On Estimating the Face Probabilities of Shaved Dice With Partial Data

Marios G. PAVLIDES and Michael D. PERLMAN

As noted by Persi Diaconis, the apparently simple problem of calculating the face probabilities of a shaved die defies exact mathematical analysis. However, this problem can be used to provide amusing and instructive examples of sufficiency, efficiency, information, ancillarity, and the value of statistical design.

KEY WORDS: Ancillarity; Design; Efficiency; Estimation; Face probability; Fisher information; Multinomial distribution; Shaved die; Sufficiency.

1. INTRODUCTION

Our friend Persi Diaconis has fascinated many audiences with a probability problem that is easy to state yet defies mathematical analysis. What are the face probabilities of a shaved die?

A standard gaming die is a cube whose six faces are numbered from 1 to 6 with all pairs of opposing faces summing to 7. By symmetry, each face of the die has probability $1/6$ when the die is rolled. When one face of the die is shaved, the shaved face and its opposite retain the same surface area while the surface areas of the other four faces are reduced by the same amount, hence have lower probability than the shaved face and its opposite.

For simplicity, assume that each edge of the cubical die has unit length 1. If face 1 is shaved by a uniform amount s ($0 < s < 1$), then face 1 and its opposite, face 6, continue to have area 1 but faces 2, 3, 4, and 5 now have area $1 - s$. Thus, by symmetry, the probabilities r_k of the six faces are now of the form

$$r_1 = r_6 = a(s), \qquad r_2 = r_3 = r_4 = r_5 = b(s), \qquad (1.1)$$

where $2a(s) + 4b(s) = 1$ and $a(s) \geq b(s) > 0$; equivalently, $1/6 \leq a(s) < 1/2$. Notice that the same probabilities would be obtained if face 6 were shaved instead.

It is one of the more surprising facts of probability theory that, for this very simple problem, the function $a(s)$ remains unknown (see *http://news.stanford.edu/news/2004/june9/diaconis-69.html*). The dynamics of a rolling, bouncing die are so complex that no exact analytic expression for $a(s)$

seems possible, leaving empirical experimentation as the sole approach to this question. This is a standard problem of estimating the binomial probability $p = 2a(s) = P[1 \text{ or } 6]$ (see, e.g., Dunn 2003). Persi actually enlisted teams of energetic undergraduates to gather enough data to distinguish between competing dynamic models of a rolling, shaved die.

Unfortunately, Persi found that many millions of dice rolls would be needed for this purpose, well beyond the reach of his enthusiastic students, so the probabilistic dynamics of a rolling die remain unresolved. As a consolation, however, we have found an amusing formulation of this estimation problem that nicely illustrates the fundamental statistical ideas of information, efficiency, sufficiency, and ancillarity, and along the way provides a few surprises concerning experimental design.

For the rest of this article, we usually will abbreviate $a(s)$ and $b(s)$ to a and b, respectively. Note that $b = (1 - 2a)/4$.

2. A HYPOTHETICAL SCENARIO

The problem can be motivated by the following scenario:

In order to accumulate the massive amounts of data needed, Persi decided to enlist the aid of the Eagle and Dove Casino. The craps tables of the casino are electronically equipped to record the result of each roll of the dice. The management agreed to stage a charity event, donating the use of its craps tables to gather empirical data for shaved dice. As he was to be out of town, Persi asked us to collect the data for him.

When we arrived at the casino, however, a complication arose: In craps, two dice are rolled simultaneously and the outcome of the game depends only on the *sum* $X + Y$ showing on the pair. (*We assume that the two dice are shaved identically.*) Therefore, the electronic equipment was designed to record only $X + Y$, not the individual values of X and Y. Thus, the formulation as a standard binomial estimation problem would not apply.

"Oh well," we thought, "surely rolling a pair of identical dice once should be equivalent to rolling a single die twice for the purpose of estimation." To be safe, however, we sent a message to Persi:

Question 1. For estimating the face probability $a(s)$ of a shaved die, based on repeated rolls of a pair of identically shaved dice, is a loss of efficiency incurred if only partial data, namely the sum of the two dice, are recorded at each roll?

Persi soon texted a reply to us:

"Well, boys, this is a little exercise with sufficient statistics and Fisher information. First, let's introduce some notation: If a pair of identically shaved dice is rolled n times, your data will consist of n pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ that comprise $2n$ independent and identically distributed (iid) rolls of a single shaved

(a) The sums $X + Y$ for regular (six-sided and 7-sum) dice X and Y.

Y \ X	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

(b) $\Pr(X = x, Y = y)$ for case $(1, 6)$.

Y \ X	1	2	3	4	5	6
1	a^2	ab	ab	ab	ab	a^2
2	ba	b^2	b^2	b^2	b^2	ba
3	ba	b^2	b^2	b^2	b^2	ba
4	ba	b^2	b^2	b^2	b^2	ba
5	ba	b^2	b^2	b^2	b^2	ba
6	a^2	ab	ab	ab	ab	a^2

Figure 1. (a) Shows the sum $X + Y$ for all 36 outcomes of (X, Y), where X and Y represent rolls of a regular (six-sided, 7-sum) die. (b) Shows the probabilities of the 36 outcomes of (X, Y) where X and Y are iid rolls of a regular die whose face 1 has been shaved and 6 is its opposite. Here, a represents the probability of the shaved face, and that of its opposite while b represents the probability of the four remaining FACES.

die. However, you will only observe $Z_1, \ldots, Z_n$, where each $Z_i := X_i + Y_i$ has range $\{2, \ldots, 12\}$. Since face 1 is shaved [recall (1.1)], the probability distribution $(p_2(a), \ldots, p_{12}(a))$ of Z_i can be found from Figure 1:

$$p_2(a) = p_{12}(a) = a^2, \qquad p_3(a) = p_{11}(a) = 2ab,$$
$$p_4(a) = p_{10}(a) = 2ab + b^2,$$
$$p_5(a) = p_9(a) = 2ab + 2b^2, \tag{2.1}$$
$$p_6(a) = p_8(a) = 2ab + 3b^2, \qquad p_7(a) = 2a^2 + 4b^2.$$

