf_s}{b+s}.$$

Problem 3 *Show that $N_s = J_1 + \cdots + J_N$ and $N_b = N - N_s$ are independent Poisson variables with mean s and b.*

The Liikelihood Function and MLE. Suppose that we observe N and $X_1, \cdots, X_N$. Then the likelihood function, log likelihood function, and score function is

$$L(s|n, \mathbf{x}) = \frac{e^{-(b+s)}}{n!} \prod_{i=1}^{n} [bf_b(x_i) + sf_s(x_i)],$$

$$\ell(s|n, \mathbf{x}) = \sum_{i=1}^{n} \log[bf_b(x_i) + sf_s(x_i)] - (b+s) - \log(n!),$$

and

$$\ell'(s|n, \mathbf{x}) = \sum_{i=1}^{n} \frac{f_s(x_i)}{bf_b(x_i) + sf_s(x_i)} - 1.$$

Observe that $\ell'(s|n, \mathbf{x})$ is decreasing.

Since $\ell'(s|n, \mathbf{x})$ is decreasing, the maximum likelihood estimator can be found by a bisection algorithm. It may also be found by a simple application of the EM Algorithm. If $J_1, \cdots, J_N$ were also observed, then the likelihood function, log-likelihood function, and score function, and MLE would be

$$\tilde{L}(s|n, \mathbf{x}, \mathbf{j}) = \frac{e^{-(b+s)}}{n!} \prod_{i=1}^{n} [bf_b(x_i)]^{1-j_i} [sf_s(x_i)]^{j_i},$$

$$\begin{aligned}
\tilde{\ell}(s|n, \mathbf{x}, \mathbf{j}) &= -(b+s) + \sum_{i=1}^{n} \{(1-j_i)\log[bf_b(x_i)] + j_i \log[sf_s(x_i)]\} - \log(n!) \\
&= -s + (j_1 + \cdots + j_n)\log(s) + C,
\end{aligned}$$

$$\frac{\partial \tilde{\ell}(s|n, \mathbf{x}, \mathbf{j})}{\partial s} = -1 + \frac{j_1 + \cdots + j_n}{s},$$

where C does not depend on s, and

$$\tilde{s} = j_1 + \cdots + j_n.$$

Moreover

$$E_s(J_i|n, \mathbf{x}) = \frac{sf_s(x_i)}{bf_b(x_i) + sf_s(x_i)},$$

and the EM Algorithm becomes: Starting with an initial guess $\hat{s}^0$, let

$$\hat{s}^{k+1} = \sum_{i=1}^{n} \frac{\hat{s}^k f_s(x_i)}{bf_b(x_i) + \hat{s}^k f_s(x_i)}$$

for $k = 0, 1, 2, \cdots$.

Problem 4 *Show that if $0 < \hat{s}^0 \le n$, then $\hat{s} := \lim_{k\to\infty} \hat{s}^k$ is the MLE. Then compute the MLE when f_b is Uniform on $[-1, 1]$, $f_s(x) = 1 - |x|$ for $|x| \le 1$, $n = 13$, and* $\mathbf{x} = \{\pm.8, \pm.6, \pm.4, \pm.3, \pm.2, \pm.1, 0\}$.

Observe that $E_s(J_i|n, \mathbf{x})$ is the conditional probability that the ith event is a signal, given n and $\mathbf{x}$, and that this is estimated by $\hat{s}f_s(x_i)/[bf_b(x_i) + \hat{s}f_s(x_i)]$.

Regions of High Likelihood. In principle, one can compute

$$\Lambda_s = \Lambda_s(n, \mathbf{x}) = \frac{L(s|n, \mathbf{x})}{L(\hat{s}|n, \mathbf{x})}$$

and c_s, the largest values of c for which

$$P_s\left[\Lambda_s(N, \mathbf{X}) \ge c\right] \ge 1 - \alpha$$

for each s. Then

$$\mathcal{C}_{n,\mathbf{x}} = \{s : \Lambda_s \ge c_s\}$$

is a level $1 - \alpha$ confidence set for s. In practice, c_s may be estimated by simulation for a grid of s.

