# Modeling Uncertainties in Performance of Object Recognition[1]

Suresh Kumar, Bir Bhanu, and Ninad S. Thakoor
*Center for Research in Intelligence Systems*
*University of California, Riverside, CA 92521-0138, USA*
and
Subir Ghosh[2]
Department of Statistics
*University of California, Riverside, CA 92521-0138, USA*

**Abstract**

Efficient probability modeling is indispensable for uncertainty quantification of the recognition data. If the model assumptions do not reflect the intrinsic nature of data and associated random variables, then a strong performance measure will most likely fail to come up with a correct match for recognition. This paper proposes the probability models for two kinds of data obtained with two distinct goals of recognition : identification and discovery. Both frequentist and Bayesian approaches are considered for drawing inferences from the data.

AMS Subject Classifications 2010: .

**Keywords:** Hyper-Geometric, Estimate, Match, Model, Multinomial, Negative-Binomial, Performance, Prior, Uncertainty.

---

[1]In loving memory of our colleague and friend Suresh Kumar (Full Name: Suresh Kumar Ramachandran Nair) who passed away on June 27, 2012 at the end of his third year while working on this research as a Ph.D. student of Distinguished Professor Bir Bhanu.

[2]Corresponding author: Tel: 951-827-3781; E-mail address: subir.ghosh@ucr.edu

# 1 Introduction

The uncertainty quantification (UQ) for the available data on a recognition system is probabilistic in nature and therefore can be performed by statistical modeling. The model for UQ is dependent on the data and the response variables of interest representing the different aspects of recognition. The probability distributions for these variables are functions dependent on the variables and some unknown parameters and are called the probability models or just models. In matching a probe object with many gallery objects, the system recognition as a *match* could be a *true match* or a *false match*, while a *non-match* could be a *true non-match* or a *false non-match*. For the performance modeling, two categories of recognition are considered : *true match (TM)* and *not-true match (NTM)*, where *NTM* includes *false match* and *non-match*. On the one hand, the number of *TM* in a recognition process is a variable of interest. On the other hand, the number of *NTM* preceding the *first TM* is another variable of interest. These two variables have different probability distributions.

The UQ using a binomial probability model was considered in the work of Wang and Bhanu (2007) for describing the probability that the match score is at rank $r$ and then calculating the probability that the match scores are within rank $r$ and its expression when the correct matches happen above a threshold $t$. Bhanu and Tan (2003) presented a model based approach of an accurate and efficient indexing of fingerprint images and performed scientific experiments to compare the performance of their approach with another prominent indexing approach and demonstrated that the performance of their approach is better for both the live scan database and the ink based database NIST-4. Boshra and Bhanu (2005, 2000) presented a theoretical framework for prediction of lower and upper bounds on the probability of correctly recognizing model objects from scene data considering data distortion factors such as uncertainty (noise in feature locations), occlusion (missing features), and clutter (spurious features) as well as the structural similarity between model objects. Their two stage approach calculated a measure of the structural similarity between every pair of objects in the model set in the first stage, and the model similarity information is used along with statistical models of the data distortion factors to determine bounds on the probability of correct recognition in the second stage. Daugman (2003) analyzed the statistical variability that is the basis of iris recognition using new large databases. Daugman (1993) had previously proposed a method for rapid visual recognition of personal identitity based on the failure of a statistical test of hypothesis. Grother and Phillips (2004) proposed binomial models of open- and closed-set identification recognition performance, giving formulae for identification and false match rates as functions of database size, match rank and operating threshold. Johnson, Sun, and Bobick (2003) gave a method to estimate recognition performance for large galleries of individuals using data from a significantly smaller gallery by mathematically modeling a cumulative match characteristic (CMC) curve. The similarity scores of the smaller gallery were used to estimate the parameters of the model and then the rank 1 point (nearest neighbor) of the modeled CMC curve is used as our measure of recognition performance. This method is non-parametric in the sense that it does not make any assumptions about the gallery distribution. Olson (1995) used the probabilistic picking effect to discriminate between likely and unlikely matches in object recognition.

