Sharpening Buffon's Needle[1]

MICHAEL D. PERLMAN AND MICHAEL J. WICHURA*

1. *Introduction*

The famous needle experiment of Buffon and its variations, which provide empirical estimates of the value of π, continue to yield amusing and educational applications of statistical theory. The reader will find thorough discussion of the historical development of this problem as well as recent contributions in [6, 7, 9, 10, 11, 13, 14, 16, 17]. After reviewing Buffon's original experiment in Section 2 of this article, we go on in Sections 3 and 4 to present two variations of the experiment in which statistical theory is utilized to "sharpen" Buffon's needle by obtaining estimators for π with dramatically reduced sampling variance. Our concern has been to (i) design the experiments to increase the efficiency of the estimators (i.e., reduce their variances) and (ii) apply the concepts of sufficiency and completeness, efficiency, and ancillarity, in the guise of the Rao-Blackwell-Lehmann-Scheffe theorems [4, 12], the Cramer-Rao lower bound [15], and the principle of conditionality [1, 2, 3, 5], to obtain alternate estimators which utilize the available statistical information as fully as possible.

2. *Buffon's Experiment—The Single Grid*

In Buffon's original formulation of the problem a needle of length l is thrown at random onto a plane grid of parallel lines separated by a common distance d (see Figure 2.1).

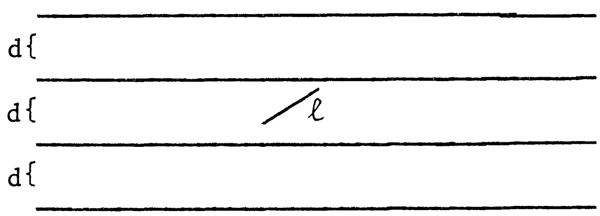

Figure 2.1

If $l \leq d$, the probability p that the needle crosses a line is readily found to be (cf. [6, 17])

$$p = \frac{2l}{\pi d} = 2r\theta \qquad (2.1)$$

where $r = l/d$ and $\theta = 1/\pi$. From now on, however, we forget that $\theta = 1/\pi$ and treat θ as an unknown parameter to be estimated. The condition $0 \leq p \leq 1$ imposes the constraint $0 \leq \theta \leq 1/2r$ in this experiment, while tighter constraints arise in Sections 3 and 4.

If n independent throws of the needle result in N crossings, then $N \sim$ Binomial (n, p) and

$$\hat{\theta}_1 = \frac{N}{2rn} \qquad (2.2)$$

is an unbiased estimator of θ. Since N is a complete, sufficient statistic for θ, the Rao-Blackwell-Lehmann-Scheffe theorems [4, Chapter 3; 12, p. 321] imply that $\hat{\theta}_1$ is the uniformly minimum variance unbiased estimator (UMVUE) of θ. Furthermore, $\hat{\theta}_1$ is the maximum likelihood estimator (MLE) of θ and therefore has 100% asymptotic efficiency in this experiment (cf. Section 4 and [15, Chapter 5]). The variance of $\hat{\theta}_1$ is

$$\operatorname{Var}(\hat{\theta}_1) = \frac{p(1-p)}{4r^2n} = \frac{\theta}{n}\left(\frac{1}{2r} - \theta\right) = \frac{\theta^2}{n}\left(\frac{1}{p} - 1\right) \qquad (2.3)$$

so the efficiency of $\hat{\theta}_1$, as measured by the reciprocal of its variance, is maximized by taking p as close to 1 as possible regardless of the value of θ, i.e., by choosing needle length $l = d$ (this choice of needle length will be seen to be optimal in our subsequent examples as well). In this case $p = 2\theta$, $n\operatorname{Var}(\hat{\theta}_1) = \theta^2[(1/2\theta) - 1]$, and an application of the δ-method [15, p. 385] shows that Buffon's estimator $\hat{\pi}_1 = 1/\hat{\theta}_1$ is an asymptotically unbiased 100% efficient estimator of $1/\theta$ with asymptotic variance

$$\operatorname{AVar}(\hat{\pi}_1) = \pi^4(\operatorname{Var}\hat{\theta}_1) = \frac{\pi^2}{n}\left(\frac{\pi}{2} - 1\right) = \frac{5.63}{n} \qquad (2.4)$$

when $\theta = 1/\pi$. Here, as in the rest of the paper, the asymptotic variance has been numerically evaluated at the true value 3.1416 of π.

3. *Laplace's Experiment—The Double Grid*

In a recent paper Schuster [16] considered a variation of the Buffon experiment, originally studied by Laplace,

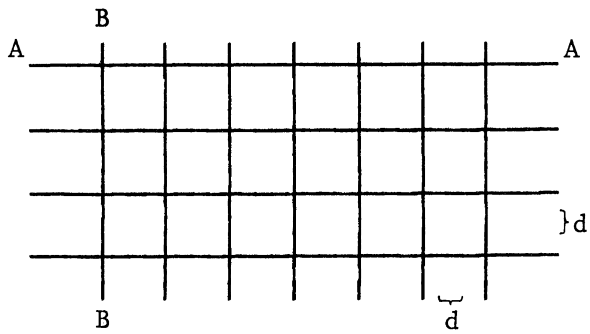

Figure 3.1

[1] This research supported in part by Research Grant No. NSF-GP 32037X from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation.

