A GENERALIZED MIXED MODEL FRAMEWORK FOR ASSESSING FINGERPRINT INDIVIDUALITY IN PRESENCE OF VARYING IMAGE QUALITY[1]

BY SARAT C. DASS[*], CHAE YOUNG LIM[†] AND TAPABRATA MAITI[†]

Universiti Teknologi PETRONAS[*] and Michigan State University[†]

Fingerprint individuality refers to the extent of uniqueness of fingerprints and is the main criteria for deciding between a match versus nonmatch in forensic testimony. Often, prints are subject to varying levels of noise, for example, the image quality may be low when a print is lifted from a crime scene. A poor image quality causes human experts as well as automatic systems to make more errors in feature detection by either missing true features or detecting spurious ones. This error lowers the extent to which one can claim individualization of fingerprints that are being matched. The aim of this paper is to quantify the decrease in individualization as image quality degrades based on fingerprint images in real databases. This, in turn, can be used by forensic experts along with their testimony in a court of law. An important practical concern is that the databases used typically consist of a large number of fingerprint images so computational algorithms such as the Gibbs sampler can be extremely slow. We develop algorithms based on the Laplace approximation of the likelihood and infer the unknown parameters based on this approximate likelihood. Two publicly available databases, namely, FVC2002 and FVC2006, are analyzed from which estimates of individuality are obtained. From a statistical perspective, the contribution can be treated as an innovative application of Generalized Linear Mixed Models (GLMMs) to the field of fingerprint-based authentication.

1. Introduction.

1.1. *Fingerprint features*, *matching and individuality.* Fingerprint individuality is the study of the extent of uniqueness of fingerprints and is the central premise for expert testimony in court. Assessment of individuality is made based on comparing features from two fingerprint images such as in Figure 1. A fingerprint image consists of alternating dark and light smooth flow lines termed as ridges and valleys. The fingerprint feature formed when a ridge occasionally either terminates or bifurcates is called a minutia; the minutia type "ending" and "bifurcation" correspondingly relate to the type of ridge anomaly that occurs. The location of a

FIG. 1. *A total of* 35 *and* 49 *minutiae were detected in left and right images, respectively, and* $w = 7$ *correspondences (i.e., matches) were found. The white squares and lines, respectively, represent the minutia location and direction. Note that for a minutia ending, the white line points away from the ridge flow, whereas for a minutia bifurcation, the white line points along the direction of the merged ridge. Images taken from the publicly available database FVC2002 DB2.*

minutia is denoted by x where $x \in D$, D is the image domain. Figure 1 identifies all minutiae as white squares in the two fingerprint images; the images are from the publicly available database FVC2002. Each white line emanating from the center of the square represents the minutia direction, denoted by u with $u \in (0, 2\pi]$, which is (i) the direction of the merged ridge flow for a minutia bifurcation, and (ii) the direction pointing away from the ridge flow for a minutia ending; see Figure 1. Minutia information of a fingerprint consists of the collection of locations and directions, (x, u), of all minutiae in the image.

Minutiae in fingerprint images are extracted using pattern recognition algorithms. In this paper, we used the algorithm described in Zhu, Dass and Jain (2007) for minutia extraction. Minutia information is easy to extract, and believed to be permanent and unique; that is, minutiae information is believed to stay the same over time and different individuals have distinct minutia patterns, thus making it a popular method for identifying individuals in the forensics community. For a pair of prints, a minutia (x, u) in one print is said to match a minutia (y, v) in the another print if

$$(1.1) \qquad |x - y|_d < r_0 \quad \text{and} \quad |u - v|_a < u_0$$

for prespecified small positive numbers r_0 and u_0, where $|x - y|_d$ and $|u - v|_a$, respectively, denote the Euclidean distance in R^2 and the angular distance

$$(1.2) \qquad |u - v|_a = \min\{|u - v|, 2\pi - |u - v|\}.$$

We subsequently assume that a set of minutiae has already been extracted (or detected) for every fingerprint image under study. Fingerprint-based authentication proceeds by determining the highest possible number of minutia matches between a pair of prints. This is achieved by an optimal rigid transformation that brings the two sets of minutiae as close to each other as possible and then counting the number of minutia pairs (x, u) and (y, v) that satisfy (1.1). For example, the number of minutia matches between the prints in the left and right panels in Figure 1 is $w \equiv 7$.

A reasonably high degree of match (high w) between the two sets of minutiae leads forensic experts to testify irrefutably that the owner of the two prints is one and the same person. Central to establishing an identity based on fingerprint evidence is the assumption of discernible uniqueness; fingerprint minutiae of different individuals are observably different and, therefore, when two prints share many common minutiae, the experts conclude that the two different prints are from the same person. The primary concern here is what constitutes a "reasonably high degree of match?" For example, in Figure 1, is $w = 7$ large enough to conclude that the two prints come from the same person? When fingerprint evidence is presented in a court, the testimony of experts is almost always included and, on cross-examination, the foundations and basis of this testimony are rarely questioned. However, in the case of Daubert vs. Merrell Dow Pharmaceuticals (1995), the U.S. Supreme Court ruled that in order for expert forensic testimony to be allowed in courts, it had to be subject to the criteria of scientific validation [see Pankanti, Prabhakar and Jain (2002) for details]. Following the Daubert ruling, forensic evidence based on fingerprints was first challenged in the case of U.S. vs. Byron C. Mitchell (1999) and, subsequently, in 20 other cases involving fingerprint evidence. The main concern with an expert's testimony as well as admissibility of fingerprint evidence is the problem of individualization. The fundamental premise for asserting the extent to which the prints match each other (i.e., the extent of uniqueness of fingerprints) has not been scientifically validated and matching error rates are unknown [Pankanti, Prabhakar and Jain (2002), Zhu, Dass and Jain (2007)]; see also the National Academy of Sciences (NAS) report (2009).

The central question in a court of law is "What is the uncertainty associated with the experts' judgement when matches are decided by fingerprint evidence?" How likely can an erroneous decision be made for the given latent print? The main issue with expert testimony is the lack of quantification of this uncertainty in the decision. To address these concerns, several research investigations have proposed measures that characterize the extent of fingerprint individuality; see Pankanti, Prabhakar and Jain (2002), Zhu, Dass and Jain (2007) and the references therein. The primary aim of these measures is to capture the inherent variability and uncertainty (in the expert's assessment of a "match") when an individual is identified based on fingerprint evidence.

A measure of fingerprint individuality is given by the probability of a random correspondence (PRC), which is the probability that two sets of minutiae, one from

the query containing m_1 minutiae and the other from the template fingerprint containing m_2 minutiae, randomly correspond to each other with at least w matches. Since large (resp., small) w is a measure of the extent of similarity (resp., dissimilarity) between a pair of fingerprints, the PRC should be a decreasing function of w. Mathematically, the PRC corresponding to w matches is given by

$$(1.3) \qquad \text{PRC}(w|m_1, m_2) = P(\mathcal{S} \geq w|m_1, m_2),$$

where the random variable $\mathcal{S}$ denotes the number of minutia matches that result when two arbitrary fingerprints from a target population are paired with each other; the notation used in (1.3) also emphasizes the dependence of the PRC on the total number of minutiae detected in the two prints, that is, m_1 and m_2. It is clear from the above formula that $\text{PRC}(w|m_1, m_2)$ decreases as w increases for fixed m_1 and m_2. The distribution of $\mathcal{S}$ is governed by the extent to which minutia matches occur randomly, an unknown quantity which has been modeled in a variety of ways in the literature. The main aim of research in fingerprint individuality is to obtain reliable inference on the PRC, which is also unknown as a result of the unknown extent of random minutia matches.

1.2. *Minutia matching models.* Several fingerprint minutia matching models and expressions for the PRC have been developed from a completely theoretical perspective [see Pankanti, Prabhakar and Jain (2002) for a detailed account of these models]. These models are distributions elicited for minutia occurrences, that is, the distributions that arise from viewing each minutia (x, u) as a random occurrence in the fingerprint, occurring independently of each other. A major drawback of these works is that the models for minutia occurrences, and consequently matching probabilities and PRCs, are not validated based on actual fingerprint images. This effort was first carried out in Pankanti, Prabhakar and Jain (2002) based on real fingerprint databases, albeit for a simple minutia matching model. Subsequent improvements on validating matching models based on actual fingerprint databases have been carried out in Zhu, Dass and Jain (2007) and several other works. Zhu, Dass and Jain (2007), for example, demonstrated that when the number of minutiae in the prints are moderate to large, the distribution of $\mathcal{S}$ in (1.3) can be approximated by a Poisson distribution with mean λ, the expected number of random matches, given by

$$(1.4) \quad \lambda = \lambda(f_1, f_2) = m_1 m_2 \left(2\pi r_0^2 u_0 \int_D \int_{(0, 2\pi]} f_1(x, u) f_2(x, u) \, dx \, du \right),$$

based on the minutia distributions f_1 and f_2 for prints 1 and 2, respectively; in (1.4), m_1 and m_2 are, respectively, the number of minutiae in prints 1 and 2, and r_0 and u_0 are as in (1.1). The above formula for the mean number of random matches can be understood in the following way: The total number of pairings available for random matching from m_1 minutiae in print 1 and m_2 minutiae in

print 2 is $m_1 m_2$. The expression $2\pi r_0^2 u_0 \int_D \int_{(0,2\pi]} f_1(x,u) f_2(x,u)\, dx\, du$ represents the probability of a single random match when r_0 and u_0 are small compared to the image size. It follows from the expected value of a binomial distribution that λ in (1.4) is the mean number of successes (i.e., random matches). The contribution of Zhu, Dass and Jain (2007) was to characterize the distribution of $\mathcal{S}$ (the extent of random matching in the target population) by a single number, namely, the probability of a (single) random match between two minutiae, one from each of the prints in question. Since the probability of a random match

$$(1.5) \qquad p(f_1, f_2) = 2\pi r_0^2 u_0 \int_D \int_{(0,2\pi]} f_1(x,u) f_2(x,u)\, dx\, du$$

depends on the minutia distributions f_1 and f_2, Zhu, Dass and Jain (2007) inferred these distributions from actual fingerprint databases by representing them as a finite mixture of normals.