"We can represent $Z := (Z_1, \ldots, Z_n)$ equivalently by a multinomial random vector $\mathbf{N} \equiv (N_2, \ldots, N_{12})$ as follows:

$$Z_i \leftrightarrow (I\{Z_i = 2\}, \ldots, I\{Z_i = 12\}), \tag{2.2}$$
$$Z \leftrightarrow \left(\sum_{i=1}^n I\{Z_i = 2\}, \ldots, I\{Z_i = 12\} \right) \tag{2.3}$$
$$=: (N_2, \ldots, N_{12}), \tag{2.4}$$
$$\mathbf{N} \sim \text{Multinomial}(n; p_2(a), \ldots, p_{12}(a)). \tag{2.5}$$

This constitutes a one-parameter statistical model, which we denote by $\mathbf{M}_{1,6}$, with multinomial probability mass function given by

$$f_{1,6}(a) = \frac{n!}{N_2! \cdots N_{12}!} \cdot \prod_{j=2}^{12} p_j(a)^{N_j}$$
$$= c \cdot a^{T_1} b^{T_2} (2a + b)^{T_3} (a + b)^{T_4}$$
$$\times (2a + 3b)^{T_5} (a^2 + 2b^2)^{T_6}$$
$$= \tilde{c} \cdot a^{T_1} (1 - 2a)^{T_2} (6a + 1)^{T_3} (2a + 1)^{T_4}$$
$$\times (2a + 3)^{T_5} (12a^2 - 4a + 1)^{T_6},$$

where

$$T_1 = 2N_2 + N_3 + N_{11} + 2N_{12},$$
$$T_2 = N_3 + N_4 + N_5 + N_6 + N_8 + N_9 + N_{10} + N_{11},$$
$$T_3 = N_4 + N_{10},$$
$$T_4 = N_5 + N_9,$$
$$T_5 = N_6 + N_8,$$
$$T_6 = N_7.$$

Here c and $\tilde{c}$ denote varying factors possibly depending on the N_j or the T_j, but not on a. The subscript $(1, 6)$ on $\mathbf{M}_{1,6}$ and $f_{1,6}(a)$ indicates that face 1 is shaved and 6 is its opposite face, and n denotes the number of trials.

"By the Fisher–Neyman Factorization Theorem (cf. Lehmann and Casella 1998, theorem 1.6.5),

$$\mathbf{T}_{1,6} := (T_1, T_2, T_3, T_4, T_5, T_6)$$

is a sufficient statistic for $\mathbf{M}_{1,6}$. Because $\mathbf{T}_{1,6}$ satisfies the linear relation

$$T_1 + T_2 + T_3 + T_4 + T_5 + 2T_6 = 2n,$$

$\mathbf{T}_{1,6}$ is actually five-dimensional, and is, in fact, minimal sufficient."

Exercise 1. Show that $\mathbf{T}_{1,6}$ is a minimal sufficient statistic for $\mathbf{M}_{1,6}$.

Hint. Because $\mathbf{M}_{1,6}$ is an exponential family, the Lehmann–Scheffé Theorem can be readily applied—cf. Lehmann and Casella [1998, corollary 1.6.16(ii)].

"To answer Question 1, we just need to find the Fisher information for the parameter a in the model $\mathbf{M}_{1,6}$, since the inverse of the Fisher information determines the efficiency of the maximum likelihood estimator (MLE) $\hat{a}_{1,6}$ as measured by its asymptotic variance (Lehmann and Casella 1998, theorem 6.3.10). Because we expect only a small shaving, we are most interested in evaluating the Fisher information in a neighborhood of $a = 1/6$, where the die is unshaved (fair). To avoid the added complication of a truncated parameter space, we now remove the constraint that $a \geq b$ (equivalently $a \geq 1/6$) which corresponds to allowing elongation of face 1 as well as shaving. Thus, the only constraint on a is $0 < a < 1/2$. (The sufficiency properties of $\mathbf{T}_{1,6}$ are unchanged.)

"We can now compute the Fisher information *per trial*, denoted by $I_{1,6}(a)$, for the multinomial model $\mathbf{M}_{1,6}$:

$$\frac{d \log f_{1,6}(a)}{da} = \frac{T_1}{a} - \frac{2T_2}{1-2a} + \frac{6T_3}{6a+1} + \frac{2T_4}{2a+1}$$
$$+ \frac{2T_5}{2a+3} + \frac{4T_6(6a-1)}{12a^2-4a+1},$$

$$\frac{d^2 \log f_{1,6}(a)}{da^2} = -\frac{T_1}{a^2} - \frac{4T_2}{(1-2a)^2} - \frac{36T_3}{(6a+1)^2} - \frac{4T_4}{(2a+1)^2}$$
$$- \frac{4T_5}{(2a+3)^2} - \frac{8T_6(36a^2-12a-1)}{(12a^2-4a+1)^2},$$

$$I_{1,6}(a) = -\frac{1}{n}\mathrm{E}\left[\frac{d^2 \log f_{1,6}(a)}{da^2}\right]$$
$$= \frac{2a+1}{a} + \frac{10a+3}{1-2a} + \frac{9(1-2a)}{2(6a+1)} + \frac{1-2a}{2a+1}$$
$$+ \frac{1-2a}{2(2a+3)} + \frac{2(36a^2-12a-1)}{12a^2-4a+1}$$
$$= \frac{1}{a} + \frac{8}{1-2a} + \frac{6}{6a+1} + \frac{2}{2a+1}$$
$$+ \frac{2}{2a+3} - \frac{8}{12a^2-4a+1},$$

where we have used Mathematica and the relations

$$\frac{1}{n}\mathrm{E}T_1 = 2p_2 + p_3 + p_{11} + 2p_{12} = a(2a+1),$$

$$\frac{1}{n}\mathrm{E}T_2 = p_3 + p_4 + p_5 + p_6 + p_8 + p_9 + p_{10} + p_{11}$$
$$= \frac{1}{4}(10a+3)(1-2a),$$

$$\frac{1}{n}\mathrm{E}T_3 = p_4 + p_{10} = \frac{1}{8}(6a+1)(1-2a),$$

$$\frac{1}{n}\mathrm{E}T_4 = p_5 + p_9 = \frac{1}{4}(2a+1)(1-2a),$$

$$\frac{1}{n}\mathrm{E}T_5 = p_6 + p_8 = \frac{1}{8}(2a+3)(1-2a),$$

$$\frac{1}{n}\mathrm{E}T_6 = p_7 = \frac{1}{4}(12a^2-4a+1),$$

with $p_j = p_j(a)$ given in (2.1). The likelihood equation

$$\frac{d \log f_{1,6}(a)}{da} = 0$$

is equivalent to a sixth-degree polynomial equation in a. By the classical theory of maximum likelihood estimation for regular (smooth) parametric statistical models (Lehmann and Casella 1998, sections 6.3 and 6.4; Perlman 1983, theorems 3.1 and 4.3), the MLE $\hat{a}_{1,6}$ is a consistent root of this equation and is asymptotically normal and asymptotically efficient:

$$\sqrt{n}(\hat{a}_{1,6} - a) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I_{1,6}(a)}\right), \quad \text{as } n \to \infty. \quad (2.6)$$