* Dept. of Statistics, Univ. of Chicago, 1118 E. 58th St., Chicago, IL 60637.

wherein a second grid (B) of parallel lines, also separated by the common distance d, is constructed at right angles to the first grid (A) (see Figure 3.1).
The needle is thrown n times, resulting in N_A crossings of the A-lines and N_B crossings of the B-lines. Then both

$$\hat{\theta}_A = \frac{N_A}{2rn} \quad \text{and} \quad \hat{\theta}_B = \frac{N_B}{2rn} \tag{3.1}$$

are unbiased estimators of θ and have the same distribution as $\hat{\theta}_1$. Schuster proposed the combined estimator

$$\hat{\theta}_2 = \frac{\hat{\theta}_A + \hat{\theta}_B}{2} = \frac{N_A + N_B}{4rn} \tag{3.2}$$

and posed the interesting question of whether the efficiency of $\hat{\theta}_2$ (based on n throws of the needle onto the double grid) is twice that of $\hat{\theta}_1$ (based on n throws onto the single grid). This would indeed be the case if the event that the needle crosses an A-line were independent of the event that it crosses a B-line for then $\hat{\theta}_A$ and $\hat{\theta}_B$ would be independent. A little reflection, however, shows that these events, and therefore $\hat{\theta}_A$ and $\hat{\theta}_B$, are negatively correlated, so in fact the efficiency of the combined estimator $\hat{\theta}_2$ will be *greater* than twice the efficiency of $\hat{\theta}_1$. This idea of combining *antithetic* (i.e., negatively correlated) *variates* to obtain an estimator with reduced variance is well-known to statisticians; see for example [8, 9].

The variance of $\hat{\theta}_2$ is readily calculated, since the necessary crossing probabilities were obtained long ago by Laplace. Denote by p_A the probability of crossing an A-line, by p_{AB} the probability of simultaneously crossing an A-line and a B-line, by $p_{A\bar{B}}$ the probability of crossing an A-line but not crossing a B-line, and similarly define p_B, $p_{\bar{A}B}$, $p_{\overline{AB}}$. Then from [17, p. 255–6] we have that

$$p_{\overline{AB}} = 1 - 4r\theta + r^2\theta, \tag{3.3}$$

while $p_A = p_B = 2r\theta$ from (2.1), so

$$p_{AB} = p_A + p_B + p_{\overline{AB}} - 1 = r^2\theta. \tag{3.4}$$

Introduce the indicator random variables

$$I_i^A = \begin{cases} 1 & \text{if an } A\text{-line is crossed on the } i\text{th throw} \\ 0 & \text{if not,} \end{cases}$$

and similarly define I_i^B, so that

$$N_A = \sum_{i=1}^{n} I_i^A, \qquad N_B = \sum_{i=1}^{n} I_i^B,$$

$$\hat{\theta}_2 = \frac{1}{4rn}\sum_{i=1}^{n}(I_i^A + I_i^B).$$

The n pairs $(I_1^A, I_1^B), \ldots, (I_n^A, I_n^B)$ are independent but I_i^A and I_i^B are dependent, so

$$\begin{aligned}\operatorname{Var}(\hat{\theta}_2) &= \frac{1}{16r^2n}[\operatorname{Var}(I_1^A) + \operatorname{Var}(I_1^B) + 2\operatorname{cov}(I_1^A, I_1^B)] \\ &= \frac{1}{16r^2n}[p_A(1-p_A) + p_B(1-p_B) \\ &\quad + 2(p_{AB} - p_Ap_B)] \\ &= \frac{\theta}{n}\left[\frac{1}{4r} + \frac{1}{8} - \theta\right].\end{aligned} \tag{3.5}$$

It is again apparent that the estimator $\hat{\theta}_2$ has greatest efficiency when $r = 1$, i.e., when $l = d$, for all values of θ. When $l = d$ we find $p_{\overline{AB}} = 1 - 3\theta$, which imposes the tighter constraint $0 \le \theta \le \frac{1}{3}$ (and shows, incidentally, that $\pi \ge 3$). Also, $n\operatorname{Var}(\hat{\theta}_2) = \theta^2[(3/8\theta) - 1]$ so

$$\operatorname{AVar}(\hat{\pi}_2) = \frac{\pi^2}{n}\left(\frac{3\pi}{8} - 1\right) = \frac{1.76}{n} \tag{3.6}$$

when $\theta = 1/\pi$, where $\hat{\pi}_2 = 1/\hat{\theta}_2$. Comparing (2.4) and (3.6) it is seen that by introducing the second set of grid lines we have obtained an estimator $\hat{\pi}_2$ which is $5.63/1.76 = 3.20$ times as efficient as $\hat{\pi}_1$ when $\theta = 1/\pi$. One throw of the needle onto the double grid contains *at least* 3.20 times the statistical information about the value of π as one throw onto the single grid.

We have stressed the phrase "at least" in the preceding sentence to raise an important question not considered by Schuster. Does the estimator $\hat{\theta}_2$ itself fully utilize all the information about θ provided by the double grid experiment? The answer is No, for we shall see that a complete and sufficient statistic exists for this experiment, but $\hat{\theta}_2$ is not a function of this statistic.

The full information obtained from n throws of the needle onto the double grid is summarized by the statistic $\mathbf{N} = (N_{AB}, N_{A\bar{B}}, N_{\bar{A}B}, N_{\overline{AB}})$, where N_{AB} is the number of times the needle simultaneously crosses an A-line and a B-line, etc. Clearly $\mathbf{N}$ has the multinomial distribution with cell probabilities $(p_{AB}, p_{A\bar{B}}, p_{\bar{A}B}, p_{\overline{AB}})$. The values of p_{AB} and $p_{\overline{AB}}$ are given by (3.3) and (3.4), while

$$p_{\bar{A}B} = p_{A\bar{B}} = p_A - p_{AB} = r(2-r)\theta. \tag{3.8}$$

Thus the probability distribution of $\mathbf{N}$ is given by

$$\begin{aligned}P_\theta(\mathbf{N} = \mathbf{n}) &= c(\mathbf{n})(p_{AB})^{n_{AB}}(p_{A\bar{B}})^{n_{A\bar{B}}}(p_{\bar{A}B})^{n_{AB}}(p_{\overline{AB}})^{n_{\overline{AB}}} \\ &= c(\mathbf{n})h(\mathbf{n})\theta^{(n_{AB}+n_{A\bar{B}}+n_{\bar{A}B})}[1 - m\theta]^{n_{\overline{AB}}}\end{aligned} \tag{3.9}$$

where

$$\begin{aligned}c(\mathbf{n}) &= \frac{n!}{(n_{AB})!(n_{A\bar{B}})!(n_{\bar{A}B})!(n_{\overline{AB}})!}, \\ h(\mathbf{n}) &= r^{2n_{AB}}[r(2-r)]^{n_{A\bar{B}}+n_{\bar{A}B}}, \\ m &= 4r - r^2.\end{aligned} \tag{3.10}$$