Factors other than the minutia distributions also govern the extent of random matches. In the case of forensic testimony, for example, it is reasonable to believe that expert matching is more prone to error if the latent prints are of poor quality. Here, poor quality images mean poor resolution (or clarity) of the ridge-valley structures (e.g., in the presence of smudges, sweaty fingers, cuts and bruises) due to which the detection of true minutiae can be missed and spurious minutiae can be detected; see Figure 2 for examples of good, moderate and poor quality fingerprint images. In other words, the extent of individualization should be smaller when the underlying image qualities are poor. One critical issue, therefore, is to be able to quantify the increase in PRC with respect to quality degradation for each matching number w. This quantification is crucial, for example, for latent prints lifted
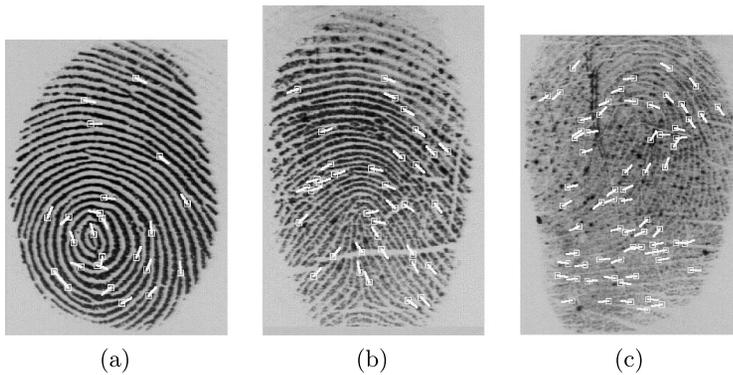

FIG. 2.   *Fingerprint images with* (a) *good* (b) *moderate and* (c) *poor quality. White squares and lines indicate locations and directions of detected minutiae, and possibly not true ones in panel* (c); *images taken from the publicly available FVC2002 databases.*

from crime scenes which are known to have inferior image quality. In the presence of poor quality, automatic as well as manual extraction of fingerprint features are prone to more errors through (i) increased likelihood of detecting spurious minutiae and (ii) missing true ones. Previous work has not quantified the effects of (i) and (ii) on fingerprint individuality; see Dass (2010).

1.3. *Objective and contributions.* The aim of this paper is to quantify PRC assessment (and hence fingerprint individuality) in the presence of varying image quality. The methodology involves two main steps. (i) First, a class of generalized linear mixed models (GLMMs) is proposed that consists of two levels. At the top level, the Poisson model is elicited as the distribution for the number of random minutia matches, following the derivation of Zhu, Dass and Jain (2007). At the second level, image quality of the two prints in question are incorporated as covariates. Also, inter-finger variability is modeled using variance components in the second level. (ii) Second, an inference procedure for the PRC is developed to accommodate a large number of images from different fingers as is typical in real fingerprint databases. Efficient computational algorithms are essential for arriving at reliable reports of fingerprint individuality in practice. Since the number of fingerprint images in real databases is typically large, inference based on Bayesian computational algorithms such as the Gibbs sampler is extremely slow due to increased dimensions of the parameter space. To alleviate this problem, the asymptotic Laplace approximation of the GLMM likelihood is used instead. To summarize, *contributions of this paper are as follows*: For fingerprint-based authentication, a procedure is developed for obtaining the PRC based on explicitly modeling the occurrence of spurious minutia on image quality degradation. Statistical contributions include (i) an innovative application of GLMM to fingerprint-based authentication, and (ii) the development of computationally fast algorithms achieved by approximating the GLMM likelihood (with associated theoretical and numerical validation). Inference results (point estimates and credible intervals) on the PRC are also given (see the experimental results section—Section 6).

The rest of this paper is organized as follows: Section 3 discusses the log-linear GLMM model that is used to study how PRCs change as a function of the underlying image quality. Section 4 presents the procedure for fitting GLMMs in a Bayesian framework. Section 5 develops the inference procedure (point estimates and credible intervals) on the PRC. Section 6 obtains the estimates of fingerprint individuality based on the FVC2002 and FVC2006 databases and gives an analysis of the numerical results. Section 7 validates the Laplace approximation and various other approximations used in this paper for efficient computation. Section 8 presents the summary, conclusions and other relevant discussions.

2. Fingerprint databases and empirical findings.

2.1. *Fingerprint databases.*   Fingerprint databases that have been used in fingerprint analysis include the FVC (e.g., FVC2000, 2002, 2004 and 2006) and NIST databases (e.g., NIST Special Database 4, 9, 27, etc.) which are publicly available for download and analysis. These fingerprint databases typically consist of images acquired from F different fingers, obtained by placing the finger onto the sensing plate of an image acquisition device. Each finger is sensed multiple times, possibly a different number of times for the different fingers. The databases have a common number of multiple acquisitions, say, L, for each finger, resulting in total of FL images in the database. Due to the different sensors used as well as varying placement of the finger, per individual, onto the sensing plate, the images acquired exhibit significant variability (even for the same finger); see Figure 3 for several examples from two of the databases used in this paper, namely, FVC2002 (DB1 and DB2) and FVC2006 (DB3). For FVC2002 DB1 and DB2, $F = 100$ and $L = 8$, whereas for FVC2006 DB3, $F = 150$ with $L = 15$. Figure 3 gives an idea of the
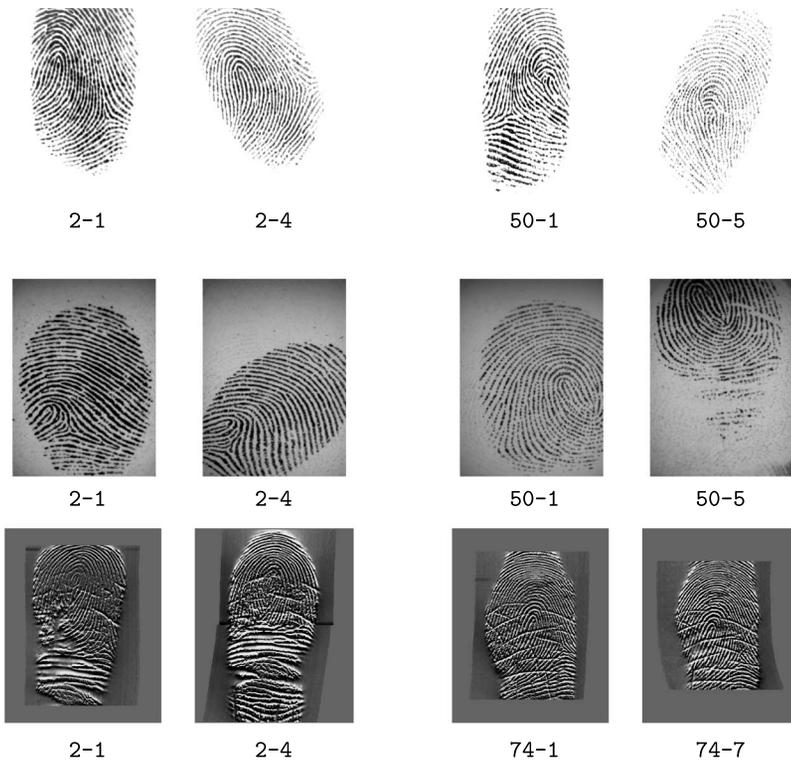

2−1	2−4	50−1	50−5
2−1	2−4	50−1	50−5
2−1	2−4	74−1	74−7

FIG. 3.   *Sample images from the FVC2002 DB1, DB2 and FVC2006 DB3 databases; impression f–l refers to the lth impression of the fth finger. Top, middle and bottom rows are images from the FVC2002 DB1, DB2 and FVC2006 DB3 databases, respectively; all images are publicly available.*

intra-finger variability (variability due to multiple impressions) that has to be taken into account in addition to image quality variability. Intra-finger variability gives different matching numbers when different impressions are used for the matching between two fingers.

2.2. *Empirical findings for varying image quality.* In practice, the quality of an image can be either ordinal (i.e., taking values in an ordered label set) or quantitative (i.e., taking values in a continuum, usually in a bounded interval $[a, b]$). We have considered one specific choice of each type of quality measure: (i) The categorical quality extractor "NFIQ," which is an implementation of the "NIST Image Quality" algorithm based on neural networks described in Tabassi, Wilson and Watson (2004), and (ii) the minutia quality extractor obtained from the feature extraction algorithm `mindtct` from NIST (see the Home Office Automatic Fingerprint Recognition System [HOAFRS (1993)]). These algorithms are all publicly available. For the NFIQ quality extractor, the output Q_0 is a quality label numbered $1, 2, \ldots, 5$, with 1 and 5 corresponding to the best and worst quality images, respectively. For the continuous quality extractor by `mindtct`, a real number Q_{con} between 0 and 1 is obtained with higher values indicating better quality images. To maintain consistency between the two quality measures, we relabel the categorial measure as $Q_{cat} = Q_{max} + 1 - Q_0$, where Q_{max} is the maximum label for a given database, so that higher labels indicate better quality images.

For a pair of prints, 1 and 2 say, with m_1 and m_2 minutiae, respectively, let Y denote the observed number of minutia matches between them. We consider only impostor matches (matches between impressions of different fingers) which are equivalent to the notion of random matching when assessing fingerprint individuality. The total number of impostor pairs in a database with F fingers and L impressions per finger is $F(F - 1)L^2/2$ (since the order of the prints is immaterial). Suppose (Q_1, Q_2) are the labels for the categorical measure Q_{cat}. Table 1 gives the average number (and standard deviations) of matches $\bar{Y}$ for each quality bin pair based on FVC2002 DB1 and DB2. Note that the average number of matches is an increasing function of the quality labels, that is, the average number of random matches increases as the image quality becomes better. This seems counterintuitive initially, but we note that the average increases because the total number of minutiae extracted from the two prints also increases as quality gets better. With more minutiae available for random pairings, the number of matches based on such pairings should also increase on the average. Table 2 gives the mean number (and standard deviations) of minutiae for each quality bin for the two databases considered to illustrate this trend. The minutia extraction algorithm `mindtct` detects (or extracts) a minutia only if its computed reliability measure (computed within the program) is above a certain threshold. For better quality images, more minutiae have a reliability index above the threshold, which explains the higher number extracted in better quality images.