Bayesian Analysis of the Marked Poisson Model. First observe that

$$L(s|n, \mathbf{x}) = Ke^{-s}\prod_{i=1}^{n}[1 + \frac{s}{b}r(x_i)].$$

where

$$K = \frac{b^n e^{-b}}{n!}\prod_{i=1}^{n} f_b(x_i) = L(0|n, \mathbf{x})$$

and

$$r(x) = \frac{f_s(x)}{f_b(x)}.$$

If s has prior density g, say, then

$$\begin{aligned}\int_0^\infty L(s|n, \mathbf{x})g(s)ds &= K\int_0^\infty \prod_{i=1}^{n}[1 + \frac{s}{b}r(x_i)]e^{-s}g(s)ds \\ &= K\sum_{k=0}^{n}\frac{C_{n,k}}{b^k}\mu_k \\ &= K\bar{L}_g(n, \mathbf{x}), \text{ say}\end{aligned}$$

where

$$C_{n,k} = \sum_{j_1+\cdots+j_n=k} \prod_{i=1}^{n} r(x_i)^{j_i}$$

and

$$\mu_k = \int_0^\infty s^k e^{-s} g(s) ds.$$

Here g can be an improper prior, provided that μ_k is finite for all k. For example if $g(s) = 1$, then $\mu_k = k!$. The posterior density of s is then

$$g(s|n, \mathbf{x}) = \frac{1}{\bar{L}_g(n, \mathbf{x})} \prod_{i=1}^{n} [1 + \frac{s}{b} r(x_i)] e^{-s} g(s)$$

Bayesian Credible Intervals. If $s \sim$ Uniform, then

$$\int_t^\infty L(s|n, \mathbf{x}) ds = K \sum_{k=0}^{n} \frac{C_{n,k}}{b^k} \int_t^\infty s^k e^{-s} ds = K \sum_{k=0}^{n} k! \frac{C_{n,k}}{b^k} Q_t(k).$$

Let $G(\cdot|n, \mathbf{x})$ denote the posterior distribution function of s. Then

$$1 - G(s|n, \mathbf{x}) = \frac{\sum_{k=0}^{n} k! C_{n,k} b^{-k} Q_s(k)}{\sum_{k=0}^{n} k! C_{n,k} b^{-k}}.$$

As above, upper Bayesian credible limits are determined by

$$1 - G(u_{n,\mathbf{x}}|n, \mathbf{x}) = P[s > u_{n,\mathbf{x}}|n, \mathbf{x}] = \alpha.$$

An efficient algorithm for computing the $C_{n,k}$ and $\bar{C}_{n,k} := C_{nk}/\binom{n}{k}$ are

$$C_{n,k} = C_{n-1,k} + C_{n-1,k-1} r(x_n).$$

and

$$\bar{C}_{n,k} = \frac{(n-1-k)\bar{C}_{n-1,k} + k\bar{C}_{n-1,k-1}}{n-1}.$$

The Discovery Problem. The discovery problem is to determine whether $s > 0$. This is sometimes called looking for a needle in a haystack, because the signal is small compared to the background. Moreover, an extremely high degree of confidence is required for claiming a discovery, significance at the 5σ level, roughly $\alpha = 10^{-6}$.

The (Convential) Bayesian View. The conventional way to formulate this question is as a testing problem $H_0 : s = 0$. Letting G denote the prior distribution function, π_0 be the prior probability that $s = 0$, and $\tilde{g}$ the conditional density of s given $s > 0$,

$$\begin{aligned}\int_0^\infty L(s|n, \mathbf{x}) G\{ds\} &= \pi_0 L(0|n, \mathbf{x}) + (1 - \pi_0) \int_0^\infty L(s|n, \mathbf{x}) \tilde{g}(s) ds \\ &= K\pi_0 + K(1 - \pi_0) \bar{L}_{\tilde{g}}(n, \mathbf{x}).\end{aligned}$$

The posterior probability that $s = 0$ is

$$\pi^* = \frac{\pi_0}{\pi_0 + (1-\pi_0)\bar{L}_{\tilde{g}}(n, \mathbf{x})},$$

and the posterior odds are

$$\frac{\pi_0^*}{1-\pi_0^*} = \frac{\pi_0}{1-\pi_0} \frac{1}{\bar{L}_{\tilde{g}}(n, \mathbf{x})}.$$

Unfortunately, this depends crucially on π_0. The (so called) Bayes Factor $1/\bar{L}_{\tilde{g}}(n, \mathbf{x})$ represents the amount of change in the odds. It does not depend on π_0, but only on $\tilde{g}$.