Since $n_{\overline{AB}} = n - (n_{AB} + n_{A\bar{B}} + n_{\bar{A}B})$, (3.9) and the Factorization Criterion for sufficiency [4, p. 115] imply that $N_{AB} + N_{A\bar{B}} + N_{\bar{A}B}$ is a sufficient statistic for θ. If we define N_j to be the number of times in n throws that the needle crosses exactly j lines ($j = 0, 1, 2$), so that $N_0 = N_{\overline{AB}}$, $N_1 = N_{A\bar{B}} + N_{\bar{A}B}$, $N_2 = N_{AB}$, and $\sum N_j = n$, then the sufficient statistic can be expressed as $N_1 + N_2$, the number of times in n throws that the needle crosses *at least one line*.

By (3.3), $N_1 + N_2 \sim \text{Binomial}(n, p^*)$ where $p^* = m\theta$, so $N_1 + N_2$ is a complete, as well as sufficient, statistic for θ. By the Rao-Blackwell-Lehmann-Scheffe theorems,

$$\hat{\theta}_3 = \frac{N_1 + N_2}{mn} \tag{3.11}$$

is the UMVUE and, being the MLE of θ, has 100% asymptotic efficiency in the double grid experiment. Its variance is

$$\text{Var}(\hat{\theta}_3) = \frac{\theta}{n}\left(\frac{1}{m} - \theta\right) \tag{3.12}$$

which, by (3.10), is minimized by the needle length $l = d$, regardless of the value of θ. In this case $m = 3$, $p^* = 3\theta$, $n\text{Var}(\hat{\theta}_3) = \theta^2[(1/3\theta) - 1]$, and

$$\text{AVar}(\hat{\pi}_3) = \frac{\pi^2}{n}\left(\frac{\pi}{3} - 1\right) = \frac{0.466}{n} \tag{3.13}$$

when $\theta = 1/\pi$, where $\hat{\pi}_3 = 1/\hat{\theta}_3$. Comparing (3.6) and (3.13) we see that the fully efficient estimator $\hat{\pi}_3$ is $1.76/.466 = 3.77$ times as efficient as Schuster's estimator $\hat{\pi}_2$ when $\theta = 1/\pi$, reflecting the fact that $\hat{\theta}_2$ is based on

$$N_A + N_B = N_{AB} + N_{A\bar{B}} + N_{AB} + N_{\bar{A}B} = N_1 + 2N_2,$$

which is not a function of the sufficient statistic $N_1 + N_2$. A moral here is that the method of antithetic variates, advocated in [8] and [9] for a wide variety of problems, should not be applied before a careful search for a sufficient statistic. Furthermore, comparing (2.4) and (3.13) our analysis shows that one throw of the needle onto the double grid contains not 3.20 but *exactly* $5.63/.466 = 12.08$ times the statistical information about the value of π as one throw onto the single grid, reflecting the fact that the maximum achievable value $p^* = 3/\pi = .955$ obtained here is much closer to 1 than the maximum achievable value $p = 2/\pi = .637$ obtained in Section 2.

4. *Uspensky's Experiment—The Triple Grid*

Enticed by the hefty 12-fold increase in efficiency provided by the second set of grid lines, we now introduce a third set. To maintain symmetry consider three sets of grid lines, labelled C, D, E, lying at 60° angles to each other and having common grid width d, so that the plane is covered by equilateral triangles of altitude d (see Figure 4.1). A needle of length $l \leq d$ is thrown at random onto this triple grid n times. This setup appears in problem 8 of Uspensky [17, p. 258] and accordingly we refer to it as the "Uspensky experiment."

The full data is summarized by the statistic

$$\mathbf{N} = (N_{CDE}, N_{CD\bar{E}}, N_{C\bar{D}E}, N_{\bar{C}DE}, N_{C\overline{DE}}, N_{\bar{C}D\bar{E}}, N_{\overline{CD}E}, N_{\overline{CDE}})$$

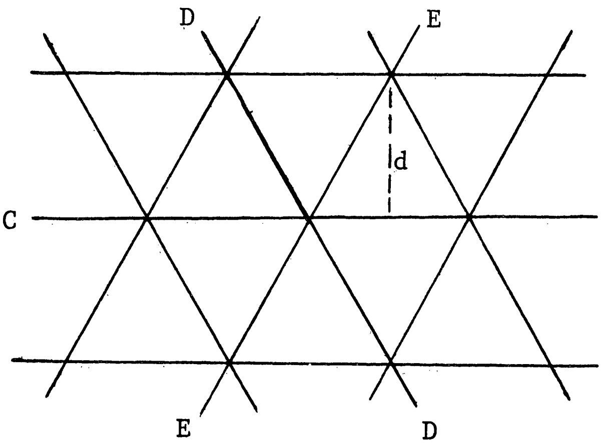

Figure 4.1

where N_{CDE} is the number of times the needle simultaneously crosses a C-line, a D-line, and an E-line, etc. The eight crossing probabilities $p_{CDE}, p_{CD\bar{E}}, \ldots, p_{\overline{CDE}}$ are computed as follows. From [17, p. 258, problem 8],

$$p_{\overline{CDE}} = 1 + \frac{r^2}{2} - \frac{3r\theta}{2}\left(4 - \frac{\sqrt{3}r}{2}\right),$$

where $r = l/d$. Also $p_C = p_D = p_E = 2r\theta$ as in Section 1, while an argument similar to that of Schuster [16, p. 28] shows that

$$p_{CD} = p_{CE} = p_{DE} = r^2\left(\frac{1}{12} + \frac{\sqrt{3}\theta}{2}\right).$$