TABLE 1

Mean and standard deviations (*in parenthesis*) *of Y* (*the observed number of random matches*) *for* Q_{cat} *pairs for FVC2002 DB1 and DB2 in panels* (a) *and* (b)

Q_1/Q_2	1	2	3	4
1	2.20	3.02	3.36	3.76
	(0.42)	(0.76)	(0.90)	(1.00)
2	2.96	3.91	4.21	5.12
	(0.75)	(1.30)	(1.39)	(1.71)
3	3.19	4.26	4.71	5.77
	(0.82)	(1.26)	(1.32)	(1.57)
4	3.74	5.26	5.96	7.47
	(0.99)	(1.51)	(1.58)	(1.91)

Q_1/Q_2	1	2	3
1	3.36	3.43	4.15
	(1.11)	(1.14)	(1.36)
2	3.73	4.02	4.94
	(1.13)	(1.19)	(1.41)
3	4.57	4.96	6.38
	(1.40)	(1.43)	(1.71)

(a)

(b)

Consequently, a better quantity to model for varying image quality is $Y/(m_1 m_2)$, which is an estimate of the probability of a random match based on the Poisson model of Zhu, Dass and Jain (2007) [see (1.4) and (1.5)]. Table 3 gives the average value of $Y/(m_1 m_2)$ (and standard deviations) for the different quality bins. Note that now the averages are *decreasing* as a function of Q_{cat}. Intuitively, this is expected since it is less likely to obtain a random match when the image quality is high for a pair of impostor fingerprints. More importantly, we can attribute the larger values of $Y/(m_1 m_2)$ for poor quality images to the extraction of spurious minutiae. The probability of a random match based on true minutiae is intrinsic to the two prints in question and depends only on the distributions f_1 and f_2. Thus, this probability should not depend on quality if all true minutiae are correctly identified. The probability of a random match, p, in the presence of noisy (both

TABLE 2

Mean and standard deviations (*in parenthesis*) *of* $m_1 m_2$ (*the total number of possible random pairings*) *for* Q_{cat} *pairs for FVC2002 DB1 and DB2 in panels* (a) *and* (b)

Q_1/Q_2	1	2	3	4
1	177	366	482	705
	(53)	(144)	(148)	(203)
2	338	684	835	1295
	(178)	(395)	(445)	(634)
3	414	845	1036	1596
	(138)	(368)	(350)	(485)
4	650	1342	1682	2546
	(192)	(527)	(491)	(655)

Q_1/Q_2	1	2	3
1	357	403	641
	(239)	(219)	(336)
2	490	594	947
	(260)	(221)	(327)
3	768	921	1459
	(383)	(301)	(440)

(a)

(b)

TABLE 3
Mean and standard deviations (*in parenthesis*) *of* $Y/(m_1m_2)$ (*the probability of a random pairing*) *for* Q_{cat} *pairs for FVC*2002 *DB*1 *and DB*2 *in panels* (a) *and* (b)

Q_1/Q_2	1	2	3
1	0.0127	0.0105	0.0077
	(0.0083)	(0.0057)	(0.0036)
2	0.0092	0.0075	0.0056
	(0.0046)	(0.0030)	(0.0019)
3	0.0069	0.0058	0.0046
	(0.0029)	(0.0018)	(0.0013)

(a)

Q_1/Q_2	1	2	3	4
1	0.0131	0.0090	0.0074	0.0056
	(0.0032)	(0.0029)	(0.0024)	(0.0021)
2	0.0104	0.0068	0.0058	0.0044
	(0.0042)	(0.0028)	(0.0023)	(0.0017)
3	0.0083	0.0056	0.0049	0.0038
	(0.0028)	(0.0021)	(0.0016)	(0.0012)
4	0.0060	0.0043	0.0037	0.0030
	(0.0017)	(0.0014)	(0.0010)	(0.0009)

(b)

spurious and true) minutiae is the sum of four component probabilities:

$$(2.1) \qquad p = p^{(0,0)} + p^{(0,1)} + p^{(1,0)} + p^{(1,1)},$$

where $p^{(u,v)}$ is the probability of a random match with minutia of type u from print 1 and type v from print 2; $u = 0, 1$ for true or spurious minutia from print 1, and $v = 0, 1$ for true or spurious minutia from print 2. From earlier discussion, it follows that $p^{(0,0)}$ does not depend on (Q_1, Q_2), whereas $p^{(0,1)}$, $p^{(1,0)}$ and $p^{(1,1)}$ all increase as either or both quality labels (Q_1, Q_2) decrease. In the GLMM framework, the dependence of $p^{(0,1)}$, $p^{(1,0)}$ and $p^{(1,1)}$ on quality is modeled explicitly. The above discussion is presented for the categorical quality measure Q_{cat}, but similar empirical findings are also obtained for Q_{con} by binning the values in the range $[0, 1]$.

3. GLMM framework for fingerprint individuality. Let a fingerprint database consists of F fingers and L impressions per finger. Each fingerprint image in the database corresponds to a fingerprint impression denoted by the index pair (f, l) with $1 \leq f \leq F$ and $1 \leq l \leq L$; here f and l, respectively, are the indices of the finger and the impression. An impostor pair of fingerprint images (i, j) with $i \equiv (f, l)$ and $j \equiv (f', l')$ arise from different fingers, that is, $f \neq f'$. The collection of all impostor pairs arising from the database is denoted by $\mathcal{I}$ with $N \equiv F(F - 1)L^2/2$. For each impostor pair $(i, j) \equiv ((f, l), (f', l'))$, the matching algorithm presented in Section 1 [see (1.1) and (1.2)] computes the observed number of minutia matches, Y_{ij}, between impressions i and j. The image quality of i and j is obtained by a quality extractor Q that outputs the ordered pair (Q_i, Q_j) which for the moment is taken to be continuous (the categorical quality case is discussed later).

Based on the discussion in the previous section, the total number of matches Y_{ij} is expressed as a sum of four components:

$$(3.1) \qquad Y_{ij} = \sum_{u=0}^{1} \sum_{v=0}^{1} Y_{ij}^{(u,v)},$$

where $Y_{ij}^{(u,v)}$ is the number of matches obtained based on minutia of type u for impression i and minutia of type v for impression j [see (2.1) and the ensuing discussion in Section 2.2]. The rest of the GLMM model specification is

$$(3.2) \qquad Y_{ij}^{(u,v)} \sim \text{Poisson}(\lambda_{ij}^{(u,v)}),$$

independently for each combination of (u, v),

$$(3.3) \qquad \lambda_{ij}^{(u,v)} = m_i m_j \exp\{b_f + b_{f'} + \eta_{ij}^{(u,v)}\}$$

with $\eta_{ij}^{(u,v)}$ for various combinations (u, v) given as

$$(3.4) \qquad \eta_{ij}^{(0,0)} = 2\beta_0,$$

$$(3.5) \qquad \eta_{ij}^{(0,1)} = \beta_0 + \theta_0 + \theta_1 Q_j,$$

$$(3.6) \qquad \eta_{ij}^{(1,0)} = \beta_0 + \theta_0 + \theta_1 Q_i \quad \text{and}$$

$$(3.7) \qquad \eta_{ij}^{(1,1)} = 2\theta_0 + \theta_1(Q_i + Q_j).$$

In (3.3), m_i and m_j are, respectively, the number of extracted minutiae from i and j. The GLMM model of (3.1)–(3.7) consists of unknown fixed effects parameters $(\theta_0, \theta_1, \beta_0)$ and random effect parameters b_f with $b_f \sim N(0, \sigma^2)$ independently for $f = 1, 2, \ldots, F$ for unknown variance $\sigma^2 > 0$.

Comparing (1.4) and (3.3), we note that the probability of a random match $p(f_1, f_2)$ [see (1.5)] is modeled as $\exp\{b_f + b_{f'} + \eta_{ij}^{(u,v)}\}$ with random effects b_f and $b_{f'}$ in the GLMM framework. The reason for this is as follows: The assessment of fingerprint individuality is typically carried out for a target population with different individuals. Hence, f_1 and f_2 are random realizations of prints from the target population. While each f_1 and f_2 is modeled as a mixture of normals, Zhu, Dass and Jain (2007) subsequently proceed with a clustering of these estimated f_1 and f_2 for a given database. The assumption made is that the target population (and, hence, the database which is considered a representative of the target population) consists of unknown K different clusters of hyperdistributions (a distribution on the mixtures) from which f_1 and f_2 are realized. Subsequent development of the clustering of mixtures of normals is reported in Dass and Li (2009) where the uncertainty of estimating the hyperdistribution is accounted for in the assessment of PRC. In the present context, the random effects b_f (and $b_{f'}$) account for variability due to different fingers (each finger f has a distribution on its minutiae)

which are assumed to be realizations from the target population. The fixed effect parameters $(\theta_0, \theta_1, \beta_0, \sigma^2)$ are target population specific: Their values can change when we move from one database to another.

It follows that the parameters $(\theta_0, \theta_1, \beta_0)$ should be all negative. More elaborate restrictions can be placed on the random and fixed effects parameters jointly by requiring that $b_f + b_{f'} + \eta_{ij}^{(u,v)} \leq 0$ for all (u, v) and (i, j), but we choose not to pursue estimation and subsequent Bayesian inference in this complicated feasibility region. The posterior estimates of $(\theta_0, \theta_1, \beta_0, \sigma^2)$ obtained in the experimental results section demonstrate that the restrictions are satisfied naturally: Estimates of $(\theta_0, \theta_1, \beta_0)$ are found to be negative for all the databases we worked with. Further, we found the estimate of σ to be so small relative to $(\theta_0, \theta_1, \beta_0)$ that the restrictions $b_f + b_{f'} + \eta_{ij}^{(u,v)} \leq 0$ satisfied automatically for all realized values of b_f and $b_{f'}$ when computing the PRC.

For a categorical quality measure labeled in increasing order of image quality, (3.4)–(3.7) in the GLMM framework are replaced by the following four equations for $\eta_{ij}^{(u,v)}$:

$$(3.8) \qquad \eta_{ij}^{(0,0)} = 2\beta_0,$$

$$(3.9) \qquad \eta_{ij}^{(0,1)} = \beta_0 + \theta_0 + \theta_1 + \cdots + \theta_{(Q_j-1)},$$

$$(3.10) \qquad \eta_{ij}^{(1,0)} = \beta_0 + \theta_0 + \theta_1 + \cdots + \theta_{(Q_i-1)} \quad \text{and}$$

$$(3.11) \qquad \eta_{ij}^{(1,1)} = 2\theta_0 + \theta_1 + \cdots + \theta_{(Q_i-1)} + \theta_1 + \cdots + \theta_{(Q_j-1)},$$

where the fixed effects parameters are now $(\theta_0, \theta_1, \ldots, \theta_{(Q_{\max}-1)})$ and β_0 which are all negative.

Equations (3.4)–(3.7) for a continuous quality measure and (3.8)–(3.11) for a categorical quality measure imply that the matching between type u minutia from print 1 and type v minutia from print 2 are independent of each other. To see this, note that the probability of a random match in (2.1), $p = p^{(0,0)} + p^{(0,1)} + p^{(1,0)} + p^{(1,1)}$, is the sum of four component probabilities depending on the type of minutiae being matched. It follows that each normalized term $p^{(u,v)}/p$ gives the multinomial probability that type u minutia is paired with type v minutia, conditional on the fact that a random pairing has occurred. For the GLMM framework, we get $p^{(u,v)}/p = e^{\eta_{ij}^{(u,v)}} / \sum_{u=0}^{1} \sum_{v=0}^{1} e^{\eta_{ij}(u,v)}$. Consequently, the odds ratio $\frac{p^{(1,1)}p^{(0,0)}}{p^{(0,1)}p^{(1,0)}} = 1$ for both the continuous and categorical quality measures. In other words, spurious minutia locations are dispersed "evenly" in between the true minutiae. No region in the print is more prone to spurious detection in relation to the distribution of true minutiae in the fingerprint impression. We further note that $\frac{p^{(u,1)}}{p^{(u,0)}} = e^{\theta_0 + \theta_1 Q_j - \beta_0}$ for the continuous and $\frac{p^{(u,1)}}{p^{(u,0)}} = e^{\theta_0 + \theta_1 + \cdots + \theta_{(Q_j-1)} - \beta_0}$ for the categorical quality measures. These expressions suggest that θ_1 for the continuous

and $\theta_1, \ldots, \theta_{(Q_{\max}-1)}$ for categorical quality measures should all be negative: As Q_j increases (better quality image), the odds of pairing with spurious minutiae ($v = 1$ versus $v = 0$) should decrease for each minutia type $u = 0, 1$.

