Optimal Strategies for Second Guessers

J. MICHAEL STEELE and JAMES ZIDEK*

A model is given for a class of contests in which the participants try to guess (or estimate) unknown quantities, and the objective of each player is to come closer to the unknown quantities than an adversary. A general optimality result is proved that gives the best guessing rules for the second guesser. These rules are first calculated exactly in a certain hierarchical linear model, and then simpler approximate rules are given.

KEY WORDS: Guessing; Optimal strategies; Hierarchical linear model; Stein estimator; Posterior median.

1. INTRODUCTION

The goal in many activities or contests is not necessarily to do well in any absolute sense, but merely to outperform an adversary. The objective of this article is to provide a model for such a contest, establish the optimality of certain procedures, and provide suitable approximations to these optimal procedures. But before yielding to the mathematics of the model, we wish to fix ideas with an anecdote.

Two statisticians, Bob and Mike, engaged in a contest to guess weights of people at a party. They agreed that Bob would always guess first. Mike would then guess, and finally the person in question would say who is closer. For example, for person number one Bob guessed 137 pounds. Mike then guessed 137.01 pounds, and the guest declared Mike the victor. The contest continued in a similar vein, and to Bob's dismay he won barely a quarter of the time.

It is intuitively clear that the second guesser has an advantage, and one of the results of Section 2 shows that this advantage is typically as large as the 75 percent obtained by Mike in the anecdote.

To continue the story, Bob was so stunned by defeat and eager for revenge, that he elicited the assistance of a professional weight guesser. Mike agreed that since the new team was so powerful it should be willing to make all its guesses about the weights of the guests before Mike had to state any of his guesses. The team agreed to the proposed rule change, and Mike then proceeded to win even more convincingly than before.

The strategy used by Mike in the second case is naturally more sophisticated than the one he used when he was matched against an equal. This second strategy derives from a hierarchical linear model like that studied in Lindley and Smith (1972). It is also closely connected with the James-Stein estimator and was originally motivated by the "Batting Average" example of Efron and Morris (1973).

Our program begins by establishing in Section 2 a formal theory of guessing contests. We also give a simple but very general optimality result that forms the basis for the rest of the article.

The third section determines the exact optimal strategy for second guessing under a certain linear model. Practical approximations to this optimal strategy are worked out in Section 4. The final section gives a critical discussion of the various sources of difficulties inherent in applying this theory of guessing contests. While the main point of this article is to provide a tractable theory of guessing contests, we feel that the largest single point established is the approximate optimality of the simple rule given by (4.1).

2. HOTELLING'S STRATEGY

The structure of our guessing model can be described by a system of four p vectors.

Target values:	$(\theta_1, \theta_2, \ldots, \theta_p) = \boldsymbol{\theta}$
First guess:	$(X_1, X_2, \ldots, X_p) = \mathbf{X}$
Second guesser's hunch:	$(Y_1, Y_2, \ldots, Y_p) = \mathbf{Y}$
Second guess:	$(G_1, G_2, \ldots, G_p) = \mathbf{G}$

The θ_i represent the real values to be guessed. The X_i are guesses made by the person who goes first, and all these are assumed to be available to the second guesser before he acts. The Y_i represent the second guesser's best estimate of the θ_i. Finally, the G_i are the guesses to be announced by the second guesser. Our principal task is to determine how $\mathbf{G}$ should be based on $\mathbf{X}$ and $\mathbf{Y}$.

The objective of each player is to come closer to $\boldsymbol{\theta}$ than his opponent, so we begin by setting

$$V(\mathbf{G}, \boldsymbol{\theta}) = \sum_{j=1}^{p} V_j(\mathbf{G}, \boldsymbol{\theta}) , \tag{2.1}$$

where

$$\begin{aligned} V_j(\mathbf{G}, \boldsymbol{\theta}) &= 1 \quad |G_j - \theta_j| \le |X_j - \theta_j| \\ &= 0 \quad \text{otherwise} . \end{aligned}$$

* J. Michael Steele is Assistant Professor, Department of Statistics, Stanford University, Stanford, CA 94305. James Zidek is Professor, Department of Mathematics, University of British Columbia, Vancouver, BC V6T1W5, Canada. This work was supported in part by the Office of Naval Research under Grant N00014-76-C-0475 (NR-042-267) and the Army Research Office Grant DAAG-29-77-G-0031. The authors wish to thank R. Chacon, P. Diaconis, J. Kadane, I. Olkin, A. Pittinger, and D.-C. Wu for their comments on earlier drafts of this article.

September 1980, Volume 75, Number 371
Theory and Methods Section

The strategic objective of the second guesser is therefore to maximize $EV(\mathbf{G}, \boldsymbol{\theta})$; that is, the second guesser wishes to maximize the expected number of times his guesses come closer to the true values.

The only probabilistic assumptions to be made now are that $\boldsymbol{\theta}$, $\mathbf{X}$, and $\mathbf{Y}$ have a joint distribution that is continuous. This assumption is made for convenience and avoids the ad hoc conventions required for dealing with ties.