This paper presents UQ with two recognition system data: Data-I and Data-II for the probe and gallery matching. It also proposes the probability modeling for

Data-I and Data-II with hyper-geometric, multinomial, and negative binomial distributions. Both frequentist and Bayesian approaches are considered for estimating the model parameters. The proposed models do not depend on the type or nature of the objects but depend on *TM scores* and *NTM scores* obtained from the ordered similarity scores between probe and gallery objects. The concentration measure *McDiarmid-Hoeffding-Azuma (MHA) Inequality* is not meaningful for the multinomial model considered here because the random variables are not independent. This opens up possibilities for developing the new concentration measures in UQ research.

The remainder of the paper is organized as follows. Section 2 presents two kinds of recognition : one for identification and another for discovery. Section 3 presents two kinds of data, Data-I and Data-II, for uncertainty modeling. Section 4 proposes the hyper-geometric probability model for UQ from Data-I. Section 5 proposes the multinomial and negative binomial probability models for UQ from Data-II. Section 6 discusses the analytical methods and issues in performance evaluation while Section 7 considers the outcome of experimental evaluation. Section 8 concludes with possible future research directions.

# 2 Recognition System and its Performance

## 2.1 Recognition−Identification

Consider an object recognition scenario with $m$ probe objects and $n$ gallery objects where $m_t$ probe objects are present in the gallery (the subscript $t$ represents "true"). The $m_t$ $(0 < m_t \leq \text{Min}(m, n))$ is numerically known in advance and it is strictly positive confirming the presence of at least one probe object in the gallery. The probe objects could be $m$ individuals. In a photo (scene data) of $n$ individuals representing the gallery, $m_t$ individuals in the probe are present facing at different angles while being engulfed by the other noise factors. The purpose of an object recognition system is to identify the probe objects in the gallery. The system may or may not correctly perform this task of identifying the $m_t$ probe objects in the gallery. The system compares between the $i^{th}$ probe object and the $j^{th}$ gallery object with respect to a set of specified features to obtain a *similarity score* and declares it as the *similarity score* for the pair $(i, j)$. For the $i^{th}$ probe object, the system obtains $n$ *similarity scores* from the $n$ gallery objects and determines at most one *similarity score* for the pair $(i, j_i)$ as the *match score*, where $j_i$ is one of the $n$ gallery objects. If the system identifies correctly a pair $(i, j_i)$, then the $i^{th}$ probe object and the $j_i^{th}$ gallery object are identical. If the system fails to identify correctly, then they are not identical. Suppose that the system identifies $m_c$ such pairs $(i, j_i)$ correctly and $0 \leq m_c \leq m_t$. The system generates $nm$ *similarity scores*, $m_c$ of them are *TM*, $(nm - m_c)$ of them *NTM*, and the $m_c$ pairs $(i, j_i)$. The recognition system performance is then measured by

$$\mathbf{P_I} = m_c/m_t, \qquad 0 \leq \mathbf{P_I} \leq 1. \tag{1}$$

Naturally, $\mathbf{P_I} = 1$ means the recognition system performs a perfect identification and $\mathbf{P_I} = 0$ means the system is totally defective. Uncertainties present make no system perfect in reality. A totally defective system is definitely unacceptable.

## 2.2 Recognition−Discovery

In another object recognition scenario with $m$ probe objects and $n$ gallery objects, the probe objects may or may not be present in the gallery. The number of probe objects present in the gallery $m_t$ ($0 \leq m_t \leq \text{Min}(m, n)$) is therefore unknown and it can be equal to zero confirming the possibility of absence of probe objects in the gallery. The recognition system discovers the presence or absence of probe objects in the gallery. If the system discovers correctly $\widehat{m_c}$ probe objects present in the gallery, then $\widehat{m_c}$ objects are $TM$ and $(nm - \widehat{m_c})$ are $NTM$. The recognition system performance is measured by

$$\mathbf{P_D} = \widehat{m_c}/m_t, \qquad 0 \leq \mathbf{P_D} \leq 1. \tag{2}$$

When $\widehat{m_c} = m_t$, or equivalently $\mathbf{P_D} = 1$, the system discovery becomes complete. The unknown $m_t$ makes the evaluation of system discovery completeness challenging.