For both the categorical and continuous quality measures, equation (3.3) can be rewritten in the general log-linear form with respect to the fixed and random effects parameters. For each fixed (u, v), this is given by

$$(3.12) \qquad\qquad \log \lambda_{ij}^{(u,v)} = K_{ij} + \delta_{ij}(u, v),$$

where $K_{ij} = \log(m_i m_j)$ and $\delta_{ij}(u, v) = x_{ij}'(u, v)\theta + z_{ij}'\mathbf{b}$ for appropriate choices of the $1 \times p$ row vector $x_{ij}'(u, v)$ and $1 \times F$ row vector z_{ij}' [which is independent of (u, v)]; θ and $\mathbf{b} = (b_1, b_2, \ldots, b_F)'$, respectively, are $p \times 1$ and $F \times 1$ column vectors representing the collection of fixed and random effects parameters. For the continuous quality measure, the parameter vector $\theta = (\theta_0, \theta_1, \beta_0)'$ with $p = 3$, whereas $\theta = (\theta_0, \theta_1, \ldots, \theta_{(Q_{\max}-1)}, \beta_0)'$ with $p = Q_{\max} + 1$ for the categorical quality measure. We also denote $\tau = (\theta', \sigma^2)'$ to be the $(p + 1) \times 1$ vector consisting of all unknown fixed effects parameters.

In matrix notation, the GLMM for each fixed (u, v) is given by

$$(3.13) \quad Y_{ij}^{(u,v)} \sim \text{Poisson}(e^{K_{ij}+\delta_{ij}(u,v)}) \qquad \text{independently for each } (i, j) \in \mathcal{I}$$

and

$$(3.14) \qquad\qquad \boldsymbol{\delta}(u, v) = \mathbf{X}(u, v)\theta + \mathbf{Z}\mathbf{b},$$

where $\boldsymbol{\delta}(u, v)$ is the $N \times 1$ vector consisting of $\delta_{ij}(u, v)$s, $\mathbf{X}(u, v)$ is the $N \times p$ matrix with rows consisting of $x_{ij}'(u, v)$, and $\mathbf{Z}$ is the $N \times F$ matrix with rows comprising of z_{ij}'.

4. Inference methodology. The subsequent subsections develop inference methodology for $\tau \equiv (\theta', \sigma^2)'$ in a Bayesian framework.

4.1. *Exact GLMM likelihood.* The following notation is developed for the ensuing discussion. Let $\mathbf{Y}_{uv} = \{Y_{ij}^{(u,v)} : (i, j) \in \mathcal{I}\}$ denote the missing data component comprised of minutia matches of type (u, v) for all impostor pairs. We also denote $\mathbf{Y}_{\text{mis}}$ and $\mathbf{Y}_{\text{obs}}$ to be the collection of all missing and observed matching numbers, that is, $\mathbf{Y}_{\text{mis}} = (\mathbf{Y}_{00}, \mathbf{Y}_{01}, \mathbf{Y}_{10}, \mathbf{Y}_{11})$ and $\mathbf{Y}_{\text{obs}} = \{Y_{ij}, (i, j) \in \mathcal{I}\}$, respectively. The set of feasible values for $\mathbf{Y}_{\text{mis}}$ is given by the set

$$(4.1) \qquad \mathcal{M} = \left\{ \mathbf{Y}_{\text{mis}} : \sum_{u=0}^{1} \sum_{v=0}^{1} Y_{ij}^{(u,v)} = Y_{ij} \text{ for all } (i, j) \in \mathcal{I} \right\}.$$

All subsequent analysis is conditional on the random effects $\mathbf{b}$. The missing data component $\mathbf{Y}_{uv}$ has a likelihood $\ell_{uv}(\mathbf{Y}_{uv}|\theta, \mathbf{b}) = e^{-h_{uv}(\mathbf{Y}_{uv}, \theta, \mathbf{b})}$, where

$$h_{uv}(\mathbf{Y}_{uv}, \theta, \mathbf{b}) = -\sum_{(i,j) \in \mathcal{I}} \left\{ (x_{ij}'(u, v)\theta + z_{ij}'\mathbf{b} + K_{ij})Y_{ij}^{(u,v)} \right.$$

$$(4.2)$$

$$\left. - \exp(x_{ij}'(u, v)\theta + z_{ij}'\mathbf{b} + K_{ij}) - \log(Y_{ij}^{(u,v)}!) \right\}$$

based on the GLMM model. Since the (u, v) pairs are independent of each other, the complete likelihood, or the likelihood of $\mathbf{Y}_{\mathrm{mis}}$, is

$$(4.3) \qquad \ell_c(\mathbf{Y}_{\mathrm{mis}}|\boldsymbol{\theta}, \mathbf{b}) = \exp\left(-\sum_{u=0}^{1}\sum_{v=0}^{1} h_{uv}(\mathbf{Y}_{uv}, \boldsymbol{\theta}, \mathbf{b})\right).$$

The observed likelihood, or the likelihood of $\mathbf{Y}_{\mathrm{obs}}$, is the marginal of ℓ_c summing over $\mathbf{Y}_{\mathrm{mis}} \in \mathcal{M}$. Thus, we have

$$(4.4) \qquad \ell_{\mathrm{obs}}(\mathbf{Y}_{\mathrm{obs}}|\boldsymbol{\theta}, \mathbf{b}) = \sum_{\mathbf{Y}_{\mathrm{mis}}\in\mathcal{M}} \ell_c(\mathbf{Y}_{\mathrm{mis}}|\boldsymbol{\theta}, \mathbf{b}) = e^{-h(\boldsymbol{\theta},\mathbf{b})}$$

with

$$h(\boldsymbol{\theta}, \mathbf{b}) = - \sum_{(i,j)\in\mathcal{I}} \left\{\left(H_{ij}(\boldsymbol{\theta}) + z_{ij}'\mathbf{b} + K_{ij}\right)Y_{ij}\right.$$
$$\left. - \exp(x_{ij}'(u, v)\boldsymbol{\theta} + z_{ij}'\mathbf{b} + K_{ij}) - \log(Y_{ij}!)\right\}$$

and

$$(4.5) \qquad H_{ij}(\boldsymbol{\theta}) = \log\sum_{u=0}^{1}\sum_{v=0}^{1}\exp\{x_{ij}'(u, v)\boldsymbol{\theta}\};$$

the last equality in (4.4) results from simplifications based on the well-known multinomial formula. Finally, marginalizing over $\mathbf{b}$, the marginal likelihood for $\mathbf{Y}_{\mathrm{obs}}$ given $\boldsymbol{\tau} = (\boldsymbol{\theta}', \sigma^2)'$, denoted by $\ell(\boldsymbol{\tau})$, becomes

$$(4.6) \quad \ell(\boldsymbol{\tau}) = \int_{R^F} \ell_{\mathrm{obs}}(\mathbf{Y}_{\mathrm{obs}}|\boldsymbol{\theta}, \mathbf{b})\frac{e^{-\mathbf{b}'\mathbf{b}/(2\sigma^2)}}{(2\pi)^{F/2}\sigma^F}\,d\mathbf{b} = \int_{R^F}\exp\{-g(\boldsymbol{\tau}, \mathbf{b})\}\,d\mathbf{b},$$

where

$$(4.7) \qquad g(\boldsymbol{\tau}, \mathbf{b}) = h(\boldsymbol{\theta}, \mathbf{b}) + h_1(\sigma^2, \mathbf{b}),$$

$$(4.8) \qquad h_1(\sigma^2, \mathbf{b}) = \frac{1}{2\sigma^2}\mathbf{b}'\mathbf{b} + \frac{F}{2}\log\sigma^2 + \frac{F}{2}\log(2\pi)$$

and the dependence of $\ell(\boldsymbol{\tau})$ on $\mathbf{Y}_{\mathrm{obs}}$ is suppressed for convenience of the subsequent presentation.

4.2. *Laplace approximation of the likelihood and Bayesian inference.* The typical approach for obtaining inference on $\boldsymbol{\tau}$ in a Bayesian framework is to utilize the Gibbs sampler. The Gibbs sampler augments the random effects vector $\mathbf{b}$ as additional parameters to be estimated and, on convergence, gives samples from the posterior distribution of $(\boldsymbol{\tau}, \mathbf{b})$. This parameter augmentation step avoids computing the integral in the likelihood (4.6). However, in the case of fingerprint databases, F, the number of fingers in a database, is typically large. As a result, parameter augmentation schemes such as the Gibbs sampler take a considerably long time to run until convergence and hinder any possibility of real time inference.

We avoid using any parameter augmentation scheme for the inference on τ. Our approach is to derive an approximation of the GLMM likelihood $\ell(\tau)$ based on the Laplace expansion given by

$$(4.9) \qquad \ell_a(\tau) = e^{-g(\tau, \hat{\mathbf{b}}(\tau))} \left[\det\left(\frac{1}{2\pi} \frac{\partial^2 g(\tau, \hat{\mathbf{b}}(\tau))}{\partial \tau^2} \right) \right]^{-1/2},$$

where $g(\tau, \mathbf{b})$ is the function defined in (4.6), $\hat{\mathbf{b}}(\tau)$ is the maximum likelihood estimate of $\mathbf{b}$ for fixed τ, and $\partial^2 g(\tau, \hat{\mathbf{b}}(\tau))/\partial \tau^2$ is the matrix of second order partial derivatives with respect to $\mathbf{b}$ evaluated at $\mathbf{b} = \hat{\mathbf{b}}(\tau)$, see Shun and McCullagh (1995). In the supplemental article [Dass, Lim and Maiti (2014)], we show that

$$(4.10) \qquad \log(\ell(\tau)) = \log(\ell_a(\tau))\left(1 + O\left(\frac{1}{F}\right)\right)$$

as $F \to \infty$, meaning that the Laplace approximation $\log(\ell_a(\tau))$ is accurate up to order $O(1/F)$ as $F \to \infty$. Equation (4.10) justifies the use of the Laplace-based approximate likelihood in place of (4.6) when F is large.