Now let $\nu_i(\mathbf{X}, \mathbf{Y})$ denote the median of the conditional distribution of θ_i given $\mathbf{X}$ and $\mathbf{Y}$. A key role in our guessing theory is played by the following strategy:

$$\begin{aligned} G_i^\epsilon &= X_i + \epsilon \quad \text{if} \quad X_i < \nu_i(\mathbf{X}, \mathbf{Y}) \\ &= X_i - \epsilon \qquad \text{otherwise} . \end{aligned}$$

These strategies will subsequently be called *Hotelling strategies* since they were essentially put forward in Hotelling (1929, p. 51). There are broad differences between the present model and Hotelling's problem in location economics, but the relationship seems close enough to justify (or even require) the name. The main fact in this section is the following simple result:

Theorem 1: The Hotelling strategies are ϵ optimal; that is,

$$\lim_{\epsilon \to 0} EV(\mathbf{G}^\epsilon, \boldsymbol{\theta}) = \sup_{\mathbf{G}} EV(\mathbf{G}, \boldsymbol{\theta}) .$$

Proof: Since any guess G_i must be on one side or the other of X_i, we have

$$\begin{aligned} P^{\mathbf{X},\mathbf{Y}}(|G_i - \theta_i| &< |X_i - \theta_i|) \\ &\leq \max\{P^{\mathbf{X},\mathbf{Y}}(\theta_i < X_i), P^{\mathbf{X},\mathbf{Y}}(\theta_i \geq X_i)\} . \end{aligned}$$

The basic observation about G_i^ϵ is that

$$\begin{aligned} \lim_{\epsilon \to 0} P^{\mathbf{X},\mathbf{Y}}(|G_i^\epsilon - \theta_i| &< |X_i - \theta_i|) \\ &= \max\{P^{\mathbf{X},\mathbf{Y}}(\theta_i \leq X_i), P^{\mathbf{X},\mathbf{Y}}(\theta_i \geq X_i)\} . \end{aligned}$$

Taking expectations in the two preceding relations and summing over $1 \leq i \leq p$, the theorem is proved.

A compelling impediment to the use of Hotelling strategies is that they require the knowledge of the joint distribution of $\boldsymbol{\theta}$, $\mathbf{X}$, and $\mathbf{Y}$, or at least the knowledge of $\nu_i(\mathbf{X}, \mathbf{Y})$. The key task of the remainder of this article is to isolate some feasible circumstances in which this impediment can be overcome.

To begin, consider the strategies

$$\begin{aligned} \tilde{G}_i &= X_i + \epsilon \quad \text{if} \quad X_i < Y_i \\ &= X_i - \epsilon \quad \text{if} \quad X_i > Y_i , \end{aligned}$$

where the second guesser places his guess just a bit to the side of the first guess in the direction of his own "hunch" Y_i.

In some cases one can show that these hunch-guided guesses are in fact Hotelling strategies. Certainly, if the vectors (θ_i, Y_i, X_i), $1 \leq i \leq p$ are independent and $\theta_i | Y_i \sim N(Y_i, \sigma_Y^2)$, $X_i | \theta_i, Y_i = X_i | \theta_i \sim N(\theta_i, \sigma_X^2)$, then $\nu_i(\mathbf{X}, \mathbf{Y})$ is on the same side of X_i as Y_i. This immediately implies that the Hotelling and hunch-guided strategies will then coincide.

Without distributional assumptions on $\boldsymbol{\theta}$, one can no longer speak of the optimality of a guessing strategy, but the following result points out a case in which the second guesser can still realize a substantial advantage.

Theorem 2—Three-Quarter Theorem[1]: If $\tilde{X} = X - \theta$ and $\tilde{Y} = Y - \theta$ are identically distributed, independent and symmetric about zero, then the hunch-guided guess has probability $\frac{3}{4}$ of winning as $\epsilon \to 0$.

Proof: As $\epsilon \to 0$, the probability that the hunch-guided guesser loses is $P(\tilde{Y} < \tilde{X} < 0) + P(0 < \tilde{X} < \tilde{Y})$. By symmetry and exchangeability this probability also equals

$$2P(0 < \tilde{Y} < \tilde{X}) = P(0 < \tilde{X} \text{ and } 0 < \tilde{Y}) = \tfrac{1}{4} .$$

In a practical application of the three-quarter theorem the assumption of identical distributions might seem to pose some difficulties. It is reassuring that the result is quite robust. For example, assuming unbiased jointly normal guesses, the second guesser still wins with probability greater than .68 when $\operatorname{var}\tilde{Y}/\operatorname{var}\tilde{X} = 2.5$ and wins with probability greater than .59 when $\operatorname{var}\tilde{Y}/\operatorname{var}\tilde{X} = 10$. (These probabilities are easily confirmed by tables of the bivariate normal, e.g., Owen 1956.) The more detailed assessment of robustness in guessing competitions will be dealt with in a subsequent report, but one should note an obvious aspect of nonrobustness under gross changes in the model of Theorem 1 is that the probability of the second guesser winning will tend to $\frac{1}{2}$ or 1 according as $\operatorname{var}\tilde{Y}/\operatorname{var}\tilde{X}$ tends to ∞ or 0.

