Notes on statistical tests*

Daniel Osherson
Princeton University

Scott Weinstein
University of Pennsylvania

September 20, 2005

We attempt to provide simple proofs for some facts that ought to be more widely appreciated. None of the results that follow are original with us.

Null hypothesis testing

Guaranteed rejection

Let an unbiased coin be used to form an ω-sequence $\mathcal{S}$ of independent tosses. Let N be the positive integers. The finite initial segment of length $n \in N$ is denoted by $\mathcal{S}_n$ (thus, $\mathcal{S}_1$ holds exactly the first toss). For $n \in N$, let H_n be the proportion of heads that show up in $\mathcal{S}_n$.

Pick a positive real number D, to be thought of as a number of standard deviations from the mean of the standard normal distribution. For $n \in N$, call $\mathcal{S}_n$ *newsworthy* just in case:

(1) $$\frac{2H_n - n}{\sqrt{n}} \geq D$$

Recall the following result from (Ross 1988, p. 170) where Φ is the cumulative standard normal function.

(2) THEOREM:

$$\textit{Prob}\left(\frac{2H_n - n}{\sqrt{n}} \geq D\right) \to 1 - \Phi(D) \quad \text{as } n \to \infty.$$

*Contact information: osherson@princeton.edu, weinstein@cis.princeton.edu

If D is 3, for example, then it follows from (2) that the probability of long newsworthy events approaches $.001$. We want to show:

(3) PROPOSITION: With probability 1, there are infinitely many newsworthy initial segments in $\mathcal{S}$.

We provide two proofs of the proposition. The first is swifter but the second isolates the active ingredients in the argument.

First proof of Proposition (3)

Let r be the probability that a point drawn according to a standard normal distribution falls $2D$ or more standard distributions above zero. To prove (3) we show that $\mathcal{S}$ can be partitioned into countably many nonempty, connected regions $R_1, R_2, \ldots$ such that for all $i \in N$, R_{i+1} comes directly after R_i, and the probability is at least r that:

(4) the proportion of heads in R_i is large enough to guarantee that the initial segment $R_1 \ldots R_i$ of $\mathcal{S}$ is newsworthy (no matter how many tails appear in $R_1 \ldots R_{i-1}$).

Since the R_i form a partition, they are independent. Moreover, by construction, $\Sigma_i P(E_i) = \infty$ where E_i is the event that R_i satisfies (4). (Each E_i is a well defined event since it is the union of finitely many basic open sets in the underlying Cantor topology.) It follows from the Second Borel-Cantelli lemma (Billingsley 1986, p. 55) that the probability is one that infinitely many of the R_i satisfy (4), which directly implies (3). We define the desired partition by induction.

For R_1, choose $q \in N$ large enough so that the probability that $\mathcal{S}_q$ is newsworthy is close to $1 - \Phi(D)$. For this q, the probability that $\mathcal{S}_q$ is newsworthy exceeds $1 - \Phi(2D) = r$. So, defining R_1 as $\mathcal{S}_q$ satisfies (4) for $i = 1$ with probability at least r.

Now suppose that $R_1 \ldots R_j$ have been defined. Let m be the last index appearing in R_j. Thus, m is the length of the initial segment of $\mathcal{S}$ determined by $R_1 \ldots R_j$, and is the maximum number of tails that can appear in $\mathcal{S}$ prior to R_{j+1}. It is easy to see that with probability 1,

$$\frac{2H_n - n}{\sqrt{n}} \to \frac{2(H_n - m) - n}{\sqrt{n}} \quad \text{as } n \to \infty.$$

Hence, with probability 1,

$$\lim_{n\to\infty} \textit{Prob} \left(\frac{2(H_n - m) - n}{\sqrt{n}} \geq 2D \right) = 1 - \Phi(2D) = r.$$

So we can choose $q \in N$ such that

$$\textit{Prob}\left(\frac{2(H_q - m) - q}{\sqrt{q}} \geq D\right) > r.$$

Define R_{j+1} to begin at position $m+1$ in $\mathcal{S}$ and finish at position $m+q$. Then with probability at least r, R_{j+1} satisfies (4) for $i = j+1$. ■

Second proof of Proposition (3)

