WHEN SEVERAL BAYESIANS AGREE THAT THERE WILL BE NO REASONING TO A FOREGONE CONCLUSION

JOSEPH B. KADANE, MARK J. SCHERVISH, AND TEDDY SEIDENFELD†‡

Carnegie Mellon University

When can a Bayesian investigator select an hypothesis H and design an experiment (or a sequence of experiments) to make certain that, given the experimental outcome(s), the posterior probability of H will be lower than its prior probability? We report an elementary result which establishes sufficient conditions under which this reasoning to a foregone conclusion cannot occur. Through an example, we discuss how this result extends to the perspective of an onlooker who agrees with the investigator about the statistical model for the data but who holds a different prior probability for the statistical parameters of that model. We consider, specifically, one-sided and two-sided statistical hypotheses involving i.i.d. Normal data with conjugate priors. In a concluding section, using an "improper" prior, we illustrate how the preceding results depend upon the assumption that probability is countably additive.

1. Expected Conditional Probabilities and Reasoning to Foregone Conclusions. Suppose that an investigator has his or her designs on rejecting, or at least making doubtful, a particular statistical hypothesis H. To what extent does basic inductive methodology insure that, without violating the total evidence requirement, this intent cannot be sure to succeed? We distinguish two forms of the question:

(1) Can the investigator plan an experiment so that *he* or *she* is certain it will halt with evidence that disconfirms H?
(2) Can the investigator plan an experiment so that *others* are certain that the investigator will halt with evidence that disconfirms H?

That is, are foregone conclusions, viewed either (1) in the first-person or (2) in the third-person perspective, precluded by fundamental principles in the design of experiments? (In both versions of the question, we understand the judgment of disconfirmation to be the investigator's.)

Related to these questions is the familiar controversy whether an experimenter's stopping rule is relevant to the analysis of his or her experimental data. Savage writes (1962, 18),

> The [*likelihood*] principle has important implications in connection with optional stopping. Suppose the experimenter admitted that he had seen 6 red-eyed flies in 100 and had then stopped because he felt that he had thereby accumulated enough data to overthrow some popular theory that there should be about 1 per cent red-eyed flies. Does this affect the interpretation of 6 out

†This research was supported by ONR contract N00014–89-J-1851, and NSF grants DMS-9303557 and SES-9123370.

‡Department of Philosophy, Carnegie Mellon University, Schenley Park, Pittsburgh, PA 15213-3890.

Philosophy of Science, 63 (Proceedings) pp. S281–S289. 0031-8248/96/63supp-0033$2.00

> of 100? Statistical tradition emphasizes, in connection with this question, that if the sequential properties of his experimental programme are ignored, the persistent experimenter can arrive at data that nominally reject any null hypothesis at any significance level, when the null hypothesis is in fact true. These truths are usually misinterpreted to suggest that the data of such a persistent experimenter are worthless or at least need special interpretation; see, for example, Anscombe (1954), Feller (1940), Robbins (1952). The likelihood principle, however, affirms that the experimenter's intention to persist does not change the import of his experience.

Here is a simple example of what Savage refers to as "statistical tradition." Consider the null hypothesis H_0: $\theta = 0$, that the mean of *i.i.d.* normal data is 0. The data have known unit variance. Fix k_α so that $k_\alpha/\sqrt{n}$ corresponds to the nominal rejection point in an α-level UMPU (Uniformly Most Powerful, Unbiased) test of H_0 versus the composite alternative hypothesis H_0^c:$\theta \neq 0$, based on a sample of size *n*. Let H_t denote the simple hypothesis that $\theta = t$. Continue observing data until the sample average, $\bar{x}_n = (x_1 + \ldots + x_n)/n$, satisfies the inequality (1.1), then halt:

$$|\bar{x}_n| > k_\alpha/\sqrt{n} \tag{1.1}$$

The likelihood principle entails that the statistician's intent to stop only when (1.1) obtains is irrelevant to the "evidential import" of the data for hypotheses about θ (for a recent view, see Berger 1985, §7.7). However, the statistician has here designed an experiment that, provided it stops, yields data with a very low likelihood for H_0 versus a rival hypothesis, H_t (for $t = \bar{x}_n$). If, contrary to traditional (Neyman-Pearson) theory, the significance level is calculated independent of the stopping rule for the experiment—a mistake by that traditional theory—then when the inquiry halts H_0 has achieved an observed significance of α, or less. Moreover, given the truth of H_0, by the law of the iterated logarithm, with probability 1 the experiment terminates, i.e., almost surely the inequality (1.1) is eventually satisfied.