A further approximation to $\log(\ell_a(\tau))$ is obtained by observing that

$$\log(\ell_a(\tau)) = -g(\tau, \hat{\mathbf{b}}(\tau)) - \frac{1}{2} \log \det\left(\frac{1}{2\pi} \frac{\partial^2 g(\tau, \hat{\mathbf{b}}(\tau))}{\partial \tau^2} \right)$$

$$(4.11)$$

$$\equiv (A) + (B)$$

is the sum of two terms: the first term (A) involves a summation over $F(F - 1)L^2/2$ terms and, hence, is of that order as $F \to \infty$. The order of the second term (B) is $F \log(F - 1 + 1/\sigma^2)$ which follows from $\partial^2 g(\tau, \hat{\mathbf{b}}(\tau))/\partial \tau^2 \sim (F - 1 + 1/\sigma^2)\mathbf{I}_F$, where $\mathbf{I}_F$ is the $F \times F$ identity matrix. This is because each diagonal entry of $\partial^2 g(\tau, \hat{\mathbf{b}}(\tau))/\partial \tau^2$ involves a sum over $F - 1$ terms together with one term involving $1/\sigma^2$, whereas the off-diagonal entries are each of order 1. We retain the term $1/\sigma^2$ in the case when $1/\sigma^2$ is of the same order as F for small σ^2. Nevertheless, $F(F - 1)L^2/2$ still dominates over $F \log(F - 1 + 1/\sigma^2)$ for large F and small σ^2 (where $\sigma^2 \sim F^{-1}$). This implies that (A) dominates (B) for large F and small σ^2 which motivates the further approximation of $\log(\ell_a(\tau))$ by (A). In Section 6, we show that the above approximation is valid based on the estimated values of σ^2 and choices of F and L for each database we worked with. The reader is also referred to further details on the approximation presented in the supplemental article [Dass, Lim and Maiti (2014)].

We assume that the maximum likelihood estimate of τ, $\hat{\tau}$, is available for the moment. Expanding (A) in a Taylor's series around $\hat{\tau}$, we get

$$(4.12) \quad g(\tau, \hat{\mathbf{b}}(\tau)) \approx g(\hat{\tau}, \hat{\mathbf{b}}(\hat{\tau})) + \frac{1}{2}(\tau - \hat{\tau})'\left(\frac{\partial^2 g(\hat{\tau}, \hat{\mathbf{b}}(\hat{\tau}))}{\partial \tau^2} \right)(\tau - \hat{\tau}),$$

where $\partial^2 g(\hat{\tau}, \hat{\mathbf{b}}(\hat{\tau}))/\partial \tau^2$ is the matrix $\partial^2 g(\tau, \hat{\mathbf{b}}(\tau))/\partial \tau^2$ evaluated at $\tau = \hat{\tau}$. It is challenging to numerically evaluate $\hat{\tau}$ and $\partial^2 g(\hat{\tau}, \hat{\mathbf{b}}(\hat{\tau}))/\partial \tau^2$ in real time. This problem is addressed in greater detail in Section 4.3 subsequently.

For inference in the Bayesian framework, we consider a prior π_0 on $\boldsymbol{\tau}$ of the product form: $\pi_0(\boldsymbol{\tau}) = \pi_0(\boldsymbol{\theta})\pi_0(\sigma^2)$ with $\pi_0(\boldsymbol{\theta}) \propto 1$ and $\pi_0(\sigma^2) \propto (1/\sigma^2)I(\sigma^2 > 0)$. These are standard noninformative priors used on location and scale parameters in Bayesian literature. Based on the likelihood and prior specifications, the exact and approximate posteriors of $\boldsymbol{\tau}$ (conditional on $\mathbf{Y}_{\text{obs}}$) are

$$(4.13) \qquad \pi(\boldsymbol{\tau}|\mathbf{Y}_{\text{obs}}) = \frac{\ell(\boldsymbol{\tau}) \times \pi_0(\boldsymbol{\tau})}{\int_{\boldsymbol{\tau}} \ell(\boldsymbol{\tau}) \times \pi_0(\boldsymbol{\tau})\, d\boldsymbol{\tau}}$$

$$(4.14) \qquad \approx \frac{\ell_a(\boldsymbol{\tau}) \times \pi_0(\boldsymbol{\tau})}{\int_{\boldsymbol{\tau}} \ell_a(\boldsymbol{\tau}) \times \pi_0(\boldsymbol{\tau})\, d\boldsymbol{\tau}}$$

$$(4.15) \qquad = \pi_a(\boldsymbol{\tau}|\mathbf{Y}_{\text{obs}}),$$

say, based on equation (4.10). The first likelihood $\ell(\boldsymbol{\tau})$ is difficult to evaluate due to the integral with respect to $\mathbf{b}$, whereas the second likelihood $\ell_a(\boldsymbol{\tau})$ is easier to evaluate for each $\boldsymbol{\tau}$ but difficult to simulate from. However, based on equations (4.11) and (4.12), we obtain samples of $\boldsymbol{\tau}$ from $\tilde{\pi}(\boldsymbol{\tau})$, the multivariate normal distribution with mean $\hat{\boldsymbol{\tau}}$ and covariance matrix $\partial^2 g(\hat{\boldsymbol{\tau}}, \hat{\mathbf{b}}(\hat{\boldsymbol{\tau}}))/\partial \boldsymbol{\tau}^2$. An importance sampling scheme is then used to convert these realizations to posterior samples from $\pi_a(\boldsymbol{\tau}|\mathbf{Y}_{\text{obs}})$. More specifically, suppose $\boldsymbol{\tau}_h^*$, $h = 1, 2, \ldots, H$ are H samples from $\tilde{\pi}$. Define the H weights w_h^*, $h = 1, 2, \ldots, H$ by

$$(4.16) \qquad w_h^* = \frac{\pi_a(\boldsymbol{\tau}_h^*)}{D\tilde{\pi}(\boldsymbol{\tau}_h^*)},$$

where $D = \sum_{h=1}^{H} \pi_a(\boldsymbol{\tau}_h^*)/\tilde{\pi}(\boldsymbol{\tau}_h^*)$ is the normalizing constant so that $\sum_{h=1}^{H} w_h^* = 1$. To obtain a sample from $\pi_a(\boldsymbol{\tau}|\mathbf{Y}_{\text{obs}})$ in (4.15), we resample from the collection $\boldsymbol{\tau}_h^*$ with weights w_h^* for $h = 1, 2, \ldots, H$. This procedure is repeated R times to obtain R samples from $\pi_a(\boldsymbol{\tau}|\mathbf{Y}_{\text{obs}})$. Numerical analysis showed that the weights w_h^* are uniformly distributed with value around $1/H$, which indicates the effectiveness of approximating the target posterior density π_a using the Gaussian approximation $\tilde{\pi}$.

4.3. *Obtaining $\hat{\boldsymbol{\tau}}$ using EM algorithm.* The previous sections assume the availability of the maximum likelihood estimate $\hat{\boldsymbol{\tau}}$ of $\boldsymbol{\tau}$ which we now describe how to obtain. The estimator $\hat{\boldsymbol{\tau}}$ is obtained based on the maximizing function $-g(\boldsymbol{\tau}, \hat{\mathbf{b}}(\boldsymbol{\tau}))$. From equations (4.10) and (4.11), it is clear that $\hat{\boldsymbol{\tau}}$ also approximately maximizes the log-likelihood function $\log(\ell_a(\boldsymbol{\tau}))$ and, subsequently, $\log(\ell(\boldsymbol{\tau}))$, for large F. Ignoring constant terms, we note that

$$(4.17) \qquad -g(\boldsymbol{\tau}, \hat{\mathbf{b}}(\boldsymbol{\tau})) = -h(\boldsymbol{\theta}, \hat{\mathbf{b}}(\boldsymbol{\tau})) - h_1(\boldsymbol{\tau}, \hat{\mathbf{b}}(\boldsymbol{\tau}))$$

$$(4.18) \qquad = \log\left(\sum_{\mathbf{Y}_{\text{mis}} \in \mathcal{M}} \ell_c(\mathbf{Y}_{\text{mis}}|\boldsymbol{\theta}, \hat{\mathbf{b}}(\boldsymbol{\tau}))\right) - h_1(\sigma^2, \hat{\mathbf{b}}(\boldsymbol{\tau})),$$

where the function $h(\boldsymbol{\theta}, \mathbf{b})$ and ℓ_c are as defined in (4.4) and

$$(4.19) \qquad h_1(\sigma^2, \mathbf{b}) = \frac{1}{2\sigma^2}\mathbf{b}'\mathbf{b} + \frac{F}{2}\log \sigma^2.$$

Equation (4.18) sets the stage for an EM algorithm to be used: Start with an initial estimate $\boldsymbol{\tau} = \boldsymbol{\tau}_0 = (\boldsymbol{\theta}_0, \sigma_0^2)$ and obtain $\boldsymbol{\tau} = \boldsymbol{\tau}_k = (\boldsymbol{\theta}_k, \sigma_k^2)$ at the end of the kth step. At step $(k+1)$, the E-step is carried out by noting that the conditional distribution of $(Y_{ij}^{(u,v)}, u = 0, 1, v = 0, 1)$ is multinomial with total number of trials Y_{ij} and multinomial probabilities

$$p_{ij,k}^{(u,v)} = \frac{e^{x_{ij}(u,v)' \boldsymbol{\theta}_k}}{\sum_{u=0}^{1} \sum_{v=0}^{1} e^{x_{ij}(u,v)' \boldsymbol{\theta}_k}}$$

independently for each $(i, j) \in \mathcal{I}$. Subsequently, we plug in the expected value of each $Y_{ij}^{(u,v)}$, $Y_{ij,k}^{(u,v)} \equiv Y_{ij} p_{ij,k}^{(u,v)}$ in place of $Y_{ij}^{(u,v)}$ in (4.2) and (4.3). The M-step now entails maximizing the objective function $-g_c(\boldsymbol{\tau}, \hat{\mathbf{b}}(\boldsymbol{\tau}))$ with respect to $\boldsymbol{\tau}$, where

(4.20) $$g_c(\boldsymbol{\tau}, \mathbf{b}) = h_c(\boldsymbol{\theta}, \mathbf{b}) + h_1(\sigma^2, \mathbf{b})$$

and $h_c(\boldsymbol{\theta}, \mathbf{b})$ is given by

$$h_c(\boldsymbol{\theta}, \mathbf{b}) = \sum_{u=0}^{1} \sum_{v=0}^{1} \sum_{(i,j) \in \mathcal{I}} \{(x_{ij}'(u,v)\boldsymbol{\theta} + z_{ij}'\mathbf{b} + K_{ij})Y_{ij,k}^{(u,v)}$$

(4.21)
$$+ \exp(x_{ij}'(u,v)\boldsymbol{\theta} + z_{ij}'\mathbf{b} + K_{ij})\}.$$

This maximization yields $\boldsymbol{\tau} = \boldsymbol{\tau}_{k+1}$. Proceeding with $k = 1, 2, \ldots$ gives $\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}$ at convergence.

