

# Asymptotic Biometric Analysis for Large Gallery Sizes

Manas Baveja, Hongsong Yuan, and Lawrence M. Wein

**Abstract**—Motivated by the need to predict the future biometric performance of the U.S. Visitor and Immigrant Status Indicator Technology program as it increases the size of its watchlist database, we use extreme-value theory (which is an asymptotic theory for the maximum of a large number of independent and identically distributed random variables) to analyze biometric performance as the gallery size gets very large. Due to the lack of published data for open-set fingerprint systems (where some users of the system are not on the watchlist), we assess the accuracy of our approach using the rank-one identification probability for closed-set fingerprint systems (where all users are on the watchlist). Consistent with earlier empirical observations, we find that the relationship between the rank-one identification probability and gallery size is log-linear to first-order and has a quadratic correction term, at least under our specific distributional assumptions. We also find that the probabilistic biometric model provides a good fit to empirical fingerprint data only when the genuine and impostor similarity scores are allowed to depend on the quality of the fingerprint images, which leads to genuine and impostor scores that are mixtures of distributions. Finally, we use the extreme-value approach to derive the receiver operating characteristic curve for open-set systems.

**Index Terms**—Biometrics, extreme-value theory, fingerprint recognition, probability.

## I. INTRODUCTION

THE U.S. Government may greatly expand the scope of several of its large biometric programs in the coming years. In the area of homeland security, the U.S. Visitor and Immigrant Status Indicator Technology (US-VISIT) Program [1] is considering increasing its gallery size (i.e., the number of biometric samples in its database) by several orders of magnitude (from  $\approx 10^6$  to  $\approx 10^8$ ). In the field of forensics, the idea has been posed to construct a national ballistics database that

Manuscript received April 07, 2010; revised June 18, 2010; accepted June 21, 2010. Date of publication July 15, 2010; date of current version November 17, 2010. This work was supported by the Center for Social Innovation, Graduate School of Business, Stanford University and by a grant from the John D. and Catherine T. MacArthur Foundation (Award 02-69383-000-GSS) in support of a fellowship at the Center for International Security and Cooperation, Stanford University. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Davide Maltoni.

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors. This includes an Online Appendix in pdf format.

M. Baveja is with Comac Capital LLP, London, W1S 2YZ, U.K. (e-mail: MBaveja@comaccapital.com).

H. Yuan is with the Management Science and Engineering Department, Stanford University, Stanford, CA 94305-5015 USA (e-mail: hsyuan@stanford.edu).

L. M. Wein is with the Graduate School of Business, Stanford University, Stanford, CA 94305-5015 USA (e-mail: lwein@stanford.edu).

Digital Object Identifier 10.1109/TIFS.2010.2058105

would use computerized ballistic imaging technology to compare crime scene evidence to images from firings of all newly manufactured and imported firearms [2]. Similar large-scale projects are being undertaken throughout the world, e.g., the iris recognition immigration system at airports in the United Kingdom [3], and the Unique Identification (UID) card project, where each Indian citizen will have a unique identification number with associated biometric data [4].

In each of these examples, it is desirable to estimate the performance of the proposed system in advance of its implementation. The performance analysis of large biometric systems is analytically intractable except in very special cases, mainly because it requires the probability distribution of the maximum of many independent and identically distributed (iid) random variables, which represent similarity scores between a visitor's biometric sample and the biometric sample of each person in the gallery. Consequently, analysts have typically resorted to computer simulation (using an actual—but smaller—biometric database) to assess the performance of these systems. While computer simulations can generate useful results, it would be valuable—as in any field of science or engineering—to have an underlying mathematical theory to shed additional insights into system behavior. Moreover, given the analytical intractability and the very large gallery size, it is natural in this case to seek an asymptotic theory that is based on the gallery size approaching infinity.

In this study, we use extreme-value theory [5], [6], which is an asymptotic theory for analyzing the maximum of many iid random variables, to assess the performance of biometric systems with large gallery sizes. As noted above, the appropriateness of extreme-value theory stems from the fact that the most difficult part of a biometric analysis involves the maximum similarity score between a visitor's biometric sample and the biometric sample of each person in the gallery.

Our main goal in this paper is to use extreme-value theory to derive asymptotic results for the main performance measures for biometric systems and to assess the accuracy of these asymptotic results using existing data. To explain how we attempt to achieve this, we first note that biometric systems can be classified as either an open-set system in which not all visitors are in the biometric database, or a closed-set system in which all visitors are in the database. Although virtually all real-life systems are open-set systems (e.g., a system designed to be closed can be visited by an imposter), the only published fingerprint data that we found that are sufficiently detailed to assess the accuracy of our approach are from closed-set systems. Consequently, Section II is devoted to an analysis of closed-set systems.

During biometric identification in a closed-set system [7], a person's biometrics (e.g., fingerprints) are compared to a database that contains the biometrics of a set of enrolled individ-

TABLE I  
LIST OF NOTATION

Notation	Description	First Location
$n$	gallery size	§II-A
$X$	genuine score, with pdf $f(x)$ , cdf $F(x)$	§II-A
$Y_i$	impostor score, with pdf $g(y)$ , cdf $G(y)$	§II-A
$Z_n$	$\max(Y_1, \dots, Y_n)$ , with pdf $h_n(z)$ , cdf $H_n(z)$	§II-A
$d(n)$	rank-one identification probability	equation (1)
$a_n, b_n$	normalizing constants in extreme-value theory	equation (2)
$r$	degree of Taylor series approximation ( $r = 1$ )	equation (6)
$\mu_i$	$i^{\text{th}}$ moment of standard Type I distribution ( $\mu_0 = 1, \mu_1 = 0.5772$ )	equation (9)
$x^*, k, \lambda$	impostor parameters in §II-D	equation (10)
$\alpha, \beta$	genuine gamma parameters in §II-D	equation (11)
$p(i)$	probability mass function for image quality	§II-E
$\alpha_i, \beta_i$	genuine gamma parameters in §II-E	equation (14)
$c_i, \tau_i$	impostor Weibull parameters in §II-E	equation (15)
$Q_p$	probe's image quality	§II-E
$\Omega_i$	cdf for impostor score given $Q_p = i$	equations (17)-(18)
$Z_{in}$	$\max(Y_1, \dots, Y_n)$ given $Q_p = i$ , with cdf $H_{in}$	§II-E
$d_i(n)$	$d(n)$ given $Q_p = i$	equations (19)-(20)
$\theta_i, \sigma_i$	impostor gamma parameters for quality $i$	equation (22)

TABLE II  
SUMMARY OF ASYMPTOTIC RESULTS FOR CLOSED-SET SYSTEMS. THE ACCURACY AND VALIDITY ARE BASED UPON OUR EXPERIMENTAL RESULTS

Assumptions			Results		
Image Quality	Impostor Scores	Genuine Scores	Location	Accuracy	Validity
aggregated	generic	generic	equation (9)	< 0.005	$n > 100$
aggregated	exponential tail	gamma	equation (12)	< 0.005	$n > 100$
disaggregated	Weibull	gamma	eqn (36) in Online App.	0.03-0.08	$n > 10^4$

uals, in the hope of identifying this person from among the entire set of individuals. This procedure is typically performed by computing a similarity score between the biometric sample of the person to be identified and the biometric sample of each of the enrolled individuals; note that the person who is undergoing identification has offered two different biometric samples, one that is in the database and one taken during the identification process. A key relationship is commonly plotted to characterize the performance of an identification system: on the vertical axis is the rank-one identification probability, which is the probability that the similarity score between the two biometric samples of the person to be identified is larger than the similarity score between his biometric sample given during identification and the biometric sample of each of the other people in the database, and on the horizontal axis is the number of individuals in the database, which is referred to as the gallery size.