Let C_ω be the collection of possible outcomes of an ω-sequence of tosses of an unbiased coin, with heads recorded as 1's and tails recorded as 0's. Let C_i be the set of finite sequences of 0's and 1's of length i and let $C = \bigcup_{i \in N} C_i$. For each $\sigma \in C$, let $B_\sigma \subseteq C_\omega$ be the set of infinite sequences with finite initial segment σ. These sets form a basis for the Cantor-set topology on C_ω. Note that a set $E \subseteq C_\omega$ is clopen with respect to this topology, if and only if, for some $i \in N$ and $a \subseteq C_i$ $E = \bigcup_{\sigma \in a} B_\sigma$; we say that such an i *secures* E. The natural probability measure on the Borel subsets of C_ω is obtained as the unique extension of the function which assigns B_σ measure 2^{-i} for every $\sigma \in C_i$. We write *Prob* (E) for the probability of measurable $E \subseteq C_\omega$.

(5) DEFINITION: Let $E_i \subseteq C_\omega$, for all $i \in N$. We say the sequence of events $\langle E_i \mid i \in N\rangle$ is *tolerant*, if and only if, there is a real number $r > 0$, such that for every $\sigma \in C$, $\lim_{i\to\infty} \textit{Prob}\,(E_i \mid B_\sigma) > r$.

(6) LEMMA: Suppose $\langle E_i \mid i \in N\rangle$ is a tolerant sequence of events, and for all $i \in N, E_i$ is clopen in C_ω. Let

$$X_k = \{\alpha \mid (\forall i > k)\alpha \notin E_i\}.$$

Then, for all $k \in N$, *Prob* $(X_k) = 0$.

Proof: Since $\langle E_i \mid i \in N\rangle$ is a tolerant sequence of events, we may choose a real number $r > 0$ such that for every $\sigma \in C$,

(7) $\lim_{i\to\infty} \textit{Prob}\,(E_i \mid B_\sigma) > r$.

We call r the *tolerance* of $\langle E_i \mid i \in N\rangle$. Let k be a natural number. In order to prove the lemma, it suffices to show that there is an increasing sequence $\langle m_i \mid i \in N\rangle$ such that $m_0 \geq k$ and for all $j \in N$,

(8) $\textit{Prob}\,(\bigcap_{i=0}^{j} E_{m_i}^c) < (1-r)^{j+1}$.

We construct simultaneously by induction the sequence $\langle m_i \mid i \in N\rangle$ and an ancillary increasing sequence $\langle n_i \mid i \in N\rangle$ such that for all $j \in N$,

(9) n_j secures E_{m_j}.

By (7), we may choose n_0 large enough so that $\textit{Prob}\,(E_{m_0}^c) < (1-r)$. By hypothesis, E_{m_0} is clopen, so we may choose an n_0 which secures E_{m_0}. Now suppose we have constructed the first j terms of each our sequences $m_0, \ldots, m_{j-1}$ and $n_0, \ldots, n_{j-1}$ so as to satisfy the above conditions. We construct m_j and n_j to satisfy conditions (8) and (9) as follows. Let $X = \bigcap_{i=0}^{j-1} E_{m_i}^c$. Note that X is clopen. Observe that since the first j terms of the sequence n_i thus far constructed are increasing, n_{j-1} secures all E_{m_i} for $i < j$, and thus secures X. Thus, we may choose $a \subseteq C_{n_{j-1}}$ with $X = \bigcup_{\sigma \in a} B_\sigma$. Now, since $\langle E_i \mid i \in N\rangle$ has tolerance r, for each $\sigma \in a$ there is an l_σ such that for all $l \geq l_\sigma$, $\textit{Prob}\,(E_l^c \mid B_\sigma) < (1-r)$. Let m_j be the maximum among the l_σ for $\sigma \in a$. It is now manifest that $\textit{Prob}\,(X \cap E_{m_j}) < (1-r) \cdot \textit{Prob}\,(X)$. We complete the construction by choosing an $n_j > n_{j-1}$ which secures E_{m_j}. ■

Bayes ratios

As before, let $\mathcal{S}$ be an ω-sequence of independent tosses of an unbiased coin. Let h_0 be the (accurate) hypothesis that the probability of heads is $.5$. Fix $q \neq .5$, and let h_1 be the hypothesis that the probability of heads is q. Assume that h_0 and h_1 each have prior probability one-half. Then for any $n \in N$, $\textit{Prob}\,(\mathsf{h}_0 \mid \mathcal{S}_\mathsf{n}) > \textit{Prob}\,(\mathsf{h}_1 \mid \mathcal{S}_\mathsf{n})$ if and only if

$$\frac{\textit{Prob}\,(\mathcal{S}_n \mid \mathsf{h}_0)}{\textit{Prob}\,(\mathcal{S}_n \mid \mathsf{h}_1)} > 1.$$

The following proposition may be compared to (3).

