DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Enhanced Security Checks at Airports: Minimizing Time to Detection or Probability of Escape?

The Harvard community has made this article openly available.
Please share how this access benefits you. Your story matters.

(Article begins on next page)

Enhanced Security Checks at Airports:
Minimizing Time to Detection or Probability of Escape?

Xiao-Li Meng[*]

January 21, 2011

1 A Turbulent Reading

As a statistician frequently in the friendly skies, I was intrigued by the article "To catch a terrorist: can ethnic profiling work?" in *Significance* (Press, 2010; hereafter "the article", and all page numbers below refer to this article). As a matter of fact, I read it during a flight, on my way to a conference honoring Larry Brown's 70th birthday. The statement that (p. 167) "Surprisingly, and bizarrely, this turns out to be the most efficient way of catching the terrorist." particularly caught my eye. I didn't know it was possible to formulate statistically or probabilistically "the most efficient way" to catch a terrorist. The indication that something is rather unusual of course further enticed me to study it carefully.

The optimality formulation turned out to be possible because the article adopted a simple mathematical setting. Specifically, the article assumed a population of N individuals passing through a network of airport checkpoints, and exactly one of them is a terrorist. It also assumed that "perfect actuarial information" (p. 166) is available in the form of p_i, that is, the prior probability that individual i is the terrorist. The article then asked: at airport checkpoints, what is the optimal probability q_i to pull out the ith individual for an enhanced security screening (e.g., a pat-down or a full body scan), given the (expected) resource constraint that $\sum_{i=1}^{N} q_i = M(\leq N)$? The optimality criterion used was to minimize the expected number of times that the terrorist goes through checkpoints "before he happens to be selected for screening" (p. 166):

$$\mu = \sum_{i=1}^{N} \frac{p_i}{q_i}. \tag{1}$$

The aforementioned "the most efficient way" then refers to the "optimal probabilities":

$$q_i = \frac{M\sqrt{p_i}}{\sum_{k=1}^{N} \sqrt{p_k}}, \quad i = 1, \ldots, N. \tag{2}$$

I felt a bit of turbulence. Earlier on page 166, it was stated that without resource constraint the optimal strategy is to pull out "*every single passenger*" (emphasis is original), i.e., to set $q_i = 1$ for all i. However, (2) does not lead to this conclusion when $M = N$. Nor does it lead to the obvious strategy at the other extreme, that is, when $M = 1$. If only one person can be selected, what else can be better than pulling out the person with the highest prior probability, when the prior information is "perfect"?

---

[*]Xiao-Li Meng is Whipple V. N. Jones Professor and Chair of Statistics, Harvard University.

An astute reader may have also noticed that in general the $q_i's$ of (2) cannot possibly be the optimal *probabilities* under any criterion, because they are not guaranteed to be between zero and one. The same problem occurred when the article argued (p. 166) that "strong profiling", by which it meant $q_i = Mp_i$, has the same expected value μ as uniform sampling, because again $q_i = Mp_i$ can exceed one unless $M \leq 1$. But here "$M = \sum_{i=1}^{N} q_i$ is the number of passengers out of total N that we can afford to select for secondary screening" (p. 166) and hence $M \geq 1$; it would make little practical sense to assume that airport checkpoints can only afford pulling out one person for an enhanced check.

The source for the mathematical oversight seems to be an unfortunate mix of notation. Before presenting its equation (4) (same as (2) above), the article stated that (p. 166) "subject to the constraint that M, the proportion of people overall that we check, is held constant," suggesting a redefinition of M as M/N. However, if this redefinition was intended (for which I cannot think of a good reason), then the constraint needs to be reset as $\sum_{i=1}^{N} q_i = MN$. Otherwise the problem discussed in the article after its equation (4) would assume $1/N$ times the resource as the one before its equation (4), making the mathematical formulation irrelevant for the reason mentioned above. It would also make the subsequent comparisons meaningless. For example, the assertion (p. 167) that its equation (5), the expected time under its equation (4), is less than its equation (3), the expected time under uniform sampling—which was expressed as N/M, not $1/M$—is true mathematically only when the two M's are the same.