"We can now compare (2.6) to the asymptotic distribution of the MLE $\tilde{a}_{1,6}$ based on the complete data $(X_1, Y_1), \ldots, (X_n, Y_n)$. Because

$$U_{1,6} := \sum_{i=1}^{n}[I\{X_i = 1 \text{ or } 6\} + I\{Y_i = 1 \text{ or } 6\}]$$
$$\sim \text{Binomial}(2n; p),$$

where $p = 2a = P[1 \text{ or } 6]$, the MLE $\tilde{p}_{1,6} \equiv U_{1,6}/2n$ of p is the usual binomial estimator. By the De Moivre–Laplace Theorem (Feller 1968, theorem VII.3.2),

$$\sqrt{2n}(\tilde{p}_{1,6} - p) \xrightarrow{d} \mathcal{N}(0, p(1-p)), \quad \text{as } n \to \infty.$$

Because $\tilde{p}_{1,6} = 2\tilde{a}_{1,6}$, it follows that

$$\sqrt{n}(\tilde{a}_{1,6} - a) \xrightarrow{d} \mathcal{N}\left(0, \frac{a(1-2a)}{4}\right), \quad \text{as } n \to \infty. \quad (2.7)$$

Thus, using Mathematica, the relative asymptotic efficiency of $\hat{a}_{1,6}$ compared to $\tilde{a}_{1,6}$, as measured by the inverse of the ratio of their asymptotic variances, is

$$e_{1,6}(a) = \frac{29}{12} - \frac{1}{3(6a+1)} - \frac{1}{2(2a+1)}$$
$$- \frac{3}{2a+3} - \frac{2a+1}{3(12a^2-4a+1)}$$

for $0 < a < 1/2$. The graph of this function is shown in Figure 3 (Section 4). Because we expect the alteration (shaving or elongation) to be small, we evaluate this efficiency ratio at $a = 1/6$ (the case of a fair die) to find that

$$e_{1,6}(1/6) = 37/120 = 0.3083\ldots.$$

"So, boys, I'm sorry to tell you that observing only the sum of each pair of altered dice is just 31% as efficient for estimating $a \equiv a(s)$ as observing the individual outcomes of each die. A lot of information will be lost so it will take a lot longer to achieve the accuracy we need, but look on the bright side—the drinks in the casino are free!"

"Goodness," we thought, "how can he text so fast?" Wiser but undaunted, we turned to the task of collecting the data from the craps tables.

3. THE PLOT THICKENS

A moment later, however, we received another text message from Persi:

"I bet you've already anticipated the next question:

Question 2. Will the efficiency change if face 2 or face 3 is altered (shaved or elongated) instead of face 1?"

"No," we replied, "why would that make any difference? Since we're starting with a symmetric cube, why would it matter which face is altered?"

"It doesn't matter if you observe the outcomes of each individual die," Persi texted back, "but it matters a great deal if you observe only the sum of each pair. The statistical model and its properties change significantly.

"Suppose that face 2 is altered. As before, your data will consist of n pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ comprising $2n$ iid rolls of a single die and you will again only observe $Z_1, \ldots, Z_n$ with

(a) $\Pr(X = x, Y = y)$ for case $(2, 5)$.

X \ Y	1	2	3	4	5	6
1	b^2	ba	b^2	b^2	ba	b^2
2	ab	a^2	ab	ab	a^2	ab
3	b^2	ba	b^2	b^2	ba	b^2
4	b^2	ba	b^2	b^2	ba	b^2
5	ab	a^2	ab	ab	a^2	ab
6	b^2	ba	b^2	b^2	ba	b^2

(b) $\Pr(X = x, Y = y)$ for case $(1, 2)$.

X \ Y	1	2	3	4	5	6
1	a^2	a^2	ab	ab	ab	ab
2	a^2	a^2	ab	ab	ab	ab
3	ba	ba	b^2	b^2	b^2	b^2
4	ba	ba	b^2	b^2	b^2	b^2
5	ba	ba	b^2	b^2	b^2	b^2
6	ba	ba	b^2	b^2	b^2	b^2

Figure 2. The probabilities of the 36 outcomes of (X, Y) where X and Y are iid rolls of a standard die whose face k has been altered (= shaved or elongated) and face l is its opposite. Here, a represents the probability of the altered face and of its opposite. (a) $(k, l) = (2, 5)$. (b) $(k, l) = (1, 2)$ (see Section 4) while b represents the probability of the four remaining FACES.

each $Z_i := X_i + Y_i$ having range $\{2, \ldots, 12\}$. Because face 2 is altered, however, (1.1) is changed. The probabilities r_k of the six faces are now of the form

$$r_2 = r_5 = a, \qquad r_1 = r_3 = r_4 = r_6 = b,$$

where again $2a + 4b = 1$ and $a \geq b > 0$. Also, as before, we drop the latter constraint, allowing a possible elongation of face 2, so that simply $0 < a < 1/2$. (Note that the same probabilities would be obtained if face 5 were altered instead.) The probability distribution $(p_2(a), \ldots, p_{12}(a))$ of Z_i can be found from Figure 2(a):

$$
\begin{aligned}
&p_2(a) = p_{12}(a) = b^2, \qquad p_3(a) = p_{11}(a) = 2ab, \\
&p_4(a) = p_{10}(a) = a^2 + 2b^2, \\
&p_5(a) = p_9(a) = 2ab + 2b^2, \\
&p_6(a) = p_8(a) = 4ab + b^2, \qquad p_7(a) = 2a^2 + 4b^2.
\end{aligned}
\tag{3.1}
$$