Then by symmetry and the method of inclusion-exclusion,

$$p_{CDE} = 1 - 3p_C + 3p_{CD} - p_{\overline{CDE}} = \tfrac{1}{4}r^2(3\sqrt{3}\theta - 1),$$

$$p_{CD\bar{E}} = p_{CD} - p_{CDE} = \frac{r^2}{3} - \frac{\sqrt{3}r^2\theta}{4},$$

$$p_{C\overline{DE}} = p_C - p_{CD} - p_{CE} + p_{CDE} = \left(2r - \frac{\sqrt{3}r^2}{4}\right)\theta - \frac{5r^2}{12}.$$

A partial reduction of the full data $\mathbf{N}$ can be achieved via sufficiency. By writing out the multinomial probability distribution of $\mathbf{N}$, noting that $p_{CD\bar{E}} = p_{C\bar{D}E} = p_{\bar{C}DE}$ and $p_{C\overline{DE}} = p_{\bar{C}D\bar{E}} = p_{\overline{CD}E}$ by symmetry, and applying the Factorization Criterion, we find that

$$\mathbf{N}' \equiv (N_0, N_1, N_2, N_3)$$

is a sufficient statistic for θ, where N_j is the number of times the needle crosses exactly j lines, i.e., $N_0 = N_{\overline{CDE}}$, $N_1 = N_{C\overline{DE}} + N_{\bar{C}D\bar{E}} + N_{\overline{CD}E}$, etc. The statistic $\mathbf{N}'$ has the multinomial distribution with cell probabilities given by

$$p_0(\theta) = p_{\overline{CDE}} = 1 + \frac{r^2}{2} - a\theta,$$

$$p_1(\theta) = 3p_{C\overline{DE}} = -\frac{5r^2}{4} + a\theta,$$

$$p_2(\theta) = 3p_{CD\bar{E}} = r^2 - b\theta,$$

$$p_3(\theta) = p_{CDE} = -\frac{r^2}{4} + b\theta, \quad (4.1)$$

where

$$a = \frac{3r}{2}\left(4 - \frac{\sqrt{3}r}{2}\right), \quad b = \frac{3\sqrt{3}r^2}{4}. \quad (4.2)$$

The requirement that each $p_i(\theta) \geq 0$, $i = 0, \ldots, 3$, imposes the restriction

$$\max\{5r^2/4a, 1/3\sqrt{3}\} \leq \theta \leq \min\{(1 + r^2/2)/a, 4/3\sqrt{3}\} \quad (4.3)$$

on the range of the parameter θ.

No further reduction by sufficiency is possible, for the Lehmann-Scheffe criterion [12, Theorem 6.1] can be applied to show that $\mathbf{N}'$ is a *minimal* sufficient statistic. However, $(N_2 + N_3) - (3nr^2)/4$ is a non-trivial unbiased estimator of zero, so $\mathbf{N}'$ is not complete and the Rao-Blackwell apparatus cannot be used to generate a UMVUE of θ, unlike the single-grid or double-grid cases.

In fact no UMVUE exists. The Cramer-Rao lower bound [15, eqn. (5a.2.15)] for the variance of an unbiased estimator of θ based on $\mathbf{N}'$ is $1/nI(\theta)$, where $I(\theta)$ is the Fisher information number

$$I(\theta) = \sum_{i=0}^{3}\left[\frac{\partial}{\partial\theta}\log p_i(\theta)\right]^2 p_i(\theta)$$

$$= a^2\left[\frac{1}{p_0(\theta)} + \frac{1}{p_1(\theta)}\right] + b^2\left[\frac{1}{p_2(\theta)} + \frac{1}{p_3(\theta)}\right]. \quad (4.4)$$

For a fixed value of θ_0 in the range (4.3), consider the linear function

$$S(\theta_0) = \sum_{i=0}^{3} N_i \frac{\partial}{\partial\theta}[\log p_i(\theta)]_{\theta_0}.$$

A straightforward computation shows that for all θ,

$$E_\theta[S(\theta_0)] = n(\theta - \theta_0)I(\theta_0),$$

while for $\theta = \theta_0$,

$$\mathrm{Var}_{\theta_0}[S(\theta_0)] = nI(\theta_0).$$

Hence for each θ_0, the linear estimator

$$\theta_0 + S(\theta_0)/nI(\theta_0) \quad (4.5)$$

is an unbiased estimator of θ and its variance achieves the Cramer-Rao lower bound of $1/nI(\theta_0)$ when $\theta = \theta_0$. Since the MVUE at θ_0 is essentially unique [15, p. 318(b)] and the estimator (4.5) depends non-trivially upon θ_0, this shows that a UMVUE cannot exist.

Even though no UMVUE for θ exists, we will be able to find asymptotically efficient estimators of θ. In this multinomial problem the standard regularity assumptions are satisfied (see [15], Chapter 5) so the ratio

$$\mathrm{AEff}(\hat{\theta}) \equiv \frac{1/nI(\theta)}{\mathrm{AVar}(\hat{\theta})} \leq 1 \quad (4.6)$$

measures the asymptotic efficiency of a consistent asymptotically normal (CAN) estimator $\hat{\theta}$ based on n trials. In view of (4.6), we first design the experiment (i.e., choose a needle length/grid width ratio $r \equiv l/d$) to maximize the information number $I(\theta)$ given by (4.4) so that our asymptotically efficient estimators will have the smallest possible asymptotic variance. Differentiation shows that the lower bound in (4.3) increases with r while the upper bound decreases with r, $0 \leq r \leq 1$, so θ must in fact satisfy the restriction

$$0.26590 \doteq 1.25/a \leq \theta \leq 1.5/a \doteq 0.31908 \quad (4.7)$$

obtained by setting $r = 1$ in (4.3). (Incidentally, the upper bound in (4.7) implies the interesting and accurate lower bound $4 - \sqrt{3}/2 = 3.1340$ for π.) We claim that for each θ satisfying (4.7), $I(\theta)$ is maximized when $r = 1$. This follows in a straightforward manner by differentiating each of the quantities