This mathematical oversight is easy to fix, and the correct solution will partially invalidate the assertion that the optimal strategy under (1) would "underplay" prior information (p. 167); see Section 3 below. However, something more profound than this oversight is needed to explain why (2) does not deliver the aforementioned obvious strategy even in the case of $M = 1$. Before I present the reason, however, let me emphasize that no one likes to be pulled out simply because of his or her ethnicity or anything of that sort. Therefore, the overall conclusion of the article that ethnic profiling is unacceptable can be argued meaningfully in many other ways, as the article presented. What I am discussing here is the article's probabilistic and statistical arguments. In the current setting, once the prior probabilities p_i's are given, then whatever criterion we adopt, the resulting optimal strategy will only depend on the *values* of these p_i's, not on their meaning or how they are derived. This potential for more general applicability makes it critical to ensure the correctness of theoretical results and to explicate when they are applicable in scientific publications. Below I will discuss why minimizing (1) is not an appropriate criterion in the context of airport enhanced security checks.

2   Catching a Terrorist or Stopping an Imminent Threat?

The article arrived at (1) by casting the "pulling out" problem in the classic framework of sampling with replacement, which it argued (p. 167) is justifiable because any particular checkpoint is "only one of many airport security checkpoints through which terrorists pass." Whereas this statement is useful in

reminding us to be prepared at each security checkpoint, we must also remind ourselves that the whole purpose of airport security checks is to prevent terrorists from getting on any flight where *they will carry out an act of terrorism.* Given that a terrorist will only have one chance to carry out his act on an airplane, because he will be either dead or detained after acting out, we clearly need to maximize the probability of pulling him out the *first time* he comes to an airport with a (detectable) device that he intends to use. We would have a far more serious problem than worrying about airport checkpoints if any terrorist is allowed to be "sampled with replacement," that is, he carried out his plot on a plane, but somehow escaped and shows up at another airport checkpoint!

To put it crisply, the ultimate goal of an enhanced screening is not to catch a terrorist—if a terrorist comes to an airport without bringing anything or acting in any way that would increase his probability of being identified as such, then placing the task of identifying him through airport checkpoints is not a meaningful goal. The ultimate goal of an enhanced screening is to *minimize imminent threat,* i.e., to stop any individuals who *will carry out* a harmful act once they get on a plane. We therefore do not have the luxury to wait for another chance to catch them. A cartoon posted outside a colleague's office illustrates this point vividly. It shows an instructor at a terrorist training camp who is ready to demonstrate how to detonate a suicide bomb, remarking "Look very carefully, because I can only do this once."

A predecessor to Press (2010), Press (2009), provided me with a plausible explanation of why this distinction was overlooked. There, the "profiling problem" was formulated as a governmental hunting of a malfeasor among N individuals, where the government has a "meaningful prior probability" p_i that the ith individual is its target. Both "authoritarian strategy" and "democratic strategy" were discussed there. The "authoritarian strategy" is that the government enumerates all its citizens, and screens them one by one *without replacement* until the malfeasor is found. In the presence of moral or practical constraints, a government may adopt a "democratic strategy" determined by some public policies, where screening processes are memory-less, that is, if an individual passes them, s/he will still be liable to be sampled again because negative screening results are not recorded. (If the screening result is positive, then of course the game is over, unless the screening is not 100% accurate, a complication also discussed in Press, 2009.) Airport security checkpoints were cited in Press (2009) as such a memory-less screening process. Hence, *in that context,* the purpose of airport checkpoints becomes helping the government to catch a malfeasor, deviating from its original goal of minimizing imminent threat to flight safety.

But even for this "malfeasor hunting" problem, there is a more subtle mathematical reason why criterion (1) is inappropriate. Let us assume now that the sampling with replacement is appropriate, and that the government has resources to conduct L national searches *sequentially,* and each time it can afford pulling out (on average) M individuals to conduct a memory-less screening. Given the probability q_i of pulling out the ith individual during each search (the memory-less property implies that q_i stays the same for all searches), the probability that the ith person is never pulled out is $e_i = (1 - q_i)^L$. Consequently, unless we set $q_i = 1$ for all i's such that $p_i > 0$, there is at least one positive probability

$e_i p_i$, however tiny, such that the malfeasor is the one who escapes every screening. The corresponding *expected time to screening* therefore is not defined (other than treating it as infinity).

Press (2009, 2010) avoided this problem by (implicitly) using $L = \infty$, which led to (1). This, however, does not avoid the *inherent* sensitivity of the expected time to extreme events, a well-known non-robustness of expectation. One can see from (1) that it will go to infinity if any *single* q_i goes to zero unless the corresponding p_i is zero. Therefore, the optimal strategy under (1) will necessarily allocate resource to any individual with nonzero p_i no matter how tiny it might be. The aggregated effect is that such a "most efficient" strategy will necessarily reduce the maximal possible probability of pulling out the actual terrorist at a given checkpoint, and the reduction can be substantial when M/N is not too close to one and the prior probabilities are reasonably informative, as shall be demonstrated in Section 3.