"With Z and $\mathbf{N} \equiv (N_2, \ldots, N_{12})$, as defined above, again

$$\mathbf{N} \sim \text{Multinomial}(n; p_2(a), \ldots, p_{12}(a)),$$

but with $p_j(a)$ given by (3.1) rather than (2.1). This is again a one-parameter model, denoted by $\mathbf{M}_{2,5}$, with probability mass function given by

$$
\begin{aligned}
f_{2,5}(a) &= c \cdot a^{T_1} b^{T_2} (a + b)^{T_3} (4a + b)^{T_4} (a^2 + 2b^2)^{T_5} \\
&= \tilde{c} \cdot a^{T_1} (1 - 2a)^{T_2} (2a + 1)^{T_3} (14a + 1)^{T_4} \\
&\quad \times (12a^2 - 4a + 1)^{T_5},
\end{aligned}
$$

where the T_i are defined differently than in the case $(1, 6)$:

$$T_1 = N_3 + N_{11},$$
$$T_2 = 2N_2 + N_3 + N_5 + N_6 + N_8 + N_9 + N_{11} + 2N_{12},$$
$$T_3 = N_5 + N_9,$$
$$T_4 = N_6 + N_8,$$
$$T_5 = N_4 + N_7 + N_{10}.$$

The subscript $(2, 5)$ on $\mathbf{M}_{2,5}$ and $f_{2,5}(a)$ indicates that face 2 is altered and 5 is its opposite face.

"By the Factorization Theorem,

$$\mathbf{T}_{2,5} := (T_1, T_2, T_3, T_4, T_5)$$

is a sufficient statistic for $\mathbf{M}_{2,5}$ and, in fact, minimally sufficient (recall Exercise 1), and is actually four-dimensional because $\mathbf{T}_{2,5}$ satisfies the linear relation

$$T_1 + T_2 + T_3 + T_4 + 2T_5 = 2n.$$

This is a nontrivial change from model $\mathbf{M}_{1,6}$ where the minimal sufficient statistic $\mathbf{T}_{1,6}$ was five-dimensional.

"The Fisher information *per trial*, $I_{2,5}(a)$, for the multinomial model $\mathbf{M}_{2,5}$ is found as before:

$$I_{2,5}(a) = \frac{1}{a} + \frac{8}{1 - 2a} + \frac{2}{2a + 1} + \frac{28}{14a + 1} - \frac{16}{12a^2 - 4a + 1}.$$

In this case, the likelihood equation (LEQ) is

$$
\begin{aligned}
\frac{d \log f_{2,5}(a)}{da} = \frac{T_1}{a} &- \frac{2T_2}{1 - 2a} + \frac{2T_3}{2a + 1} \\
&+ \frac{14T_4}{14a + 1} + \frac{4T_5(6a - 1)}{12a^2 - 4a + 1} = 0,
\end{aligned}
$$

which is equivalent to a fifth-degree polynomial equation in a, one degree less than the LEQ for model $\mathbf{M}_{1,6}$—another nontrivial change. The MLE, $\hat{a}_{2,5}$, satisfies

$$\sqrt{n}(\hat{a}_{2,5} - a) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I_{2,5}(a)}\right), \quad \text{as } n \to \infty.$$

It is easy to see that $\tilde{a}_{2,5}$ has the same distribution as $\tilde{a}_{1,6}$. Hence, it follows from (2.7) that the relative asymptotic efficiency of $\hat{a}_{2,5}$ compared to $\tilde{a}_{2,5}$ (the MLE of a based on the complete data $(X_1, Y_1), \ldots, (X_n, Y_n)$) is

$$
\begin{aligned}
e_{2,5}(a) &= \frac{1}{4} a(1 - 2a) I_{2,5}(a) \\
&= \frac{167}{84} - \frac{1}{2(2a + 1)} \\
&\quad - \frac{4}{7(14a + 1)} - \frac{2(2a + 1)}{3(12a^2 - 4a + 1)}
\end{aligned}
$$

for $0 < a < 1/2$. Again, we evaluate this efficiency at $a = 1/6$ to find that

$$e_{2,5}(1/6) = 13/120 = 0.1083\dots.$$

"Thus, when we observe only the sum of each pair of altered dice but alter face 2 instead of face 1 for each die, the efficiency is reduced by another factor of 3. So, you see, planning the design of a statistical experiment in advance can make a significant difference in efficiency.

"Here are the results if face 3 is altered—you can check the details yourselves. (Also see Figure 3 and Table 1 in Section 4.) The minimal sufficient statistic is four-dimensional and the LEQ is equivalent to a sixth-degree polynomial equation in a. The Fisher information *per trial*, $I_{3,4}(a)$, and relative asymptotic efficiency $e_{3,4}(a)$ of the MLE $\hat{a}_{3,4}$ compared to $\tilde{a}_{3,4}$ are

$$I_{3,4}(a) = \frac{2}{a} + \frac{6}{1-2a} + \frac{6}{6a+1} - \frac{4}{4a^2+1} - \frac{8}{12a^2 - 4a + 1}$$

and

$$e_{3,4}(a) = \frac{5}{3} - \frac{1}{3(6a+1)} - \frac{1+2a}{2(4a^2+1)} - \frac{1+2a}{3(12a^2-4a+1)}$$

for $0 < a < 1/2$. The relative efficiency of $\hat{a}_{3,4}$ versus $\tilde{a}_{3,4}$, when $a = 1/6$, is

$$e_{3,4}(1/6) = 7/30 = 0.2333\dots.$$

Thus, $e_{3,4}(1/6)$ falls between $e_{2,5}(1/6) = 0.1083$ and $e_{1,6}(1/6) = 0.3083$."

4. TRUTH IN RELABELING

"Wow, this is a lot more complicated than we imagined," we thought, "but a lot more interesting too. Well, at least we can proceed to collect the data knowing that altering face 1 provides the most efficient design for estimating the face probabilities of a shaved or elongated die based on observing only the sums of pairs." But just then we received another message from Persi:

"Ok, guys, you must have realized by now that there is an even more efficient design, one that is two-and-a-half times as efficient as altering face 1, in fact 75% as efficient as the complete-data case where each individual die is observed. Can you tell me what it is?"

"Um, well, actually, . . . , ok, we have no clue."

"All right, let's think out of the box, or, you might say, out of the cube. Suppose we start with an *unlabeled* shaved or elongated die. However, instead of the standard labeling of the six faces of the die where each opposing pair of faces sums to 7, suppose we allow a nonstandard labeling where any of the $\binom{6}{2} = 15$ ordered pairs from $\{1, 2, 3, 4, 5, 6\}$ can appear as the labels k and l of the altered face and its opposite face.

Question 3. Can the efficiency be improved if nonstandard labelings are allowed?