# 3 Data

## 3.1 Data-I

Suppose that $n > m$ and the $\binom{n}{m}$ possible samples of m objects from the gallery are chosen. In the $i^{th}$ sample, the recognition system discovers correctly the presence of $d_i$ probe objects from the $m^2$ *similarity scores* based on the extracted features of probe and gallery objects, where $0 \leq d_i \leq m$ for $i = 1, \ldots, \binom{n}{m}$. So, the $d_i$ objects are $TM$ and the $m^2 - d_i$ objects are $NTM$.

## 3.2 Data-II

The extracted features of probe and gallery objects are the primary data that are first collected and then used to obtain the $nm$ *similarity scores* $s_{ij}$ from the $nm$ pairs

$(i, j)$ presented below in the matrix form:

$$\begin{pmatrix} s_{11} & \dots & s_{1j} & \dots & s_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{i1} & \dots & s_{ij} & \dots & s_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{m1} & \dots & s_{mj} & \dots & s_{mn} \end{pmatrix},$$

or equivalently

$$\begin{pmatrix} s_{11} & \dots & s_{1j} & \dots & s_{1n}, & \dots & , & s_{m1} & \dots & s_{mj} & \dots & s_{mn} \end{pmatrix}.$$

Assume that the lower the *similarity score*, the higher the match between the probe and gallery objects. After arranging the *similarity scores* in ascending order, each of the $mn$ *similarity scores* indexed below

$$\begin{pmatrix} 1, & \dots & , j, & \dots & , n; & \dots & ; & (m-1)n+1, & \dots & , (m-1)n+j, & \dots & , mn \end{pmatrix},$$

represents one of the $mn$ pairs $(i, j)$. The $n$ could be greater or smaller than $m$ or even be equal to $m$. Assume the existence of $d$ pairs giving $TM$, their indexes are denoted by $r_1$, $r_2$, ... , and $r_d$, where $1 \le r_1 < ... < r_d \le mn$. The scores of the remaining $mn - d$ pairs giving $NTM$ can be classified as: the number $x_1$ of *similarity scores* below the *matched similarity score* indexed by $r_1$; the number $x_i$ of *similarity scores* between two *matched similarity scores* indexed by $r_{i-1}$ and $r_i$; and the number $x_{d+1}$ of *similarity scores* above the *matched similarity score* indexed by $r_d$ with $r_1 < ... < r_i < ... < r_d$. Clearly $r_i - i = x_1 + ... + x_i$ for $i = 1, ..., d$ and $x_1 + ... + x_i + ... + x_{d+1} = mn - d$. For a given $d$, the known values of $d$ and $x_1$, ... , $x_d$ make the values of $r_1$, ... , $r_d$ known and vice versa.

# 4 Modeling Uncertainties from Data-I

What is the chance of determining the $TM$ pairs with Data-I by the recognition system? The number of *true match scores* $T_i$ for the $i^{th}$ sample is a random variable representing the number of $TM$ pairs discovered in the $i^{th}$ sample by the identification system. The probability $P(T_i = d_i)$ can be described by the hyper-geometric distribution

$$P(T_i = d_i) = \binom{m_t}{d_i}\binom{n-m_t}{m-d_i}\bigg/\binom{n}{m},\ max(0, m-n+m_t) \leq d_i \leq min(m, m_t). \quad (3)$$

The expectation and variance of $T_i$ (Bain and Engelhardt (1992), page 97) are

$$E(T_i) = m\left(\frac{m_t}{n}\right),\ Var(T_i) = m\left(\frac{m_t}{n}\right)\left(1 - \frac{m_t}{n}\right)\left(\frac{n-m}{n-1}\right). \quad (4)$$