3. GAUSSIAN GUESSING

Since Hotelling strategies have been shown to be optimal, one would naturally like to provide a class of models in which the strategies can be determined explicitly. The main result of this section is to give such explicit strategies under a multivariate normal model studied by Lindley (1971) and Lindley and Smith (1972).

We write $\mathbf{U}|\mathbf{V}$ for the conditional distribution of $\mathbf{U}$ given $\mathbf{V}$, $\mathbf{1}_p$ for the row p vector $(1, 1, \ldots, 1)$, and I_p for the $p \times p$ identity matrix.

Our Gaussian model assumptions are the following:

$$\begin{aligned} \boldsymbol{\theta} | \mu &\sim N(\mu \mathbf{1}_p, \sigma_\theta^2 I_p) \\ \mu &\sim N(\mu_0, \sigma_\mu^2) \end{aligned} \tag{3.1}$$

and

$$\mathbf{X}, \mathbf{Y} | \theta, \mu \sim N(\boldsymbol{\theta}(I_p, I_p), \Gamma) ,$$

[1] A result equivalent to that given here was told to the first author in 1975 by R. Chacon and was known much earlier to R. Chacon and S. Kochen. The result was also known earlier to T. Cover in the form: Between two "equally matched" basketball teams the odds are 3 to 1 in favor of the team leading at the half.

where

$$\Gamma = \begin{pmatrix} \sigma_X{}^2 I_p & 0 \\ 0 & \sigma_Y{}^2 I_p \end{pmatrix} .$$

The physical motives behind this model are that the true weights θ of the persons we see are viewed as independent realizations of a single fixed random process that was itself once drawn from a population of random processes. For example, the parameter μ can be viewed as a geographically fixed quantity determined at an earlier time by (random) immigrations. The assumption of normality is made partially out of traditional convenience, but also because it seems justifiable in the weight-guessing example. The structural model together with the normality lead uniquely to the Gaussian model specified in (3.1). The promised explicit determination of the Hotelling strategy is now possible.

Theorem 3: Under the preceding Gaussian model the Hotelling strategy is

$$\begin{aligned} G_i{}^* = X_i + \epsilon \quad \text{if} \quad & X_i < \gamma\mu_0 + (1-\gamma)[\beta\bar{X} + (1-\beta)\bar{Y}] \\ & + (1-\alpha)[\beta(X_i - \bar{X}) \\ & + (1-\beta)(Y_i - \bar{Y})] \qquad (3.2) \\ = X_i - \epsilon \qquad & \text{otherwise} , \end{aligned}$$

where

$$\gamma = \sigma^{-2}(\sigma^{-2} + \sigma_X{}^{-2} + \sigma_Y{}^{-2})^{-1} ,$$

$$\sigma^2 = \sigma_\theta{}^2 + p\sigma_\mu{}^2$$

$$\alpha = \sigma_\theta{}^{-2}(\sigma_\theta{}^{-2} + \sigma_X{}^{-2} + \sigma_Y{}^{-2})^{-1}$$

and

$$\beta = \sigma_X{}^{-2}(\sigma_X{}^{-2} + \sigma_Y{}^{-2})^{-1} .$$

The proof of the preceding theorem depends on a multivariate calculation that we have deferred to the Appendix in order to take up directly the problem of interpreting the result.

The basic part is the mixture of means,

$$\gamma\mu_0 + (1-\gamma)[\beta\bar{X} + (1-\beta)\bar{Y}] ,$$

which is perturbed on trial i by the "mixture" of residuals $\alpha \cdot 0 + (1-\alpha)[\beta(X_i - \bar{X}) + (1-\beta)(Y_i - \bar{Y})]$. The coefficient $\beta = \sigma_X{}^{-2}(\sigma_X{}^{-2} + \sigma_Y{}^{-2})^{-1}$ appearing in these mixtures is near 0, $\frac{1}{2}$, or 1 accordingly as $\sigma_X{}^2\sigma_Y{}^{-2}$ is near ∞, 1, or 0. This ratio is one natural measure of the relative abilities of the two guessers, and this interpretation is reinforced by considering the extreme cases. When $\sigma_X{}^2\sigma_Y{}^{-2} \sim \infty$ the first guess is essentially ignored and when $\sigma_X{}^2\sigma_Y{}^{-2} \sim 0$ it is the hunch that is ignored. This last case is of particular interest since it corresponds to trying to outguess a far better informed adversary.

4. STEIN-GUIDED GUESSING

The strategies just derived have the drawback that they are functions of μ_0, $\sigma_\mu{}^2$, $\sigma_\theta{}^2$, $\sigma_X{}^2$, and $\sigma_Y{}^2$. Although the magnitude of μ_0 and of the relevant variance ratios may be sufficiently understood for some applications, the exact values of these quantities cannot generally be assumed to be known. The next objective is thus to derive reasonable estimates to the unknown mean and variances. One benefit of this analysis is a clearer understanding of the empirical fact (Efron and Morris 1973) the Stein estimator performs well with respect to the reward function V.