"We've already considered the ordered pairs $(k, l) = (1, 6)$, $(2, 5)$, and $(3, 4)$, which correspond to the standard labeling, so twelve nonstandard-labeled pairs remain. However, these twelve pairs occur in six isomorphic couples, since the probability mass function of the model $\mathbf{M}_{k,l}$ is identical to that of $\mathbf{M}_{7-l,7-k}$ under the inversion $(N_2, \dots, N_{12}) \to (N_{12}, \dots, N_2)$, so each member of the couple attains the same efficiency. For example, $(1, 2)$ and $(5, 6)$ yield isomorphic models with the same efficiency. Thus, we need only consider the six cases $(k, l) = (1, 2), (1, 3), (1, 4), (1, 5), (2, 3)$, and $(2, 4)$.

Case $(1, 2)$. Face 1 is altered and its opposite face is labeled '2.'

"The probabilities r_k of the six faces are

$$r_1 = r_2 = a, \qquad r_3 = r_4 = r_5 = r_6 = b, \qquad 0 < a < 1/2.$$

The probability distribution $(p_2(a), \dots, p_{12}(a))$ of $X_i + Y_i$ is now asymmetric [see Figure 2(b)]:

$$p_2(a) = a^2, \qquad p_3(a) = 2a^2, \qquad p_4(a) = a^2 + 2ab,$$
$$p_5(a) = 4ab, \qquad p_6(a) = 4ab + b^2,$$
$$p_7(a) = 4ab + 2b^2, \qquad p_8(a) = 2ab + 3b^2, \qquad (4.1)$$
$$p_9(a) = 4b^2, \qquad p_{10}(a) = 3b^2,$$
$$p_{11}(a) = 2b^2, \qquad p_{12}(a) = b^2.$$

"The random vector $\mathbf{N}$ follows a one-parameter multinomial model $\mathbf{M}_{1,2}$ with probability mass function given by

$$f_{1,2}(a) = c \cdot a^{T_1} b^{T_2} (a + 2b)^{T_3} (4a + b)^{T_4}$$
$$\times (2a + b)^{T_5} (2a + 3b)^{T_6}$$
$$= \tilde{c} \cdot a^{T_1} (1 - 2a)^{T_2} (14a + 1)^{T_4}$$
$$\times (6a + 1)^{T_5} (2a + 3)^{T_6} \qquad (4.2)$$

and

$$T_1 = 2N_2 + 2N_3 + N_4 + N_5,$$
$$T_2 = N_5 + N_6 + N_7 + N_8 + 2N_9$$
$$+ 2N_{10} + 2N_{11} + 2N_{12},$$
$$T_3 = N_4,$$
$$T_4 = N_6,$$
$$T_5 = N_7,$$
$$T_6 = N_8,$$
$$2n = T_1 + T_2 + T_3 + T_4 + T_5 + T_6,$$

so $\mathbf{T}_{1,2} := (T_1, T_2, T_3, T_4, T_5, T_6)$ is a five-dimensional minimal sufficient statistic for $\mathbf{M}_{1,2}$. Note that T_3 does not appear in (4.2) because $a + 2b = 1/2$.

"The Fisher information *per trial*, $I_{1,2}(a)$, for $\mathbf{M}_{1,2}$ is given by

$$I_{1,2}(a) = \frac{3}{2a} + \frac{7}{1-2a} + \frac{14}{14a+1} \frac{6}{6a+1} + \frac{1}{2a+3}. \qquad (4.3)$$

The LEQ is now equivalent to a fourth-degree polynomial in a:

$$\frac{d \log f_{1,2}(a)}{da} = \frac{T_1}{a} - \frac{2T_2}{1-2a} + \frac{14T_4}{14a+1}$$
$$+ \frac{6T_5}{6a+1} + \frac{2T_6}{2a+3} = 0 \qquad (4.4)$$

and the MLE $\hat{a}_{1,2}$ has relative asymptotic efficiency, compared to $\tilde{a}_{1,2}$, given by

$$e_{1,2}(a) = \frac{1}{4}a(1-2a)I_{1,2}(a)$$

$$= \frac{251}{168} - \frac{a}{4} - \frac{2}{7(14a+1)} - \frac{1}{3(6a+1)} - \frac{3}{2(2a+3)}$$

for $0 < a < 1/2$. When $a = 1/6$ this efficiency becomes

$$e_{1,2}(1/6) = 3/4 = 0.75 > 0.3083 = e_{1,6}(1/6). \quad (4.5)$$

"A surprising fact emerges: re-labeling the faces of the altered die in this way increases the efficiency by a factor of almost 2.5, and is now 75% as efficient as the complete-data case. Now, that is a nice example of statistical design in action!

"I'll let you work out the Fisher information and relative efficiencies for the five remaining cases $(1, 3)$, $(1, 4)$, $(1, 5)$, $(2, 3)$, and $(2, 4)$. The relative efficiencies $e_{k,l}(a)$ for all nine distinct cases (labelings) are shown in Figure 3. Because the nonstandard-labeled case $(1, 2)$ [equivalently, the isomorphic case $(5, 6)$] clearly dominates all others over most of the range $0 < a < 1/2$, it is worth noting two distinguishing features of its probability mass (likelihood) function $f_{1,2}(a)$.

"First, the LEQ (4.4) is equivalent to a fourth-degree polynomial equation, the lowest degree among all nine distinct cases (see Table 1). Second, and more significantly, each factor in $f_{1,2}(a)$ [see (4.2)] has the linear-based form $(\delta a + \epsilon)^{T_j}$. Thus, $f_{1,2}(a)$ is log-concave so can have at most one relative extremum in $(0, 1/2)$, necessarily a relative maximum, and the LEQ can have at most one root in $(0, 1/2)$. Furthermore, when $T_1 > 0$ and $T_2 > 0$, $\log f_{1,2}(a) \to -\infty$ as $a \to 0$ and $a \to 1/2$

Table 1. For each of the nine distinct cases (k, l) of an altered (shaved or elongated) six-sided die, the relative efficiency $e_{k,l}(1/6)$, the dimension $\dim_{k,l}$ of the minimal sufficient statistic $\mathbf{T}_{k,l}$, and the polynomial degree $\deg_{k,l}$ of the likelihood equation for the one-parameter multinomial model $\mathbf{M}_{k,l}$, are shown. The three cases marked * occur for the standard labeling where each pair of opposing sides of the die sum to 7.