When $m_t = 25$, $m = 30$, and $n = 100$, $E(T_i) = 7.50$, $Var(T_i) = 3.98$, and $\frac{m_t}{E(T_i)} = 3.3\bar{3}$. So the geometric probability model describes that on the average less than one-third of probe objects present in the gallery is greater than the number of *true matched pairs* in each of the samples for $m = 25$ and $n = 100$. When $n = m$, $E(T_i) = m_t$ and $Var(T_i) = 0$. Naturally

$$m_t \geq max(d_1,\ \ldots,\ d_{\binom{n}{m}}). \quad (5)$$

The maximum likelihood estimator of $m_t$ for the $i^{th}$ sample is given by

$$\widehat{m}_t^{(i)} = \left[\frac{d_i(n+1)}{m}\right],\ for\ i = 1,\ \ldots,\ \binom{n}{m}, \quad (6)$$

where $[x]$ denotes the largest integer less than or equal to $x$. If $d_i = 7$ for the $i^{th}$ sample, then $\frac{d_i(n+1)}{m} = 23.566$ and hence $\widehat{m}_t^{(i)} = 23$. When $d_i = 8$, $\frac{d_i(n+1)}{m} = 26.93$ and thus $\widehat{m}_t^{(i)} = 26$. If $\frac{d_i(n+1)}{m}$ is an integer, then both $\frac{d_i(n+1)}{m}$ and $\frac{d_i(n+1)}{m} - 1$ are maximum likelihood estimators of $m_t$. If the gallery size $n$ and the number of $TM$ pairs $m_t^{(i)}$

become very large satisfying $\frac{m_t^{(i)}}{n} = \pi_i$, then the hyper-geometric distribution for $P(T_i = d_i)$ becomes the binomial distribution (Feller (1968), page 59 and Bain and Engelhardt (1992), page 97)

$$P(T_i = d_i) = \binom{m}{d_i} {\pi_i}^{d_i} (1 - \pi_i)^{m-d_i}, \; d_i = 0, \; \dots, m, \; i = 1, \; \dots, \binom{n}{m}. \tag{7}$$

The maximum likelihood estimator of $\pi_i$ is $\widehat{\pi}_i^{ML} = \frac{d_i}{m}$. Assume the beta prior density of $\pi_i$ expressed in terms of gamma functions as

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} {\pi_i}^{\alpha-1} (1 - \pi_i)^{\beta-1}, \; 0 \leq \pi_i \leq 1, \alpha > 0, \beta > 0. \tag{8}$$

The expectation and variance of this prior density are

$$E_{pr}(\pi_i) = \frac{\alpha}{\alpha+\beta}, \; Var_{pr}(\pi_i) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{9}$$

The posterior mean and variance are

$$E_{po}(\pi_i|data) = \frac{\alpha+d_i}{\alpha+\beta+m}, \; Var_{po}(\pi_i|data) = \frac{(\alpha+d_i)(\beta+m-d_i)}{(\alpha+\beta+m)^2(\alpha+\beta+m+1)}. \tag{10}$$

The Bayesian estimator of $\pi_i$ (Ghosh, Delampady, and Samanta (2006), pages 32−33; Hoff (2009), pages 38−39) is

$$\widehat{\pi}_i^{Bayes} = \frac{\alpha+d_i}{\alpha+\beta+m} = \frac{\alpha+\beta}{\alpha+\beta+m} E_{pr}(\pi_i) + \frac{m}{\alpha+\beta+m} \widehat{\pi}_i^{ML}. \tag{11}$$

If $m$ is very large compared to $\alpha+\beta$, then $\widehat{\pi}_i^{Bayes} \doteq \widehat{\pi}_i^{ML}$.