Traditional hypothesis testing sidesteps this forgone conclusion (of a low significance level) only by incorporating the experimenter's intention, of when to terminate sampling, as part of the relevant evidence. This is contrary to the likelihood principle. By contrast, the Bayesian answer to the first-person version of our question (1) is straightforward and elementary. In short, with a countably additive probability, the law of total probability ensures that conditional probabilities cannot lead to a foregone conclusion. Thus, when the investigator uses Bayes' rule for updating, he or she cannot plan an experiment that, by his or her own lights leads to a posterior opinion surely below (or surely above) his or her prior opinion. This argument has been reported before. (See, e.g., D. Kerridge's 1963 note.) Kadane et al. have discussed it elsewhere (1995) and we include the bare details here for completeness of our presentation.

Let $(S, \mathcal{A}, P)$ be a (countably additive) probability space, which we think of as the underlying joint space for all quantities of interest. Expectations are with respect to the probability P. Unconditional expectation is denoted by E(•) and conditional expectation given a random variable X is denoted by E(•|X). Let $(\mathcal{X}, \mathcal{B})$ and (Ω, τ) be measurable spaces where,

$X: S \to \mathcal{X}$ is a random quantity to be learned,
$\Theta: S \to \Omega$ is any random quantity,
and $h: \Omega \to \Re^*$ is an (extended) real-valued function whose expectation $E(h)$ exists.

Then the familiar law of total probability implies that:

$$E[E(h(\Theta) \mid X)] = E(h(\Theta)). \tag{1.2}$$

(See, e.g., Ash 1972, T. 6.5.4, p.257.)

This result says that there can be no experiment with outcome X designed (almost surely with respect to P) to drive up or designed to drive down the conditional expectation of h, given X. Equation (1.2) has no special logical dependence on Bayes's theorem, except that non-Bayesian statistical methods often begin with the claim that neither the "prior" expectation $E(h(\Theta))$, nor the "posterior" expectation $E(h(\Theta) \mid X)$, has objective status.

For example, suppose that h is the indicator for an hypothesis H (an unobserved "event") in Ω, i.e., $h(\Theta) = 1$ if $\Theta \in H$, $h(\Theta) = 0$ otherwise. Thus, $E(h(\Theta))$ is the agent's "prior" probability of H, denoted by $P(H) = p$. Let $X_1, X_2, \ldots$ be observations which become available sequentially. In order to consider experimental designs which mandate a minimum sample size, $k \geq 0$, define $N = \inf\{n \geq k: P(H \mid X_1, \ldots, X_n) \geq q\}$ where $N = \infty$ if the set is empty. That is, N identifies the first point after the kth in the sequence of X_i-observations when the agent's "posterior" probability of H reaches q, at least. The event $N = \infty$ obtains when, after k-many observations, the sequence of conditional probabilities, $P(H \mid X_1, \ldots, X_n)$ $(n \geq k)$, all remain below q. Then,

$$P(N < \infty) \leq p/q < 1. \tag{1.3}$$

Thus, when $p < q$, the agent's prior probability is less than 1 that, using Bayes' rule for updating, he or she will halt the sequence of experiments and conclude that the posterior probability of H has risen to q, at least.

The same argument provides a bound for the conditional probability of terminating the experiment in finite time, given that H is false:

$$P(N < \infty \mid \neg H) \leq p(1-q)/q(1-p). \tag{1.4}$$

For example, assume $0 < p \leq .1$ and let $q/p = 10$. That is, by the choice of the stopping rule, whenever the experiment terminates the posterior probability for H increases ten-fold (at least). Then, the inequality (1.4) asserts that, given that H is false, the conditional probability of terminating the experiment is no more than $(.1-p)/(1-p) < .1$.