The mathematical difficulty with a finite L itself is telling us that we are not looking at the right criterion. When L is permitted to grow without limit, then the probability of *eventually* pulling out the terrorist is one, and hence minimizing the time to pull him out is meaningful. But with any given finite L, there is always a chance that the terrorist will never be pulled out, and hence we have to focus on minimizing this probability of escape, which, under the setting of Press (2009), is

$$P_L = \sum_{i=1}^{N} (1 - q_i)^L p_i. \tag{3}$$

We can then minimize (3) subject to the constraint $\sum_{i=1}^{N} q_i = M$. For the "malfeasor hunting" problem in Press (2009), which does not cap L, we will see in Section 4 that the optimal strategy for minimizing (3) indeed will converge to uniform sampling $q_i = M/N$ as $L \to \infty$, though this does not mean that the impact of the prior will vanish completely at the limit. In terms of stopping an imminent threat, however, we have to set $L = 1$, and hence the optimality criterion is to minimize P_1, as shown below.

3 Correctly Minimizing, and Minimizing the Correct Criterion

To simplify the notation without loss of generality, we will assume the p_i's are sorted in descending order, namely, $1 > p_1 \geq p_2 \geq \cdots \geq p_N > 0$; recall $\sum_{i=1}^{N} p_i = 1$ under the article's assumption of a single terrorist. Furthermore, to avoid mathematical complications that are too involved to discuss in this article, all the optimal solutions presented below should be understood as representing an equivalence class of solutions, including randomized ones (e.g., if a solution calls for pulling out the first M individuals for certain, but $p_{M+1} = p_M$, then any randomized procedure that shares the pulling probability q_M by the Mth and the $(M+1)$st individuals will obviously also be optimal).

To minimize (1) correctly, we let

$$\tilde{q}_i = \frac{[M - (i-1)]\sqrt{p_i}}{\sum_{k=i}^{N} \sqrt{p_k}}, \quad \text{for } i = 1, \dots, M, \tag{4}$$

and let I be *the largest i* such that $\tilde{q}_i \geq 1$; if no such i exists, we let $I = 0$. (A subtlety here: although $\tilde{q}_i$ is not necessarily decreasing in i, simple algebra shows that $\tilde{q}_j \geq 1$ for all $j \leq I$.) Here I is the number of

individuals our resource can afford pulling out with certainty. When $M = N$, $\tilde{q}_N = 1$ and hence $I = N$, that is, we will pull out everyone, confirming the obvious optimal strategy discussed in Section 1. When $M < N$, $\tilde{q}_M = \sqrt{p_M}/\sum_{k=M}^{N} \sqrt{p_k} < 1$, and hence $I \leq M - 1$, always leaving resources to be distributed to the remaining $N - I$ individuals.

Given I as defined above, the correct optimal q_i for minimizing (1) is given by

$$q_i = \begin{cases} 1, & i = 1, \ldots, I; \\ \frac{(M-I)\sqrt{p_i}}{\sum_{k=I+1}^{N} \sqrt{p_k}}, & i = I+1, \ldots, N. \end{cases} \tag{5}$$

This result can be obtained by adding the inequality constraint $0 \leq q_i \leq 1$ to the derivation that led to (2). An elementary proof is given in the Appendix, using the Cauchy-Schwarz inequality.

This correct solution indicates that as long as $I > 0$, even under (1), the optimal strategy tend to "overplay" prior information (p. 167), because it will pull out for certain the I individuals with the highest prior probabilities. A simple example illustrates the interplay between the informativeness of the prior probabilities and the resource constraint in determining the value of I. Suppose N individuals consist of two groups, N_0 of them with the same high value p_1 and the remaining $N - N_0$ with the same lower value p_N, and $M > N_0$. Let $r = p_N/p_1$, and $\rho = (M - N_0)/(N - N_0)$. Then (4) and (5) imply that when the prior is discriminative enough relative to the resource constraint, in the sense that $\sqrt{r} \leq \rho$, the optimal strategy will pull out everyone in the high-value group for certain, and pull out everyone in the other group uniformly with the sampling rate ρ. Indeed, *in this case*, this strategy also minimizes the escaping probability P_1, demonstrating the influence of strong prior information. However, when $\sqrt{r} > \rho$, $I = 0$ and hence (2) provides the optimal strategy under (1), differing from the one that minimizes P_1.