A Bayesian approach to the estimations given before can be made along the lines suggested in Lindley and Smith (1972), but such a procedure can prove quite complex (cf. discussion by V. D. Barnett in Lindley and Smith 1972). The estimators considered here are based on an empirical Bayes procedure that seems both simple and sensible.

As before, we write $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ and begin by transforming $\mathbf{Z}$ into a canonical form. Next we recall that $1_p{}^T 1_p = pP^T\Delta P$, where P denotes the $p \times p$ Helmert orthogonal matrix (cf. Bennett and Franklin 1954, p. 102) and Δ is the $p \times p$ matrix with 1 in the (1, 1) position and all other entries zero. We define $(\mathbf{U}, \mathbf{V})$ by

$$(\mathbf{U}, \mathbf{V}) = 2^{-\frac{1}{2}}(\mathbf{X}, \mathbf{Y}) \begin{pmatrix} I_p & I_p \\ I_p & -I_p \end{pmatrix} \begin{pmatrix} P^T & 0 \\ 0 & P^T \end{pmatrix} .$$

A straightforward calculation shows

$$(\mathbf{U}, \mathbf{V}) \sim N(\mu^*, \Sigma^*) ,$$

where

$$\mu^* = (2p)^{\frac{1}{2}}\mu_0(\mathbf{e}, 0) ,$$

$$\mathbf{e} = (1, 0, \ldots, 0) ,$$

$$\Sigma^* = \begin{bmatrix} \sigma_+{}^2 I_p & \sigma_-{}^2 I_p \\ \sigma_-{}^2 I_p & \sigma_+{}^2 I_p \end{bmatrix} + 2\sigma_\theta{}^2 \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix} + 2p\sigma_\mu{}^2 \begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} ,$$

and

$$\sigma_+{}^2 = \tfrac{1}{2}(\sigma_X{}^2 + \sigma_Y{}^2) , \quad \sigma_-{}^2 = \tfrac{1}{2}(\sigma_X{}^2 - \sigma_Y{}^2) .$$

From the canonical form shown one notes that $\sigma_\mu{}^2$ cannot be meaningfully estimated since there is only one degree of freedom available for its estimation.

We now turn to the analysis of important special cases that correspond to qualitatively different contexts.

Case A—Known Variances: We need to estimate only μ_0, and this is done by maximizing the Type II likelihood (cf. Good 1965). This calculation follows easily from equation (A.1) of the Appendix, and the estimator obtained is

$$\hat{\mu}_0 = (1_{2p}\Sigma_{ZZ}{}^{-1}Z^T)(1_{2p}\Sigma_{ZZ}{}^{-1}1_{2p}{}^T)^{-1} .$$

This simplifies further to just

$$\hat{\mu}_0 = \beta\bar{X} + (1-\beta)\bar{Y} ,$$

so the estimated Hotelling strategy becomes

$$\begin{aligned} G_i{}^* = X_i + \epsilon \quad \text{if} \quad & X_i < \alpha[\beta\bar{X} + (1-\beta)\bar{Y}] \\ & + (1-\alpha)[\beta X_i + (1-\beta)Y_i] \\ = X_i - \epsilon \qquad & \text{otherwise} , \end{aligned}$$

where the parameters α and β are as specified in Theorem 2.

Case B: $\sigma_\mu^2 = 0$; $\sigma_X^2 = \sigma_Y^2 = \delta^2$ and σ_θ^2 unknown. The canonical model simplifies to

$$(\mathbf{U}, \mathbf{V}) \sim N\left((2p)^{\frac{1}{2}}\mu_0(e, 0),\ \delta^2\begin{pmatrix} I_p & 0 \\ 0 & I_p\end{pmatrix} + 2\sigma_\theta^2\begin{pmatrix} I_p & 0 \\ 0 & 0\end{pmatrix}\right),$$

and this time estimators are easily found without carrying out the likelihood maximization. We take the estimators of μ_0, $\tau = (\delta^2 + 2\sigma_\theta^2)^{-1}$ and δ^2, respectively, given by

$$\hat{\mu}_0 = (2p)^{-\frac{1}{2}}U_1$$

$$\hat{\tau} = p(\sum_{i=2}^{p} U_i^2)^{-1}$$

and

$$\hat{\delta}^2 = p^{-1}\mathbf{V}\mathbf{V}^T .$$

In terms of $\mathbf{X}$ and $\mathbf{Y}$ we then get

$$\hat{\mu}_0 = \tfrac{1}{2}(\bar{X} + \bar{Y})$$

$$\hat{\tau} = p\left(\sum_{i=1}^{p}\left[\frac{1}{\sqrt{2}}(X_i + Y_i) - \frac{1}{\sqrt{2}}(\bar{X} + \bar{Y})\right]^2\right)^{-1}$$

and

$$\hat{\delta}^2 = (p)^{-1}\sum_{i=1}^{p}\left[\frac{1}{\sqrt{2}}(X_i - Y_i)\right]^2 .$$