Savage (1962, 72–73) offers a similar conclusion and illustrates it with a simple case of two Binomial hypotheses. D. Kerridge (1963) derives the same bounds as in (1.4) for the case of a uniform prior ($p = .5$). And Cornfield (1970, 20–21) uses Kerridge's inequality to argue that as a Bayesian, you cannot be sure to defeat a true "null" hypothesis. However, these results offer answers only to the first version of our opening question. In the following section, we extend the analysis to cover the second version of our question, involving the perspective of an onlooker.

2. Hypothesis Testing and Sampling to A Foregone Posterior Probability: a Second Perspective. The results of section 1 are "internal" to the Bayesian agent who is designing the sequential experiment. For example, the probability bound (1.4) is for the investigator's personal probability who designed the sequential experiment. It tells the experimenter who is prepared to stop at the first observation when there is a rise of at least q-p in his (her) probability of H that, given ¬H, it is not certain that the experiment ever halts. Apart from this "internal" check on reasoning to a foregone conclusion, what can be said from the standpoint of an onlooker who ponders whether the experimenter will come to assign a high posterior probability to a false hypothesis?

2.1. The Probability of Halting When H is False One way to answer this question is to consider a second point of view that agrees with the first about the likelihood (or sampling model) of the data, but which differs with the experimenter's prior probability of the hypotheses H and ¬H. This is our approach in what follows. We consider two varieties of hypotheses, where ¬H is a one-sided or a two-sided alternative. Also, when ¬H is two-sided, there is the distinction between an interval-valued and a (traditional) point-valued null hypothesis H. The experimenter designing the sequential study stops gathering data once his or her posterior probability for H rises to q, at least. His (her) prior probability for H is p (which is $<$q), so that the experiment terminates on the first occasion when H has gained at least $q-p > 0$ over its prior probability.

For our illustration, the possible observations, $X_1, X_2, \ldots$ are *i.i.d.* Normal $\mathbf{N}(\theta,1)$ data. Let the experimenter have a conjugate Normal prior over Θ, $\mathbf{N}(\mu, 1/\lambda)$, with specified mean μ and precision λ. (We make the obvious adjustment and follow Jeffreys's 1939 analysis for the case where H is a point-valued hypothesis.) We do not restrict the onlooker's prior over Θ, except that we assume it is a countably additive probability. Given ¬H, does the onlooker believe, along with the experimenter, that with positive probability the latter will assign H probability less than q at each stage of an unending inquiry? Will both parties agree that, given H is false, it is *not* a certainty the experimenter will arrive at the foregone conclusion he (she) aims for?

One Sided Testing: For a specified quantity θ_0, H is the hypothesis that $-\infty < \theta \leq \theta_0$ and ¬H is the complementary hypothesis that $\theta_0 < \theta < \infty$. We show that no matter what the onlooker's prior probability for ¬H, in parallel with (1.4), he (she) too believes that, given ¬H, it is *less than certain* that the experimenter will ever conclude the study. We attack this problem by demonstrating a result that makes sense from a Classical point of view:

Theorem 2.1 For each $\theta > \theta_0$, P_θ(experimenter stops inquiry) < 1.❑

(The proof of this theorem is given in the Appendix.)

This theorem establishes that, *no matter how H fails,* still it is not a certainty that the experimenter comes to assign H (posterior) probability q, at least. Then, we have as an easy consequence:

Corollary From the onlooker's perspective, P(experimenter halts | ¬H) < 1.❑

Two Sided Testing: Next, we examine two cases: where H is an interval with non-empty interior, and the traditional point-null hypothesis (as presented in Jeffreys's work). Let H be the interval hypothesis that $-\infty < \theta_a \leq \theta \leq \theta_b < \infty$ ($\theta_a < \theta_b$) and ¬H is the complementary hypothesis that $\theta < \theta_a$ or $\theta_b < \theta$. By reasoning similar to the case of one-sided testing, we show that the onlooker agrees with the experimenter that, given ¬H, there is positive probability the study never terminates.

Theorem 2.2 For each $\theta > \theta_b$, and for each $\theta < \theta_a$, P_θ(experimenter stops inquiry) < 1.❑

Therefore, the onlooker agrees, given ¬H, there is positive probability the experimenter does not arrive at the forgone conclusion that H has increased its (posterior) probability to q.

