Strong profiling is not mathematically optimal for discovering rare malfeasors

William H. Press[1]

Department of Computer Science and School of Biological Sciences, University of Texas, Austin, TX 78703; and Los Alamos National Laboratory, Los Alamos, NM 87545

Contributed by William H. Press, December 23, 2008 (sent for review December 4, 2008)

The use of profiling by ethnicity or nationality to trigger secondary security screening is a controversial social and political issue. Overlooked is the question of whether such actuarial methods are in fact mathematically justified, even under the most idealized assumptions of completely accurate prior probabilities, and secondary screenings concentrated on the highest-probablity individuals. We show here that strong profiling (defined as screening at least in proportion to prior probability) is no more efficient than uniform random sampling of the entire population, because resources are wasted on the repeated screening of higher probability, but innocent, individuals. A mathematically optimal strategy would be ''square-root biased sampling,'' the geometric mean between strong profiling and uniform sampling, with secondary screenings distributed broadly, although not uniformly, over the population. Square-root biased sampling is a general idea that can be applied whenever a ''bell-ringer'' event must be found by sampling with replacement, but can be recognized (either with certainty, or with some probability) when seen.

screening | square-root biased sampling | rare events

In a large population of individuals labeled $j = 1,2,\ldots, N$, governments attempt to find the rare malfeasor $j = j_*$ (terrorist, for example, refs. 1–3)[†] by assigning prior probabilities p_j to individuals j, in some manner estimating the chance that each is a malfeasor. Societal resources for secondary security screening are then concentrated against individuals with the largest priors. We may call this "strong profiling" if the concentration is at least proportional to p_j for the largest values of p_j. Secondary screening may take the form of airport luggage search, police investigation, physical search, or other societally sanctioned but personally intrusive actions.

In general, police strategies that use such priors are termed actuarial methods (4). Racial profiling, as commonly defined (5), is one such actuarial method. It occurs when an individual's prior is explicitly conditioned on his or her race, ethnicity, nationality, or religion. What distinguishes racial profiling, and actuarial methods generally, from investigational methods often perceived as more acceptable is that the prior probabilities are associated with the individual a priori, and not associated with evidence of any actual criminal conduct.

This article looks at the first-order efficiency of profiling methods: How much screening must we do, on average, to catch a malfeasor. There are also second-order effects, not addressed here. Groups may change their behavior in response to being profiled (or not). Indeed, it is a matter of debate as to whether second-order effects are net positive or negative (4, 6, 7), because nonprofiled groups may (under lower scrutiny) increase their antisocial behaviors, even as such behaviors by profiled groups may decrease. Given such second-order ambiguities, elucidation of the first-order problem seems useful, especially because (as we will see) it has some nonobvious features.

Authoritarian vs. Democratic Strategies

For simplicity, assume that there is only a single malfeasor $j = j_*$. (Below, we will indicate why this assumption is not actually necessary to what follows.) An omnipotent authoritarian government can enumerate all of its citizens j, and then screen each in turn, that is, by sampling *without replacement*. If the government knows nothing else about its citizens, then it must simply screen all in an arbitrary order until it finds the malfeasor j_*. On average, this will occur after $N/2$ samples.

But what if the government can assign a meaningful prior probability p_j to each individual $j = 1,\ldots, N$? Then the optimal strategy is to sort the p_j's from largest to smallest value, and then screen individuals in the population, visiting each just once in decreasing order of their probability. This "authoritarian" strategy can easily be seen to find the malfeasor with the smallest possible average number of tests, because any other screening order can be improved by the pairwise exchange of any 2 out-of-order individuals. The smallest possible average number of tests is thus

$$\mu_A \equiv \sum_{i=1}^{N} i p_{(i)} \qquad [1]$$

where $p_{(i)}$ is the order statistic; that is, $p_{(i)}$ is the ith largest value among the p_j's.