(10) PROPOSITION: With positive probability,

$$\frac{\textit{Prob}\,(\mathcal{S}_n \mid \mathsf{h}_0)}{\textit{Prob}\,(\mathcal{S}_n \mid \mathsf{h}_1)} > 1 \quad \text{for all } n \in N. \quad (*)$$

To prove (10), for even $i \in N$, let F_i be the event:

$$\mathcal{S}_i = \underbrace{HTHT\cdots HT}_{i \text{ times}}.$$

and for odd $j \in N$, let E_j be the event:

$$|H_m - .5| < \left|\frac{.5 - q}{2}\right| \quad \text{for all } m \geq j.$$

These events are unions of basis elements in the underlying topology, hence well defined. Note that E_1 implies (*) in (10), and also for all odd $j > 2$, $E_j \cap F_{j-1}$ implies (*). It therefore suffices to show that for some odd $j_0 \in N$, $\textit{Prob}\,(E_{j_0}) > 0$. For, in this case F_{j-1} and E_i are independent, and $\textit{Prob}\,(F_i) > 0$; so $\textit{Prob}\,(E_j \cap F_{j-1}) > 0$. For a contradiction, suppose that for all odd $j \in N$, $\textit{Prob}\,(E_{j_0}) = 0$. Then by the first Borel-Cantelli lemma (Billingsley 1986, p. 55), the probability is one that only finitely many of the E_j occur. Hence, the probability is one that *none* of the E_j occur (since if any of the E_j occur then cofinitely many of them occur). Hence the probability that $\mathcal{S}$ converges to $.5$ is zero, contradicting the strong law of large numbers (Billingsley 1986, p. 80). ■

Remarks (very preliminary)

A possible rejoinder to the results presented above is that they show only that hypothesis-testers should announce their sample sizes N ahead of time. Only newsworthy events that are discovered within the announced N are "truly" newsworthy.

In response, imagine a sequence of researchers studying the same ω-sequence of coin flips, each announcing a different sample size. Almost surely, one of them will be entitled to make an announcement, according to the foregoing suggestion. This person P would presumably be justified in believing that something newsworthy had happened. But all the other researchers know just what P knows, so wouldn't they also be justified in believing that a newsworthy event has transpired? Otherwise, we must reject the principle:

(11) If two people have the same information (and know this), and one of them is justified in believing a statement S then both are justified in believing S.

If we reject (11) in this context it could only be because P performed a certain act, namely, announcing her sample-size. How could such an act change the epistemological situation of P? (Believing that the act would make a difference seems to embrace something magical.) The point, of course, is that (properly) accepting (11) implies that with probability one a "truly" newsworthy event will be on all the researchers' minds at some point in the experiment.

Confidence intervals

We consider the following problem, relying on the discussion in (Baird 1992, p. 273). An urn is filled with balls numbered from 1 to L (no gaps, no repeats). We assume that $L > 1$. The urn is sampled with replacement k times. We wish to investigate the idea of a "confidence interval around L."

Let $X_{L,k}$ be the set of possible samples of k balls that can be drawn from the urn with L balls. We think of such samples as ordered sequences. It is therefore clear that $X_{L,k}$ is finite, and its members have uniform probability of being drawn [namely $(1/L) \exp k$]. Let f be a mapping of $X_{L,k}$ into the set of intervals of the form $[i, j]$, where i, j are positive integers ($i \leq j$).

(12) DEFINITION: Let a percentage r be given. Call f r-*reliable* just in case for every $L, k > 0$, and for at least r% of $x \in X_{L,k}$, $L \in f(x)$.

Of course, some r-reliable functions are more interesting than others in virtue of situating L in narrower intervals. Here our concern is to show:

(13) PROPOSITION: There are 95%-reliable f, g such that for some $L, k > 0$ and $x \in X_{k,L}$, $f(x) \cap g(x) = \emptyset$.

The proposition rules out assigning 95% "personal confidence" to $L \in f(x)$ simply on the basis that f is reliable. This is because g is also reliable and it is incoherent to have 95% confidence that both $L \in f(x)$ and $L \in g(x)$ for the L, k, x that witness (13).