Third, we adapt the experiment's "prior" to address the traditional case of a point-valued null hypothesis using Jeffreys's (1939, Chapter 5) model of Bayesian hypothesis testing. Let H assert that $\theta = \theta_0$. The experimenter's "prior" for θ satisfies:

$$P(H) = p \quad \text{and} \quad \text{given } \neg H, \theta \sim \mathbf{N}(\theta_0, 1/\lambda).$$

Theorem 2.3 For each $\theta \neq \theta_0$ P_θ(experimenter stops inquiry) $< 1.$ □

We have considered three cases for the "null" hypothesis H: where H is either (1) a one-sided interval, $\theta \leq \theta_0$; or (2) H is a two-sided interval, $\theta_a \leq \theta \leq \theta_b$; or (3) H is a point $\theta = \theta_0$. In each case, an onlooker who shares the same statistical model as the experimenter but who has a different prior over θ, also believes that if H fails then it is not certain that the experimenter will ever conclude sampling.

2.2 The Probability of Halting When H Obtains. Consider, now, the dual question, when H is true. Will the experimenter him/herself be sure of halting, given that H obtains? Will the onlooker, with a different prior for H, agree that the experimenter is sure to halt if H is true? The answer is Yes to both questions, as straight-forward reasoning establishes:

One sided testing Lemma 2.1: Given θ, $\theta \leq \theta_0$, P_θ(experiment halts) $= 1.$ □

Two sided testing Lemma 2.2: Given θ, $\theta_a \leq \theta \leq \theta_b$, then P_θ(experiment halts) $= 1.$ □

Thus, both the experimenter and the onlooker are certain (and they know the other is certain) that, given H, the study stops. They share a common expectation that, given H, the former comes to assign H a posterior probability of q, at least: P(experimenter halts the study | H) = 1.

Remark: These lemmas imply that, by continuity, unfortunately, $\lim \sup_{\theta \in \neg H} P_\theta$(experimenter halts) $= 1$.

Thus, the conditional probability of stopping when H fails cannot be bounded away from 1.

2.3 Summary of the Discussion about Hypothesis Testing. We have examined the case of a second Bayesian, an onlooker, including as a special instance a so-called "frequentist" (whose "prior" may be concentrated on a single value of the parameter), who considers the same problem as the Bayesian experimenter. We have given details for the case of testing the mean of a normal distribution. In Section 1, we saw that an experimenter who assigns an hypothesis probability p believes that the conditional probability, given that the hypothesis is false, is less than 1 that he or she will stop sampling and declare that the posterior probability is at least $q > p$. In this section, for the special case of testing a normal mean, we saw that an onlooker also believes that, given the hypothesis is false, the probability is less than 1 that the experimenter will stop and declare the probability of the hypothesis to be at least q.

We also saw that, given the truth of the hypothesis, the onlooker (and the experimenter) believe that the experimenter will stop and declare the probability of the hypothesis to be at least q. These results applied to the cases of hypotheses of the form $\theta \leq \theta_0$, $\theta_a \leq \theta \leq \theta_b$, and $\theta = \theta_0$, using natural conjugate or related priors. Since the lemma (Lemma A of the appendix) on which these results rest applies in greater generality, it stands to reason that these results can be extended. Our purpose with these examples is to illustrate where the "internal" conclusions (of Section 1) extend to the perspective of an onlooker.

3. Summary and a Caveat about "Improper" Priors. In Section 1, we displayed some elementary reasoning, based on the law of total probability, about the protection simple Bayesian theory affords against sampling to a forgone conclusion. In Section 2, using a problem of inference about a normal mean, we extended these results to the perspective of a second investigator and to some results that apply from a traditional (Neyman-Pearson) point of view as well, i.e., results that apply regardless of the unknown value of the parameter.

We conclude, however, with a warning that the foregoing analysis depends upon the assumption that personal probability is countably additive, rather than being merely finitely additive. (Kadane et al. (1995) examine the connection between Bayesian reasoning to a foregone conclusion and the use of merely finitely additive probability.) This is particularly important when the Bayesian investigator uses a so-called "improper" prior to capture a merely finitely additive personal probability. Here, then, is an illustration of how improper priors can produce Bayesian inference to a foregone conclusion, even from the first-person perspective of the investigator.