# 5 Modeling Uncertainties from Data-II

## 5.1 Model 1

Given the number of $TM$ scores $T = d$ with Data-II, can the uncertainties present in the $NTM$ scores be adequately explained by a probability model? The numerical values of ordered indexes $r_1, \ldots, r_d$ $(r_1 < \ldots < r_d)$ of *true match scores* in the $mn$ positions are unknown. The $(mn-d)$ *similarity scores* can be in any of $(d+1)$ categories in the sense that the index can be below $r_1$, in between $r_{i-1}$ and $r_i$ for $i = 2, \ldots, d$, and above $r_d$. Let $p_i$ be the probability that a *similarity score* falls between two $TM$ *similarity scores* indexed by $r_{i-1}$ and $r_i$ for $i = 2, \ldots, d$, $p_1$ the probability that a *similarity score* falls below the $TM$ *similarity score* indexed by $r_1$, and $p_{d+1}$ the probability that a *similarity score* falls above the $TM$ *similarity score* indexed by $r_d$; $0 < p_i < 1$ and $p_1 + \ldots + p_{d+1} = 1$. Let $X_i, i = 1, \ldots, d+1$, be random variables having their realizations $x_i, i = 1, \ldots, d+1$, with probabilities $p_i, i = 1, \ldots, d+1$. Assume the multinomial probability model

$$P\left(X_1 = x_1, \ldots, X_{d+1} = x_{d+1} \,\middle|\, T = d, \sum_{i=1}^{d+1} x_i = (mn-d)\right) = \frac{(mn-d)!}{x_1! \ldots x_{d+1}!} p_1^{x_1} \ldots p_{d+1}^{x_{d+1}}. \tag{12}$$

The expectation and variance of $X_i$ are $E(X_i) = (mn-d)p_i$ and $Var(X_i) = (mn-d)p_i(1-p_i)$. The covariance between $X_i$ and $X_j$ is $Cov(X_i, X_j) = -(mn-d)p_ip_j$. The random variables $X_i, i = 1, \ldots, d+1$ are all correlated. The natural estimators of $p_1$, $p_2, \ldots, p_{d+1}$ are $\widehat{p_i} = \frac{x_i}{mn-d}$, $i = 1, \ldots, d+1$ (Rice (1995), page 259). If the recognition system is perfect with probability one all the time, then $x_1 = x_2 = \ldots = x_d = 0$ and $x_{d+1} = mn - d$ or equivalently, $\widehat{p_i} = 0$ for $i = 1, \ldots, d$ and $\widehat{p}_{d+1} = 1$.

If $p_i = p_i(\theta)$ for $i = 1, \ldots, d+1$ are known functions of an unknown parameter $\theta$, then the maximum likelihood estimator $\hat{\theta}_{ML}$ of $\theta$ if it exists in the admissible set of values of $\theta$ (Rao (1957), page 140) is

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}}\left(x_1 \, log \, p_1(\theta) + \ldots + x_{d+1} \, log \, p_{d+1}(\theta)\right). \tag{13}$$

The $\hat{\theta}_{ML}$ satisfies the equation obtained by setting the *score function* to zero:

$$\frac{x_1}{p_1(\theta)}\frac{dp_1(\theta)}{d\theta} + \ ... \ + \frac{x_{d+1}}{p_{d+1}(\theta)}\frac{dp_{d+1}(\theta)}{d\theta} = 0. \tag{14}$$

Then the maximum likelihood estimator of $p_i = p_i(\theta)$ is $\widehat{p_i}^{ML} = p_i\left(\hat{\theta}_{ML}\right)$ for $i = 1, \ ... \ , d+1$. Assume the Dirichlet prior density of $p_1, \ ... \ , p_{d+1}$ expressed in terms of gamma functions as

$$\frac{\Gamma\left(\sum_{i=1}^{d+1}\alpha_i\right)}{\prod_{i=1}^{d+1}\Gamma(\alpha_i)}\prod_{i=1}^{d+1}p_i^{\alpha_i-1}, \tag{15}$$

where $\alpha_i > 0$ for $i = 1, \ ... \ , d+1$. Denote $W = \sum_{i=1}^{d+1}\alpha_i$ and $W_{(-i)} = \sum_{j=1, j\neq i}^{d+1}\alpha_j$. The expectation, variance and covariance of this prior density are

$$E_{pr}(p_i) = \frac{\alpha_i}{W}, \ Var_{pr}(p_i) = \frac{\alpha_i W_{(-i)}}{W^2\left(W+1\right)}, \ Cov_{pr}(p_i, p_j) = \frac{-\alpha_i\alpha_j}{W^2\left(W+1\right)}. \tag{16}$$

