International Standard

ISO/IEC 5152

First edition
2024-07

Information technology — Biometric performance estimation methodologies using statistical models

Technologies de l'information — Méthodologies d'estimation des performances biométriques à l'aide de modèles statistiques

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

Reference number
ISO/IEC 5152:2024(en)

© ISO/IEC 2024

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2024

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

Introduction

This document provides a methodology for measuring the accuracy of biometric verification systems based on the statistics categorized as the extreme value theory.[1] The methodology is particularly useful when estimating the false match rate with a relatively small sample set. The methodology is an alternative to empirical accuracy measurement.

In order to measure the false match rate of biometric verification systems, evaluators need to prepare a dataset with a sufficiently large number of non-mated attempts in order to observe a sufficient number of false match cases for a reliable estimation of the false match rate. For highly accurate systems the quantity of attempts required to test the false match rate is likely to be extremely large. As performance of biometric verification systems improves dramatically, acquiring representative data of non-mated attempts in sufficient quantity becomes increasingly difficult in terms of the time, cost and practicality of creating datasets. Policy considerations that apply to biometric data collection and use can pose further constraints.

If no false match case is found within the evaluation samples, metrics based on statistics known as “the rule of three” (as is defined in ISO/IEC 19795-1) are widely used in the biometric industry. However, the rule of 3 is only applicable when no false match case is observed within the tested sample set and do not give any indication of the accuracy and confidence levels expected if more than zero false matches were tested. Only if at least 30 false matches were observed, the “rule of thirty” applies, i.e. the true error rate is with 90 % confidence within ± 30 % of the observed error rate.

In this document, two major statistical methods are introduced to estimate the false match rate with a relatively small number of samples. Both methods are widely used in a variety of industries including civil engineering, meteorology, hydrology and financial engineering. Both methods are proven to be highly reliable techniques to estimate the probability of the occurrence of rare, extreme events such as maximum wind velocity or tsunami heights. These statistical methods are applied to similarly rare events of false match cases in biometrics and used to estimate the probability of occurrence of such cases if a larger non-mated sample set is not available. The estimated false match rate is available in the form of cumulative distribution function (CDF) and its interval of confidence.

This document defines procedures for extrapolating performance metrics in technology evaluations. These procedures can also be applied in scenario evaluations and operational evaluations if comparison scores are obtained. This document defines the methodology to be used by evaluators to reliably estimate the false match rate in case of a limited number of false match cases or even no false match case at all. This document does not address certification or conformance.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

Information technology — Biometric performance estimation methodologies using statistical models

1 Scope

This document provides statistical methodologies to estimate false match rates (FMRs) from small biometric sample sets.

This document intends to:

— lay out a methodology for biometric performance estimation based on extrapolation using extreme value statistical models;

— provide statistical methodologies to estimate FMRs of biometric verification systems;

— be applicable to systems that include algorithms that produce likelihood dissimilarity or similarity scores;

NOTE Throughout the document, if not otherwise specified, scores refer to similarity scores.

— specify the methodology for data recording and result reporting;

— introduce metrics for the estimated biometric performance.

The following are not within the scope of this document.

— Estimation of false positive identification rates for one-to-many implementations.

— Estimation of false accept rates for verification transactions.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 19795-1:2021, *Information technology — Biometric performance testing and reporting — Part 1: Principles and framework*

ISO/IEC 2382-37, *Information technology — Vocabulary — Part 37: Biometrics*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 2382-37 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
extrapolated false match rate
extrapolated FMR
false match rate (FMR) that is estimated by using any statistical models such as those used in extreme value theory

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

3.2
quantile-quantile plot
Q-Q plot
quantile-quantile comparison of two distributions, either or both of which may be empirical or theoretical[2]

4 Symbols and abbreviated terms

CDF cumulative distribution function

FMR false match rate

FNMR false non-match rate

GEV generalized extreme value

GP generalized Pareto

MLE maximum likelihood estimation

PDF probabilistic distribution function

σ scale parameter

ξ shape parameter

5 Conformance

To conform to this document, a biometric performance test shall be executed and reported in accordance with the requirements contained in Clauses 6 through 9.

6 Details of estimation

6.1 Estimation of biometric performance based on extreme value theory

Extreme value theory is used to properly estimate the tails of the distribution of comparison scores from different data subjects. By appropriately capturing the tail of the comparison score distribution, it becomes possible to extrapolate scores outside the observed score range, i.e. the score range larger than the maximum empirical score.