Further, our goal is to construct "natural" functions f and g for Proposition (13). For the proposition follows from crafty modification of a single reliable function. This is shown later.

Existence of reliable functions.

We first establish the existence of reliable functions.

(14) PROPOSITION: For every $0 \leq r < 1$, there is an $r\%$-reliable function.

Proof: Fix r with $0 \leq r < 1$. Let $X_k = \bigcup_{1 \leq L} X_{L,k}$. For $x \in X_k$, we write $\mathsf{max}(x)$ for the largest member of x.

(15) FACT: Let $L_0 > 1$ and $1 \leq m \leq L_0$ be given. For all $L > L_0$, the proportion of samples y from $X_{L,k}$ with $\mathsf{max}(y) \leq m$ is less than or equal to $(m/L_0)^k$.

Now we define an $r\%$-reliable function, f. For $x \in X_{k},$, $f(x) = [\mathsf{max}(x), L_0]$ where:

(16) L_0 is the least integer greater than or equal to $\mathsf{max}(x)$ such that for all $L > L_0$, the proportion of samples y from $X_{L,k}$ with $\mathsf{max}(y) \leq \mathsf{max}(x)$ is less than or equal to $1-r$.

That such an L_0 exists for each $x \in X_k$ (and can be calculated from x) follows from Fact (15).

To show that f is $r\%$-reliable, let $L > 1$, $k > 0$, and $x \in X_{L,k}$ be given. Then $L \in f(x)$ iff $L \in [\mathsf{max}(x), L_0]$ where L_0 satisfies (16). Since $L \geq \mathsf{max}(x)$, $L \notin [\mathsf{max}(x), L_0]$ iff $L > L_0$, which implies that $x \in A = \{y \in X_{L,k} \mid \mathsf{max}(y) \leq \mathsf{max}(x)\}$. But by (16), *Prob*$(A) \leq 1 - r$. ■

(17) EXAMPLE: Following (Baird 1992, p. 275), suppose we draw the following sample x of size 16.

92	93	94	95	96	97	98	99
101	102	103	104	105	106	107	108

Then, $\mathsf{max}(x) = 108$. To form a confidence interval using the 95%-reliable function f defined by (16), we seek the least L_0 such that the probability of drawing 16 balls labeled 108 or less from an L_0-urn is no greater than 5%. By Fact (15), L_0 is the least integer satisfying:

$$\left(\frac{108}{L_0}\right)^{16} < .05.$$

Calculation reveals that $L_0 = 130$. Hence $f(x) = [108, 130]$.

Another reliable function

Let U_L be an urn with L balls labelled $1, \ldots, L$. We think of U_L as a probability space with the uniform distribution, that is, we consider the experiment of drawing a ball from the urn under the condition that each ball is equally likely to be drawn. Define the random variable Z on this space which measures the value of each outcome of our experiment, that is, Z applied to the ball with label i gives the value i. We first compute $E(Z)$, the *expected value* (also called the *mean*) of Z.

(18) $E(Z) = 1/L \cdot \Sigma_{i=1}^{L} i = (L+1)/2$.

Let μ denote $E(Z) = (L+1)/2$. Next, we compute $\sigma^2(Z)$, the variance of Z, that is, the expectation of the squared deviation of Z from its mean. Making use of the well-known identity

(19) $\Sigma_{i=1}^{L} i^2 = L(L+1)(2L+1)/6$

we have

$$\sigma^2(Z) = E((Z - E(Z))^2) = E(Z^2 - 2E(Z)Z + E(Z)^2)$$

$$= E(Z^2) - 2E(Z)E(Z) + E(Z)^2 = E(Z^2) - E(Z)^2 = \frac{1}{L}\Sigma_{i=1}^{L} i^2 - \frac{(L+1)^2}{4}$$

$$= \frac{(L+1)(2L+1)}{6} - \frac{(L+1)^2}{4} = (L+1)\frac{(4L+2) - 3(L+1)}{12} = \frac{L^2-1}{12}.$$

Let $\sigma = \sqrt{(L^2-1)/12}$, that is, the *standard deviation* of Z.

Let $Z_i, i \in N$ be a sequence of countably many independent copies of Z, and let $S_n = \Sigma_{i=1}^{n} Z_i$. By the *Central Limit Theorem* (Billingsley 1986, p. 367), the distribution of $(S_n - n\mu)/\sigma\sqrt{n}$ converges to the unit normal as $n \to \infty$. We rely on the following fact [see (Ross 1988, p. 164)].