The posterior density is also Dirichlet with parameters $(x_i + \alpha_i)$. The posterior mean is

$$E_{po}(p_i|data) = \frac{x_i + \alpha_i}{mn - d + W}. \tag{17}$$

The Bayesian estimator of $p_i$ is expressed as (Lindley (1964), Good(1965))

$$\widehat{p_i}^{Bayes} = \frac{x_i + \alpha_i}{mn - d + W} = \frac{W}{mn - d + W}E_{pr}(p_i) + \frac{mn - d}{mn - d + W}\widehat{p_i}. \tag{18}$$

If $(mn - d)$ is very large with respect to $W$, then $\widehat{p_i}^{Bayes} \doteq \widehat{p_i}$.

Consider the $i^{th}$ ordered index $r_i$ of *true match score* with $(r_i - i)$ *non-matched scores* below it and $((mn - d) - (r_i - i))$ *non-matched scores* above it. Denote $p_{(i)} = p_1 + ... + p_i$ and $Z_i = X_1 + ... + X_i$. Then, for $i = 1, ..., d$,

$$P\left(Z_i = (r_i - i) \middle| T = d, \sum_{i=1}^{d+1} x_i = (mn - d)\right) = \binom{mn-d}{r_i - i} p_{(i)}^{r_i - i}\left(1 - p_{(i)}\right)^{(mn-d)-(r_i-i)}. \tag{19}$$

## 5.2 Model 2

Assume that the probability of a *similarity score* to be a $TM$ score is $p$, the probability of a similarity score to be a $NTM$ score is $q = 1 - p$, and moreover, the turning out of similarity scores for pairs $(i, j)$ to be $TM$ or $NTM$ are independent. What is the probability of the $i^{th}$ *match score* rank to be $r_i$? The distribution of the random variable $R_i$ representing the rank of the $i^{th}$ *match score* can be described by the negative binomial model (Feller (1968))

$$P(R_i = r_i) = \binom{r_i - 1}{i - 1} p^i q^{r_i - i}, \qquad r_i = 1, 2, ..., \ 0 \leq p \leq 1, \ p + q = 1. \tag{20}$$

The expectation and variance of $R_i$ are

$$E(R_i) = \frac{i}{p}, \ Var(R_i) = \frac{iq}{p^2}. \tag{21}$$

The maximum likelihood estimator of $p$ is

$$\widehat{p}^{ML} = \frac{i}{r_i}. \tag{22}$$

Assume the beta prior density of $p$ as

$$\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) + \Gamma(\alpha_2)} p^{\alpha_1 - 1} q^{\alpha_2 - 1}, \ \alpha_1 > 0, \alpha_2 > 0. \tag{23}$$

The posterior distribution of $p$ given $R_i = r_i$ is the beta distribution with parameters $\alpha_1 + i$ and $\alpha_2 + r_i - i$. The mean of this posterior distribution is the Bayesian estimator of $p$ and is given by

$$\widehat{p}^{Bayes} = \frac{\alpha_1 + i}{\alpha_1 + \alpha_2 + r_i} = \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + r_i} E_{pr}(\widehat{p}) + \frac{r_i}{\alpha_1 + \alpha_2 + r_i} \widehat{p}^{ML}. \tag{24}$$

A question of importance: How many $NTM$ scores would precede the first $TM$ score? This can be answered by examining the random variable $R_1$ with the geometric distribution, a special case of the negative binomial distribution when $i = 1$. The expectation and variance of $R_1$ are $E(R_1) = 1/p$ and $Var(R_1) = q/p^2$.

# 6 Performance Evaluation-Analytical

The concentration inequality measures are used for assessing the performance uncertainties. The celebrated McDiarmid-Hoeffding-Azuma (MHA) inequality (McDiarmid (1989), Azuma (1967), Hoeffding (1963)) is stated below.