$$\frac{a^2}{p_0(\theta)}, \frac{a^2}{p_1(\theta)}, b^2\left[\frac{1}{p_2(\theta)} + \frac{1}{p_3(\theta)}\right]$$

with respect to r. Therefore we set $r = 1$ (i.e., $l = d$) for the remainder of this section, in which case the cell probabilities for the distribution of $\mathbf{N}'$ take the form

$$p_0(\theta) = \tfrac{3}{2} - a\theta, \quad p_1(\theta) = -\tfrac{5}{4} + a\theta$$

$$p_2(\theta) = 1 - b\theta, \quad p_3(\theta) = -\tfrac{1}{4} + b\theta \quad (4.8)$$

where now

$$a = 6 - \frac{3\sqrt{3}}{4}, \quad b = \frac{3\sqrt{3}}{4}, \quad (4.9)$$

and θ is constrained by (4.7).

We shall obtain asymptotically efficient estimators through the methods of maximum likelihood, scoring, minimum modified chi-square, and the principle of conditionality. The standard regularity conditions being satisfied, the maximum likelihood estimator $\hat{\theta}_4$ is a CAN estimator with 100% asymptotic efficiency for all possible values of θ. The MLE $\hat{\theta}_4$ is that value of θ maximizing the likelihood function $\prod_{i=0}^{3}[p_i(\theta)]^{N_i}$. By differentiation it is easily seen that the log likelihood function is concave in θ, so $\hat{\theta}_4$ is the unique root of the likelihood equation

$$0 = \sum_{i=0}^{3} N_i \frac{\partial}{\partial\theta}\log p_i(\theta).$$

This amounts to a cubic equation which can be solved numerically for $\hat{\theta}_4$. Since $\mathrm{AEff}(\hat{\theta}_4) = 1$, we find from (4.6), (4.4), (4.8), and (4.9) that for $\hat{\pi}_4 = 1/\hat{\theta}_4$,

$$\mathrm{AVar}(\hat{\pi}_4) = \pi^4\,\mathrm{AVar}(\hat{\theta}_4) = \frac{\pi^4}{nI(1/\pi)}, \quad (4.10)$$

when $\theta = 1/\pi$. The numerical value of $\mathrm{AVar}(\hat{\pi}_4)$ is given, along with comparisons with (3.13) and (2.4), at the end of this section (see (4.32)).

Asymptotically efficient estimators are not unique, and a computationally simpler estimator with 100% efficiency can be obtained from (4.5) by the method of scoring (see [15], Chapter 5 and [18], Theorem 5.5.4). Let $\{\theta_n'\}$ be any consistent sequence of estimators such that

$$n^{1/4}(\theta_n' - \theta) \to 0 \text{ in probability.} \tag{4.11}$$

Then the estimator

$$\hat{\theta}_5 = \theta_n' + S(\theta_n')/nI(\theta_n') \tag{4.12}$$

obtained from (4.5) is also CAN with 100% asymptotic efficiency. A convenient choice for θ_n' is either $\hat{\theta}^*$ given by (4.19) or $\hat{\theta}_9$ given by (4.28), each of which converges to θ at the rate $n^{-1/2}$.

Another estimator with 100% asymptotic efficiency is easily obtained via the minimum modified chi-square method [15, p. 352]. This estimator $\hat{\theta}_6$ is defined to be that value of θ which minimizes the expression

$$\sum_{i=0}^{3} \frac{[N_i - np_i(\theta)]^2}{N_i}.$$

Differentiation with respect to θ shows that

$$\hat{\theta}_6 = \frac{a\left(\frac{3}{2N_0} + \frac{5}{4N_1}\right) + b\left(\frac{1}{N_2} + \frac{1}{4N_3}\right)}{a^2\left(\frac{1}{N_0} + \frac{1}{N_1}\right) + b^2\left(\frac{1}{N_2} + \frac{1}{N_3}\right)}.$$

It is of interest to apply the principle of conditionality to obtain alternate estimators with 100% asymptotic efficiency. From (4.8) notice that

$$p_0(\theta) + p_1(\theta) = \tfrac{1}{4}, \qquad p_2(\theta) + p_3(\theta) = \tfrac{3}{4} \tag{4.13}$$

for all values of θ, which implies that the (binomial) distributions of $N_0 + N_1$ and $N_2 + N_3 = n - (N_0 + N_1)$ do not depend on the unknown parameter θ. A statistic whose marginal distribution is completely known but which is not independent of the minimal sufficient statistic is called an *ancillary statistic* [1, 2, 3, 5, 15], a notion introduced by Fisher. Cox [3] writes: "An ancillary statistic by itself gives no information about the parameter of interest, but specifies the precision with which inference is possible in the sample under analysis. It is fairly compelling that the probability distribution used in inference should be conditional on the observed value of the ancillary statistic." This "fairly compelling" dictum is often referred to as the *principle of conditionality.* A quantitative justification for the application of this principle in estimation problems stems from the fact [see [3, eqn. (15)] and (4.18)] that the (unconditional) Fisher information $nI(\theta)$ is the expectation of the conditional information given the value of the ancillary statistic. Hence one expects that an optimal conditional estimator should be nearly optimal unconditionally, at least for large samples. This is indeed the case in the multinomial estimation problem under consideration, as we shall now show.