Let $X \sim N(\theta, \tau)$ and, for a first case without reasoning to a foregone conclusion, assume that the precision τ ($=$ variance^{-1}) is known. Suppose the prior for θ is conjugate, $\theta \sim N(\mu_0, \kappa_0)$. Then the posterior distribution of θ given $X = x$ is normal with mean $\mu_0\left(\frac{\kappa_0}{\kappa_0+\tau}\right) + x\left(\frac{\tau}{\kappa_0+\tau}\right)$ and precision $\kappa_0 + \tau$. If the prior is, on a sequence, getting more diffuse, i.e., when $\kappa_0 \to 0$, this posterior approaches a normal distribution with mean x and precision τ. This posterior can be recovered by a formal Bayes calculation with an improper, uniform distribution for θ, uniform on $(-\infty, \infty)$.

Now, for a second version where reasoning to a foregone conclusion occurs, suppose that τ is unknown as well. The conjugate prior for θ, given τ, is normal with mean μ_0 and precision $\kappa_0\tau$, and the prior on τ is the Gamma (α_0, β_0) distribution, a proper distribution! The posterior distribution is of the same form, with posterior hyperparameters:

$$\mu_1 = \left(\frac{\kappa_0\mu_0 + x}{\kappa_0 + 1}\right), \text{ precision } (\kappa_0+1)\tau,\ \alpha_1 = \alpha_0 + 1/2$$

$$\text{and } \beta_1 = \beta_0 + \frac{\kappa_0(x-\mu_0)^2}{2(\kappa_0+1)}. \quad (3.1)$$

Again, the prior on θ is taken to be diffuse, so we allow that $\kappa_0 \to 0$ in (3.1). Then, the posterior remains in the conjugate family, with:

$$\mu_1 = x, \text{ precision } \tau,\ \alpha_1 = \alpha_0 + 1/2, \text{ and } \beta_1 = \beta_0. \quad (3.2)$$

This posterior has the following peculiar implication: The prior on τ is Gamma (α_0, β_0) while the posterior (as calculated) is Gamma $(\alpha_0 + 1/2, \beta_0)$. The posterior on τ does not depend on the datum $X = x$ and is different from its prior. Thus, prior to observing $X = x$ the investigator believes τ is Gamma (α_0, β_0). However, regardless what value of X is observed, he or she knows (even in advance) that afterwards the posterior will give τ a Gamma $(\alpha_0 + 1/2, \beta_0)$ distribution. Thus, the investigator will have reasoned to a foregone conclusion about τ.

One can show, using the results of Regazzini (1987) or of Berti, Regazzini, and Rigo (1991), that the limit of priors, limit of posteriors, and likelihood in this example are jointly coherent, in de Finetti's (1974) sense. That is, there is no finite collection of pairs of events $(E_1, A_1), \ldots, (E_n, A_n)$ and constants $c_1, \ldots, c_n$ such that $\sum_{i=1}^{n} c_i I_{A_i}[I_{E_i} - \Pr(E_i|A_i]$ is uniformly negative. Thus, de Finetti's standard of *coherence* is not sufficient to ward off reasoning to a foregone conclusion.

Moreover, there exists a joint improper prior where the posterior (3.2) can be obtained by a naïve application of Bayes' theorem, namely the product of the (improper) conditional density for θ given τ, $\tau^{.5}d\theta$, and the (proper) marginal Gamma (α_0, β_0) density for τ.

$$\tau^{\alpha_0 - .5} e^{-\beta_0 \tau} d\theta d\tau \tag{3.3}$$

Remark: Inspection of (3.3) shows that the improper joint density for (θ,τ) is also the product of the uniform, improper density for θ, dθ, and the proper Gamma ($\alpha_0 + .5$, β_0) distribution for τ—for which there is no reasoning to a foregone conclusion about τ based on X. A direct elicitation of the agent's prior (marginal) opinion about the parameter τ will distinguish between these two different priors. That information cannot be recovered, however, by considering the predictive distribution of the data.