Several researchers have observed that this relationship is approximately log-linear with a negative slope, i.e., the rank-one identification probability is linearly decreasing in the logarithm of the gallery size, both for fingerprint matching for gallery sizes up to 600 k ([8, Fig. 3], [9, Figs. 19 and 21]) and for face recognition for gallery sizes up to 40 k ([10, Fig. 4]). However, these are empirical findings and no theory has been put forth for why this log-linear relationship should hold. In Section II, we use extreme-value theory to derive an explicit analytical expression for the rank-one identification probability that predicts this log-linear relationship. An application of this result to existing fingerprint data reveals the importance of explicitly mod-

eling the image quality of the fingerprints. While our approach applies to other biometrics, such as face or iris, we leave the application of our methods to other biometrics as a topic for future research.

In Section III, we use extreme-value theory to derive the receiver operating characteristic (ROC) curve, which is the most common performance measure for open-set systems. The ROC curve plots the true match rate (abbreviated by TMR), which is the probability that a visitor to the system who has a mate in the database will be correctly identified as being in the database, versus the false match rate (abbreviated by FMR), which is the probability that a visitor to the system who is not in the database will be incorrectly identified as being in the database. Concluding remarks appear in Section IV.

## II. ANALYSIS OF CLOSED-SET SYSTEMS

After defining the rank-one probability in Section II-A, we introduce extreme-value theory in Section II-B and use it to approximate the rank-one probability in Section II-C. The results in Section II-C are applied to two fingerprint examples—one with aggregated image quality and one with disaggregated image quality—in Sections II-D and II-E. The notation for all functions and parameters is listed in Table I and the asymptotic results from this section are summarized in Table II.

### A. The Problem Statement

Suppose we have a gallery of  $n + 1$  unique biometric records (e.g., fingerprint images). Now consider a new biometric

sample, called the probe, which has a mate (biometric sample from the person to whom the probe belongs) in the gallery. A biometric system matches the probe against the  $n + 1$  records in the gallery. Assume that the identification rule followed by the biometric system is that the record with which the probe produces the highest similarity score is considered the probe’s mate. The rank-one identification probability, defined as the probability that the genuine similarity score (sometimes called the match score) from the probe-mate match is greater than the  $n$  impostor similarity scores (sometimes called the nonmatch scores) produced, is then a measure of the identification performance of the system.

Let  $X$  denote the random similarity score generated by the probe-mate match and taking values on the positive real axis. Let  $Y_i$  denote the random similarity score, again taking values on the positive real axis, generated when the system matches the probe to the  $i$ th nonmate in the gallery. Let  $Z_n = \max(Y_1, Y_2, \dots, Y_n)$  be the maximum of the  $n$  random impostor similarity scores. Then the rank-one identification probability for a gallery size of  $n$  is defined as  $d(n) = \mathbf{P}(X \geq Z_n)$ .

Let  $f(x)$  denote the probability density function (pdf) of  $X$ ,  $g(y)$  denote the pdf of  $Y_i$  (assuming the  $Y_i$ ’s are iid),  $h_n(z)$  denote the pdf of  $Z_n$  and  $F, G, H_n$  denote the corresponding cumulative distribution functions (cdf). We assume that the random variables  $Y_i$  are independent of  $X$ . Later, we show how this framework accommodates the case in which the  $Y_i$ ’s and  $X$  are conditionally independent given the probe’s image quality. Then the rank-one identification probability is given by

$$d(n) = \int_0^\infty H_n(x)f(x)dx \tag{1}$$

where  $H_n(z) = G^n(z)$  because the  $Y_i$ ’s are iid. In Section II-C, we develop an explicit analytical expression for  $d(n)$  in terms of the gallery size  $n$  and the parameters characterizing the distributions  $F$  and  $G$ .

### B. Extreme Value Theory

Perhaps the most celebrated result in probability and statistics is the central limit theorem, which states that the sum of iid random variables, when properly normalized (i.e., subtract its mean and divide by its standard deviation), converges in distribution to a standard normal random variable. This powerful theorem is used in a variety of applications by approximating the distribution of the normalized sum by the standard normal distribution.

Extreme-value theory is an analogous set of results for the maximum, rather than the sum, of many iid random variables. It has been applied in a wide variety of fields, including weather, equipment reliability, material strength, pollution, and insurance. Referring to our earlier notation, we can use extreme-value theory to approximate the probability distribution of  $Z_n = \max(Y_1, \dots, Y_n)$ , where the  $Y_i$ ’s are iid with cdf  $G$ . Classical extreme-value theory [5], [6] states that if for some sequence of normalizing constants  $a_n > 0$  and  $b_n$ , the random variable  $(Z_n - a_n)/b_n$  has a nondegenerate limiting distribution, then

the limiting distribution must have one of three forms, referred to simply as Type I, Type II, and Type III distributions. If this occurs, then we say that the impostor cdf  $G$  is in the domain of attraction of one of these three distributions. Hence, whereas the normalized sum of iid random variables converges to a normal random variable if it converges, the normalized maximum of iid random variables converges to one of three random variables, depending upon the tail characteristics of the cdf  $G$  of the iid random variables. Loosely speaking, distributions without fat tails, such as the exponential, Weibull, normal, lognormal, and gamma, are in the domain of attraction of Type I, fat-tailed distributions (e.g., Pareto, Cauchy) are in the domain of attraction of Type II, and distributions with upper bounds (e.g., the uniform) are in the domain of attraction of type III; in all cases, the normalizing constants  $a_n$  and  $b_n$  depend upon the cdf  $G$ .

Based on an empirical analysis of impostor fingerprint similarity scores in Fig. 6 of the Online Supporting Information of [11], which suggests that these scores (when not disaggregated by image quality) have an exponential tail and have an upper bound that is not approached for practical gallery sizes (see §1 of the Online Appendix), we focus in this study on the Type I extreme-value distribution. Under the assumption that  $G$  is in the domain of attraction of the Type I distribution, we have that [5], [6]

$$H_n(x) \Rightarrow \exp(-e^{-(x-a_n)/b_n}) \quad \text{as } n \rightarrow \infty \tag{2}$$

where  $\Rightarrow$  denotes weak convergence.