The optimal solution under $P_1 = \sum_{i=1}^{N}(1 - q_i)p_i$ is of course well known: set $q_i = 1$ for $i = 1, \ldots, M$, and $q_i = 0$ for the rest; we will label this strategy Q_1. The minimal value of P_1 therefore is $P_1^{(Q_1)} = \sum_{i=M+1}^{N} p_i$, which can be considerably smaller than the probability of escape under uniform sampling or under (5). As an illustrative example, suppose $N = 880$ and $M = 220$, and the prior is given by

$$p_i = \begin{cases} \frac{1}{100}, & i = 1, \ldots, 40; \\ \frac{1}{400}, & i = 41, \ldots, 240; \\ \frac{1}{6400}, & i = 241, \ldots, 880. \end{cases} \tag{6}$$

Although the individual prior probabilities might seem rather low with their maximum being 1%, under the optimal strategy Q_1, $P_1^{(Q_1)} = 15\%$, compared with $P_1 = 75\%$ under uniform sampling.

This finding is inconsistent with a main assertion on page 164, that even if one puts aside moral and other practical issues and "even with unrealistically perfect data it is surprisingly difficult to gain any benefit from such profiling." The "surprising difficulty" apparently is induced by the inappropriate criterion (1). Indeed, for the current example, it is easy to verify that (2) yields

$$q_i = \begin{cases} 1, & i = 1, \ldots, 40; \\ \frac{1}{2}, & i = 41, \ldots, 240; \\ \frac{1}{8}, & i = 241, \ldots, 880. \end{cases} \tag{7}$$

These values do provide the correct optimal strategy under (1), though if we set M any larger, we will need to use (5). However, under this strategy, the probability of escape is

$$P_1 = 40 \times 0 \times \frac{1}{100} + 200 \times \frac{1}{2} \times \frac{1}{400} + 640 \times \frac{7}{8} \times \frac{1}{6400} = \frac{27}{80} = 33.75\%. \tag{8}$$

Although this represents 55% reduction from uniform sampling, moving to Q_1 will lead to another (about) 55% reduction in the P_1 value, a reduction that is hard to argue as insubstantial on its own.

4   A Mathematical Justification for Uniform Sampling?

It is true, however, that the reduction in P_L of (3) from either uniform sampling or from (5) to optimal P_L tend to (but not always) become less as L grows. When $L > 1$, we can define

$$\breve{q}_i = 1 - \frac{(i - M)p_i^{-(L-1)^{-1}}}{\sum_{k=1}^{i} p_k^{-(L-1)^{-1}}}, \quad \text{for } i = M, \dots, N. \tag{9}$$

Let K be the *largest* i such that $\breve{q}_i > 0$; note such K must exist because $\breve{q}_M = 1$. (Again, $\breve{q}_i$ is not necessarily decreasing, but $\breve{q}_i > 0$ for all $i \leq K$.) Here K is the number of non-zero q_i's in Q_L, the optimal solution for minimizing P_L of (3), which is given by

$$q_i = \begin{cases} 1 - \frac{(K-M)p_i^{-(L-1)^{-1}}}{\sum_{k=1}^{K} p_k^{-(L-1)^{-1}}}, & i = 1, \dots, K; \\ 0, & i = K+1, \dots, N. \end{cases} \tag{10}$$

The optimality of (10) can be easily established using Hölder's inequality; see the Appendix.

The solution (10) indicates that when the formulation in Press (2009) is *relevant*, the optimal strategy for maximizing the probability of capturing the malfeasor is "complementary harmonic sampling" as long as $L > 1$. Although I doubt its practical value, (10) provides an important theoretical insight into the interplay between prior information and resource constraint, *on purely mathematical grounds,* as Press (2009, 2010) intended. This is because Q_L of (10) turns out to connect the two cases of interests, namely Q_1 and uniform sampling. Perhaps expected, Q_1 is the limit of Q_L when $L \downarrow 1$. The more interesting case is that when $L \to \infty$, Q_L converges to uniform sampling $\{q_i = M/N, \text{ for } 1 \leq i \leq N\}$, which therefore can be denoted as Q_∞. These facts are not hard to see formally by inspecting (10), but a rigorous proof takes a bit of algebra; see the Appendix. The left panel of Figure 1 demonstrates this convergence for the numerical example in Section 2 (note that there are only three distinct q values for that example.)