For moral or practical reasons, democratic governments employ strategies not requiring the enumeration of all individuals and the availability of their individual dossiers at every checkpoint. Thus, the only "democratic" strategies available involve sampling *with replacement*: Individuals may be sampled with some individualized profile sampling probability q_j determined by a public policy. But, the sampling process is memory-less in that an individual is liable to be sampled more than once, according to his profile—for example, whenever he goes through an airport security checkpoint.

Square-Root Biased Sampling. In the democratic case, the probability of not finding the malfeasor on exactly $m \geq 0$ looks, and finding him on the $m + 1$st is $(1 - q_{j^*})^m q_{j^*}$. So the mean number of looks required is

$$\sum_{m=0}^{\infty} (m + 1)(1 - q_{j^*})^m q_{j^*} = 1/q_{j^*} \qquad [2]$$

(an answer that we could have written down by inspection). We can take the expectation of this over the remaining random variable, namely, which value j is j_*. This expectation, which we want to minimize subject to $\Sigma q_i = 1$, is thus

Author contributions: W.H.P. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

Freely available online through the PNAS open access option.

[1]E-mail: wpress@cs.utexas.edu.

[†]Also see Siggins P, Racial profiling in an age of terrorism. Presentation at Markkula Center for Applied Ethics, March 20, 2002, Santa Clara, CA. Available at http://www.scu.edu/ethics/publications/ethicalperspectives/profiling.html.

$$\mu_D = \sum_{j=1}^{N} p_j/q_j. \qquad [3]$$

A straightforward minimization with a Lagrange multiplier gives the optimal choice for the q_j's

$$q_j = p_j^{1/2} \Big/ \sum_{i=1}^{N} p_i^{1/2} \qquad [4]$$

and the mean number of tests per found malfeasor,

$$\mu_D = \left(\sum_{j=1}^{N} p_j^{1/2} \right)^2. \qquad [5]$$

In words, Eq. **4** says that individuals should be selected for screening in proportion to the *square root* of their prior probability. This does use the priors, but only weakly; it results in secondary screening being distributed over a much larger segment of the population than would be the case with strong profiling. Although Eq. **4** should be a well-known result, we are not aware of any published reference earlier than Abagyan and collaborators in a completely different context (8, 9).

Comparison with Naive Sampling Strategies. It is informative to compare Eq. **5**, the optimal result, with the corresponding results for 2 naive (but still democratic) sampling strategies. First, uniform sampling with replacement (ignoring the p_i's):

$$q_i = \frac{1}{N}, \qquad \mu = \sum_i \frac{p_i}{1/N} = N. \qquad [6]$$

Second, sampling in proportion to p_i [what would be called *importance sampling* in the context of Monte Carlo integration (10)]. This seems like a natural way to sample likely malfeasors more heavily, and is an example of what we have termed "strong profiling." However, as long as no p_i's are exactly zero, it gives

$$q_i = p_i, \qquad \mu = \sum_i \frac{p_i}{p_i} = N \qquad [7]$$

exactly the same as uniform sampling, Eq. **6**. The reason that this strong profiling strategy is inefficient is that, on average, it keeps retesting the same innocent individuals who happen to have large p_j values. The optimal strategy is optimal precisely because it avoids this oversampling.

Efficiency of Democratic Strategy for Model Distributions

A figure of merit for the optimal democratic sampling, Eq. **4**, is its efficiency with respect to the best authoritarian strategy, Eq. **1**. Define the efficiency as $\varepsilon = \mu_A/\mu_D$, a value between 0 (ineffectual) and 1 (as good as best authoritarian). To get a sense of how square root biased sampling performs, we can compute ε for various assumptions about the distribution of p_j's. For example, if the prior probability is concentrated uniformly in some number of individuals N_0 (out of N), and negligible (but not strictly zero) in the remaining $N - N_0$, then $\varepsilon \approx 1/2$, independent of N_0. That is, the optimal democratic sampling is just a factor of 2 less efficient than the authoritarian strategy; this is entirely due to its repeated sampling of some individuals. In this case, both uniform sampling and importance sampling (above) would have much smaller efficiencies, $\varepsilon \approx (1/2)N_0/N$.