(k, l)	$e_{k,l}(1/6)$	$\dim_{k,l}$	$\deg_{k,l}$
$(1, 2)$	0.75	5	4
$(1, 3)$	0.4375	6	8
$(2, 3)$	0.3375	7	8
$(1, 6)^*$	0.3083	5	6
$(1, 4)$	0.2708	6	8
$(3, 4)^*$	0.2333	4	6
$(1, 5)$	0.2083	6	8
$(2, 4)$	0.1708	6	9
$(2, 5)^*$	0.1083	4	5

so the LEQ has exactly one root, necessarily the unique consistent root and the MLE (Lehmann and Casella 1998, corollary 6.3.8; Perlman 1983, theorems 3.1 and 4.3).

"By contrast, in the eight other distinct cases, $f_{k,l}(a)$ has at least one irreducible quadratic-based factor $(\gamma a^2 + \delta a + \epsilon)^{T_j}$. Thus, when $T_j > 0$, the likelihood function $f_{k,l}(a)$ may not be log-concave, may be multimodal, and the LEQ may have multiple roots."

"Gosh, Persi," we wrote, "this is fascinating—you've sure given us a lot to think about. We hope you won't mind if we ask more questions later, but we have one that we're anxious to ask now. How'd you get to be such a whiz at texting?"
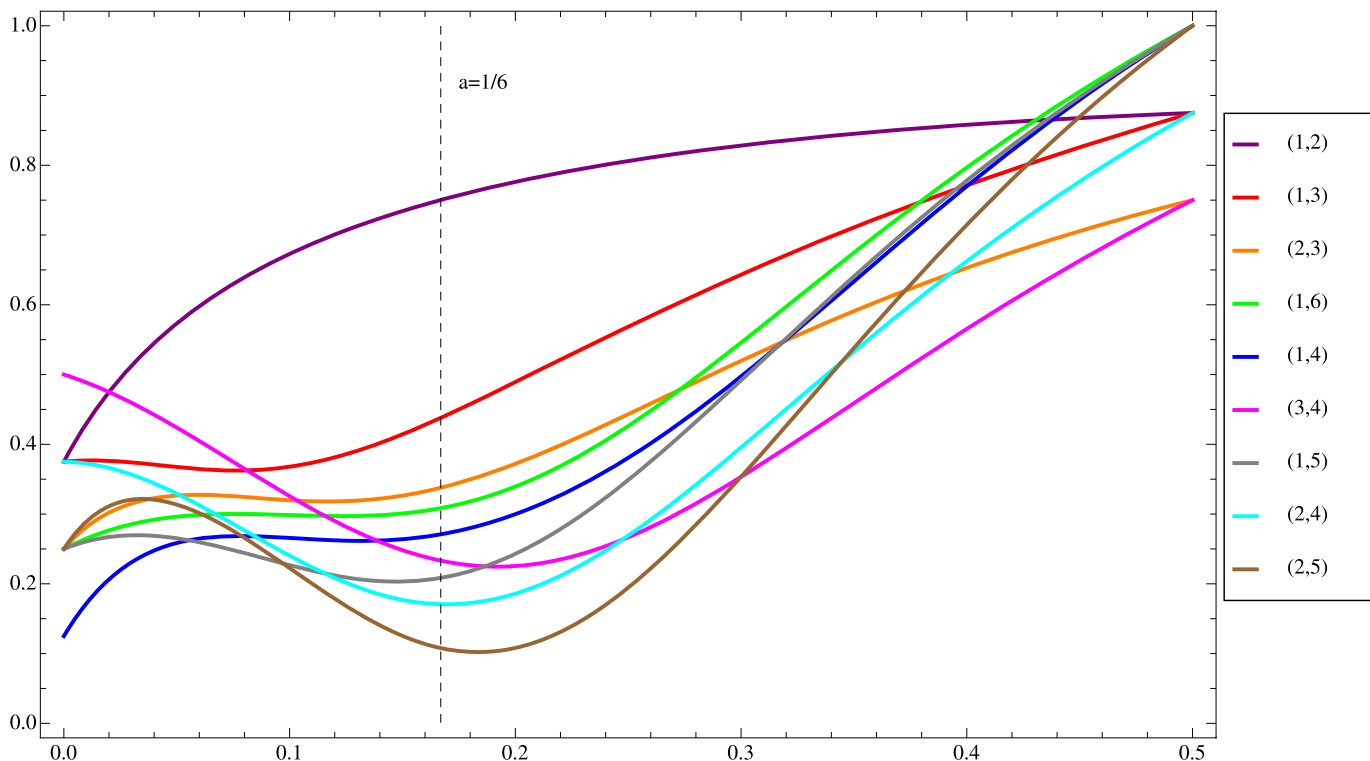

Figure 3. The plot of the nine relative efficiencies $e_{k,l}(a)$, as functions of $0 < a < 0.5$.

"Well, boys," Persi replied, "I suppose you could say it's magic!"

5. TOPICS FOR FURTHER INVESTIGATION

Instructors may find the following exercises of interest to their students. These exercises range from straightforward to challenging/unsolved. Some partial solutions can be found in the report by Pavlides and Perlman (2009).

5.1 Two Identically Altered (Shaved or Elongated), "Three-Sided Dice" With Identical Labelings Chosen From the Integers {1, 2, 3}

Select any of the $\binom{3}{2} = 3$ ordered pairs (k, l) from $\{1, 2, 3\}$ and let m denote the nonselected integer. Let X and Y be iid random variables, each with range $\{1, 2, 3\}$ and probabilities

$$r_k = r_l = a, \qquad r_m = b,$$

where $2a + b = 1$ and $0 < a < 1/2$. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be n iid pairs with each $(X_i, Y_i) \sim (X, Y)$, and let $Z_i = X_i + Y_i$. Find the relative asymptotic efficiency $e_{k,l}(a)$ of $\hat{a}_{k,l}$ (the MLE of a based on $Z_1, \ldots, Z_n$) compared to $\tilde{a}_{k,l}$ (the MLE of a based on $(X_1, Y_1), \ldots, (X_n, Y_n)$).

Comments: As in (2.2)–(2.5) define a five-cell multinomial random vector

$$(N_2, \ldots, N_6) \sim \text{Multinomial}(n; p_2(a), \ldots, p_6(a)) =: \mathbf{M}_{k,l}$$

based on $Z_1, \ldots, Z_n$. The models $\mathbf{M}_{1,2}$ and $\mathbf{M}_{2,3}$ are isomorphic under the inversion $(N_2, \ldots, N_6) \to (N_6, \ldots, N_2)$, hence only cases $(k, l) = (1, 2)$ and $(1, 3)$ need be considered. When $a = 1/3$, corresponding to a fair three-sided die, the relative efficiencies for these two cases are $e_{1,2}(1/3) = 5/6$ and $e_{1,3}(1/3) = 1/3$. [Note that the distribution of $\tilde{a}_{k,l}$ is unchanged from the six-sided case, hence (2.7) remains valid.]