Given the value of the ancillary statistic $N_0 + N_1$, the conditional distribution of the minimal sufficient statistic (N_0, N_1, N_2, N_3) is described as follows: (N_0, N_1) and (N_2, N_3) are (conditionally) independent,

$$N_0 \sim \text{Binomial}(N_0 + N_1, 4p_0(\theta)) \tag{4.14}$$

$$N_2 \sim \text{Binomial}(N_2 + N_3, 4p_2(\theta)/3). \tag{4.15}$$

The (conditional) information numbers based on (4.14) and (4.15) separately are computed to be

$$I^*(\theta \mid N_0 + N_1) = 4(N_0 + N_1)a^2\left[\frac{1}{p_0(\theta)} + \frac{1}{p_1(\theta)}\right] \equiv 4(N_0 + N_1)I^*(\theta), \tag{4.16}$$

$$I^{**}(\theta \mid N_0 + N_1) = \tfrac{4}{3}(N_2 + N_3)b^2\left[\frac{1}{p_2(\theta)} + \frac{1}{p_3(\theta)}\right] \equiv \tfrac{4}{3}(N_2 + N_3)I^{**}(\theta), \tag{4.17}$$

respectively. By the additivity of information numbers based on independent experiments [5, p. 329 (i)], the (conditional) information number based on (4.14) and (4.15) together is

$$I(\theta \mid N_0 + N_1) = 4(N_0 + N_1)I^*(\theta) + \tfrac{4}{3}(N_2 + N_3)I^{**}(\theta).$$

Since $N_0 + N_1 \sim \text{Binomial}(n, \frac{1}{4})$, notice that

$$E_\theta I(\theta \mid N_0 + N_1) = n(I^*(\theta) + I^{**}(\theta)) = nI(\theta), \tag{4.18}$$

the unconditional information number given by (4.4). The estimators

$$\hat{\theta}^* = \frac{1}{4a}\left(6 - \frac{N_0}{N_0 + N_1}\right) \text{ and } \hat{\theta}^{**} = \frac{3}{4b}\left(\frac{4}{3} - \frac{N_2}{N_2 + N_3}\right) \tag{4.19}$$

are (conditionally) independent, unbiased, minimum variance estimators of θ based on (4.14) and (4.15), respectively, with (conditional) variances $1/I^*(\theta \mid N_0 + N_1)$ and $1/I^{**}(\theta \mid N_0 + N_1)$ (we ignore the exceptional cases $N_0 + N_1 = 0$, $N_0 + N_1 = n$, which occur with negligible probability when n is large). Among all convex combinations $\alpha\hat{\theta}^* + (1-\alpha)\hat{\theta}^{**}$, $0 \le \alpha \le 1$, the minimum (conditional) variance $1/I(\theta \mid N_0 + N_1)$ is attained when

$$\alpha = \alpha(\theta, N_0 + N_1) = I^*(\theta \mid N_0 + N_1)/[I^*(\theta \mid N_0 + N_1) + I^{**}(\theta \mid N_0 + N_1)].$$

Since $N_0 + N_1 = n/4 + 0_p(n^{1/2})$, we approximate $\alpha(\theta, N_0 + N_1)$ by

$$\bar{\alpha}(\theta) = \alpha(\theta, n/4) = I^*(\theta)/[I^*(\theta) + I^{**}(\theta)], \tag{4.20}$$

and in turn we estimate $\bar{\alpha}(\theta)$ by $\bar{\alpha}(\theta_n')$, where θ_n' is any consistent sequence of estimators for θ satisfying

$$\theta_n' \to \theta \text{ in probability,} \tag{4.21}$$

which is a slightly weaker requirement than (4.11) needed for the method of scoring. Either $\hat{\theta}^*$ in (4.19) or

$\hat{\theta}_9$ in (4.28) is a convenient choice for θ_n'. Thus we are led by the principle of conditionality to the estimator

$$\hat{\theta}_7 = \bar{\alpha}(\theta_n')\hat{\theta}^* + (1 - \bar{\alpha}(\theta_n'))\hat{\theta}^{**}. \qquad (4.22)$$

To evaluate the asymptotic efficiency of $\hat{\theta}_7$, first apply the multivariate δ-method [15, p. 388 (iii)] and the asymptotic normality of the multinomial statistic (N_0, N_1, N_2, N_3) [15, p. 383 (ii)] to obtain that

$$\sqrt{n}[(\hat{\theta}^*, \hat{\theta}^{**}) - (\theta, \theta)] \rightarrow N\left[0, \begin{pmatrix} I^*(\theta) & 0 \\ 0 & I^{**}(\theta) \end{pmatrix}^{-1}\right] \qquad (4.23)$$

in distribution. Then by (4.18), (4.20), and (4.23),

$$\sqrt{n}[\{\bar{\alpha}(\theta)\hat{\theta}^* + (1 - \bar{\alpha}(\theta))\hat{\theta}^{**}\} - \theta] \rightarrow N[0, 1/I(\theta)] \qquad (4.24)$$

in distribution. Since $\bar{\alpha}(\theta)$ is continuous in θ, (4.21) implies that $\bar{\alpha}(\theta_n') \rightarrow \bar{\alpha}(\theta)$ in probability, so that

$$\sqrt{n}[\hat{\theta}_7 - \{\bar{\alpha}(\theta)\hat{\theta}^* + (1 - \bar{\alpha}(\theta))\hat{\theta}^{**}\}] = \sqrt{n}(\hat{\theta}^* - \hat{\theta}^{**})(\bar{\alpha}(\theta_n') - \bar{\alpha}(\theta)) \rightarrow 0 \qquad (4.25)$$

in probability. We conclude from (4.24) and (4.25) that

$$\sqrt{n}(\hat{\theta}_7 - \theta) \rightarrow N[0, 1/I(\theta)] \qquad (4.26)$$

in distribution, so $\hat{\theta}_7$ is indeed a 100% asymptotically efficient CAN estimator.