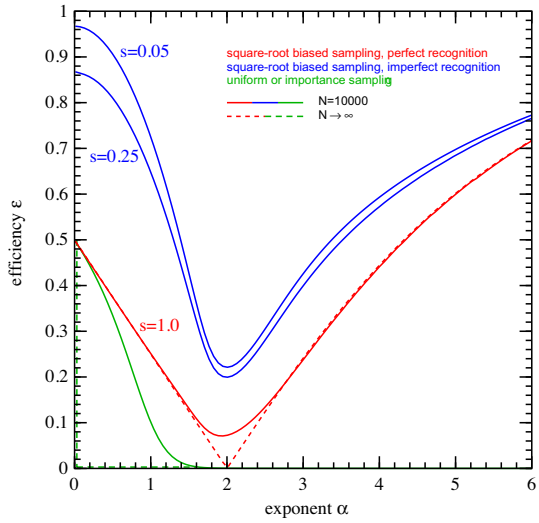

Fig. 1. Efficiency of various "democratic" sampling strategies with respect to the "perfect authoritarian" strategy, for power-law distributions of prior probability $p_j \propto j^{-\alpha}$ with exponent α. Solid curves are calculated for population size $N = 10{,}000$; dashed curves show asymptotic results for $N \to \infty$. The red and blue curves are for optimal square-root biased sampling and (except for the special case $\alpha \approx 2$) maintain finite efficiency even as N becomes large. The green curves represent both uniform importance sampling (a type of strong profiling). Their efficiency is always suboptimal, and goes to zero for large N or increasing exponent α. Blue and red curves differ by whether the malfeasor can always be recognized ($s = 1.0$), or can be recognized only with probabilities $s = 0.25$ or 0.05. As the recognition probability becomes small, the advantage of an authoritarian strategy over a democratic strategy decreases.

Another interesting case is the "scale-free" power-law distribution $p_j \propto 1/j^{\alpha}$. This yields (see *Appendix*)

$$\varepsilon \approx \begin{cases} (2-\alpha)/4 & \text{for } 0 \le \alpha < 2, \\ \zeta(\alpha-1)/\zeta(\alpha/2)^2 & \text{for } \alpha > 2, \end{cases} \qquad [8]$$

where ζ is the Riemann Zeta function. For $\alpha \approx 2$, both cases are approximately $|\alpha - 2|/4$. For large α, $\varepsilon \approx 1$. In all of these cases, for any fixed value α not near 2, the democratic strategy is within a constant efficiency factor of the authoritarian strategy. The singular case $\alpha \to 2$ gives a result approaching zero with large N, $\varepsilon \to 1/\log N$, which is small only logarithmically.

The red curves in Fig. 1 show these behaviors, both for the asymptotic case of $N \to \infty$ (Eq. **8**), and for a finite case with $N = 10{,}000$ (that is, $1/\ln(N) \approx 0.11$, solid curve). Shown in green are the corresponding efficiencies for the democratic, but suboptimal, strategies of uniform sampling or importance sampling (i.e., strong profiling), the identical results of Eqs. **6** and **7**. Although the efficiency is finite for finite $N = 10{,}000$ (solid green curve), as $N \to \infty$ it goes to zero except at $\alpha = 0$ (dotted green curve).

A third test case of interest is $p_j \propto \exp(-\gamma j^{1/\beta})$. This occurs in cases where the j's are ordered by radius from the origin in a (say) high-dimensional space, and the probability decreases away from the origin either exponentially or as a multivariate normal distribution. It also applies to a mixture of such distributions, and thus to Gaussian mixture models, in general. In all such cases β is related to (and increases with) the dimension of the space. One readily calculates (see *Appendix*),

$$\varepsilon = \Gamma(1+2\beta)/[2^{2\beta+1}\Gamma(1+\beta)^2] \qquad [9]$$

with the limiting cases $\approx 1/2$ as $\beta \to 0$ and $\approx (1/2)(\pi\beta)^{-1/2}$ as β becomes large, a surprisingly modest increase for what might have been thought to be a dimensional explosion of volume.