The case $(1, 2)$ for a three-sided die is similar to case $(1, 2)$ for the six-sided die in that $f_{1,2}(a)$ again is log-concave. When $T_1 := 2N_2 + 2N_3 + N_4 + 2N_5 > 0$ and $T_2 := N_5 + 2N_6 > 0$, $\log f_{1,2}(a) \to -\infty$ as $a \to 0$ and $a \to 1/2$, so the LEQ has exactly one root, necessarily the unique consistent root and the MLE.

Kuindersma and Blais (2007) discussed this problem from the viewpoint of model selection; they used the terminology "three-sided coin" rather than our "three-sided die."

5.2 Two Identically Altered, "Three-Sided Dice" With Non-identical Labelings Chosen From the Integers {1, 2, 3}

Select two ordered pairs (k, l) and (k', l') from $\{1, 2, 3\}$ and let m and m' denote the corresponding nonselected integers. Let X and X' be independent random variables, each with range $\{1, 2, 3\}$ and probabilities

$$r_k = r_l = a, \qquad r_m = b$$

and

$$r'_{k'} = r'_{l'} = a, \qquad r'_{m'} = b,$$

respectively, where $2a + b = 1$ and $0 < a < 1/2$. Let $(X_1, X'_1), \ldots, (X_n, X'_n)$ be n iid pairs with each $(X_i, X'_i) \sim (X, X')$

and let $Z_i = X_i + X'_i$. Find the relative asymptotic efficiency $e_{(k,l),(k',l')}(a)$ of $\hat{a}_{(k,l),(k',l')}$ (the MLE of a based on $Z_1, \ldots, Z_n$) compared to $\tilde{a}_{(k,l),(k',l')}$ (the MLE of a based on $(X_1, X'_1), \ldots, (X_n, X'_n)$).

Comments: There are $3 \times 3 = 9$ such pairs $((k, l), (k', l'))$ of ordered pairs from $\{1, 2, 3\}$. Because the cases $((k, l), (k', l'))$ and $((k', l'), (k, l))$ are equivalent, only six of these nine cases need be considered. Of these, the three cases where $(k, l) = (k', l')$ have already been treated in Section 5.1, leaving the three cases $((1, 2), (1, 3))$, $((1, 2), (2, 3))$, and $((1, 3), (2, 3))$. However, the pairs $((1, 2), (1, 3))$ and $((1, 3), (2, 3))$ are isomorphic under the inversion $(N_2, \ldots, N_6) \to (N_6, \ldots, N_2)$, so only the first two cases need be considered.

When $a = 1/3$, the relative efficiencies are

$$e_{(1,2),(1,3)}(1/3) = 5/24 \qquad \text{and} \qquad e_{(1,2),(2,3)}(1/3) = 3/4,$$

both less than $e_{1,2}(1/3) = 5/6$ given in Section 5.1. (Again, (2.7) remains valid for $\tilde{a}_{(k,l),(k',l')}$.)

5.3 Two Identically Altered, Six-Sided Dice With Nonidentical Labelings Chosen From the Integers {1, 2, 3, 4, 5, 6}

Select two ordered pairs (k, l) and (k', l') from $\{1, 2, 3, 4, 5, 6\}$. Let X and X' be independent random variables, each with range $\{1, 2, 3, 4, 5, 6\}$ and with probabilities

$$
\begin{aligned}
r_k &= r_l = a, \\
r_m &= b \quad \text{for } m \in (1, 2, 3, 4, 5, 6) \setminus (k, l), \\
r'_{k'} &= r'_{l'} = a, \\
r'_{m'} &= b \quad \text{for } m' \in (1, 2, 3, 4, 5, 6) \setminus (k', l'),
\end{aligned}
$$

respectively, where $2a + 4b = 1$ and $0 < a < 1/2$. Let $(X_1, X'_1), \ldots, (X_n, X'_n)$ be n iid pairs with each $(X_i, X'_i) \sim (X, X')$ and let $Z_i = X_i + X'_i$. Find the relative asymptotic efficiency $e_{(k,l),(k',l')}(a)$ of $\hat{a}_{(k,l),(k',l')}$ (the MLE of a based on $Z_1, \ldots, Z_n$) compared to $\tilde{a}_{(k,l),(k',l')}$ (the MLE of a based on $(X_1, X'_1), \ldots, (X_n, X'_n)$).

Comments: There are $15 \times 15 = 225$ such pairs $((k, l), (k', l'))$ of ordered pairs from $\{1, 2, 3, 4, 5, 6\}$. Because the cases $((k, l), (k', l'))$ and $((k', l'), (k, l))$ are equivalent, only 120 of these 225 cases need be considered. Of these, the 15 cases where $(k, l) = (k', l')$ have already been treated in Sections 2, 3, and 4, leaving 105 cases. However, these 105 cases include 66 isomorphic pairs under the inversion $(N_2, \ldots, N_{12}) \to (N_{12}, \ldots, N_2)$; for example, the pairs $((1, 2), (1, 6))$ and $((1, 6), (5, 6))$ are isomorphic, as are the pairs $((1, 2), (1, 3))$ and $((4, 6), (5, 6))$. Thus, only $105 - 33 = 72$ cases need be considered.

For $a = 1/6$, we do not expect that the relative efficiencies of any of these 72 cases exceed that for $((1, 2), (1, 2))$, namely $e_{1,2}(1/6) = 3/4$ as found in (4.5). (Again, (2.7) remains valid for $\tilde{a}_{(k,l),(k',l')}$.)

5.4 Two Identically Altered, q-Sided Dice With Nonidentical Labelings Chosen From the Integers $\{1, 2, \ldots\}$

If arbitrary integer labelings of the faces are allowed, such labelings can be chosen so that there is a 1–1 correspondence between the range of $X + Y$ and that of (X, Y), so no information will be lost if only the sums $X_i + Y_i$ are observed. The labelings $(1, 2, \ldots, q)$ and $(1, q + 1, 2q + 1, \ldots, (q - 1)q + 1)$ provide the simplest example.