While there exist error bounds for  $|H_n(x) - \exp(-e^{-(x-a_n)/b_n})|$ , they are case-specific and more complicated than the analogous Berry–Esseen bounds for the central limit theorem (although one simple, but possibly loose, bound from [12, eq. (2.4.10)] is  $3/\ln n$ ), and we simply refer readers to [5, Th. 2.10.1] and to [13, Sec. 2.4.2]. In addition, useful guidelines for the minimum  $n$  (as a function of the coefficient of variation of  $G$ ) required for an effective extreme-value approximation appear in [14].

### C. Generic Analysis

Our generic analysis of (1) involves an application of extreme-value theory to approximate  $H_n(x)$  by the right side of (2), followed by a Taylor series expansion to compute the integral in (1).

The integrand in (1) involves the product of a cdf and a pdf, which have very different scales. Consequently, it is more convenient to treat  $n$  as a continuous random variable (recall that  $n$  is very large), and differentiate (1) with respect to  $n$ , which yields

$$d'(n) = \int_0^\infty \frac{\partial H_n(x)}{\partial n} f(x) dx. \tag{3}$$

Now we use the limit theorem in (2) to approximate the derivative of  $H_n(x)$  with respect to  $n$  by the derivative of the right side of (2), i.e.,

$$\frac{\partial H_n(x)}{\partial n} \approx -h_n(x) \left[ a'_n + \left( \frac{x - a_n}{b_n} \right) b'_n \right] \tag{4}$$

where  $a'_n$  and  $b'_n$  are derivatives of the normalizing constants  $a_n$  and  $b_n$  with respect to  $n$ , and  $h_n(x)$  is the pdf of the Type I distribution and is given by

$$h_n(x) = \frac{1}{b_n} \exp(-e^{-(x-a_n)/b_n}) e^{-(x-a_n)/b_n}. \quad (5)$$

Motivated by the key observation (see Fig. 1 of the Online Appendix, which uses data from [9, Sec. 5.13]) that  $f(x)$  is much more diffuse than  $g(x)$  and  $h_n(x)$ , and indeed is nearly linear in the domain where  $h_n(x)$  has most of its mass for large values of  $n$ , we approximate  $f(x)$  by a low-degree Taylor series that is expanded about the mode of  $h_n(x)$ , which is  $a_n$  [15]. That is, we assume

$$f(x) \approx \sum_{i=0}^r \frac{f^{(i)}(a_n)(x-a_n)^i}{i!} \quad (6)$$

where  $f^{(i)}$  denotes the  $i$ th derivative of  $f$  and  $r$  is a small integer ( $r = 1$  suffices in our numerical investigations).

Taken together, we approximate the rank-one identification probability by equations (7)–(9), shown at the bottom of the page, where  $\mu_i$  is the  $i$ th moment of the standard (i.e.,  $a_n = 0$ ,  $b_n = 1$ ) Type I distribution (i.e.,  $\mu_0 = 1$  and  $\mu_1 = \gamma \approx 0.5772$  is the Euler constant [16]).

Simplifying our expression for the rank-one identification probability beyond (9) requires us to assume that the genuine and impostor similarity scores follow parametric probability distributions; that is, we cannot take a nonparametric approach based solely on the empirical data to determine the normalizing constants  $a_m$  and  $b_m$  in (9). The choice of distributions and parameter values will depend upon the type of biometric, the specific matching software, and the population-wide characteristics of the individuals in the gallery. To make further progress, we look at two fingerprint examples in Sections II-D and II-E, one where the fingerprints are disaggregated by image quality, and one where they are not.

#### D. Fingerprint Example Aggregated by Image Quality

The National Institute of Standards and Technology (NIST) study in [9, Sec. 5.13] allows us to assess (9) because this study has data on  $>4$  M impostor similarity scores and  $>40$  k genuine similarity scores (using the Bozorth98 matcher on the DHS2 fingerprint database), and it reports the rank-one identification probability versus gallery size (i.e.,  $d(n)$  versus  $n$ ) in its Fig. 19. As described in §1 of the Online Appendix, we use these data to model the impostor similarity score cdf by

$$G(x) = 1 - ke^{-\lambda x} \quad \forall x \geq x^* \quad (10)$$

with  $x^* = 6$ ,  $k = 3.7042$  and  $\lambda = 0.3394$ . Fig. 4 in the Online Supporting Information of [11] shows that the gamma, Weibull, and log-normal each provide a reasonable fit to the genuine similarity scores from [9, Sec. 5.13], and we use the gamma pdf

$$f(x) = \frac{e^{-x/\beta} x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \quad (11)$$

with maximum likelihood estimates  $\alpha = 1.3671$  and  $\beta = 34.2575$ , where  $\Gamma(x)$  is the gamma function [17]. Because the coefficient of variation of this distribution is  $\alpha^{-1/2} = 0.855$ , the results in [14] suggest that the extreme-value approximation should be highly accurate for practical values of  $n$ . Fig. 1 of the Online Appendix shows  $g(x)$  and  $f(x)$  from (10) and (11).

In §3 of the Online Appendix, we show that under the distributional assumptions in (10) and (11) and setting  $r = 1$ , (9) simplifies to

$$d(n) \approx \frac{(\lambda\beta - \gamma)(\ln(kn))^{\alpha-1}}{(\lambda\beta)^\alpha \Gamma(\alpha)(kn)^{1/\lambda\beta}} + \frac{(\alpha-1)\Gamma\left(\alpha-1, \frac{\ln(kn)}{\lambda\beta}\right)}{\Gamma(\alpha)} \quad (12)$$

where  $\Gamma(\alpha, x)$  is the upper incomplete gamma function [17]. Equation (12) is an extremely accurate approximation of (1) for this set of distributional assumptions: the absolute difference between the  $d(n)$  values in (1) and (12) is  $< 2 \times 10^{-3} \forall n \in [10^2, 10^9]$ .