The M-step, or maximization of $-g_c(\boldsymbol{\tau}, \hat{\mathbf{b}}(\boldsymbol{\tau}))$ with respect to $\boldsymbol{\tau}$, is carried out using the Newton–Raphson procedure involving the first and second order partial derivatives of g_c with respect to $\boldsymbol{\tau}$. At step $(k+1)$ of the EM algorithm, we start with the initial value $\boldsymbol{\tau} \equiv \boldsymbol{\tau}_{k+1}^0 \equiv \boldsymbol{\tau}_k$ as defined above. At step $(l+1)$, the current value $\boldsymbol{\tau}_{k+1}^{l+1}$ is obtained from $\boldsymbol{\tau}_{k+1}^l$ using the equation

(4.22) $$\boldsymbol{\tau}_{k+1}^{l+1} = \boldsymbol{\tau}_{k+1}^l - \left[\frac{\partial^2 g_c}{\partial \boldsymbol{\tau}^2}(\boldsymbol{\tau}_{k+1}^l, \hat{\mathbf{b}}(\boldsymbol{\tau}_{k+1}^l)) \right]^{-1} \frac{\partial g_c}{\partial \boldsymbol{\tau}}(\boldsymbol{\tau}_{k+1}^l, \hat{\mathbf{b}}(\boldsymbol{\tau}_{k+1}^l)),$$

where $\frac{\partial^2 g_c}{\partial \boldsymbol{\tau}^2}(\boldsymbol{\tau}_{k+1}^l, \hat{\mathbf{b}}(\boldsymbol{\tau}_{k+1}^l))$ and $\frac{\partial g_c}{\partial \boldsymbol{\tau}}(\boldsymbol{\tau}_{k+1}^l, \hat{\mathbf{b}}(\boldsymbol{\tau}_{k+1}^l))$ are, respectively, the first and second order partial derivatives of g_c evaluated at $\boldsymbol{\tau} = \boldsymbol{\tau}_{k+1}^l$. The explicit expressions for the first and second order partial derivatives of $g_c(\boldsymbol{\tau}, \hat{\mathbf{b}}(\boldsymbol{\tau}))$ with respect to $\boldsymbol{\tau}$ are given in the supplemental article Dass, Lim and Maiti (2014). These expressions, in turn, involve the first and second order partial derivatives of $\hat{\mathbf{b}}(\boldsymbol{\tau})$ with respect to $\boldsymbol{\tau}$. Since no analytical form of $\hat{\mathbf{b}}(\boldsymbol{\tau})$ as a function of $\boldsymbol{\tau}$ is available, one has to resort to numerical methods to estimate $\hat{\mathbf{b}}(\boldsymbol{\tau})$, its first and second order partial derivatives at each iterative step of $k \geq 0$ and $l \geq 0$. A fast and effective way of obtaining these numerical estimates is outlined in the supplemental article Dass,

Lim and Maiti (2014) for the interested reader. On convergence at $\tau = \hat{\tau}$, one obtains the numerical estimate of the matrix $\partial^2 g(\hat{\tau}, \hat{b}(\hat{\tau}))/\partial \tau^2$ by a similar method. The reader is referred to the supplemental article Dass, Lim and Maiti (2014) for details.

5. Bayesian inference for the PRC. Suppose w minutia matches are observed for a fingerprint pair with total number of detected minutiae m_1 and m_2, and with quality measures Q_1 and Q_2 (assume continuous for the moment). Due to varying image quality, not all of the w matches correspond to matches between genuine (true) minutiae. The model developed in Section 3 gives the number of true minutia matches, $Y^{(0,0)}$, to be binomially distributed with parameters w and $p_{00} \equiv e^{\eta^{(0,0)}}/\sum_{u=0}^{1}\sum_{v=0}^{1} e^{\eta^{(u,v)}}$ for the number of trials and success probability, respectively, where $\eta^{(u,v)}$ are as defined in (3.4)–(3.7). The binomial distribution for $Y^{(0,0)}$ results by observing that the conditional distribution of independent Poisson random variables $(Y^{(u,v)}, u, v = \{0, 1\})$ given their sum $Y = \sum_{u=0}^{1}\sum_{v=0}^{1} Y^{(u,v)} = w$ is multinomial with total number of trials w and probabilities $p_{uv} \equiv e^{\eta^{(u,v)}}/\sum_{u=0}^{1}\sum_{v=0}^{1} e^{\eta^{(u,v)}}$ summing to one. It follows that the marginal distribution of each $Y^{(u,v)}$ is binomial for each $(u, v) = (0, 0), (0, 1), (1, 0)$ and $(1, 1)$. Assuming $Y^{(0,0)}$ is known, the PRC corresponding to $Y^{(0,0)}$ matches is given by

$$(5.1) \qquad \text{PRC}^*\big(Y^{(0,0)}|b_1, b_2, m_1, m_2, \tau\big) = P\big(\mathcal{S} \geq Y^{(0,0)}\big),$$

where $\mathcal{S} \sim \text{Poisson}(m_1 m_2 e^{\eta})$ and $\eta = 2\beta_0 + b_1 + b_2$ with b_1 and b_2 distributed as independent $N(0, \sigma^2)$ random variables. The notation of $\text{PRC}^*(\cdot | \cdots)$ in (5.1) emphasizes its dependence on the GLMM parameters τ, the unobserved matches between genuine minutiae, $Y^{(0,0)}$, and the random effects parameters b_1 and b_2. Since $(Y^{(0,0)}, b_1, b_2)$ are unknown, the unconditional PRC is obtained by integrating out all unknown random parameters, that is,

$$(5.2) \quad \text{PRC}(w|m_1, m_2, \tau) = E_{b_1, b_2, Y^{(0,0)}}\big[\text{PRC}^*\big(Y^{(0,0)}|b_1, b_2, m_1, m_2, \tau\big)\big],$$

where the expectation is taken over the joint distribution of $(Y^{(0,0)}, b_1, b_2)$ given τ. The random variables b_1 and b_2 are independent from each other based on our modeling assumptions, but note further that $Y^{(0,0)}$ is independent of (b_1, b_2). This is because the p_{00} parameter of the binomial distribution of $Y^{(0,0)}$, conditional on b_1 and b_2, is independent of b_1 and b_2 (they cancel out from the numerator and denominator expressions of p_{00}). Thus, for given τ, it is easy to sample from the joint distribution $(Y^{(0,0)}, b_1, b_2)$ and estimate the expectation using a Monte Carlo sum. To obtain inference for the PRC, assume M samples from the posterior $\pi_a(\tau|\mathbf{Y}_{\text{obs}})$, $\tau_1, \tau_2, \ldots, \tau_M$, are available. For $r = 1, 2, \ldots, R$, we obtain the rth sample of the PRC, $\text{PRC}_r \equiv \text{PRC}(w|m_1, m_2, \tau_r)$, by plugging $\tau = \tau_r$ in (5.2). These R posterior samples are then used to obtain the PRC posterior mean, standard deviation and the $100(1 - \alpha)\%$ credible intervals (mean $\pm z_{1-\alpha/2}$ sd) in Section 6. Note that $\text{PRC}(w|m_1, m_2, \tau)$ in (5.2) should be the same value whether we

use the combination (Q_1, Q_2) or (Q_2, Q_1). Although this symmetry can be established mathematically, small deviations away from symmetry arise due to sampling error when (i) approximating the expected value in (5.2) using Monte Carlo, and (ii) using different random samples from the posterior of τ for the two combinations (Q_1, Q_2) and (Q_2, Q_1). Thus, a common estimate is reported in Section 6 which is the average of the estimates obtained for the two combinations (Q_1, Q_2) and (Q_2, Q_1).

6. Data analysis. The publicly available databases provided by the Fingerprint Verification Competitions [FVCs, Maio et al. (2002) and FVC2006] are considered here. Specifically, subsets DB1 and DB2 of FVC2002 and subset DB3 of FVC2006 are used. The FVC2002 DB1 database consists of fingerprint images of $F = 100$ different fingers and $L = 8$ impressions per finger obtained using the optical sensor "TouchView II" by Identix with image size 388×374 and resolution 500 dots per inch (dpi). The FVC2002 DB2 database consists of $L = 8$ impressions from $F = 100$ fingers collected from an optical sensor "FX2000" by Biometrika with image size 296×560 and resolution 569 dpi. Fingerprint images in the DB1 and DB2 subsets were collected under exaggerated distortions but were of good quality in general. The FVC2006 database is comprised of four subsets, DB1 through DB4, each consisting of $F = 150$ fingers with $L = 12$ impressions per finger. Fingerprint image data collection in FVC2006 was carried out using a thermal sweeping sensor with resulting image size 400×500 and resolution 500 dpi. Examples of images from the FVC2002 and FVC2006 databases are shown in Figure 3. Note the variability in the image acquisition process due to using different sensors. Our methodology is applied to each of the three subsets with two quality measures: (i) the NFIQ categorical quality measure and (ii) the continuous quality measure described in Section 2.2. Results of the parameter inference methodology for components of τ are as given in Tables 4, 5 and 6 based on $R = 200$ samples from the posterior distribution $\pi_a(\tau|\mathbf{Y}_{\text{obs}})$.