(20) FACT: If X is normally distributed with mean μ and variance σ^2 then $Y = \alpha X + \beta$ is normally distributed with mean $\alpha\mu + \beta$ and variance $\alpha^2\sigma^2$.

It follows that the distribution of S_n/n converges to the normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ as $n \to \infty$. Hence, for large n (and letting $\bar{X}$ stand for S_n/x):

(21)
$$\textit{Prob}\left(\mu - 2\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 2\frac{\sigma}{\sqrt{n}}\right) =$$

$$\textit{Prob}\left(\mu - 2\frac{\sqrt{(L^2-1)}}{\sqrt{12n}} \leq \bar{X} \leq \mu + 2\frac{\sqrt{(L^2-1)}}{\sqrt{12n}}\right) =$$

$$\textit{Prob}\left(\frac{L+1}{2} - 2\frac{\sqrt{(L^2-1)}}{\sqrt{12n}} \leq \bar{X} \leq \frac{L+1}{2} + 2\frac{\sqrt{(L^2-1)}}{\sqrt{12n}}\right) \approx$$

$$\textit{Prob}\left(\frac{L+1}{2} - 2\frac{L}{\sqrt{12n}} \leq \bar{X} \leq \frac{L+1}{2} + 2\frac{L}{\sqrt{12n}}\right) =$$

$$\textit{Prob}\left(\frac{3\sqrt{n}(2\bar{X}-1)}{3\sqrt{n}+2\sqrt{3}} \leq L \leq \frac{3\sqrt{n}(2\bar{X}-1)}{3\sqrt{n}-2\sqrt{3}}\right) \approx .95.$$

Of course, the latter interval converges on $\{2\bar{X} - 1\}$ as $n \to \infty$. The first approximation step in the foregoing chain results from suppressing the 1 in the expression $\sqrt{(L^2 - 1)}$. The second comes from the fact that 2 standard deviations around 0 in the unit normal distribution subtend a little more than 95%.

(22) EXAMPLE: We return to the example in (Baird 1992, p. 275), concerning following sample x of size 16.

92	93	94	95	96	97	98	99
101	102	103	104	105	106	107	108

In this case, $\bar{X} = 100$, and $n = 16$, so *Prob* $(154.4 \leq L \leq 279.8) \approx .95$.

The 95%-confidence interval just derived is disjoint from the 95%-confidence interval computed in Example (22), namely, $[108, 130]$. Of course, it is incoherent to attach 95% probability (even personal probability!) to two mutually exclusive events. This is because for every pair A, B of events,

$$\textit{Prob}\,(A) \geq .95 \wedge \textit{Prob}\,(B) \geq .95 \Rightarrow \textit{Prob}\,(A \wedge B) \geq .9.$$

So, the juxtaposition of the two examples reveals the obstacle to using confidence intervals to build confidence.

Notice, however, that the interval computed in Example (22) is only approximate because convergence to the unit normal is gradual, depending on the sample size. So we exhibit an example in which sample size is allowed to vary.

(23) EXAMPLE: For each $n \geq 1$, consider the sample:

$$1 \quad \underbrace{50 \cdots 50}_{n \text{ times}} \quad 60.$$

According to the $\max$-based method of constructing confidence intervals, we get $[60, X]$ where X is least such that

$$\left(\frac{60}{X}\right)^{n+2} < .05,$$

in other words, the interval:

$$(*) \qquad \left[60,\ \frac{60}{.05^{1/(n+2)}}\right].$$

As $n \to \infty$, this interval converges to 60. In contrast, as noted above, the mean-based confidence interval that relies on the normal approximation, converges to twice the sample mean minus 1, in this case 99. For large n, the two intervals thus become disjoint. Moreover, as $n \to \infty$, the mean-based interval becomes arbitrarily exact (as the distribution of the sample mean converges to the unit normal).

There is an interesting fact about samples of the form

$$1 \quad \underbrace{50 \cdots 50}_{n \text{ times}} \quad 1000.$$

The mean-based confidence interval eventually begins to exclude 1000 (since it converges to $\{99\}$). This is another reason to reject the "confidence" interpretation of the interval.