In our opening section we recalled the familiar concern that, with traditional (Neyman-Pearson) hypothesis testing, optional stopping opens the door to foregone conclusions when (fixed sample size) significance levels are used to report evidential import. We have seen how the fear of sampling to a foregone conclusion can be alleviated within the Bayesian paradigm, under the assumption that all probability is countably additive. In short, then there are bounds on how high the probability can be of sampling until the posterior probability reaches a specified level. Moreover, these safeguards are recognized also from the standpoint of an onlooker whose prior may differ from the investigator's. However, the safeguards are no longer in place when the Bayesian analysis appeals to "improper" priors (since they correspond to finitely additive probabilities that are not countably additive). Therefore, it is important, we think, to examine more carefully when the use of "improper" priors lead to foregone conclusions, even though they satisfy de Finetti's or Savage's (1954) standards of coherent opinion.

APPENDIX

For the proof of Theorem 2.1 (Section 2.1), we use a lemma, below, which is a substitution instance of Theorem 2 of Chow and Teicher 1988, 146:

Lemma A: Suppose that $Z_1, Z_2, \ldots$ are i.i.d. with $\mathbf{E}(Z_i) = \mu > 0$ and assume that the moment generating function for Z_i exists. Let $Y_n = \sum_{i=1}^{n} Z_i$, for $n = 1, 2, \ldots$. Fix $c < 0$ and let $M = \mathbf{min}\{n: Y_n \leq c\}$. Then $P(M < \infty) < 1$. □

Theorem 2.1. For each $\theta > \theta_0$, P_θ(experimenter stops inquiry) < 1.

Proof: Let $S_n = \sum_{i=1}^{n} X_i$. Then the experimenter's posterior probability distribution of Θ is

$$\Theta \sim \mathbf{N}\left(\frac{S_n + \lambda\mu}{n+\lambda}, \frac{1}{n+\lambda}\right).$$

Consider the prospect of sampling until the experimenter's posterior probability of H is at least q. The experimenter's posterior probability of H is: $\mathbf{\Phi}\left(\sqrt{n+\lambda}\left[\theta_0 - \frac{S_n + \lambda\mu}{n+\lambda}\right]\right)$, where $\mathbf{\Phi}$ is the standard normal **cdf**. This expression is at least q if and only if

$$S_n \leq n\theta_0 - [\surd(n+\lambda)]\mathbf{\Phi}^{-1}(q) + \lambda(\theta_0 - \mu). \tag{A.1}$$

Note that as p is the experimenter's prior probability of H, $\Phi^{-1}(p) = [\sqrt{\lambda}](\theta_0 - \mu)$.

To apply the lemma, choose $\theta > \theta_0$ and define $Z_i = X_i - (\theta + \theta_0)/2$. Then, $E_\theta(Z_i) > 0$. Write

$$Y_n = \sum_{i=1}^{n} Z_i = S_n - n\frac{\theta_0 + \theta}{2}.$$

Then condition (A.1) becomes:

$$Y_n \leq -[\sqrt{(n+\lambda)}]\Phi^{-1}(q) + \lambda(\theta_0 - \mu) - n(\theta\text{-}\theta_0)/2 = c_n. \quad \text{(A.2)}$$

The experimenter's stopping rule is to halt at the first $n = N$ such that $Y_n \leq c_n$. Observe that c_n is *decreasing* at rate $O(n)$. Thus,

$$\forall(\theta > \theta_0)\ \exists m_\theta \text{ (both: } c_{m\theta} < 0 \text{ and } c_j < c_{m\theta} < c_i \text{ whenever } j > m_\theta > i).$$

Lemma A applies to the Y_i and Z_i with $c = c_{m\theta}$. That is, let $M = \mathbf{min}\{n: Y_n \leq c_{m\theta}\}$. Then, by the Lemma A, $P_\theta(M < \infty) < 1$.

Remark: If $q \geq .5$ then $\Phi^{-1}(q) \geq 0$. Since $q > p$, $[\sqrt{\lambda}]\Phi^{-1}(p) - [\sqrt{(n+\lambda)}]\Phi^{-1}(q) < 0$. Thus, $\lambda(\theta_0 - \mu) - [\sqrt{(n+\lambda)}]\Phi^{-1}(q) < 0$ and $m_\theta = 1$.