In §4 of the Online Appendix and Fig. 3 of the Online Appendix, we show that under the distributional assumptions in

$$d(n) = - \int_n^\infty d'(m) dm \quad \text{because} \\ \lim_{n \rightarrow \infty} d(n) = 0 \quad \text{due to the unbounded impostor scores} \quad (7)$$

$$\approx \int_n^\infty \left( \int_0^\infty \left[ a'_m + \left( \frac{x-a_m}{b_m} \right) b'_m \right] \right. \\ \times \frac{1}{b_m} \exp(-e^{-(x-a_m)/b_m}) e^{-(x-a_m)/b_m} \\ \times \sum_{i=0}^r \frac{f^{(i)}(a_m)(x-a_m)^i}{i!} dx \Big) dm \quad \text{by equations (3)–(6)} \quad (8)$$

$$= \int_n^\infty \left( a'_m \sum_{i=0}^r \frac{f^{(i)}(a_m) b_m^i \mu_i}{i!} + b'_m \sum_{i=0}^r \frac{f^{(i)}(a_m) b_m^i \mu_{i+1}}{i!} \right) dm \quad \text{by §2 in the Online Appendix} \quad (9)$$

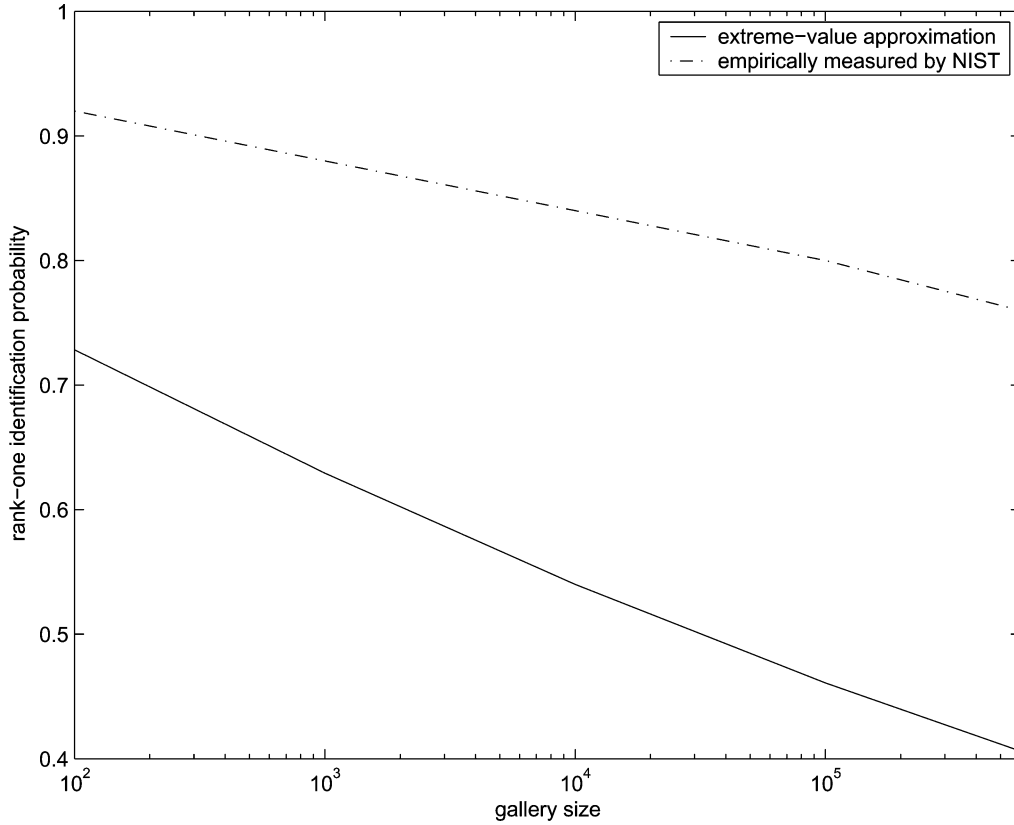


Fig. 1. Comparison of rank-one identification probability versus gallery size computed from (13) and as measured empirically in [9, Fig. 19].

(10) and (11),  $nd'(n)$  grows nearly linear (and very slowly) in  $\ln(n)$ , and that

$$d(n) \approx 0.9532 - 5.25 \times 10^{-2} \ln(n) + 8.4 \times 10^{-4} (\ln(n))^2. \quad (13)$$

That is, to first order,  $d(n)$  is indeed log-linear in  $n$ , and has a quadratic correction term that makes  $d(n)$  slightly convex on a log-linear scale, particularly when  $n > 10^6$  (Fig. 2 in the Online Appendix).

However,  $d(n)$  from (12) underestimates the empirically measured  $d(n)$  from [9, Fig. 19] by 0.18–0.35 for  $n \in [10^2, 10^6]$  (Fig. 1). This underprediction has been previously observed [10] and appears to be due to the independence assumption in our modeling framework. More specifically, because fingerprint images vary in image quality, a probe’s genuine similarity score and his  $n$  impostor similarity scores are likely to be positively correlated. We have previously modeled similarity scores that are dependent on image quality [11] and we show in Section II-E how to apply (9) in this modeling framework.

*E. Fingerprint Example With Image Quality*

As in [11], we assume each person’s fingerprints (whether the person is a visitor or on the watchlist) have an associated image quality  $Q$ , which is a random variable taking on values in  $\{1, 2, \dots, 8\}$ , where the best quality is 1 and the worst quality is 8. Based on data in [18], we estimated the probability that a person has image quality  $i$ , which is denoted by  $p(i)$  (Table II in the Online Supporting Information of [11]). We assume that the genuine similarity score depends on image

quality, and following [11], the genuine pdf for image quality  $i$  is gamma with

$$f_i(x) = \frac{e^{(-x/\beta_i)} x^{\alpha_i-1}}{\beta_i^{\alpha_i} \Gamma(\alpha_i)}. \quad (14)$$

The eight  $(\alpha_i, \beta_i)$  pairs were estimated in [11] from the eight ROC curves in [18, Fig. 12], which were generated using the Cogent matcher and the DOS database. These parameter values and the resulting pdf plots appear in Table IV and Fig. 5, respectively, of the Online Material of [11].

As in [11], we assume that an impostor similarity score depends only on the worse of the two fingerprint image qualities under comparison. Hence, we need to define only eight impostor cdfs  $G_i(x)$ . Although we modeled these distributions by lognormals in [11], the lognormal’s normalizing constants  $a_n$  and  $b_n$  from extreme-value theory are difficult to handle analytically. Instead, we use Weibull distributions, which belong to the domain of attraction of the Type I distribution but have normalizing constants that are relatively tractable. These cdfs are given by

$$G_i(x) = 1 - e^{(-c_i x^{\tau_i})}. \quad (15)$$

We estimate the parameters  $(c_i, \tau_i)$  using the exact same approach as in the Interperson Similarity Score section of the Online Material of [11] (except replacing lognormals with Weibulls), which uses the eight TMR versus FMR curves in [18, Fig. 12], and the parameter values appear in Table I of the Online Appendix, where TMR is the probability that the genuine similarity score is greater than a specific threshold,

and FMR is the probability that the maximum (over all people in the gallery) impostor similarity score is above the threshold. The eight pdf's corresponding to (15) appear in Fig. 4 in the Online Appendix.