**McDiarmid-Hoeffding-Azuma (MHA) Inequality**

*Let* $X_1, \ldots, X_n$ *be independent random variables, with* $X_i$ *taking values in some set* $A_i$ *and* $f : A_1 \times \ldots \times A_n \to R$ *(the real line) be such that there exist* $c_1, \ldots, c_n > 0$ *satisfying*

$$\sup_{x_1,\ldots,x_n,x_{i'}} \left| f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_{i'}, \ldots, x_n) \right| \leq c_i.$$

*Then for any* $\epsilon > 0$

$$P\bigg( f(x_1, \ldots, x_i, \ldots, x_n) - E\left[ f(x_1, \ldots, x_i, \ldots, x_n) \right] \geq \epsilon \bigg) \leq e^{-2\epsilon^2 / \sum\limits_{i=1}^{n} {c_i}^2},$$

*and*

$$P\bigg( f(x_1, \ldots, x_i, \ldots, x_n) - E\left[ f(x_1, \ldots, x_i, \ldots, x_n) \right] \leq -\epsilon \bigg) \leq e^{-2\epsilon^2 / \sum\limits_{i=1}^{n} {c_i}^2}. \tag{25}$$

As a special case of the MHA inequality, the Hoeffding inequality (Hoeffding (1963)) follows.

**Hoeffding (H) Inequality**

*For* $X_i \in [a_i, b_i]$, $f = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ , $c_i = b_i - a_i$,

$$P\bigg( f(x_1, \ldots, x_i, \ldots, x_n) - E\left[ f(x_1, \ldots, x_i, \ldots, x_n) \right] \geq \epsilon \bigg) \leq e^{-2\epsilon^2 / \sum\limits_{i=1}^{n} (b_i - a_i)^2}. \tag{26}$$

For Model 1 in modeling uncertainties from Data-II, $X_1, \ldots, X_{d+1}$ are not independent and therefore the MHA and H inequality measures are not applicable. However, the simple Chebyshev inequality measure is applicable

**Chebyshev Inequality**

For a random variable X with finite variance

$$P\left(|X - E(X)| \geq c\right) \leq \frac{Var(X)}{c^2},\ for\ c > 0. \tag{27}$$

How good are the proposed models to describe the two kinds of data in Data-I and Data-II to recognize correctly the probe objects in the gallery? The models are based on assumptions on the distribution of a response variable or its parameters in the Bayesian paradigm. Even the the concentration inequality measures are based on some distributional assumptions like the independence in the MHA and H inequality measures. Although a lot of research has already been done in developing robust non-parametric, semi-parametric, and Bayesian methods, the real world recognition problems with the advancement in methods, the real world recognition problems continue to remain challenging (Pinto, Cox, and DiCarlo (2008)).

# 7 Performance Evaluation-Experimental

An extensive performance evaluation was executed at different noise levels by Suresh Kumar under the guidance of Professor Bir Bhanu and with the assistance of Dr. Ninad Thakoor (Kumar et al. (2011)). Several of the models discussed were used for the evaluation. The equal values of $n$ and $m$ were taken within the range 100 to 50,000. For the size 50,000, the average success of finding the exact matching object for searching the top 194 objects was 91.11% with a variance of 0.58 demonstrating the strength of the prediction model.

The use of existing NIST-4 fingerprint database was also used to obtain the probe and gallery objects with $n = m = 2,000$. This dataset was divided randomly into two equal parts for model building with one part and the other part for validating the model. With the prediction for the need of 215 *match scores* for achieving the true match 95% of the time, the observed success rate was 93.8% after evaluating the top 215 scores. The different gallery sizes 100, 200, 500, and 1,000 were also constructed from the NIST-4 dataset to perform 1,000 trials with each of them (Kumar et al. (2011)).

The performance evaluation is based on achieving a trade off between (correct match, false match) and (correct non-match, false non-match) in the *confusion matrix* and in the *Receiver Operating Characteristics (ROC) curve*.