Given θ, Z_i is a (normal) random variable whose support is the entire real line. Hence, there is positive probability that the experiment continues at least to the m_θth trial. That is, denote $P_\theta(Y_i > c_i : i = 1, \ldots, m_\theta - 1)$ by $\mathbf{k}$, and we know that $\mathbf{k} > 0$. We use this fact as follows. The experimenter's stopping rule is to halt at the first $n = N$ such that $Y_n \leq c_n$. Thus, as $c_j < c_{m_\theta}$ for $j > m_\theta$, having reached the m_θth trial the experimenter halts no sooner than the first $n = M$ such that $Y_n \leq c_{m_\theta}$. Thus, if the m_θth trial is reached, $N \geq M$. Then

$$P_\theta(N < \infty \mid Y_i > c_i : i = 1, \ldots, m_\theta - 1) \leq P_\theta(M < \infty \mid Y_i > c_i : i = 1, \ldots, m_\theta - 1).$$

But, $P_\theta(M < \infty \mid Y_i > c_i : i = 1, \ldots, m_\theta - 1) \leq P_\theta(M < \infty \mid M \geq m_\theta) < P_\theta(M < \infty) < 1$. The first of these inequalities follows because $c_i > c_{m\theta} : i = 1, \ldots, m_\theta - 1$. The second inequality follows because $P_\theta(M < m_\theta) > 0$. And the third inequality is given by Lemma A. Let $\mathbf{k}'$ denote the conditional probability $P_\theta(N < \infty \mid Y_i > c_i : i = 1, \ldots, m_\theta - 1)$. Thus, $\max\{\mathbf{k}, \mathbf{k}'\} < 1$. By the multiplication theorem, $P_\theta(N < \infty) = 1 - \mathbf{k} + \mathbf{k}\mathbf{k}'$. Hence, $P_\theta(N < \infty) < 1$: there is positive probability that the experiment fails to terminate.□T.2.1

Theorems **2.2** and **2.3** are established in a similar fashion, using Lemma A.

Lemmas **2.1** and **2.2** (Section 2.2) are proven by cases: Use the *Strong Law of Large Numbers* whenever θ is not an endpoint of the hypothesis H, and use the *Law of the Iterated Logarithm* when θ is an endpoint of H.

REFERENCES

Anscombe, F. J. (1954), "Fixed-sample-size analysis of sequential observations", *Biometrics* 10: 89–100.

Ash, R. B. (1972), *Real Analysis and Probability.* New York: Academic Press.

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer-Verlag.

Berti, P., Regazzini, E., and Rigo, P. (1991), "Coherent Statistical Inference and Bayes' Theorem". *Annals of Statistics* 19: 366–381.

Chow, Y. S. and Teicher, H. (1988), *Probability Theory: Independence, Interchangeability, Martingales* (2nd ed.). New York: Springer-Verlag.

Cornfield, J. (1970), "The Frequency Theory of Probability, Bayes' Theorem, and Sequential Clinical Trials." In D. L.Meyer and R. O. Collier, Jr. (eds.), *Bayesian Statistics*. Itasca, IL: Peacock Publishing, pp. 1–28.

De Finetti, B. (1974), *The Theory of Probability* (2 volumes). New York: John Wiley.

Feller, W. (1940), "Statistical Aspects of E.S.P.", *J. Parapsychology* 4, 271–298.

Jeffreys, H. (1939), *Theory of Probability* (3rd edition, 1961). New York: Oxford University Press.

Kadane, J. B., Schervish, M. J., and Seidenfeld, T. (1995), "Reasoning to a Foregone Conclusion", *J. American Statistical Association*, forthcoming.

Kerridge, D. (1963), "Bounds for the Frequency of Misleading Bayes Inferences", *Annals of Math. Statistics* 34: 1109–1110.

Regazzini, E. (1987), "De Finetti's Coherence and Statistical Inference", *Annals of Statistics* 15: 845–864.

Robbins, H. (1952), "Some aspects of the sequential design of experiments", *Bull. American Mathematical Society* 58: 527–535.

Savage, L. J. (1954), *The Foundations of Statistics*. New York: John Wiley.

Savage, L. J. et al. (1962), *The Foundations of Statistical Inference*. London: Methuen.