With these parameter values in hand, we now derive the exact rank-one identification probability in our image quality framework. The key modeling assumption is that a probe's genuine similarity score and  $n$  impostor similarity scores are conditionally independent given the probe's image quality. If  $d_i(n)$  is the rank-one identification probability given that the probe image quality, which is denoted by  $Q_p$ , is  $i$ , then

$$d(n) = \sum_{i=1}^8 d_i(n)p(i). \quad (16)$$

To develop an expression for  $d_i(n)$ , we first derive the conditional cdf  $\Omega_i(x)$  for the generic  $k$ th impostor score  $Y_k$  given  $Q_p = i$ , which (recalling that the worse of the two image qualities determines the impostor similarity score distribution) is

$$\Omega_i(x) = \sum_{j=1}^8 \mathbf{P}(Y_k \leq x | Q_p = i, Q_k = j)p(j) \quad (17)$$

$$= G_i(x) \sum_{j=1}^i p(j) + \sum_{j=i+1}^8 G_j(x)p(j). \quad (18)$$

If we define  $Z_{in} = \max(Y_1, Y_2, \dots, Y_n)$  given  $Q_p = i$ , then the cdf of  $Z_{in}$ , which we denote by  $H_{in}(z) = P(Z_{in} \leq z | Q_p = i)$ , is  $\Omega_i^n(z)$  due to the conditional independence of the  $Y_k$ 's given the probe image quality. By analogy to (1), we have

$$d_i(n) = \int_0^\infty H_{in}(x) f_i(x) dx \quad (19)$$

$$= \int_0^\infty \Omega_i^n(x) f_i(x) dx \quad (20)$$

and combining (16) and (20) yields

$$d(n) = \sum_{i=1}^8 \left( \int_0^\infty \Omega_i^n(x) f_i(x) dx \right) p(i). \quad (21)$$

In §5 of the Online Appendix, we perform the generic analysis in (9), consisting of extreme-value theory and a Taylor series expansion, to approximate  $d(n)$  in (21). More specifically, using results from [19], we undertake an extreme-value analysis of  $\Omega_i(x)$ , which is a mixture of Weibulls, and obtain the normalizing constants  $a_{in}$  and  $b_{in}$  for the limit theorem  $H_{in}(x) \Rightarrow \exp(-e^{-(x-a_{in})/b_{in}})$  as  $n \rightarrow \infty$ . This extreme-value analysis reveals that the constants  $a_{in}$  and  $b_{in}$  are largely determined by the Weibull distribution in the mixture that has the dominant right tail (the right tail of a distribution  $\Phi_1(x)$  dominates the tail of  $\Phi_2(x)$  if  $\lim_{x \rightarrow \infty} 1 - \Phi_1(x)/1 - \Phi_2(x) \geq 0$  [19]), which in our example turns out to be  $G_5(x)$ , followed by  $G_8(x)$ . However, the results are more complicated than in the aggregated image quality case: rather than being able to carry out the integration in (9) and obtain the closed-form result in (12), we can only derive complicated expressions for  $d_i'(n)$  (equation (36) of the Online Appendix) and need to perform the integration in (9) numerically.

In comparing the theoretical  $d_i(n)$  in (20) to the approximate value [i.e., numerically integrating (9)], we find that the approximation is no longer extremely accurate for practical values of  $n$  because the convergence is very slow (indeed, the modes  $a_{in}$  are negative when  $n \leq 1300$  for  $i = 1, \dots, 4$ , and so the approximation is only shown for  $n \geq 10^4$ ): for  $n \in [10^4, 10^9]$ , the absolute error is  $< 0.01$  for image quality 8 (i.e.,  $d_8(n)$  because  $\Omega_8(x) = G_8(x)$  and is not a mixture distribution), but the absolute error is 0.03–0.07 for  $d_1(n), \dots, d_4(n)$ , 0.01–0.05 for  $d_5(n)$ , and 0.12–0.18 for  $d_6(n)$  and  $d_7(n)$  (Fig. 5 in the Online Appendix). Overall, the approximation overestimates the theoretical  $d(n)$  in (21) by 0.03–0.08 for  $n \in [10^4, 10^9]$  (Fig. 2). Similarly, whereas  $nd'(n)$  increases linearly in  $\ln(n)$  in the image-aggregated example,  $nd_i'(n)$  decreases for  $n \in [10^4, 10^9]$  for  $i = 1, \dots, 7$  (displaying a lack of convergence in this range), and  $nd_8'(n)$  starts increasing only at  $n \approx 10^6$  [20, Fig. 4]). The reason for the slow convergence appears to be that the two image qualities that have the dominant right tails comprise only 9.5% of the people on the watchlist. More specifically, the dominant tails ( $G_5$  or  $G_8$ ) play a large role ( $\geq 90\%$ ) in the mixture distributions ( $\Omega_5$  and  $\Omega_8$ ) that display fast convergence, but a small role ( $\leq 5\%$ ) in the other six  $\Omega_i$ 's, which display slow convergence.

Having compared the theoretically exact  $d(n)$  and the approximate  $d(n)$ , we now turn to a comparison of the theoretically exact  $d(n)$  in (21) and the empirical  $d(n)$ . As noted earlier, the rank-one identification performance depends on the database of fingerprint images and the fingerprint matcher. Because NIST did not report the rank-one identification probabilities corresponding to the data presented in the previous paragraph (i.e., for the DOS database and the Cogent matcher), we need to use an entirely different data set [21] for the theoretical versus empirical comparison. For each of 18 fingerprint matching systems that participated in the medium-scale test described in [21], the rank-one identification probability  $d(n)$  was calculated for  $n = 3240$  [21, Table 20]). Here we compare these 18 empirical values to the corresponding theoretical values. The study in [21] uses four fingerprint image qualities (where  $i = 1, \dots, 4$  correspond to qualities A, B, C, and D/F in [21]) with probability mass function  $p(1) = 0.703$ ,  $p(2) = 0.179$ ,  $p(3) = 0.105$ ,  $p(4) = 0.014$  ([21, Table 19]).

We assume the genuine and impostor cdfs for quality  $i$  are both gamma, with genuine cdf  $F_i(x)$  and pdf  $f_i(x)$  given in (14), and the impostor cdf  $G_i(x)$  and pdf

$$g_i(x) = \frac{e^{(-x/\theta_i)} x^{\sigma_i-1}}{\theta_i^{\sigma_i} \Gamma(\sigma_i)}. \quad (22)$$

In §6 of the Online Appendix, we describe the procedure for estimating the 16 parameters ( $\alpha_i, \beta_i, \sigma_i, \theta_i$ ) from the four TMR versus FMR curves (one curve for each image quality, where the FMR is the probability that the impostor similarity score exceeds the threshold because these curves consider one-to-one matching; i.e., the gallery size is 1) in [21, Figs. D20–D23]. The resulting parameters for each of the 18 matching systems are given in Tables II and III of the Online Appendix.