Alternative (Bayesian) Formulation. Another way to formulate the question is to ask is $N_s > 0$; that is, have we seen a signal event yet. The probability of $N_s = 0$ given the data may be computed as

$$\frac{\int_0^\infty e^{-s} dG(s)}{\sum_{k=0}^n C_{n,k} b^{-k} \int_0^\infty s^k e^{-s} dG(s) ds},$$

or equivalently,

$$\frac{\pi_0 + (1-\pi_0) \int_0^\infty e^{-s} \tilde{g}(s) ds}{\pi_0 + (1-\pi_0) \sum_{k=0}^n C_{n,k} b^{-k} \int_0^\infty s^k e^{-s} \tilde{g}(s) ds}, \tag{10}$$

and the optimal Bayesian decision procedure is again to decide that $N_s > 0$ if and only if this posterior probability is sufficiently small. This appears to depend much less crucially on π_0. It is possible to have large value of (10) even if $\pi_0 = 0$. For example, if $\pi_0 = 0$ and $\tilde{g}$ is the uniform density, then(10) is$1/[\sum_{k=0}^n k! C_{n,k} b^{-k}]$.

An Inequality. Let $\mathcal{G}$ denote the class of (proper) prior distributions for which $\tilde{g}$ is a decreasing function. Then it seem reasonable to suppose that $g \in \mathcal{G}$, since the signal is small if it exists. We will find a lower bound for (10) if $G \in \mathcal{G}$. First observe that the right side of (10) is an increasing function of π_0 for fixed $n, \mathbf{x}$, and $\tilde{g}$. To see this observe that we may write (10) as $[\pi_0 + (1-\pi_0)A]/[\pi_0 + (1-\pi_0)B]$, where $A < B$ and $A \le 1$. Then the derivative of (10) with respect to π_0 is

$$\frac{1-A}{\pi_0 + (1-\pi_0)A} - \frac{1-B}{\pi_0 + (1-\pi_0)B} \ge \frac{1-A}{\pi_0 + (1-\pi_0)B} - \frac{1-B}{\pi_0 + (1-\pi_0)B} \ge 0.$$

So, the infimum of π_0 is attained when $\pi_0 = 0$, in which case (10) becomes

$$\frac{\int_0^\infty e^{-s} \tilde{g}(s)}{\sum_{k=0}^n C_{n,k} b^{-k} \int_0^\infty s^k e^{-s} \tilde{g}(s) ds}.$$

Next, if $\tilde{g}$ is decreasing, then

$$\int_0^\infty s^k \tilde{g}(s) e^{-s} ds \le \int_0^\infty s^k e^{-s} ds \times \int_0^\infty e^{-s} \tilde{g}(s) ds,$$

by the correlation inequality, and the asertion follows directly. So, from a Bayesian veiwpoint,and *a necessary condition for claiming* $N_s > 0$ is that (10) be less than or equal to α, or equivalently

$$\sum_{k=0}^{n} k! C_{n,k} b^{-k} \geq \frac{1}{\alpha} \approx 10^6.,$$

Discovery: The Frequentist View. From a frequentist perspective, the problem is to test $H_0 : s = 0$ vs. $H_1 : s > 0$. There are several cases that can be considered. We may use either a likelihood ratio test or a score test (defined below), and we may condition on n or not. So far, the discussion has emphasized full likelihood (no conditioning) and the likelihood ratio test statistic. I will now consider conditional likelihood and the score function. This is a case that has been worked out in some detail and provides some variety. Also, I will replace s by $b \times s$ and (implicitly) consider large values of b.

A Conditional Score Test. Then the conditional likelihood function given n, log likelihood function, and score function are

$$L_n(s) = \prod_{i=1}^{n} \left[\frac{f_b(x_i) + s f_s(x_i)}{1+s} \right]$$

$$\ell_n(s) = \sum_{i=1}^{n} \log[f_b(x_i) + s f_s(x_i)] - n \log(1+s),$$

and

$$\ell_n'(s) = \sum_{i=1}^{n} \left[\frac{f_s(x_i)}{f_b(x_i) + s f_s(x_i)} \right] - \frac{n}{1+s}.$$

The score test rejects H_0 for large values of

$$\ell_n'(0) = \sum_{i=1}^{n} \left[\frac{f_s(x_i)}{f_b(x_i)} - 1 \right].$$

The score test maximizes the derivative of the (conditional) power function at $s = 0$, and so should have good power against small alternatives. The null distribution of $\ell_n'(0)$ may be approximated by a simple application of the Central Limit Theorem. Observe that

$$E_0 \left[\frac{f_s(x_i)}{f_b(x_i)} - 1 \right] = \int \left[\frac{f_s(x)}{f_b(x)} - 1 \right] f_b(x) dx = 0$$

and

$$E_0 \left[\frac{f_s(x_i)}{f_b(x_i)} - 1 \right]^2 = \int \frac{f_s(x)^2}{f_b(x)} dx - 1 = \sigma^2, \text{ say},$$

assumed finite. Thus

$$Z_n = \frac{\ell_n'(0)}{\sigma\sqrt{n}} \approx \Phi$$

is approximately standard normal for large n, and an *approximate* version of the score test rejects H_0 when $Z_n > 5$.