# 8 Future Direction

Considerable research progress has already been made for probability modeling of the recognition data when the random variables are independent. This paper opens up a new direction of probability modeling when the random variables are not independent. The popular concentration measure *McDiarmid-Hoeffding-Azuma (MHA) Inequality* is no longer applicable when the random variables are dependent. This paper sets the stage for the possibility of developing new concentration measures for UQ.

**Acknowledgments**

The authors are grateful to the Editor and a referee for their suggestions leading to an improved presentation. Our research was supported by the National Science Foundation Funding : NSF Award 0915270, 10/01/2009−09/30/2013.

# References


[1] Azuma, K. (1967). *Weighted sums of certain dependent random variables. Tohoku Mathematical Journal.* **19**, 357−367.

[2] Bain, L. J. and Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics.* Second Edition, Duxbury Press, California.

[3] Bhanu, B. and Tan, X. (2003). *Fingerprint indexing based on novel features of minute triplets The IEEE Transaction on Pattern Analysis and Machine Intelligence.* **25**, 616−622.

[4] Boshra, M. and Bhanu, B. (2005). *A theoretical framework for predicting performance of object recognition. Computer Vision Beyond the Visible Spectrum − Advances in Pattern Recognition.* 1−37. Springer, New York.

[5] Boshra, M. and Bhanu, B. (2000). *Predicting performance of object recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence.* **22**, 956−969.

[6] Daugman, J. (2003). *The importance of being random : statistical principles of iris recognition. Pattern Recognition.* **36**, 279−291.

[7] Daugman, J. (1993). *High confidence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence.* **15**, 1148−1161.

[8] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications. Volume 1.* Third Edition, Wiley, New York.

[9] Ghosh, J. K., Delampady, M., and Samanta, T. (2006). *An Introduction to Bayesian Analysis : Theory and Methods.* Springer, New York.

[10] Good, I. J. (1965). *The Estimation of Probabilities : An Essay on Modern Bayesian Methods.* MIT Press, Cambridge, Massachusetts.

[11] Grother, P. and Phillips, P. J. (2004). *Models of large population recognition performance. The IEEE Proceedings for the 2004 Computer Society Conference on Computer Vision and Pattern Recognition.* **2**, 68−75.

[12] Hoeffding, W. (1963). *Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association.* **58**, 13−30.

[13] Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods.* Springer, New York.

[14] Johnson, A. Y., Sun, J., and Bobick, A. F. (2003). *Using similarity scores from a small gallery to estimate recognition performance for larger galleries.* 4 pages, *The IEEE Proceedings for the 2003 Workshop on Analysis and Modelling of Faces and Gestures held in conjunction with ICCV*, Nice, France.

[15] Kumar, S., Bhanu, B., Ghosh, S., and Thakoor, N. (2011). *Prediction and validation of indexing performance for Biometrics.* 6 pages, *The IEEE Proceedings for the 2011 International Joint Conference on Biometrics(IJCB)*, Washington, DC.

[16] Lindley, D. V. (1964). *The Bayesian analysis of contingency tables. The Annals of Mathematical Statistics.* **35**, 1622−1643.

[17] McDiarmid, C. (1989). *On the method of bounded differences. Surveys in Combinatorics.* 148−188. Cambridge University Press, New York.

[18] NIST Biometric Score Set. http://www.itl.nist.gov/iad/894.03/biometricscores/.

[19] Olson, C. F. (1995). *Probabilistic indexing for Object Recognition. The IEEE Transaction on Pattern Analysis and Machine Intelligence.* bf 17, 518−522.

[20] Pinto, N., Cox, D. D., and DiCarlo, J. J. (2008). *Why is real-world visual object recognition hard? PLoS Computational Biology.* **4**, 151−156.

[21] Rao, C. R. (1957). *Maximum likelihood estimation for multinomial distribution. Sankhyā,* **18**, 139−148.

[22] Rice, J. A. (1995). *Mathematical Statistics and Data Analysis.* Second Edition, Duxbury Press, California.

[23] Wang, R. and Bhanu, B. (2007). *Predicting fingerprint biometrics performance from a small gallery. Pattern Recognition Letters.* **28**, 40−48.