The approximate Hotelling strategy for this case is therefore

$$\begin{aligned} G_i^* &= X_i + \epsilon \quad \text{if} \quad X_i < \hat{\alpha}\cdot\tfrac{1}{2}(\bar{X}+\bar{Y}) + (1-\hat{\alpha})\cdot\tfrac{1}{2}(X_i+Y_i) \\ &= X_i - \epsilon \qquad \text{otherwise} , \end{aligned}$$

where

$$\hat{\alpha} = \left[\sum_{i=1}^{p}\left[\frac{1}{\sqrt{2}}(X_i - Y_i)\right]^2\right] \cdot\left[\sum_{i=1}^{p}\left\{\frac{1}{\sqrt{2}}(X_i + Y_i) - \frac{1}{\sqrt{2}}(\bar{X} + \bar{Y})\right\}^2\right]^{-1} .$$

One should note that $\hat{\alpha}$ has a natural interpretation. It is just the ratio (between the guesser variance) $\div$ (between the trial variance). The optimal strategy favors using $\frac{1}{2}(\bar{X}+\bar{Y})$ when $\hat{\alpha}$ is close to 1 and favors $\frac{1}{2}(X_i+Y_i)$ when $\hat{\alpha}$ is close to 0. Also, the strategy can be improved slightly by replacing α by 1 if it happens that $\hat{\alpha} > 1$.

Case C: $\sigma_\mu^2 = 0$, $\sigma_Y^2 = \infty$, σ_X^2 known; σ_θ^2 unknown. This is a case we feel to be of particular interest. Calculating as before, we find that a Stein estimator determines the approximate Hotelling strategy, but that it plays a cameo role since the strategy simplifies to just "betting on the X_i side of $\bar{X}$." This simple result gives some theoretical justification to the otherwise somewhat mysterious empirical fact that $\bar{X}$ performs even better than the Stein estimator in terms of "gambler's" loss on the batting average data set (cf. Plackett's comment in Efron and Morris 1973, p. 416, and an easy computation).

The formal analysis begins as in Case B. Since $\sigma_Y^2 = \infty$ the $\mathbf{Y}$ is uninformative and the analysis must rest on $\mathbf{X}$. Also, since $\sigma_\mu^2 = 0$, the canonical form of the $\mathbf{X}$'s marginal model can be simplified to

$$\mathbf{U}^* \sim N(p^{\frac{1}{2}}\mu_0\mathbf{e} \ , \quad (\sigma_X^2 + \sigma_\theta^2)I_p) .$$

The obvious estimate of μ_0 is given by

$$\hat{\mu}_0 = \bar{X} ,$$

and if we require that the estimate, $\hat{\tau}_X$, of

$$\tau_X = (\sigma_X^2 + \sigma_\theta^2)^{-1}$$

be unbiased,

$$\hat{\tau}_X = (p - 3)(\sum_{i=1}^{p}(X_i - \bar{X})^2)^{-}$$

becomes the natural choice and the estimated Hotelling strategy is

$$\begin{aligned} G_i^* &= X_i + \epsilon \quad \text{if} \quad X_i < \hat{\alpha}^*\bar{X} + (1 - \hat{\alpha}^*)X_i \\ &= X_i - \epsilon \qquad \text{otherwise} , \end{aligned}$$

where $\hat{\alpha}^* = \min\{1, \hat{\alpha}\}$ and

$$\hat{\alpha} = \sigma_X^2(p - 3)[\sum_{i=1}^{p}(X_i - \bar{X})^2]^{-1} .$$

The direction of the guess on the side of X_i is determined by $\hat{\alpha}^*\bar{X} + (1 - \hat{\alpha}^*)X_i$, which is precisely the Stein estimator as modified by Lindley (cf. James and Stein 1961 and the discussion following Lindley and Smith 1972).

Now since the convex combination of $\bar{X}$ and X_i will always be on the same side of X_i as $\bar{X}$, the estimated Hotelling strategy can be more simply written as just

$$\begin{aligned} G_i^* &= X_i + \epsilon \quad \text{if} \quad X_i < \bar{X} \\ &= X_i - \epsilon \qquad \text{otherwise} . \end{aligned} \tag{4.1}$$

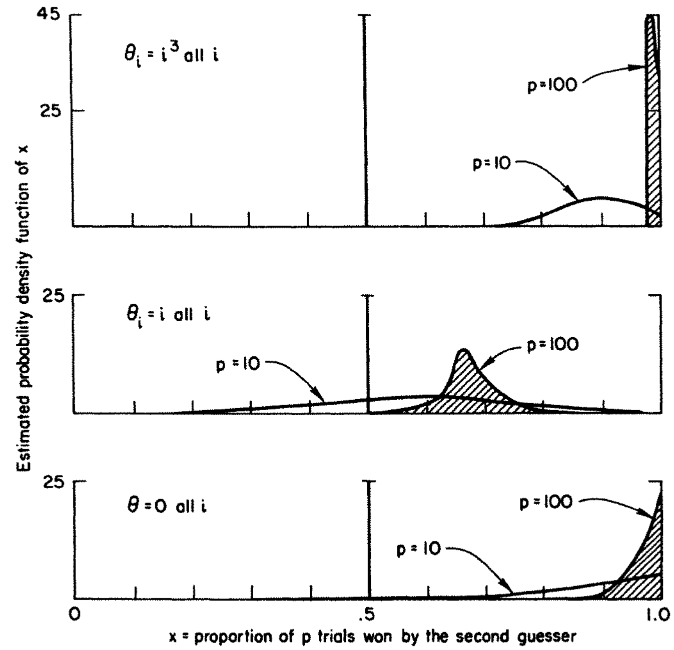
This is an extraordinarily simple procedure in a model that we feel may be realistic in several sporting and business contexts.