Using the same example, the right panel of Figure 1 compares the *log* of the probability of escape for the three strategies: dashed line for the optimal strategy (10), solid line for uniform sampling, and dotted line for the "square-root sampling" (5). Intriguingly, we see that the difference between optimal and uniform sampling remains a constant on the log scale as L grows, implying that the two probabilities are never the same but are of the same magnitude. This is not a contradiction to the fact that $Q_L \to Q_\infty$ because their corresponding P_L values converge to zero. In contrast, the probability of escape under (5) converges to zero exponentially slower compared with either optimal or uniform sampling.
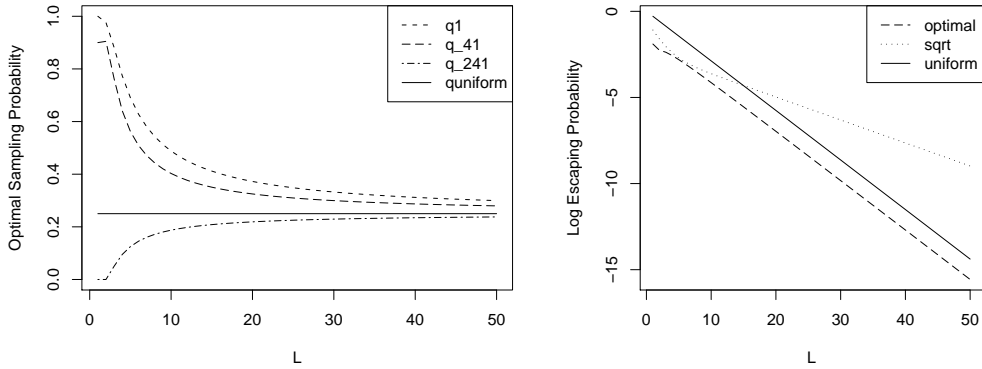
Figure 1: *Left Panel*: Demonstrating optimal sampling (10) converging to uniform sampling. *Right Panel:* Comparing log of the escaping probabilities.

Both phenomena are general. Let $R_L = P_L^{(Q_L)}/P_L^{(Q_\infty)}$, that is, the ratio of the probability of escape under the optimal sampling to that of uniform sampling, and $\Delta_L = -\log R_L$. Then, as shown in the Appendix, as soon as $K = N$, Δ_L will decrease as L increases, indicating that the impact of the prior information decreases as our resource becomes unlimited. However, the impact of the prior does not vanish even at the limit, because the limit of Δ_L, Δ_∞, turns out to be exactly the Kullback–Leibler (KL) divergence between the uniform distribution on $\{1, \ldots, N\}$ and the prior; see the Appendix. Therefore, the more informative the prior distribution, as measured by a larger KL divergence from the uniform distribution, the larger Δ_∞ is. For the numerical example in Section 2, the KL divergence is 1.16, which corresponds to the vertical distance between the solid and dashed lines (as $L \to \infty$). Consequently, the minimal probability of escape is never more than $1/3$ (since $R_\infty = e^{-\Delta_\infty} = e^{-1.16} = 0.312$) of that under uniform sampling for all L, with the minimal $R_L = 0.186$ achieved at $L = 2$. Note that the minimal R_L does not occur at $L = 1$—recall here $R_1 = 1/5$—because R_L increases with L only after K becomes N; in this example, this occurred after $L \geq 3$.

On the other hand, for large L, the probability of escape under the square-root sampling (5) is dominated by the individual(s) with the largest probability of escape, which is $e_N = (1 - q_N)^L$. Since q_N has to be smaller than the uniform sampling rate M/N unless the prior is uniform (because $Nq_N < \sum_i q_i = M$), we see that if we replace the optimal escaping probability in R_L by the escaping probability under (5), R_L will go to infinity at the rate $[(1 - q_N)/(1 - M/N)]^L$. For the numerical example here, $q_N = 1/8$, $M/N = 1/4$, and hence this rate is $(7/6)^L$, explaining the growing distance between the dotted line and the solid (and dashed) line in the right panel of Figure 1.