For the sake of some interesting comparisons, we introduce three other estimators $\hat{\theta}_8$, $\hat{\theta}_9$, $\hat{\theta}_{10}$. Numerical computations show that the minimum value of $\bar{\alpha}(\theta)$ over the range (4.7) is approximately .957, attained near $\theta = .292$, and $\bar{\alpha}(\theta) \rightarrow 1$ as θ approaches the extremes in (4.7). Thus $\hat{\theta}_7$ assigns much more weight to $\hat{\theta}^*$ than $\hat{\theta}^{**}$, which suggests consideration of the relatively simple estimator $\hat{\theta}_8 = \hat{\theta}^*$ defined in (4.19). From (4.23),

$$\sqrt{n}(\hat{\theta}_8 - \theta) \rightarrow N[0, 1/I^*(\theta)] \text{ in distribution.} \qquad (4.27)$$

Next we base $\hat{\theta}_9$, the triple-grid analog of $\hat{\theta}_3$, on the statistic $N_1 + N_2 + N_3$ which counts the number of times the needle *crosses at least one* line. Since $N_1 + N_2 + N_3 \sim \text{Binomial}(n, \tilde{p}(\theta))$ where $\tilde{p}(\theta) = p_1(\theta) + p_2(\theta) + p_3(\theta) = a\theta - \frac{1}{2}$, the simple estimator

$$\hat{\theta}_9 = \frac{1}{a}\left(\frac{N_1 + N_2 + N_3}{n} + \frac{1}{2}\right) \qquad (4.28)$$

is unbiased and

$$\sqrt{n}(\hat{\theta}_9 - \theta) \rightarrow N[0, \tilde{p}(\theta)(1 - \tilde{p}(\theta))/a^2] \text{ in distribution.} \qquad (4.29)$$

Finally consider the Schuster-type estimator $\hat{\theta}_{10}$, analogous to $\hat{\theta}_2$, based on the antithetic variates N_C, N_D, N_E, where N_C is the number of times the needle crosses a C grid line, etc. Since N_C, N_D, N_E are each Binomial$(n, 2\theta)$, the estimator

$$\hat{\theta}_{10} = \frac{N_C + N_D + N_E}{6n} = \frac{N_1 + 2N_2 + 3N_3}{6n} \qquad (4.30)$$

is unbiased; proceeding as in (3.5) we find that

$$\sqrt{n}(\hat{\theta}_{10} - \theta) \rightarrow N\left[0, \frac{1}{72} + \frac{\theta}{6}\left(1 + \frac{\sqrt{3}}{2}\right) - \theta^2\right] \text{ in distribution.} \qquad (4.31)$$

As promised earlier, here are the numerical values of the asymptotic variances of the estimators $\hat{\pi}_i = 1/\hat{\theta}_i$, $i = 4, \ldots, 10$, evaluated at $\theta = 1/\pi$ according to the relation $\text{AVar}(\hat{\pi}_i) = \pi^4 \text{AVar}(\hat{\theta}_i)$. From (4.10), (4.27), (4.29), (4.31), and the facts that $\hat{\theta}_4$, $\hat{\theta}_5$, $\hat{\theta}_6$, $\hat{\theta}_7$ are each 100% asymptotically efficient we compute that

$$\text{nAVar}(\hat{\pi}_i) = \begin{cases} \pi^4/I(1/\pi) & = 0.01577 \quad i=4,5,6,7 \\ \pi^4/I^*(1/\pi) & = 0.01580 \quad i=8 \\ \pi^4\tilde{p}(1/\pi)(1 - \tilde{p}(1/\pi))/a^2 & = 0.01597 \quad i=9 \\ \pi^4/72 + \pi^3(1 + \sqrt{3}/2) - \pi^2 & = 1.12638 \quad i=10. \end{cases} \qquad (4.32)$$

The asymptotic efficiencies of $\hat{\pi}_8$, $\hat{\pi}_9$, and $\hat{\pi}_{10}$ are thus 99.8%, 98.7%, and 1.4%, respectively. It is both surprising that the antithetic-variate estimator $\hat{\pi}_{10}$ should fare so poorly, and pleasing that the computationally simple estimators $\hat{\pi}_8$ and $\hat{\pi}_9$ should be so nearly efficient. The high efficiency of $\hat{\pi}_8$ is due to the fact that $\hat{\pi}_8$ and $\hat{\pi}_7$ are virtually identical when $\theta = 1/\pi$ since $\bar{\alpha}(1/\pi) = .998$, while that of $\hat{\pi}_9$ reflects the fact that $\tilde{p}(1/\pi) = .9964$ is extremely close to 1. Lastly, comparing (2.4), (3.13), and (4.32), one sees that the triple grid experiment is .466/.01577 = 29.5 times as efficient as the double grid experiment and a remarkable 5.63/.01577 = 357 times as efficient as the single grid experiment for the estimation of π.

With the aid of a computer we simulated $n = 10{,}000$ throws of a needle of length $l = d$ onto the triple grid. This provides as much information about the value of π as 3,570,000 throws onto Buffon's original single grid. The observed data were $N_0 = 32$, $N_1 = 2429$, $N_2 = 5864$, and $N_3 = 1675$. The estimators $\hat{\pi}_4$, $\hat{\pi}_5$, $\hat{\pi}_6$, $\hat{\pi}_7$, $\hat{\pi}_8$, $\hat{\pi}_9$ each turned out to be either 3.1407 or 3.1408 with standard error 0.0013; $\hat{\pi}_{10} = 3.1279$ with standard error 0.011 finished a weak seventh. These values are all consistent with the true value $\pi = 3.1416$.

Although the problem of estimating π is solely of pedagogical interest, the methods of estimation discussed in this section are quite generally applicable. For example, multinomial distributions whose cell probabilities are simple functions of a parameter θ commonly arise in genetical experiments (e.g. [15, p. 368]; also [1, 3, 5]). Maximum likelihood, the method of scoring, minimum modified chi-square, and conditioning on an ancillary statistic are tools which every statistician can apply profitably.

Acknowledgment

We wish to thank Professors V. P. Bhapkar, Patrick Billingsley, and David L. Wallace for helpful comments.

REFERENCES

[1] Basu, D.: "Recovery of Ancillary Information," *Sankhya*, Series A, 26 (July 1964), 3–16.

[2] Cox, D. R.: "Some Problems Connected with Statistical Inference," *Annals of Mathematical Statistics*, 29 (June 1958), 357–372.

[3] ——: "The Choice Between Alternative Ancillary Statistics," *Journal of the Royal Statistical Society*, Series B, 33 (1971), 251–255.