A comparison of the theoretically exact  $d(3240)$  from (21) with the empirical values from [21, Table 20] appears in Table III, and shows that (21) correctly captures the relative

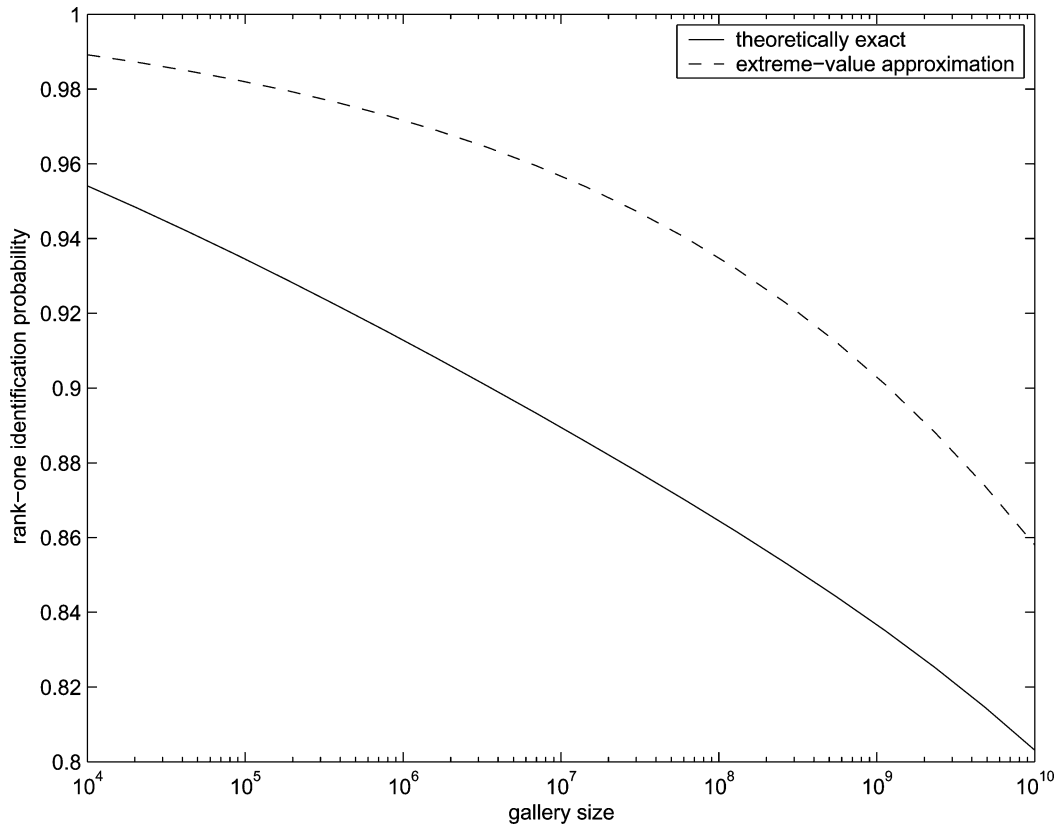


Fig. 2. Rank-one identification probability versus gallery size, when incorporating image quality. The theoretical curve is from (21) and the approximate curve is derived by numerically integrating (9). Parameters are based on TMR versus FMR curves generated by the Cogent matcher and the DOS database.

TABLE III  
FOR EACH OF 18 FINGERPRINT MATCHING SYSTEMS, A COMPARISON OF THE THEORETICAL [ACCORDING TO (21)] AND EMPIRICAL (TAKEN FROM OF [21]) RANK-ONE IDENTIFICATION PROBABILITIES FOR A GALLERY SIZE OF  $n = 3240$

Matching System	Rank-One Identification Probability		
	Empirical	Theoretical	Empirical - Theoretical
NEC	0.994	0.995	0.001
Cogent	0.989	0.988	0.001
SAGEM M2	0.984	0.985	0.001
SAGEM M1	0.978	0.978	0.000
Neurotech M1	0.951	0.940	0.011
Motorola	0.951	0.940	0.011
Identix	0.916	0.905	0.011
NIST VTB	0.900	0.860	0.040
UltraScan M1	0.894	0.869	0.025
UltraScan M2	0.894	0.870	0.024
Biolink	0.858	0.836	0.022
Antheus	0.806	0.784	0.022
Phoenix	0.781	0.741	0.040
Technomagia	0.754	0.645	0.109
123 ID M2	0.733	0.699	0.034
Golden Finger	0.721	0.689	0.032
Raytheon	0.721	0.702	0.019
Avalon	0.610	0.542	0.068

ranking of nearly all of the matching systems. The accuracy of (21) increases with the rank-one identification probability, and is very accurate for probabilities  $>0.95$ .

To further investigate the accuracy of (21), we construct cumulative match curves (CMCs), which are sometimes used to measure biometric performance of closed systems (e.g.,

[7, Sec. 6.4.3]). A CMC computes the probability that the probe-mate similarity score  $X$  ranks in the top  $k$  among the scores  $(X, Y_1, \dots, Y_n)$ , and plots this probability versus  $k$  (with the gallery size  $n$  fixed); note that this probability reduces to the rank-one identification probability when  $k = 1$ . To perform the CMC calculations, we generalize (19) and (21)

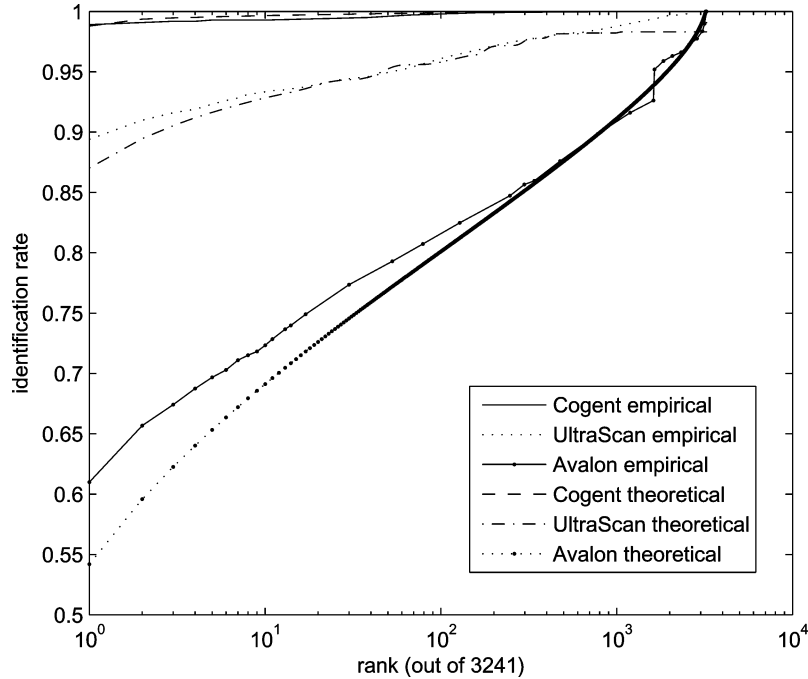


Fig. 3. For three of the 18 matching systems in Table III, a comparison of the empirical (taken from [21, Fig. 22]) and theoretical cumulative match curve (CMC), which plots the probability of identifying a probe-match among the top  $k$  versus  $k$ , for a gallery size of  $n = 3240$ .

by using [22, Proposition 4.1.2], which gives the distribution function of the  $k$ th largest out of  $n$  iid random variables. We compute this curve for three of the 18 matching systems listed in Table III: Cogent (a high-performing system), Avalon (a low-performing system and one that we did not accurately assess in Table III), and UltraScan M2 (a medium-performing system). A comparison (Fig. 3) of the theoretical CMCs to the empirical CMCs (taken from [21, Fig. 22]) reveals that the accuracy of the theoretical CMCs improves with rank (except for the extreme right end of UltraScan M2's CMC), and is very accurate for rank greater than 1, 10, and 100, respectively, for the Cogent, UltraScan M2, and Avalon systems.