The same plot shows that the relationships among the (log) escaping probabilities are more complex when L is small. That (5) did nearly as well as (10) when $L = 5$ indicates the possibility for different criteria to lead to very similar strategies in special cases. However, the main message here is that (10),

7

including its two limiting cases, confirms our intuition that the prior information generally has the most impact when our resource is most limited, and the least impact when our resource becomes unlimited. However, unlike in common cases of Bayesian inference where the impact of the prior vanishes as the size of the data goes to infinity, the impact of the prior can persist in a design-based context even with unlimited resource for collecting data, an insight that is less obvious. Therefore, theoretically, the results above indicate that uniform sampling is never optimal for minimizing the probability of escape unless the prior is already uniform or $M = N$. Practically, however, achieving this optimality requires precise knowledge of p_i. Yet remarkably, uniform sampling achieves the *optimal magnitude* without any prior knowledge (excluding of course the trivial case where the prior distribution is a singleton, that is, the government can correctly identify the malfeasor based on prior probability alone). In fact, as soon as L is *large enough* to ensure $K = N$, the ratio R_L will be bounded below by NH_p and above by NG_p, where H_p and G_p are respectively the harmonic mean and geometric mean of the prior probabilities $\{p_1, \ldots, p_N\}$; see the Appendix. Consequently, one could argue that *for the "malfeasor hunting" problem as specified in* Press (2009), this *robust optimality* of uniform sampling provides a *mathematical* reason to justify its use. However, this mathematical reasoning is not applicable when our goal is to stop an imminent threat, precisely because then $L = 1$, rendering the large-L argument above untrustworthy.

5   Minimizing Error-escaping Probability in Statistical Review

Incidentally, the theme of the Larry Brown conference was "Borrowing Strength: Theory Powering Applications". The current problem is a good demonstration of the potential impact of theoretical thinking on real-life applications, and the impact can be both positive and negative. The real-life problem of airport security enhanced checks is of course far more complex than any mathematical formulation in Press (2010) or in this article. For example, if we take its probabilistic modeling seriously, we need to consider specification of a loss function, how terrorists may game the system whatever non-uniform strategy we adopt (an issue related to the "second-order" effect alluded to in Press, 2009), how multiple terrorists may coordinate in their actions, etc. Whereas such issues may be extremely hard to model theoretically, the associated thought process itself could help prevent us from using the mathematical optimality of Q_1 under the simple criterion (3) as the argument for strong profiling, or from using the erroneous (2) or inferior criterion (1) as mathematical reasoning for the opposite. Incorrect mathematical reasoning can only serve as distractions, much like a technical oversight in a court argument is often used to distract the juries from otherwise irrefutable evidence.

These distractions could have been easily avoided if they had been identified during the review process. If there is any profession that should be particularly alerted by a larger-than-one probability or by mixing sampling with replacement with sampling without replacement, it must be ours. Ironically, therefore, the "Surprisingly, and bizarrely" sentiment also describes my feeling towards the fact that

these oversights passed through our editorial checkpoints. Sadly, this is not the first time that a larger-than-one probability was permitted to pass a statistical editorial checkpoint (see Meng, 1995), and it would be another glaring error if I were to predict that this is the last time. Whereas the consequences of false negative errors at editorial checkpoints are of no comparison to those at airport checkpoints (at least in the short run), I couldn't help but wonder if the underlying probabilistic strategies are similar. We all should exercise our due diligence to minimize the error-escaping probability for every paper we review, not merely hope to minimize the expected time to detecting the next improbable probability.

Acknowledgments

References

[1] Meng, X.-L. (1995) Comment on "A Characterization of the Poisson Distribution and the Probability of Winning a Game," by J. B. Keller. *American Statistician*. **63**. 202-210.

[2] Press, W.-H. (2009) Strong profiling is not mathematically optimal for discovering rare malfeasors. *Proceedings of National Acadamey of Sciences USA*, **106**, 1716-1719.

[3] Press, W.-H. (2010) To catch a terrorist: can ethnic profiling work? *Significance*, December 2010, **7**, 164-167.

Appendix: Mathematical Proofs

I. The Optimality of (5). We first write

$$\mu = \sum_{i=1}^{N} \frac{p_i}{q_i} = \sum_{i=1}^{I} \frac{p_i}{q_i} + \sum_{i=I+1}^{N} \frac{p_i}{q_i}. \tag{11}$$