To assess the performance of this Stein-guided strategy, guessing trials were simulated for a variety of special cases. The θ_i were chosen as $\theta_i = 0$, $\theta_i = i$, and $\theta_i = i^3$ for each of $i = 1, 2, \ldots, p$ with $p = 10$ and then with $p = 100$; thus, in all, $3 \times 2 = 6$ cases were considered.

As an illustration of the computation consider the case in which $\theta_i = i$ and $p = 100$. In this case 200 repetitions were made as follows:

1. $\mathbf{X}$ was generated as $N(\boldsymbol{\theta}, \mathbf{I})$ with $\boldsymbol{\theta} = (1, 2, \ldots, 100)$.
2. G_i^* was calculated by (4.1) with $\epsilon = 10^{-6}$.
3. V was then calculated, and the process was repeated 200 times.
4. The 200 realizations of $V/100$ were used to estimate the density of $V/100$, the percentage of times the second player wins using the $\mathbf{G}^*$ of (4.1).
5. This density was plotted in Figure B (in this case, the unshaded density in the middle graph).

A. Estimated Density of the Proportion of Second Guesser Wins Using the Stein-Guided Optimal Strategy (based on runs of 200 contests)

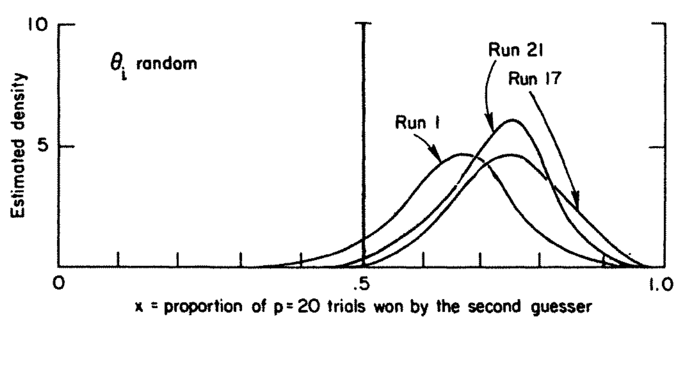


From Figure A one learns by looking at the unshaded density in the top graph that when as many as 10 parameters growing like i^3 are to be guessed the modal percentage of correct guesses made by the second guesser is about 95 percent. The general conclusions to be drawn from Figure A are

1. The more parameters to be guessed, the greater the advantage to the second guesser.
2. The more spread out the θ_i to be guessed, the more the advantage to the second guesser.

In Figure B, these conclusions are further examined by taking the θ_i themselves to be random. Here $p = 20$ was fixed throughout. First we took a realization of $\boldsymbol{\theta} \sim N(0, 4\mathbf{I}_{20})$. Then 200 of the **X**'s we generated with the same fixed underlying $\boldsymbol{\theta}$ (just as in Figure A). The density of $V/200$ was estimated as before, and altogether 25 runs were made. The 25 runs produced remarkably similar estimates of the density of $V/200$, and the estimates from runs number 1, 17, and 21 were selected as indicative of the variability in the 25 runs.

B. Estimated Density of the Proportion of Second Guesser Wins in $p = 20$ Trials Using the Stein-Guided Optimal Strategy (three runs of 100 contests and combined runs of 2,500 contests)



5. REALITIES OF APPLICATION

Almost any discussion of the preceding theory eventually turns to the problem of football betting, and it seems generally worthwhile to note why the theory is not applicable to that problem. A key reason is that the bookie (or person "setting the line") is not trying to estimate the actual point spread. The bookie is trying to produce a point spread that will produce a nearly equal number of takers on each side of the spread. The bookie is therefore not a first guesser in the sense of this article, and our theory naturally does not apply.

Consider instead two bookies of equal caliber, one of whom sets his line on Monday and the other on Tuesday (for the game on Sunday). If the Tuesday bookie wished to obtain only a more even distribution of customers on either side of his line than the Monday bookie, he should be able to do so in almost three-quarters of the games by using the hunch-guided guessing of Section 2. In this case, each bookie is a bona fide guesser of that spread s that will evenly split the pool of bettors.

Since there are actually many games each week, the Tuesday bookie could actually outperform the Monday bookie by using the Stein-guided strategy of Section 4, particularly (4.1). The assumptions of (3.1) may not be applicable to the whole set of games; but if one considers only noncharismatic games outside the bookie's city, then (3.1) seems reasonable. (This is a stratification step to obtain increased homogeneity of the spreads to be guessed.)