5.5 Three or More Identically Altered, q-Sided Dice With Identical Labelings Chosen From the Integers $\{1, 2, \ldots, q\}$

Returning to the framework of the original problem, suppose that $t \geq 3$ identically altered, q-sided dice are rolled but only their sum is recorded. Now the data consist of tn iid random variables $\{X_{hi} \mid h = 1, \ldots, t, i = 1, \ldots, n\}$ where for some ordered pair $(k, l) \subset (1, \ldots, q)$,

$$P[X_{hi} = k] = P[X_{hi} = l] = a,$$
$$P[X_{hi} = m] = b \quad \text{for } m \in (1, \ldots, q) \setminus (k, l).$$

Here, $2a + (q - 2)b = 1$ and $0 < a < 1/2$. Let

$$Z_i = X_{h1} + \cdots + X_{ht}, \qquad i = 1, \ldots, n.$$

Find the asymptotic relative efficiency $e_{k,l}(a)$ of $\hat{a}_{k,l}$ (the MLE of a based on $Z_1, \ldots, Z_n$) compared to $\tilde{a}_{k,l}$ (the MLE of a based on $\{X_{hi} \mid h = 1, \ldots, t, i = 1, \ldots, n\}$). Does $e_{k,l}(a) \to 0$ as $t \to \infty$, and if so, at what rate?

5.6 Nonunique Ancillary Statistics for Two Identically Altered, Six-Sided Dice With Identical Labelings Chosen From the Integers $\{1, 2, 3, 4, 5, 6\}$

The one-parameter multinomial models $\mathbf{M}_{k,l}$ possess multiple nontrivial ancillary statistics. Show, for example, that in the optimal case $(k, l) = (1, 2)$, each of the following three statistics is ancillary:

$$S_1 := (N_2 + N_5 + N_9, N_3 + N_6 + N_7$$
$$+ N_{10} + N_{11}, N_4 + N_8 + N_{12}), \qquad (5.1)$$

$$S_2 := (N_2 + N_6 + N_{10}, N_3 + N_5 + N_7$$
$$+ N_9 + N_{11}, N_4 + N_8 + N_{12}), \qquad (5.2)$$

$$S_3 := (N_2 + N_7 + N_{11}, N_3 + N_5 + N_6$$
$$+ N_9 + N_{10}N_4 + N_8 + N_{12}). \qquad (5.3)$$

Comment: It is often suggested that if a nontrivial ancillary statistic S exists, statistical inference about the model should be carried out by conditioning on S (see, e.g., Fisher 1973). When competing nontrivial ancillaries S_i exist, Cox (1971) proposed to select the S_i that maximizes the variance of the conditional information given S_i; also see the works by Becker and Gordon (1983), Sundberg (2003, section 2.5), and references therein. It

is straightforward to show that, when $a = 1/6$, this variance is maximized by S_1, with S_3 a close second.

5.7 Genetic Linkage Models

One-parameter multinomial models similar to our $\mathbf{M}_{k,l}$, also having polynomial likelihood equations, occur in genetic linkage models. A well-known example appears in the book by Rao (1973, pp. 368–369); see also Sundberg (2001).

5.8 Two Identically Altered Six-Sided Dice: Altering Two Non-Opposing Faces

For the standard face numbering, the face probabilities in (1.1) become

$$r_1 = r_6 = a, \qquad r_2 = r_5 = b, \qquad r_3 = r_4 = c,$$

and the entries in Figure 1(b) are changed accordingly. Since $c = (1/2) - a - b$, there are now two parameters to be estimated, a and b, and the Fisher Information number $I_{1,6}(a)$ is replaced by a Fisher Information matrix $I(a, b)$, which determines the asymptotic efficiency of the MLE $(\hat{a}, \hat{b})$. Does this asymptotic efficiency change for non-standard face numberings, and if so, which numbering is optimal when $a = b = 1/6$?

5.9 Calculation of the Fisher Information for the One-Parameter Multinomial Models $\mathbf{M}_{k,l}$ for Six-Sided Dice

Since $(N_2, \ldots, N_{12}) \sim \text{Multinomial}(n; p_2(a), \ldots, p_{12}(a))$ is a sufficient statistic for a based on $Z_1, \ldots, Z_n$, show that the Fisher information number *per trial*, $I_{k,l}(a)$, is given by

$$I_{k,l}(a) = \sum_{i=2}^{12} \left\{ \frac{[p_i'(a)]^2}{p_i(a)} \right\},$$

where the probabilities $p_2(a), \ldots, p_{12}(a)$ depend on the case (k, l). Use this result and a computer algebra system (such as Mathematica or Maple) to replicate the nine relative efficiencies shown in Table 1.

REFERENCES

Becker, N., and Gordon, I. (1983), "On Cox's Criterion for Discriminating Between Alternative Ancillary Statistics," *International Statistical Review*, 51 (1), 89–92. [44]

Cox, D. R. (1971), "The Choice Between Alternative Ancillary Statistics," *Journal of the Royal Statistical Society, Ser. B*, 33, 251–255. [44]

Dunn, P. K. (2003), "What Happens When a $1 \times 1 \times r$ Die Is Rolled?" *The American Statistician*, 57 (4), 258–264. [37]

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, Vol. I (3rd ed.), New York: Wiley. [39]

Fisher, R. A. (1973), *Statistical Methods for Research Workers* (14th ed.), New York: Hafner Publishing. [44]

Kuindersma, S. R., and Blais, B. S. (2007), "Teaching Bayesian Model Comparison With the Three-Sided Coin," *The American Statistician*, 61 (3), 239–244. [43]

Lehmann, E. L., and Casella, G. (1998), *Theory of Point Estimation* (2nd ed.), New York: Springer-Verlag. [38,39,42]

Pavlides, M. G., and Perlman, M. D. (2009), "On Estimating the Face Probabilities of Shaved Dice With Partial Data," Technical Report 566, University of Washington, Dept. of Statistics. [43]

Perlman, M. D. (1983), "The Limiting Behavior of Multiple Roots of the Likelihood Equation," in *Recent Advances in Statistics*, New York: Academic Press, pp. 339–370. [39,42]

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), New York–London–Sydney: Wiley. [44]

Sundberg, R. (2001), "Ancillarity and Conditional Inference for ML and Interval Estimates in Some Classical Genetic Linkage Trials," Research Report 2001:3, Stockholm University, Mathematical Statistics. [44]

—— (2003), "Conditional Statistical Inference and Quantification of Relevance," *Journal of the Royal Statistical Society, Ser. B*, 65 (1), 299–315. [44]