By the Cauchy-Schwarz inequality, we have

$$C_I \equiv \left(\sum_{i=I+1}^{N} \sqrt{p_i} \right)^2 = \left(\sum_{i=I+1}^{N} \sqrt{\frac{p_i}{q_i}} \sqrt{q_i} \right)^2 \leq \left(\sum_{i=I+1}^{N} \frac{p_i}{q_i} \right) \left(\sum_{i=I+1}^{N} q_i \right), \tag{12}$$

where equality holds if and only if $q_i \propto \sqrt{p_i}, i = I+1, \ldots, N$. Therefore, when $I = 0$, (2) follows from (12) immediately because the q_i's defined there make equality hold in (12) and none of them exceeds

$q_1 = \tilde{q}_1 < 1$, which holds by the definition of $\tilde{q}_I$. When $I > 0$, combining (11) and (12) together with $\sum_{i=I+1}^{N} q_i = M - \sum_{i=1}^{I} q_i$, we see that for any $0 < q_i \leq 1, i = 1, \ldots, N$,

$$\mu = \sum_{i=1}^{N} \frac{p_i}{q_i} \geq \sum_{i=1}^{I} \frac{p_i}{q_i} + \frac{C_L}{M - \sum_{i=1}^{I} q_i} \equiv f(q_1, \cdots, q_I). \tag{13}$$

For any $i = 1, \ldots, I$, we have

$$\frac{\partial f}{\partial q_i} = -\frac{p_i}{q_i^2} + \frac{C_I}{(M - \sum_{k=1}^{I} q_k)^2} \leq -p_I + \frac{C_I}{(M-I)^2}, \tag{14}$$

because $p_i \geq p_I$ and $q_i \leq 1$ for all $i = 1, \ldots, I$, and $I < M$ whenever $M < N$ (the $M = N$ case is trivial). But the right-hand side of (14) is non-positive if and only if $\tilde{q}_I \geq 1$, which holds by the definition of I. Hence $f(q_1, \ldots, q_I) \geq f(1, \ldots, 1)$ for all $0 \leq q_i \leq 1$. Consequently, (13) implies

$$\mu = \sum_{i=1}^{N} \frac{p_i}{q_i} \geq \sum_{i=1}^{I} p_i + \frac{C_I}{M - I}, \tag{15}$$

proving (5) because q_i's there satisfy $0 \leq q_i \leq 1$, $\sum_{i=1}^{N} q_i = M$, and they achieve the bound in (15).

II. The Optimality of (10) and Its Limiting Cases.   To prove the optimality of (10) for any $L > 1$, we first write

$$P_L \equiv \sum_{i=1}^{N} (1 - q_i)^L p_i = \sum_{i=1}^{K} (1 - q_i)^L p_i + \sum_{i=K+1}^{N} (1 - q_i)^L p_i. \tag{16}$$

Because $(1 - q_i) = [(1 - q_i)p_i^{1/L}][p_i^{-1/L}]$, by Hölder's inequality, we have

$$\sum_{i=1}^{K} (1 - q_i) \leq \left[\sum_{i=1}^{K} (1 - q_i)^L p_i \right]^{\frac{1}{L}} \left[\sum_{i=1}^{K} p_i^{-(L-1)^{-1}} \right]^{1 - \frac{1}{L}}, \tag{17}$$

where equality holds if and only if $1 - q_i \propto p_i^{-(L-1)^{-1}}, i = 1, \ldots, K$. Therefore, when $K = N$, (10) follows from (17) immediately because the q_i's defined there make equality hold in (17) and none of them goes below $q_N = \check{q}_N > 0$, which holds by the definition of $\check{q}_N$. When $K < N$ (recall $K \geq M$), combining (16)-(17) with $\sum_{i=1}^{K} q_i = M - \sum_{i=K+1}^{N} q_i$, we see that for any $0 \leq q_i \leq 1, i = 1, \ldots, N$,

$$P_L \geq \left[K - M + \sum_{i=K+1}^{N} q_i \right]^L D_L^{1-L} + \sum_{i=K+1}^{N} (1 - q_i)^L p_i \equiv g(q_{K+1}, \ldots, q_N), \tag{18}$$

where $D_L = \sum_{i=1}^{K} p_i^{-(L-1)^{-1}}$. Differentiating g with respect to $q_i, i \geq K + 1$, we obtain

$$\frac{\partial g}{\partial q_i} = L(K - M + \sum_{i=K+1}^{N} q_i)^{L-1} D_L^{1-L} - L(1 - q_i)^{L-1} p_i \geq L \left[(K - M)^{L-1} D_L^{1-L} - p_{K+1} \right] \tag{19}$$

because $q_i \geq 0$ and $p_{K+1} \geq p_i$ for $i \geq K + 1$. But the right-hand side of (19) is non-negative if and only if $\check{q}_{K+1} \leq 0$, which holds by the definition of K. Hence $g(q_{K+1}, \ldots, q_N) \geq g(0, \ldots, 0)$ for all $0 \leq q_i \leq 1$, $i \geq K + 1$. Combining this fact with (18) yields

$$P_L \geq \frac{(K - M)^L}{D_L^{L-1}} + \sum_{i=K+1}^{N} p_i. \tag{20}$$

10

This completes the proof because the lower bound in (20) is achieved by the q_i's of (10), which also satisfy the constraints that they are probabilities and that $\sum_{i=1}^{N} q_i = M$.