The examples put forward before are in the long tradition of *gedanken experimente,* and the problem of producing a truly telling application remains open. An intriguing aspect of a theory of this nature is that it is only necessary to find one good application.

APPENDIX: PROOF OF THEOREM 3

By Theorem 1 the problem depends on the calculation of the posterior median $\nu_i(\mathbf{X}, \mathbf{Y})$, which by the normality assumptions (3.1) coincides with the posterior mean. The argument given here for completeness is similar to those of Lindley (1971) and Lindley and Smith (1972). It depends on the well-known fact (cf. Anderson 1958, p. 27) that if **U** and **V** are jointly normal then

$$\mathbf{U} \mid \mathbf{V} \sim N(E\mathbf{U} - E\mathbf{V} \cdot \zeta + \mathbf{V}\zeta, \Sigma_{\mathbf{U}\cdot\mathbf{V}}) \ ,$$

where $\zeta = \Sigma_{\mathbf{VV}}^{-1}\Sigma_{\mathbf{VU}}$, $\Sigma_{\mathbf{U}\cdot\mathbf{V}} = \Sigma_{\mathbf{UU}} - \Sigma_{\mathbf{UV}}\Sigma_{\mathbf{VV}}^{-1}\Sigma_{\mathbf{UV}}$, $\Sigma_{\mathbf{VV}}$ denotes the covariance matrix of **V** and so on.

Setting $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ and applying the preceding identities to (3.1), we have

$$\mathbf{Z} \sim N(\mu_0(1_p, 1_p), \Sigma_{\mathbf{ZZ}}) \ , \tag{A.1}$$

where

$$\Sigma_{\mathbf{ZZ}} = \operatorname{diag}\{\sigma_X^2, \sigma_Y^2\} \otimes I_p + \sigma_\theta^2 J_2 \otimes I_p + \sigma_\mu^2 J_2 \otimes 1_p^T 1_p \ ,$$

where $\otimes$ denotes the Kronecker matrix product and J_2 is the matrix with all 1 entries. Further,

$$E(\boldsymbol{\theta}|\mathbf{Z}) = E(\boldsymbol{\theta}) - E(\mathbf{Z})\cdot\boldsymbol{\zeta} + \mathbf{Z}\cdot\boldsymbol{\zeta} \ ,$$

where

$$\boldsymbol{\zeta} = \Sigma_{ZZ}^{-1}\Sigma_{Z0}$$

with

$$\Sigma_{Z\theta} = \begin{pmatrix} I_p \\ I_p \end{pmatrix} \Sigma_{\theta\theta}$$

and

$$\Sigma_{\theta\theta} = \sigma_\theta^2 I_p + \sigma_\mu^2 1_p{}^T 1_p \ .$$

We now determine Σ_{ZZ}^{-1}. First we note that

$$1_p{}^T 1_p = pP^T\Delta P \ ,$$

where P denotes the Helmert orthogonal matrix (cf. Bennett and Franklin 1954, p. 102) and Δ is the $p \times p$ matrix with 1 in the (1, 1) position and all remaining entries zero.

One then notes that

$$(I_2 \otimes P)\Sigma_{ZZ}(I_2 \otimes P^T)$$

reduces to just

$$\operatorname{diag}\{\sigma_X^2, \sigma_Y^2\} \otimes I_p + \sigma_\theta^2 J_2 \otimes I_p + p\sigma_\mu^2 J_2 \otimes \Delta \ ,$$

which makes it straightforward to show

$$\Sigma_{ZZ}^{-1} = (I_2 \otimes P^T)\Theta(I_2 \otimes P) \ ,$$

where

$$\Theta = \begin{pmatrix} \sigma_X^{-2} & 0 \\ 0 & \sigma_Y^{-2} \end{pmatrix} \otimes \begin{pmatrix} \gamma & 0 \\ 0 & \alpha I_{p-1} \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \begin{pmatrix} \sigma_X^{-2}\sigma_Y^{-2}\sigma^2\gamma & 0 \\ 0 & \sigma_X^{-2}\sigma_Y^{-2}\sigma_\theta^2\alpha I_{p-1} \end{pmatrix},$$

where γ and α are in Theorem 2.

From these results an explicit expression for $\mathbf{Z}\cdot\boldsymbol{\zeta}$ is readily obtained. We first note

$$\mathbf{Z}\cdot\boldsymbol{\zeta} = (\mathbf{X}P^T, \mathbf{Y}P^T)\ \Theta\ (P^T, P^T)^T\Sigma_{\theta\theta}$$

and

$$\Theta(P^T, P^T)^T = (\sigma_X^{-2}, \sigma_Y^{-2})^T \otimes \left\{\begin{pmatrix} \gamma & 0 \\ 0 & \alpha I_{p-1} \end{pmatrix} P\right\}.$$