TABLE 4

Posterior means (M), standard deviations (SD) and 99.9% credible intervals (CI) for components of τ based on FVC2002 DB1

τ	Continuous, Q_{con}			τ	Categorical, Q_{cat}		
	M	**SD**	**CI**		**M**	**SD**	**CI**
θ_0	-1.2801	0.0061	$[-1.3003, -1.2599]$	θ_0	-3.4857	0.0049	$[-3.5018, -3.4697]$
θ_1	-5.8520	0.0137	$[-5.8969, -5.8070]$	θ_1	-0.7429	0.0013	$[-0.7472, -0.7386]$
$-$	$-$	$-$	$-$	θ_2	-1.6144	0.0001	$[-1.6145, -1.6143]$
β_0	-2.9047	0.0065	$[-2.9259, -2.8835]$	β_0	-2.7297	0.0064	$[-2.7507, -2.7087]$
$\log(\sigma^2)$	-4.9518	0.1460	$[-5.4321, -4.4716]$	$\log(\sigma^2)$	-4.9537	0.1231	$[-5.3588, -4.5486]$

TABLE 5
Posterior means (M), standard deviations (SD) and 99.9% credible intervals (CI) for components of τ based on FVC2002 DB2

τ	Continuous, Q_{con}			τ	Categorical, Q_{cat}		
	M	**SD**	**CI**		**M**	**SD**	**CI**
θ_0	−0.6346	0.0020	[−0.6413, −0.6280]	θ_0	−2.9255	0.0019	[−2.9319, −2.9191]
θ_1	−9.2982	0.0024	[−9.3062, −9.2901]	θ_1	−1.1496	0.0011	[−1.1532, −1.1460]
−	−	−	−	θ_2	−0.9676	0.0007	[−0.9698, −0.9653]
−	−	−	−	θ_3	−4.2827	0.0007	[−4.2850, −4.2804]
β_0	−2.8721	0.0015	[−2.8769, −2.8672]	β_0	−2.8810	0.0001	[−2.8811, −2.8810]
$\log(\sigma^2)$	−4.2974	0.0185	[−4.3581, −4.2366]	$\log(\sigma^2)$	−4.7817	0.0132	[−4.8252, −4.7382]

Inference for the PRC: We report the PRC corresponding to $w = 12$ matches using the method outlined in Section 5. Note that $w = 12$ is used for illustrative purposes only; similar inference results on the PRC can be obtained for any observed number of minutia matches w. The values of m_1 and m_2, the total number of extracted minutiae in the two prints, are fixed at mean values for each database: These are $(m_1, m_2) = (38, 38)$, $(40, 40)$ and $(84, 84)$ for FVC2002 DB1, FVC2002 DB2 and FVC2006 DB3, respectively. The mean numbers for FVC2002 DB1 and DB2 are similar, and are much smaller compared to FVC2006 DB3. Tables 7 and 8 give the results for FVC2002 DB1 and DB2 based on the categorical quality measure, whereas Tables 9 and 10 give the results for the continuous quality measure. Similar reports and conclusions are obtained for FVC2006 DB3 as in FVC2002 DB1 and DB2, so the relevant tables are not presented here but relegated to the Appendix.

The posterior mean estimates of the PRC are monotonically decreasing functions of increasing quality measures (for both NFIQ and the continuous quality

TABLE 6
Posterior means (M), standard deviations (SD) and 99.9% credible intervals (CI) for components of τ based on FVC2006 DB3

τ	Continuous, Q_{con}			τ	Categorical, Q_{cat}		
	M	**SD**	**CI**		**M**	**SD**	**CI**
θ_0	−6.1444	0.0025	[−6.1526, −6.1362]	θ_0	−5.4501	0.0014	[−5.4547, −5.4456]
θ_1	−3.6746	0.0042	[−3.6885, −3.6607]	θ_1	−0.2537	0.0009	[−0.2565, −0.2509]
−	−	−	−	θ_2	−1.1425	0.0011	[−1.1462, −1.1389]
−	−	−	−	θ_3	−2.3753	0.0032	[−2.3859, −2.3647]
−	−	−	−	θ_4	−2.9186	0.0029	[−2.9281, −2.9091]
β_0	−3.1624	0.0011	[−3.1660, −3.1587]	β_0	−3.1734	0.0009	[−3.1763, −3.1705]
$\log(\sigma^2)$	−3.9248	0.0127	[−3.9666, −3.8831]	$\log(\sigma^2)$	−4.2314	0.0065	[−4.2527, −4.2101]

TABLE 7
Inference on PRC(12|38, 38) *based on* Q_{cat} *and FVC*2002 *DB*1: *M and CI are, respectively, the posterior means and* 99.9% *credible intervals*

$Q_1 \setminus Q_2$	1		2		3	
	M	CI	M	CI	M	CI
1	0.4082	[0.3759, 0.4404]	0.2653	[0.2424, 0.2882]	0.1541	[0.1325, 0.1758]
2	0.2653	[0.2424, 0.2882]	0.1383	[0.1113, 0.1654]	0.0565	[0.0451, 0.0679]
3	0.1541	[0.1325, 0.1758]	0.0565	[0.0451, 0.0679]	0.0130	[0.0095, 0.0164]

measure), as should be expected. As quality becomes better, erroneous decisions due to spurious minutia matches are reduced. In turn, this component of uncertainty in the PRC evaluation is also decreased. Thus, for high quality images, the PRC essentially captures the inherent inter-finger variability in the population. The results reported in Tables 7–10 show the importance of having good quality images when a decision of a positive match has to be made. With the requirement of PRC ≈ 0.05, we see from the four tables that this is achieved when both prints are of the best quality. The PRC deteriorates quickly as the image quality degrades, as is indicated by all the four tables, signifying the importance of very good quality images for PRC assessment. For poor quality images, one has to be extra cautious interpreting the extent of a fingerprint match, as the uncertainty associated with a positive match is very large.

Tables 7–10 invite some comparisons. For DB1, the average values of Q_{con} for each value of $Q_{\text{cat}} = 1, 2, 3$ are $0.32, 0.45, 0.51$, respectively; for DB2, the

TABLE 8
Inference on PRC(12|40, 40) *based on* Q_{cat} *and FVC*2002 *DB*2: *M and CI are, respectively, the posterior means and* 99.9% *credible intervals*

$Q_1 \setminus Q_2$	1		2		3		4	
	M	CI	M	CI	M	CI	M	CI
1	0.7958	[0.7873, 0.8044]	0.5858	[0.5733, 0.5983]	0.4734	[0.4605, 0.4862]	0.3881	[0.3750, 0.4013]
2	0.5858	[0.5733, 0.5983]	0.2725	[0.2609, 0.2841]	0.1556	[0.1469, 0.1643]	0.0902	[0.0838, 0.0967]
3	0.4734	[0.4605, 0.4862]	0.1556	[0.1469, 0.1643]	0.0667	[0.0608, 0.0726]	0.0275	[0.0245, 0.0306]
4	0.3881	[0.3750, 0.4013]	0.0902	[0.0838, 0.0967]	0.0275	[0.0245, 0.0306]	0.0073	[0.0062, 0.0085]

TABLE 9
Inference on PRC(12|38, 38) *based on* Q_{con} *and FVC*2002 *DB*1: *M and CI are, respectively, the posterior means and* 99.9% *credible intervals*

$Q_1 \setminus Q_2$	0.3		0.4		0.5	
	M	CI	M	CI	M	CI
0.3	0.5358	[0.5126, 0.5591]	0.3908	[0.3671, 0.4146]	0.2868	[0.2650, 0.3087]
0.4	0.3908	[0.3671, 0.4146]	0.2383	[0.2174, 0.2592]	0.1455	[0.1280, 0.1631]
0.5	0.2868	[0.2650, 0.3087]	0.1455	[0.1280, 0.1631]	0.0724	[0.0610, 0.0838]

average values of Q_{con} for each value of $Q_{cat} = 1, 2, 3, 4$ are 0.24, 0.38, 0.53 and 0.58, respectively. The levels of Q_{con} chosen in Tables 9 and 10 are taken to reflect these average values. So, one would expect that the PRC results presented for each database be consistent over Q_{cat} and Q_{con}, especially when the value of PRC is small, say, less than 0.05, for example. This is generally the case when we compare the entries of Tables 7 and 9 for DB1 and Tables 8 and 10 for DB2. Minor differences in the PRCs can be attributed to the significant variability of Q_{con} for each level of Q_{cat} in the three databases. While we would expect Q_{con} values to be different corresponding to different levels of Q_{cat}, we do not find this to be the case empirically. There is significant overlap between the Q_{con} levels for the different Q_{cat} values. In the case when $Q_{cat} = 3$ for DB1, the range of Q_{con} is from 0.32 to 0.67, whereas it is 0.30 to 0.64 for $Q_{cat} = 2$; see Figure 4 for an idea of the range of these values for Q_{con}.

TABLE 10
Inference on PRC(12|40, 40) *based on* Q_{con} *and FVC*2002 *DB*2: *M and CI are, respectively, the posterior means and* 99.9% *credible intervals*

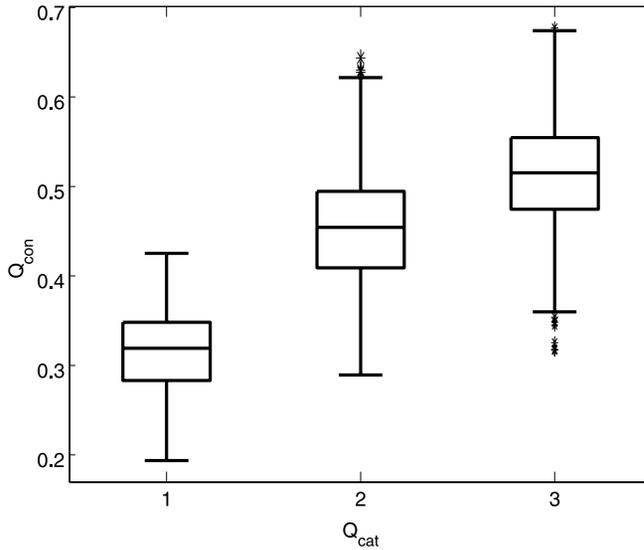| $Q_1 \setminus Q_2$	0.2		0.4		0.5		0.6	
	M	CI	M	CI	M	CI	M	CI
0.2	0.9152	[0.9087, 0.9217]	0.6967	[0.6811, 0.7123]	0.6292	[0.6134, 0.6451]	0.5965	[0.5797, 0.6134]
0.4	0.6967	[0.6811, 0.7123]	0.1956	[0.1816, 0.2096]	0.1158	[0.1051, 0.1265]	0.0873	[0.0787, 0.0959]
0.5	0.6292	[0.6134, 0.6451]	0.1158	[0.1051, 0.1265]	0.0559	[0.0489, 0.0630]	0.0371	[0.0323, 0.0420]
0.6	0.5965	[0.5797, 0.6134]	0.0873	[0.0787, 0.0959]	0.0371	[0.0323, 0.0420]	0.0228	[0.0193, 0.0264]

FIG. 4. *Boxplots showing the distribution of Q_{con} for each level of $Q_{\mathrm{cat}} = 1, 2$ and 3 for DB1.*

The PRCs corresponding to Q_{con} and Q_{cat} are much larger for DB3 [see Tables 1 and 2 in the supplemental article Dass, Lim and Maiti (2014)] compared to DB1 and DB2, because $(m_1, m_2) = (84, 84)$, the mean values for DB3, are much larger compared to the mean minutia numbers for DB1 and DB2. As a result, it is much more likely for random pairings to occur in DB3 since more minutia are available for pairings.

Analysis and interpretation for the forensic practitioner: We illustrate how PRCs can be obtained for a pair of prints under investigation in forensic practice. The pair in Figure 1 has $w = 7$ matches with $m_1 = 35$ and $m_2 = 49$, and $Q_{\mathrm{cat}} = 2$ and 3 for the left and right images, respectively. Since these images are from FVC2002 DB2, we base our inference using the results from this database. Corresponding to $w = 7$, the mean PRC is 0.6395 with 99.9% credible interval [0.6263, 0.6523]. The mean value implies that about 64% of random pairings of impostor fingerprints will have minutia matches of 7 or more. Thus, there is no evidence to suggest that the pair in Figure 1 is something other than a random match (namely, an impostor pair).