### III. ANALYSIS OF OPEN-SET SYSTEMS

Now that we have assessed our approach for closed-set systems, we turn our attention to the more interesting setting of an open-set system. Because we have no data to directly assess the accuracy of our approach for open-set systems, we restrict ourselves in this section to developing extreme-value asymptotics for key performance measures in open-set systems. The central relationship in the performance analysis of open-set biometric systems is the ROC curve, which plots the TMR versus FMR. In Section III-A and Section III-B, we derive the asymptotic ROC curve in the absence and presence of image quality. As noted

earlier, a lack of empirical data for open-set systems prevents us from comparing the asymptotic ROC curve to the empirical ROC curve.

#### A. Aggregated Image Quality

The ROC curve is generated by varying the threshold  $t$ . The TMR is the probability that the genuine similarity score is greater than the threshold, and the FMR is the probability that the maximum impostor similarity score is greater than the threshold. In terms of our previous notation,  $\text{TMR} = P(X > t) = 1 - F(t)$  and  $\text{FMR} = P(Z_n > t) = 1 - H_n(t)$ , where  $Z_n = \max(Y_1, \dots, Y_n)$  has cdf  $H_n(z) = G^n(z)$ . By extreme-value theory, we know that  $H_n(z) \Rightarrow \exp(-e^{-(z-a_n)/b_n})$  as  $n \rightarrow \infty$  for constants  $a_n$  and  $b_n$  that depend upon the distributional assumptions about the impostor similarity scores.

Hence, for a large gallery size

$$\text{FMR} \approx 1 - \exp(-e^{-(t-a_n)/b_n}). \quad (23)$$

Solving (23) for the threshold  $t$  and substituting the resulting expression into  $\text{TMR} = 1 - F(t)$  yields equations (24) and (25), shown at the bottom of the page. Equations (24) and (25) depict the TMR versus FMR curve for any large gallery size  $n$ . These equations depend on the impostor distribution only via its

$$\text{TMR} \approx 1 - F\left(a_n - b_n \ln\left(\ln\left(\frac{1}{1 - \text{FMR}}\right)\right)\right) \quad (24)$$

$$\approx 1 - F(a_n - b_n \ln(\text{FMR})) \quad \text{for small FMR (e.g., FMR} < 0.05) \quad (25)$$

tail and can be used with a parametric or nonparametric genuine distribution.

**B. Disaggregated Image Quality**

As in Section III-B, the probability that a person has image quality  $i$  is  $p(i)$ , the genuine similarity score for image quality  $i$  has cdf  $F_i(x)$ , and the impostor cdf for match quality (i.e., the worse of the two image qualities)  $i$  has the Weibull cdf  $G_i(x)$  in (15).

The TMR is the probability that the genuine similarity score exceeds the threshold  $t$ , which is

$$\text{TMR} = 1 - \sum_{i=1}^8 p(i)F_i(t). \tag{26}$$

As in Section II-E, we have that  $H_{in}(t) \Rightarrow \exp(-e^{-(t-a_{in})/b_{in}})$  as  $n \rightarrow \infty$ , where  $H_{in}(t)$  is the cdf of  $\max(Y_1, \dots, Y_n)$  given the probe quality  $Q_p = i$ , and  $a_{in}$  and  $b_{in}$  are given in §5 of the Online Appendix. Hence, FMR is the probability that the maximum impostor similarity score is greater than the threshold, and is given by

$$\text{FMR} \approx 1 - \sum_{i=1}^8 p(i) \exp(-e^{-(t-a_{in})/b_{in}}). \tag{27}$$

In contrast to the analysis in Section III-A, we cannot analytically solve for  $t$  in (27). However, (26) and (27) can be easily computed for all values of  $t$  to derive the ROC curve for any genuine cdfs  $F_i(x)$ .

**IV. DISCUSSION**

Probabilistic analysis can be useful for predicting a biometric system’s performance as it scales up in size (i.e., greatly increases its gallery size), as is happening with the US-VISIT Program [1]. Moreover, extreme-value theory [5], [6], which provides asymptotics for the largest of many iid random variables, is the natural tool to use in this situation. Surprisingly, this asymptotic theory has been rarely used in biometrics (the only explicit application to biometrics that we are aware of is in [23, Sec. 9.2.5, A.1, and A.2]).

In our basic model that does not differentiate by image quality, we find that the asymptotics are extremely accurate at estimating the theoretical rank-one identification probability and are quite explicit [e.g., (12) and (13)]. These asymptotics provide the first rigorous derivation of the empirically observed log-linear relationship between the rank-one identification probability and the gallery size [8]–[10]. However, as foreshadowed by earlier work [10], the underlying probabilistic model does not generate reliable estimates for the empirical rank-one identification probability for fingerprint matching.

Consequently, in (21), we derive the rank-one identification probability for a model that incorporates image quality [11]. The extreme-value theory asymptotics are more cumbersome in this case (e.g., equation (36) in the Online Appendix), and converge much more slowly to the theoretically exact values in (21). Indeed, using NIST data for the Cogent matcher and the DOS database [18], we find that the asymptotics are not useful when the gallery size  $n < 10^4$  and the rank-one identification probability is overestimated by 0.03–0.08 for  $n \in [10^4, 10^9]$  (Fig. 2).

Convergence is slow because the worst two image qualities (as measured by the dominance of the right tails of their impostor distributions in the extreme-value analysis) comprise <10% of the gallery population.

Nonetheless, the theoretically exact rank-one identification probability in (21) provides a good fit to the empirical rank-one identification probabilities for 18 different fingerprint matching systems (Table III), providing a validation of the key conditional independence assumption in our image quality model (i.e., a probe’s genuine similarity score and  $n$  impostor similarity scores are conditionally independent given the probe’s image quality) and suggesting that two-parameter distributions (e.g., gamma, lognormal, Weibull) provide sufficient accuracy, at least for  $n \approx 10^3 - 10^4$ .

Image quality is an active area of biometric research [24] but it is not yet clear whether differentiating by image quality plays as large a role for other biometrics (e.g., face, iris) as it does for fingerprints. Indeed, attempting to fit our basic model without image quality to data for other biometrics may provide an indication of whether or not differentiating by image quality is an important factor. On a related note, multimodal fusion schemes can be incorporated into our framework as long as the probability distribution of the fused impostor similarity score can be derived. For example, a weighted-average fusion of face and fingerprint would give a fused impostor similarity score distribution that is simply a mixture of the face score distribution and the fingerprint score distribution.