Thus

$$\begin{aligned} \mathbf{Z}\cdot\boldsymbol{\zeta} &= (\mathbf{X}P^T, \mathbf{Y}P^T) \\ &\quad \cdot\left[(\sigma_X^{-2}, \sigma_Y^{-2})^T \otimes \begin{pmatrix} \gamma & 0 \\ 0 & \alpha I_{p-1} \end{pmatrix}\right] \\ &\quad \cdot(\sigma_\theta^2 I_p + p\sigma_\mu^2\Delta)P = (\sigma_X^{-2}X + \sigma_Y^{-2}Y)A \ , \end{aligned} \tag{A.2}$$

where

$$A = P^T \begin{pmatrix} \sigma^2\gamma & 0 \\ 0 & \sigma_\theta^2\alpha I_{p-1} \end{pmatrix} P \ .$$

Now represent $\mathbf{X}$ as

$$X = \bar{X}1_p + (X - \bar{X}1_p) \ .$$

Then we have

$$\mathbf{X}P^T = \bar{X}(p^{\frac{1}{2}}, 0, \ldots, 0) + (0, \mathbf{X}_{(2)}) \ ,$$

where

$$\mathbf{X}_{(2)} = \left(\frac{X_1 - X_2}{\sqrt{1.2}}, \frac{X_1 + X_2 - 2X_3}{\sqrt{2.3}}, \ldots, \frac{X_1 + X_2 + \cdots + X_{p-1} - (p-1)X_p}{(p(1-p))^{\frac{1}{2}}}\right),$$

so

$$\begin{aligned} \mathbf{X}A &= \sigma^2\gamma\bar{X}(p^{\frac{1}{2}}, 0, \ldots, 0)P + \sigma_\theta^2\alpha(0, \mathbf{X}_{(2)})P \\ &= \sigma^2\gamma\bar{X}1_p + \sigma_\theta^2\alpha(\mathbf{X} - \bar{X}1_p) \ . \end{aligned}$$

A similar result holds for $\mathbf{Y}A$. Introducing the resulting expressions for $\mathbf{X}A$ and $\mathbf{Y}A$ in (A.2) yields

$$\begin{aligned} (\sigma_X^{-2}\mathbf{X}+\sigma_Y^{-2}\mathbf{Y})A &= (1-\gamma)[\beta\bar{X}+(1-\beta)\bar{Y}]1_p \\ &\quad +(1-\alpha)[\beta(\mathbf{X}-\bar{X}1_p)+(1-\beta)(\mathbf{Y}-\bar{Y}1_p)] \end{aligned} \tag{A.3}$$

since

$$\gamma\sigma^2(\sigma_X^{-2} + \sigma_Y^{-2}) = 1 - \gamma$$

and

$$\alpha\sigma_\theta^2(\sigma_X^{-2} + \sigma_Y^{-2}) = 1 - \alpha \ .$$

From the expression just obtained in (A.3), $E(\mathbf{Z})\cdot\boldsymbol{\zeta}$ and hence $E(\boldsymbol{\theta}) - E(\mathbf{Z})\cdot\boldsymbol{\zeta}$ are easily found. To get $E(\mathbf{Z})\cdot\boldsymbol{\zeta}$, simply substitute $\mu_0 1_p$ for both $\mathbf{X}$ and $\mathbf{Y}$. This gives

$$E(\mathbf{Z})\cdot\boldsymbol{\zeta} = \gamma\sigma^2[\sigma_X^{-2} + \sigma_Y^{-2}]\mu_0 1_p \ .$$

At the same time $E(\theta) = \mu_0 1_p$, so we have

$$E(\boldsymbol{\theta}) - E(\dot{\mathbf{Z}})\cdot\boldsymbol{\zeta} = \gamma\mu_0 1_p \ . \tag{A.4}$$

The proof of Theorem 3 is now an immediate consequence of (A.3) and (A.4).

[*Received September 1978. Revised March 1980.*]

REFERENCES

Anderson, T.W. (1958), *Introduction to Multivariate Statistical Analysis*, New York: John Wiley & Sons.

Bennett, C.A., and Franklin, N.L. (1954), *Statistical Analysis in Chemistry and the Chemical Industry*, New York: John Wiley & Sons.

Efron, B., and Morris, C. (1973), "Combining Possibly Related Estimation Problems" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 35, 379–421.

Feller, W. (1966), *An Introduction to Probability Theory and Its Applications* (Vol. 2), New York: John Wiley & Sons.

Good, I.J. (1965), *The Estimation of Probabilities*, Cambridge, Mass.: MIT Press.

Hotelling, H. (1929), "Stability in Competition," *Economic Journal*, 39, 41–57.

James, W., and Stein, C. (1961), "Estimation With Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*, 1, 361–379.

Lindley, D.V. (1971), "The Estimation of Many Parameters," in *Foundations of Statistical Inference*, eds. V.P. Godambe and D.A. Sprott, Toronto: Holt, Rinehart and Winston of Canada, 435–455.

Lindley, D.V., and Smith, A.F.M. (1972), "Bayes Estimates for the Linear Model" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

Owen, D.B. (1956), "Tables for Computing Bivariate Normal Probabilities," *Annals of Mathematical Statistics*, 27, 1075–1090.

Stein, C. (1962), "Confidence Sets for the Mean of a Multivariate Normal Distribution" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 24, 265–296.