Case of Probabilistic Recognition of Malfeasor

What if we can not always recognize the malfeasor j_*, even when we sample him? This could be because some additional random condition (not within our control) is required for recognition. Suppose that, on each look, the probability that we recognize the malfeasor is s_i, $i = 1,\ldots, N$; and that (for simplicity) each look is independently random.

Best Authoritarian Strategy. The authoritarian strategy that led to Eq. **1** is now no longer valid, because it looked at each enumerated person only once. Instead, if we have looked at person i already m_i times, then its probability of both being the malfeasor and escaping previous detection is $(1 - s_i)^{m_i} p_i$. So the total remaining probability in which the malfeasor is to be found is

$$P = \sum_i (1 - s_i)^{m_i} p_i. \qquad [10]$$

Now suppose we look next at person j. Then the change in Eq. **10** is

$$-\Delta P = (1 - s_j)^{m_j} p_j - (1 - s_j)^{m_j+1} p_j = (1 - s_j)^{m_j} s_j p_j = u_{m_j,j}. \qquad [11]$$

Thus, the greedy strategy, which can easily be seen to be also the optimal strategy, is to visit the j's according to the order statistic of the 2-dimensional lattice $u_{m,j}$, with $j = 1,\ldots, N$ and integer $m \geq 0$. Denoting that order statistic by $u_{(i)}$, we have

$$\mu_{A'} = \sum_i i u_{(i)} \qquad [12]$$

because one easily checks that

$$\sum_{m,j} u_{m,j} \equiv \sum_i u_{(i)} = 1. \qquad [13]$$

Best Democratic Strategy. We derive the best democratic strategy as before. The mean number of looks to success is $(s_{i*}q_{i*})^{-1}$, so we want to minimize

$$\mu = \sum_j \frac{p_j}{s_j q_j}. \qquad [14]$$

Now the same calculation as before gives,

$$q_j = \sqrt{\frac{p_j}{s_j}} \Big/ \sum_i \sqrt{\frac{p_i}{s_i}} \qquad [15]$$

and

$$\mu_{D'} = \left(\sum_i \sqrt{\frac{p_i}{s_i}} \right)^2. \qquad [16]$$

The optimal sampling, still square-root biased as before, is seen to expend relatively more samples on the less-likely-to-recognize cases (smaller values s_i). It is the opposite of the proverbial "looking under the lamppost." In rough terms, if you *do not* spend quite a lot of time "not under the lamppost," then you provide excessive sanctuary for the malfeasor who might be there.

We have calculated the efficiency of the best democratic strategy, now $\varepsilon = \mu_{A'}/\mu_{D'}$, for the same kinds of test distributions for the p_j's as was done above, with various assumptions about the s_j's (for example, constant values $0 < s \leq 1$). Eq. **16** is evaluated straightforwardly, whereas Eq. **12** benefits from the use of a heap data structure to iterate efficiently over the (m, j) lattice. Specifically, because $u_{m,j}$ decreases monotonically with m, we can store the next-to-use value for each j on the heap, and then efficiently retrieve the largest one (and store the next). In all cases tried, the efficiency ε increases monotonically as s (or any of the s_j's) decreases from the original case of perfect recognition, $s_j = 1$. Intuitively, the authoritarian advantage of being able to sample without replacement becomes less important when the malfeasor is more difficult to recognize. The blue curves in Fig. 1 show the efficiency ε in the case of a power-law distribution for the p_j's with $N = 10{,}000$, for 2 assumed values of $s_j = s$ (0.25 and 0.05), computed as described.

Discussion

None of the results in the article actually depend on our original assumption of a single malfeasor. To see this, note that all results apply separately to each of multiple malfeasors as if he were the only one. Because the identical sampling prescription is obtained in each case, it is optimal for minimizing the mean number of tests for *each* malfeasor. That is, there is no better strategy for *any* malfeasor.