Pilla, Loader, and Others. In an interesting recent paper Pilla, Loader, and others have extended this approach to the case that f_s depends on nuisance parameters, say $f_s = f_{s,\theta}$. In this case $\ell_n'(0)$ depends on θ, say

$$\ell_n'(0,\theta) = \sum_{i=1}^{n} \left[\frac{f_{s,\theta}(x_i)}{f_b(x_i)} - 1 \right]$$

and

$$\sigma_\theta^2 = \int \frac{f_{s,\theta}(x)^2}{f_b(x)} dx - 1.$$

In this case

$$Z_n(\theta) = \frac{\ell_n'(0,\theta)}{\sigma_\theta \sqrt{n}} \approx \Phi$$

is approximately standard normal for each θ. In fact, the process $Z_n(\theta)$ converges in distribution to a Guassian process, $Z(\theta)$ say, with mean 0 and covariance function

$$r(\theta,\omega) = \frac{1}{\sigma_\theta \sigma_\omega} \left[\int \frac{f_{s,\theta}(x) f_{s,\omega}(x)}{f_b(x)} dx - 1 \right]$$

A modified approximate score test is to reject $H_0 : s = 0$ if $\sup_\theta Z_n(\theta) > c$ where

$$P[\sup_\theta Z(\theta) > c] = \alpha.$$

The exact distribution of $\sup_\theta Z(\theta)$ can only be found in special cases, but there are approximation valid for large c in some generality. For examples, if $Z(\theta)$ is a stationary process (that is, $r(\theta,\omega) = r(\theta - \omega)$), then

$$P[\sup_{a \le \theta \le b} Z(\theta) > z] \sim C e^{-\frac{1}{2} z^2}$$

and $z \to \infty$.

Questions Concerns. The approach just outlined can be questioned on several counts. These questions arise even in the absence of nuisance parameters and will be discussed in that context.

Q_1: First, is conditioning on N really a good idea? It seems to ignore the information in N. Even in the absence of marks, one would reject H_0 for a sufficiently large value of N.

Q_2: Next, is the normal approximation really reasonable. Recall that we are using it in the extreme tail ($c = 5$). Even if one could show that

$$\sup_z |P_0[Z_n \le z] - \Phi(z)| \le 10^{-4},$$

we would only be guaranteed an α of approximately 10^{-4}, far short of the 5σ demanded by the physicists. I think that an approach based on large deviation appoximations may well be indicated.

Q_3: Finally, is the score test really better than the likelihood ratio test, or even an adequate substitute. There is some evidence that the chi-square approximations are accurate in the tails (in the sense of relative error). Unfortunately, the only real proofs that I know require exponential famililies–for example, Chuang and Lai [1].

Problem 5 *For the full (unconditional) likelihood, find the asymptotic distributions of* $\ell'(0|N, \mathbf{X})$, *properly normalized, and* $\lambda_0 = -2\log(\Lambda_0)$, *as* $b \to \infty$*; and prove your assertions.*

Research Questions 1. Develop approximations to $P_0[\Lambda_0 > c]$ that are valid when $c = c(b) \to \infty$ (at a suitable rate) as $b \to \infty$.

2. Can the type I error probability be estimated by simulation to order 10^{-8} say. This would require some sophistication.

References

[1]

[2] Chuang, C. and T.L. Lai (2000). Hybrid resampling methods for confidence intervals. *Statistica Sinica*, **10**, 1-50.

[3] Feldman, Gary and Robert Cousins (1998). Unified approach to the classical analysis of small signals. *Physics Reviews, D*,

[4] Lehmann, Erich (1991). *Testing Statistical Hypotheses.* Wiley.

[5] Roe, Byron and Michael Woodroofe (2000). Setting confidence belts. *Physics Reviews, D*, **62**, $\cdots$