[4] Ferguson, T. S.: *Mathematical Statistics*, Academic Press, New York (1967).

[5] Fisher, R. A.: *Statistical Methods and Scientific Inference*, Oliver and Boyd, London (1956).

[6] Gnedenko, B. V.: *Theory of Probability*, Chelsea, New York (1962).

[7] Gridgeman, N. T.: "Geometric Probability and the Number π," *Scripta Mathematica*, 25 (1960), 183–195.

[8] Hammersley, J. M. and Handscomb, D. C.: *Monte Carlo Methods*, Methuen, London (1964).

[9] Hammersley, J. M. and Morton, K. W.: "A New Monte Carlo Technique: Antithetic Variates," *Proceedings of the Cambridge Philosophical Society*, 52 (1956), 449–475.

[10] Kahan, B. C.: "A Practical Demonstration of a Needle Experiment to Give a Number of Concurrent Estimates of π," *Journal of the Royal Statistical Society*, Series A, 124 (1961), 227–239.

[11] Kendall, M. G. and Moran, P. A. P.: *Geometrical Probability*, Charles Griffin, London (1963).

[12] Lehmann, E. L. and Scheffe, H.: "Completeness, Similar Regions, and Unbiased Estimation—Part I," *Sankhya*, 10 (1950, part 4), 305–340.

[13] Mantel, N.: "An Extension of the Buffon Needle Problem," *Annals of Mathematical Statistics*, 24 (1953), 674–677.

[14] Ramaley, J. F.: "Buffon's Noodle Problem," *American Mathematical Monthly*, 76 (October 1969), 916–918.

[15] Rao, C. R.: *Linear Statistical Inference and its Applications*, second edition, John Wiley and Sons, New York (1973).

[16] Schuster, E. F.: "Buffon's Needle Experiment," *American Mathematical Monthly*, 81 (January 1974), 26–29.

[17] Uspensky, J. V.: *Introduction to Mathematical Probability*, McGraw-Hill, New York (1937).

[18] Zacks, S.: *The Theory of Statistical Inference*, Wiley, New York (1971).

U S POSTAL SERVICE
STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION
(Act of August 12, 1970: Section 3685, Title 39, United States Code

1 TITLE OF PUBLICATION: THE AMERICAN STATISTICIAN

2 DATE OF FILING: 9/12/75

3 FREQUENCY OF ISSUE: Quarterly - February, May, August and November

3A ANNUAL SUBSCRIPTION PRICE: $8.00

4 LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, city, county, state and ZIP code) (Not printers)*: American Statistical Association, 806 15th St., N. W., Washington, D. C. 20005

5 LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*: American Statistical Association, 806 15th St., N. W., Washington, D. C. 20005

6 NAMES AND ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and address)*: American Statistical Association, 806 15th St., N. W., Washington, D. C. 20005

EDITOR *(Name and address)*: Donald F. Morrison, Stat. & Operations Res. Dept., Wharton School, Univ. of Pennsylvania Philadelphia, PA 19174

MANAGING EDITOR *(Name and address)*: None - Business Manager, Edgar M. Bisgyer, 806 15th St., N. W., Washington, D. C. 20005

7 OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)*

NAME	ADDRESS
American Statistical Association	806 15th St., N. W., Washington, D. C. 20005

8 KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

NAME	ADDRESS
None	

9 FOR OPTIONAL COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES *(Section 132 121, Postal Service Manual)*

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in Item 1 at the reduced postage rates presently authorized by 39 U. S. C. 3626

(Signature and title of editor, publisher, business manager, or owner)

10 FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 132 122 Postal Service Manual) (Check one)*

The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes: ☒ Have not changed during preceding 12 months ☐ Have changed during preceding 12 months *(If changed, publisher must submit explanation of change with this statement.)*

11. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NUMBER OF COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A TOTAL NO. COPIES PRINTED *(Net Press Run)*	15,250	14,500
B PAID CIRCULATION 1 SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	-	-
2 MAIL SUBSCRIPTIONS	14,500	13,778
C TOTAL PAID CIRCULATION	14,500	13,778
D FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	86	86
E TOTAL DISTRIBUTION *(Sum of C and D)*	14,586	13,864
F COPIES NOT DISTRIBUTED 1 OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	664	636
2 RETURNS FROM NEWS AGENTS	-	-
G TOTAL *(Sum of E & F—should equal net press run shown in A)*	15,250	14,500

I certify that the statements made by me above are correct and complete. SIGNATURE OF EDITOR, PUBLISHER BUSINESS MANAGER, OR OWNER

PS Form Jan 1975 3526 (Page 1) *(See instructions on reverse)*

PUBLICATION POLICY FOR ARTICLES IN *THE AMERICAN STATISTICIAN*

General

The American Statistician attempts to publish articles of general interest to the statistical profession. These articles should be timely and of sufficient generality to appeal to large numbers of readers. Articles should ordinarily be of an expository, not highly technical nature. The following list illustrates topics that would be appropriate:

1. Important current national and international statistical problems and programs.
2. Public policy matters of direct interest to the statistical profession.
3. The training of statisticians and education in statistics.
4. General discussions dealing with developments affecting statistical practice.
5. Short papers on the history of statistics and probability.
6. Teaching of statistics, including classroom matters, curriculum, and program development.
7. Non-technical or tutorial state-of-the-art surveys of the discipline.

Expository Papers

The Editorial Board is particularly desirous of attracting general expository articles on topics of current importance to the statistical community. Such articles should be particularly styled for our general readership and should be authoritative and well researched. Authors with plans for possible expository papers are urged to communicate with the Editor with an outline or draft of the proposed article.

The Teacher's Corner

"The Teacher's Corner" Department carries articles on all levels of statistical education. These include new and unique approaches to curriculum development, course organization, and original technical contributions useful in the teaching of particular statistical concepts.

Letters to the Editor

Letters on matters of interest to the statistical profession will be considered. They should be concise; when necessary the Editorial Board may require a shorter version.