The idea of sampling by square-root probabilities is quite general and can have many other applications. It applies whenever a "bell-ringer" event must be found by sampling with replacement, but can be recognized when seen. For example, one can thus sample paths through a trellis or hidden Markov model when their number is too large to enumerate explicitly, but one path can be recognized (e.g., by secondary testing) as the desired bell ringer. It seems peculiar that the method is not better known.

Appendix

We here calculate the approximations for ε in Eqs. **8** and **9**. Because ε is invariant under scaling all of the p_i's by a constant, we may here use nonnormalized p_i's. For $0 \leq \alpha < 2$, we approximate the sums by integrals, which is accurate for large N,

$$\sum_i i p_{(i)} \approx \int_1^{N+1} x^{1-\alpha} dx \approx \frac{1}{2-\alpha} (N+1)^{2-\alpha}$$

$$\left(\sum_i p_i^{1/2} \right)^2 \approx \left(\int_1^{N+1} x^{-\alpha/2}\, dx \right)^2 \approx \left(\frac{2}{2-\alpha} \right)^2 (N+1)^{2-\alpha}$$

[17]

yielding one case of Eq. **8**. For $\alpha > 2$, the sums are now dominated by small values of i, so we can approximate by extending the sums to infinity. If $\zeta(\,)$ is the Riemann Zeta function, we have $p_{(i)} \approx i^{-\alpha}/\zeta(\alpha)$ and

$$\sum_i i p_{(i)} \approx \frac{\zeta(\alpha - 1)}{\zeta(\alpha)}$$

$$\left(\sum_i p_i^{1/2} \right)^2 \approx \frac{\zeta(\alpha/2)^2}{\zeta(\alpha)} \qquad [18]$$

which implies the second case of Eq. **8**.

To obtain Eq. **9**, we approximate the sums by integrals,

$$\sum_i i p_{(i)} \approx \int_0^\infty x \exp(-\gamma x^{1/\beta}) dx = \gamma^{-2\beta} \beta \Gamma(2\beta)$$

$$\left(\sum_i p_i^{1/2} \right)^2 \approx \left(\int_0^\infty \exp\left(-\frac{1}{2} \gamma x^{1/\beta} \right) dx \right)^2 = 2^{2\beta} \gamma^{-2\beta} \Gamma(\beta + 1)^2.$$

[19]

ACKNOWLEDGMENTS. I thank Martha Minow, Nozer Singpurwalla, Sallie Keller-McNulty, Richard Garwin, and an anonymous referee for helpful comments and discussion.

1. Ellmann SJ (2003) Racial profiling and terrorism. *New York Law School Law Rev* 46:675–730.
2. Lund N (2003) The conservative case against racial profiling in the war on terrorism. *Albany Law Rev* 66:329–342.
3. London H (2003) Profiling as needed. *Albany Law Rev* 66:343–347.
4. Harcourt BE (2007) *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (Univ of Chicago Press, Chicago).
5. American Civil Liberties Union (2008) *Campaign Against Racial Profiling*. Available at http://www.aclu.org. Accessed January 12, 2009.
6. Knowles J, Persico N, Todd P (2001) Racial bias in motor vehicle searches: Theory and evidence. *J Political Econ* 109:203–229.
7. Dominitz J, Knowles J (2006) Crime minimization and racial bias: What can we learn from police search data? *Econ J* 116:F368–F384.
8. Abagyan RA, Totrov M (1999) Ab initio folding of peptides by the optimal-bias Monte Carlo minimization procedure. *J Comput Phys* 151:402–421.
9. Zhou Y, Abagyan R (2002) Efficient stochastic global optimization for protein structure prediction. *Rigidity Theory and Applications*, eds Thorpe MF, Duxbury PM (Springer, New York).
10. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical Recipes: The Art of Scientific Computing* (Cambridge Univ Press, New York), 3rd Ed, section 7.9.1.