The forensic practitioner can also perform a "what if" analysis. If the quality of the images was good with Q_{cat} taking value 4 for both images in the pair, then the PRC is reduced to 0.3112, which still indicates this is more likely a random pairing. With this analysis, a practitioner can eventually conclude that the pairs in question are very likely a random match even in the best case scenario where the image qualities are perfect.

The GLMM procedure can also be utilized to provide a guideline for choices of w that will give desired PRC values for forensic testimony, for example,

TABLE 11
Smallest w required for PRC$(w|35, 49) \leq 0.01$ *based on* Q_{cat} *and DB2. * indicates no such w exists*

Q_1/Q_2	1	2	3	4
1	*	*	33	29
2	*	25	21	18
3	33	21	17	15
4	29	18	15	13

PRC $= 0.01$ means that the error we make in declaring a positive match when in fact it is not is at most 0.01. Table 11 provides the smallest values for w for guaranteeing PRC$(w|m_1, m_2) \leq 0.01$ for different Q_{cat} combinations when $m_1 = 35$ and $m_2 = 49$ based on DB2 with maximum possible number of matches w being $\min(m_1, m_2) = 35$. Based on the $(2, 3)$ [or $(3, 2)$] entry of the table, we see that the desired w to make a reliable decision for the above prints in question should be 21. Thus, $w = 7$ is too low for making a reliable positive match decision. The * entries corresponding to the lowest quality combinations indicate that even for the maximum value of $w = 35$, the PRC never goes below 0.01. This again emphasizes that positive identification decisions cannot and should not be made with very low quality images. Corresponding results can be obtained for DB1 and DB3 similarly, and for Q_{con}, and are therefore not presented here.

7. Validation based on simulation. We performed simulation and validation experiments for DB1 and DB2 with $F = 100$ fingers and $L = 8$ impressions per finger. Q_{cat} and Q_{con} levels were fixed as in the respective databases, but the total number of minutia matches Y_{ij} for each pair of prints (i, j) (see Section 3 for the GLMM model and related notation) were simulated from the GLMM model with fixed and known parameter values $\tau = (\theta', \log(\sigma^2))'$. The parameter values were fixed at the estimated values obtained based on real data for DB1 and DB2, as given in Tables 4 and 5. As noted previously, our inference procedure yields posterior mean and standard deviation estimates as well as 99.9% credible intervals as mentioned in Section 4. The coverage probabilities of each credible interval were calculated for all the true parameter values as well as the true PRC values based on 50 runs. We use the coverage probabilities as a measure of how well the Laplace method approximates the exact GLMM likelihood for large F. Based on simulations, we find that the average coverage probabilities for parameters and PRC values (averaged over all three databases) are approximately 98% and 98.5%, respectively. There is some underestimation of the coverage probability but not grossly so, indicating that the Laplace approximation to the GLMM likelihood is good even for $F = 100$.

8. Summary and discussion.   To assess the extent of fingerprint individuality for different image quality of prints, a GLMM framework and Bayesian inference procedure is developed. Our inference scheme is able to provide a point estimate as well as a credible interval for the PRC depending on the image quality and the observed number of matches between a pair of prints. Numerical reports of the PRC are obtained which can be used to validate forensic testimony. Further, we also provide the smallest number of minutia matches w needed to keep the PRC around a prespecified (small) number. These matching numbers serve as a guideline and safeguard against falsely deciding a positive match when the quality of the prints is unreliable. The best report of fingerprint individuality is obtained when both or at least one is of very good quality and the other is of good quality. No inference on individuality should be made when either print is of moderate to poor quality as observed from our experimental results; having poor quality latent prints, for example, severely hampers reliable matching results.

As far as we know, previous work has not considered including image quality in fingerprint individuality assessment in a quantitative manner and in a formal statistical framework. Our work can be used as a baseline in forensic applications for acquiring a quantitative assessment of individuality given two prints in question. Regardless of whether the prints are of poor or good quality, the methodology outlined will output a PRC value. This PRC value could serve as a baseline and as a first guide to what extent there is evidence beyond a reasonable doubt. The subsection of Section 6 titled "Analysis and inference for the forensic practitioner" highlights the kinds of analysis that may be conducted based on the GLMM methodology. The GLMM method still does not utilize other fingerprint features such as type of minutiae detected, ridge lengths and the class of fingerprint, which could potentially further decrease the PRC and increase the extent of individualization. This will be addressed in a future work.

Intra-finger correlations and inter-finger variability: The incorporation of the random effects b_f is very important in any statistical analysis involving different fingers and multiple impressions per finger. The random effects b_f serve to model both intra-finger correlations as well as inter-finger variability. Models that incorporate correlations due to multiple impressions of the same finger give very different results compared to models that do not account for this. For example, Dass and Jain (2006) show that the upper and lower confidence bounds are misleadingly narrower if correlations, when present, are not taken into account. Similarly, the upper and lower credible bounds for the PRC will be affected if one does not account for intra-finger correlations. The random effects b_f also serve another purpose. They model inter-finger variability as we move from different fingers in the target population; the database is merely a representative sample of the target population of fingers for which the true population PRC is unknown and has to be estimated. In Zhu, Dass and Jain (2007), the PRC was calculated for each pair of print based on minutia distributions on the respective fingers. However, this is not a representative

PRC for the *entire* fingerprint database. Thus, Zhu, Dass and Jain (2007) considered clustering the minutia distributions for the entire fingerprint database. A more formal approach for obtaining inference for population level PRCs was addressed in Dass and Li (2009). The random effects b_f in the present context represent deviations in the target population, thus enabling inference for this overall population PRC to be made.

Alignment of prints: The alignment prior to matching a pair of prints is not a separate issue but a function of the detected minutia. Based on the detected minutia in a print, the alignment with the other print is done by first aligning a pair of detected minutiae (one from each print), then finding the optimal translation (via the paired minutiae locations) and rotation (via the paired minutiae directions) that exactly aligns the two. So, one can see that the alignment is also a function of image quality: For poor quality images, more spurious minutiae are detected and, hence, the alignment can be more random (compared to the true alignment based on genuine minutia). The randomness in the alignment also contributes to increasing the number of minutia matches, Y_{ij}, for poor quality images. We do not model the alignment separately because of the fact that its randomness, which depends on spurious minutia, is captured in the observed Y_{ij}.

Fusion schemes: The difference in the estimates of fingerprint individuality for the three databases can be attributed to the intrinsic nature of the images in these separate databases. For the databases considered, the intrinsic variability arises due to the different sensors used, the extent of distortion due to varying skin elasticity and the composition of the subjects (manual workers versus aged population and others). Where forensic application is concerned, it is usually the practice to match a latent print to template prints in a database. The sensor used to acquire the templates is known and, therefore, it makes sense to report sensor-specific estimates of fingerprint individuality. However, sensor-specific estimates should not be too different from each other: A significant difference in the reported PRCs (using sensor-specific models) indicates systematic bias of that particular sensor toward or against random matching. As we mentioned and observed earlier, we also find significant variability (as well as overlap) in Q_{con} values corresponding to different Q_{cat} levels. This motivates us to consider GLMMs and inference for PRCs based on combining two or more quality covariates that measure different aspects of image clarity. It is possible to arrive at a single measure of fingerprint individuality by incorporating additional random effects (for different fingerprint databases and sensors) and fixed effects (for two or more quality measures) in the GLMM formulation. We will investigate these fusion issues in our future research.

SUPPLEMENTARY MATERIAL

For the supplemental article "A generalized mixed model framework for assessing fingerprint individuality in presence of varying image quality" (DOI: 10.1214/14-AOAS734SUPP; .pdf). The results quoted in the main text are proved in Section 1 and the tables of PRC results for DB3 are in Section 2.

REFERENCES

DASS, S. (2010). Assessing fingerprint individuality in presence of noisy minutiae. *IEEE Trans. of Information Forensics and Security* **5** 62–70.

DASS, S. and JAIN, A. K. (2006). Validating a biometric authentication system: Sample size requirements. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28** 1902–1319.

DASS, S. and LI, M. (2009). Hierarchical mixture models for assessing fingerprint individuality. *Ann. Appl. Stat.* **3** 1448–1466. MR2752141

DASS, S. C., LIM, C. and MAITI, T. (2014). Supplement to "A generalized mixed model framework for assessing fingerprint individuality in presence of varying image quality." DOI:10.1214/14-AOAS734SUPP.

DAUBERT VS. MERRELL DOW PHARMACEUTICALS (1995). 509 U.S. 579, 113 S. Ct. 2786, 125 L.Ed.2d 469.

FVC2006: FINGERPRINT VERIFICATION COMPETITION (2006). http://bias.csr.unibo.it/fvc2006/.

HOME OFFICE AUTOMATIC FINGERPRINT RECOGNITION SYSTEM (HOAFRS), LICENSE 16-93-0026 (1993). *Science and technology group*. Home Office, London.

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York. MR2135927

MAIO, D., MALTONI, D., CAPPELLI, R., WAYMAN, J. L. and JAIN, A. K. (2002). FVC2002: Fingerprint verification competition. In *Proceedings of the International Conference on Pattern Recognition* (*ICPR*) 744–747. IEEE Computer Society, Quebec, Canada.

NATIONAL ACADEMY OF SCIENCES COMMITTEE ON IDENTIFYING THE NEEDS OF THE FORENSIC SCIENCE COMMUNITY, NATIONAL RESEARCH COUNCIL (2009). Strengthening forensic science in the United States: A path forward. National Academies Press.

PANKANTI, S., PRABHAKAR, S. and JAIN, A. K. (2002). On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24** 1010–1025.

SHUN, Z. and MCCULLAGH, P. (1995). Laplace approximation of high-dimensional integrals. *J. Roy. Statist. Soc. Ser. B* **57** 749–760. MR1354079

TABASSI, E., WILSON, C. and WATSON, C. (2004). Fingerprint image quality. Technical Report 7151. Available at http://fingerprint.nist.gov/NBIS.

U.S. VS. BYRON C. MITCHELL (1999). Criminal Action No. 96-407, U.S. District Court for the Eastern District of Pennsylvania.

ZHU, Y., DASS, S. C. and JAIN, A. K. (2007). Statistical models for assessing the individuality of fingerprints. *IEEE Transactions on Information Forensics and Security* **2** 391–401.

S. C. DASS
DEPARTMENT OF FUNDAMENTAL & APPLIED SCIENCES
UNIVERSITI TEKNOLOGI PETRONAS
31750 TRONOH, PERAK
MALAYSIA
E-MAIL: saratcdass70@gmail.com

C. Y. LIM
T. MAITI
DEPARTMENT OF STATISTICS & PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824
USA
E-MAIL: lim@stt.msu.edu
         maiti@stt.msu.edu