Finally, if impostor similarity scores for other biometrics were to display fat tails or upper bounds, then the same approach could be used (for the rank-one identification probability in closed-set systems or for the ROC curve in open-set systems), but with the limit theorems for type II and type III distributions, respectively, in place of the limit theorems for type I distributions [5], [6]. In addition, some biometric performance measures depend on the  $k$ th largest impostor similarity score (e.g., CMCs), and extreme-value asymptotics exist for the  $k$ th largest out of  $n$  as  $n \rightarrow \infty$  ([5, Sec. 2.8]), which is appropriate for  $k \ll n$ , and also for the  $k$ th largest out of  $n$  as  $k, n \rightarrow \infty$  and  $k/n \rightarrow c$  for some  $c \in (0, 1)$  ([26, Th. 5.8]), which is appropriate for large values of  $k$ .

We conclude with several caveats. A nonparameteric approach might be more accurate than our approach. However, our ROC results in (24) and (25) are semiparametric and well-suited to the fingerprint data in [25, Figs. 1–4], where the impostor pdf is fairly smooth and the genuine pdf appears to be more difficult to estimate with a parameteric distribution. Finally, regardless of the approach taken, estimating the tail of a distribution from limited data (e.g., estimating the largest of  $10^8$  impostor similarity scores based on data from  $10^4$  similarity scores) is a challenging statistical problem and may give inaccurate results.

**REFERENCES**

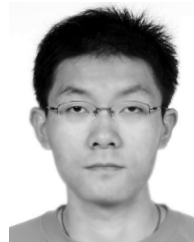
[1] First Phase of Visitor and Immigration Status Program Operating, but Improvements Needed U.S. General Accounting Office, Washington, DC, Report GAO-04-586, 2004.  
 [2] *Ballistic Imaging*, D. L. Cork, J. E. Rolph, E. S. Meieran, and C. V. Petrie, Eds.. Washington, DC: National Academies Press, 2008.

- [3] Iris Recognition Immigration System (IRIS) U.K. Border Agency, Mar. 29, 2010 [Online]. Available: <http://www.ukba.homeoffice.gov.uk/managingborders/technology/iris/>
- [4] Unique Identification Authority of India (UIDAI) Jun. 16, 2010 [Online]. Available: <http://uidai.gov.in>
- [5] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*. Malabar, FL: Krieger, 1987.
- [6] E. J. Gumbel, *Statistics of Extremes*. New York: Columbia Univ. Press, 1958.
- [7] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, *Guide to Biometrics*. New York: Springer-Verlag, 2004.
- [8] Report to the Congress: Use of Technology Standards and Interoperable Databases With Machine-Readable, Tamper-Resistant Travel Documents The Attorney General, Secretary of State, and the National Institute of Standards and Technology, Gaithersburg, MD, 2003.
- [9] C. L. Wilson, C. I. Watson, M. D. Garris, and A. Hicklin, "Studies of fingerprint matching using the NIST verification test bed (VTB)," in *National Institute of Standards and Technology Internal Report 7020*, Gaithersburg, MD, 2003.
- [10] P. P. Grother and P. J. Phillips, "Models of large population recognition performance," in *Proc. 2004 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR'04)*, 2004, pp. 68–77.
- [11] L. M. Wein and M. Baveja, "Using fingerprint image quality to improve the identification performance of the U.S. Visitor and Immigrant Status Indicator Technology Program," *PNAS*, vol. 102, pp. 7772–7775, 2005.
- [12] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*. Berlin: Springer-Verlag, 1983.
- [13] S. I. Resnick, *Extreme Values, Regular Variation, and Point Processes*. New York: Springer-Verlag, 1987.
- [14] C. S. Crow, IV, D. Goldberg, and W. Whitt, "Two-moment approximations for maxima," *Operations Res.*, vol. 55, pp. 532–548, 2007.
- [15] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*. New York: Wiley, 1995, vol. 1–2.
- [16] E. W. Weisstein, Extreme Value Distribution From MathWorld-A Wolfram Web Resource [Online]. Available: <http://mathworld.wolfram.com/ExtremeValueDistribution.html>
- [17] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.
- [18] C. L. Wilson, M. D. Garris, and C. I. Watson, "Matching performance for the US-VISIT IDENT system using flat fingerprints," in *National Institute of Standards and Technology Internal Report 7110*, Gaithersburg, MD, 2004.
- [19] S. Kang and R. F. Serfozo, "Extreme values of phase-type and mixed random variables with parallel-processing examples," *J. Appl. Prob.*, vol. 36, pp. 194–210, 1999.
- [20] M. Baveja, "An Analysis of the US-VISIT Biometric Program," Ph.D., Institute of Computational and Mathematical Engineering, Stanford University, Stanford, CA, 2007.
- [21] C. L. Wilson, R. A. Hicklin, H. Korves, B. Ulery, M. Zoepfl, M. Bone, P. Grother, R. Michaels, S. Otto, and C. Watson, "Fingerprint vendor technology evaluation 2003: Summary of results and analysis report," in *National Institute of Standards and Technology Internal Report 7123*, Gaithersburg, MD, 2004.
- [22] P. Embrechts, C. Kluppelberg, and T. Mikosch, *Modelling Extremal Events*. Berlin: Springer, 1997.
- [23] H. Jarosz and J.-C. Fondeur, "Large-scale identification system design," in *Biometric Systems: Technology, Design and Performance Evaluation*, J. Wayman, A. Jain, D. Maltoni, and D. Maio, Eds. London: Springer, 2005, ch. 9.
- [24] in *National Institute of Standards and Technology. Biometric Quality Workshop II*, 2008 [Online]. Available: <http://www.itl.nist.gov/iad/894.03/quality/workshop07/index.html>
- [25] J. C. Wu and C. L. Wilson, "Nonparametric analysis of fingerprint data on large data sets," *Pattern Recognit.*, vol. 40, pp. 2574–2584, 2007.
- [26] A. A. Balkema and L. de Haan, "Limit distributions for order statistics II," *Theory of Probability and its Applications*, vol. 23, pp. 341–358, 1978.



**Manas Baveja** received the B.Tech. degree in chemical engineering from India Institute of Technology–Delhi, in 2001, and the Ph.D. degree in computational and mathematical engineering at Stanford University, Stanford, CA, in 2007.

He currently works at a hedge fund in London.



**Hongsong Yuan** received the B.S. degree in mathematics from Peking University in 2002. He is currently working toward the Ph.D. degree in the Management Science and Engineering Department, Stanford University, Stanford, CA.

His research interests include probability theory, statistical learning, optimization, and Markov decision processes.



**Lawrence M. Wein** received the B.S. degree from Cornell University in 1979 and the Ph.D. degree from Stanford University, Stanford, CA, in 1988.

He is the Jeffrey S. Skoll Professor at the Graduate School of Business, Stanford University. His research interests include manufacturing, public health, and homeland security.