To establish the limiting cases of (10), we first note that there is nothing in the proof above that requires L to be an integer, but only that $L > 1$. Let M_1 and M_2 be the smallest and the largest i such that $p_i = p_M$; hence $M_1 \leq M \leq M_2$. Then it is easy to see that as $L \downarrow 1$,

$$\breve{q}_{M_2} = 1 - \frac{M_2 - M}{(M_2 - M_1 + 1) + \sum_{k=1}^{M_1-1} (p_M/p_k)^{(L-1)^{-1}}} \longrightarrow \frac{M - M_1 + 1}{M_2 - M_1 + 1} > 0,$$

and

$$\breve{q}_{M_2+1} = 1 - \frac{M_2 + 1 - M}{1 + \sum_{k=1}^{M_2} (p_{M_2+1}/p_k)^{(L-1)^{-1}}} \to M - M_2 \leq 0.$$

Consequently, as $L \to \infty$, $K \to M_2$ and hence (10) converges to

$$q_i = \begin{cases} 1 & i = 1, \ldots, M_1 - 1; \\ \frac{M - M_1 + 1}{M_2 - M_1 + 1}, & i = M_1, \ldots, M_2; \\ 0, & i = M_2 + 1, \ldots, N. \end{cases}$$

This limit is equivalent to Q_1 when $M_2 > M_1$ because it also reaches the minimal value of P_1, and it is identical to Q_1 when $M_2 = M_1(= M)$, that is, where there is no tie for p_M.

At the other extreme, as $L \to \infty$, $\breve{q}_N \to M/N > 0$ and hence $K = N$ when L is large enough. It is then easy to see that as $L \to \infty$, q_i of (10) will converge to M/N for all i because $p_i^{-(L-1)^{-1}} \to 1$.

III. The Monotonicity and Limit of R_L and of Δ_L

By the definition of $R_L = P_L^{(Q_L)}/P_L^{(Q_\infty)}$, we see from (10) and (3) that as soon as $K = N$,

$$R_L = \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{Np_i} \right)^{\frac{1}{L-1}} \right]^{-(L-1)}. \tag{21}$$

Using the fact that for any positive $\vec{a} = (a_1, \ldots, a_N)$, $h_{\vec{a}}(x) = [\frac{1}{N} \sum_{i=1}^{N} a_i^x]^{1/x}$ is a strictly increasing function of x unless all a_i's are the same, we see that $R_L = [h_{\vec{a}}((L-1)^{-1})]^{-1}$ is a strictly increasing function of L unless $a_i = (Np_i)^{-1}$ are all equal. It follows then, as long as $L \geq 2$, the right-hand side of (21) is bounded below by $[h_{\vec{a}}(1)]^{-1} = NH_p$, where $H_p = [\sum_{k=1}^{N} p_i^{-1}/N]^{-1}$ is the harmonic mean of $\{p_1, \ldots, p_N\}$. To get the upper bound, we let $L \to \infty$ and applying the L'Hospital's rule, which yields

$$R_\infty = \lim_{L \to \infty} R_L = N \left(\prod_{i=1}^{N} p_i \right)^{1/N} = NG_p,$$

where G_p is the geometric mean of $\{p_1, \ldots, p_N\}$. Hence, as long as the prior is not uniform, we have

$$0 < NH_p < R_L < NG_P < 1,$$

$\Delta_L = -\log(R_L)$ is strictly decreasing, and

$$\Delta_\infty = -\log R_\infty = N^{-1} \sum_{i=1}^{N} \log(N^{-1}/p_i),$$

which is the Kullback–Leibler divergency between the uniform distribution $\{N^{-1}, \ldots, N^{-1}\}$ and the prior distribution $\{p_1, \ldots, p_N\}$. [Note all results above assume $L(\geq 2)$ is large enough to ensure $K = N$.]