A more exact (but conservative) confidence interval

Appeal to Chebyshev's Inequality, in place of the Central Limit Theorem, may also be used to establish that a mean-based algorithm achieves a high level of confidence. Chebyshev's Inequality (24) gives uniform bounds on the probability of deviation from the mean for arbitrary distributions. For the current application, Chebyshev's Inequality may be stated as follows:

$$(24) \qquad \mathit{Prob}\,(\mid \bar{X} - \mu \mid \geq k\sigma/\sqrt{n}) \leq k^{-2}$$

As before, let us use $L/\sqrt{12}$ as an approximation for σ. It follows at once from (24) and a computation similar to (21) that for all k, L, and $n > (k^2)/3$, with probability at least $1 - k^{-2}$,

$$(25) \qquad \frac{\sqrt{n}(2\bar{X} - 1)}{\sqrt{n} + c \cdot k} \leq L \leq \frac{\sqrt{n}(2\bar{X} - 1)}{\sqrt{n} - c \cdot k},$$

where $c = 1/\sqrt{3}$. This gives rise, in the obvious way, to a k-stds-mean-based algorithm which projects $1 - k^{-2}$-confidence intervals. Moreover, it is clear that for each fixed k, for all large enough n, samples of the form

$$1 \quad \underbrace{50 \cdots 50}_{n \text{ times}} \quad 1000$$

cause the k-stds-mean-based algorithm to project confidence intervals that exclude 1000, and that are thus disjoint from the confidence intervals projected by max-based algorithms.

Two remarks

(a) It would seem that within a Bayesian framework, there could not be disjoint intervals each of which is predicted with 95% confidence to contain L. The impact of a prior distribution on L would render coherent all estimates of the form:

$$\textit{Prob}\,(L \in [a,b] \mid \text{data } = \{x_1 \cdots x_n\}) = r\%.$$

(b) According to (Salsburg 2002), concerns about interpreting the confidence interval were voiced as soon as the idea was formulated by Neyman. Neyman understood the worries perfectly well, and tried to interpret such intervals in purely frequentist terms (involving the limiting frequency of being right if you stuck with a single method).

Multiplicity of reliable functions.

Let g and h be 95%-reliable functions. We call g, h *incompatible* just in case for all $L > 1$ there are cofinitely many $k \in N$ such that for some $x \in X_{L,k}$, $g(x) \cap h(x) = \emptyset$; we call g, h *weakly incompatible* just in case for all $L > 1$ there are infinitely many $k \in N$ such that for some $x \in X_{L,k}$, $g(x) \cap h(x) = \emptyset$.

(26) PROPOSITION:

(a) There is a countably infinite set of pairwise incompatible 95%-reliable functions.

(b) There is a set of pairwise weakly incompatible 95%-reliable functions of cardinality the continuum.

(A simple cardinality argument shows that the proposition is best possible.)

Proof: Let f be a 96%-reliable function as guaranteed by Proposition (14). Choose k_0 big enough so that for all $L > 1$, each sample of at least k_0 balls has probability less than 1% ($k_0 = 7$ will do). For all $k \in N$, fix such a sample $s(k) = 1^{k_0+k}$, the sequence consisting of ball 1 sampled $k_0 + k$ times in a row.

To establish (a), we proceed as follows. For $i = 0$, define function g_0 to be f, and for $i > 0$, define function g_i as follows. if x is not in the range of s, let $g_i(x) = f(x)$; and for each k, let $g_i(s(k))$ be some interval chosen to be disjoint with $g_j(s(k))$ for all $j < i$. Since $s(k)$ has probability less than 1%, each g_i is 95%-reliable. It is also immediate that they are all incompatible.

To establish (b), we proceed as follows. Pick disjoint intervals I_0 and I_1. Now, for every $c \in 2^\omega$ (the 0-1 valued functions on ω) define g_c as follows: if x is not in the range of s, let $g_c(x) = f(x)$; and for all k, let $g_c(s(k)) = I_{c(k)}$. As in the argument for (a), it is clear that each g_c is 95%-reliable. Moreover, observe that there are continuum many $c \in 2^\omega$ which pairwise differ infinitely often. The result now follows at once. ■

References

BAIRD, D. (1992): *Inductive Logic: Probability and Statistics*. Prentice Hall, Englewood Cliffs NJ.

BILLINGSLEY, P. (1986): *Probability and Measure (2nd Edition)*. John Wiley, New York.

ROSS, S. (1988): *A First Course in Probability, 3rd Edition*. Macmillan, New York City.

SALSBURG, D. (2002): *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. Owl Books.