The method of extracting the tail of the comparison score distribution differs depending on whether the extreme value distribution model applied is a generalized extreme value distribution model (rGEV; the limiting joint generalized extreme value distribution for the r largest-order statistic) or generalized Pareto distribution model. More details on these models can be found in 6.3, 6.4 and Annex A.

The extreme value theory is applied to the extracted score set to estimate the optimum parameters of the distribution, and the comparison score distribution is approximated.

The result approximated by the extreme value statistic is compared with the original comparison score distribution using the diagnostic diagram. By extrapolating the estimated comparison score distribution, a distribution can be obtained for a section having no score value. Finally, the extrapolated FMR is obtained by setting a threshold value from the comparison score distribution.

This methodology can be applied if comparison scores are obtained. It can be applied not only in technology evaluations, but also in scenario evaluations and operation evaluations as long as the comparison scores are obtained.

If the scores are produced from a population where some subjects are genetically related, e.g. identical twins or siblings, the probability distribution at the extreme value domain can be higher than that without such

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

related scores, which typically results in a small secondary peak. The extrapolated FMR methodology still works for such score distributions reflecting the increase of the probability. The evaluator shall report the estimated results together with the demographic information of the population.

If the dataset has an unintended peak(s) in the extreme value domain, and if these counts are not within an ignorable error range, it can be appropriate to introduce stratified analysis. A single dataset consists of *n* different subject groups. The distribution of the dataset will have synthesized characteristics of each group, reflecting each statistical parameter (e.g. mean, variance and the number of samples). If the differences between the groups are found to be statistically significant and the proportion of such groups cannot be ignored, these groups may be separated into up to *n* sub-datasets and evaluated independently. While these sub-datasets are dependent on the biometric modalities, they are typically characterized by the test crew properties such as:

a) kinship,

b) human races and genders,

c) occupations,

d) health conditions,

e) other factors that reduce the uniqueness of the biometric features of interest.

The extrapolated FMR for each sub-dataset shall be computed in the same manner as described in 6.3 and 6.4. The details of the sub-datasets shall be reported in accordance with the requirements defined in ISO/IEC 19795-1:2021, 12.1.

6.2 Estimation design

The sample size is determined considering the target FMR to measure and the accuracy of the estimation. Measuring an FMR of 0,000 1 % ± 30 % with 90 % confidence takes 30 false matches in 30 million non-mated comparisons (“rule of 30”; see ISO/IEC 19795-1). This typically means several thousand test subjects are necessary to calculate the FMR. This number can be reduced by using appropriate statistical estimations, and by accepting some errors.

Since the extrapolated FMR is estimated by using extreme values, it is always preferred to have a reasonably large number of samples for better accuracy, with accuracy being quantified by a confidence interval giving the best and the worst case FMR values for a certain confidence level. If the extrapolated FMR values are obtained at multiple thresholds, the extrapolated FMR value with the narrower confidence interval is regarded as more reliable.

As the confidence interval and the confidence level shall be reported, it is the evaluator who decides which threshold score and its corresponding extrapolated FMR value to report. If FNMR is reported, the same score threshold shall be applied.

6.3 Generalized extreme value distribution

The rGEV estimation is based on the generalized extreme value distribution model as described in Clause A.2. The validation processes are as follows:

1) Determine *n*, the number of samples in a block.

   n needs to be large enough for each block to contain some extreme values, i.e. large non-mated scores. On the other hand, when choosing *m*, the number of blocks, the trade-off between *m* and *n* needs to be considered, as the number of extreme samples per block decreases as *m* increases.

2) Specify the number of extreme values per block, *r*.

   The *r* value determines the number of extreme samples extracted from each block and hence the number of samples used for the estimation. For similarity scores, the largest *r* scores within the block are used for estimation. For dis-similarity scores, the smallest *r* scores are used. The larger the value

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

for *r*, the more samples for estimation, which contributes to obtaining better fitness. On the other hand, if *r* is too large, non-mated scores that cannot be regarded as extreme values can be included in the samples for estimation, which deteriorate the fitness of the resulting estimated cumulative distribution function (CDF). The typical *r* range to estimate extreme natural phenomena is from 1 to 5, which is also applicable to extrapolated FMR estimation.

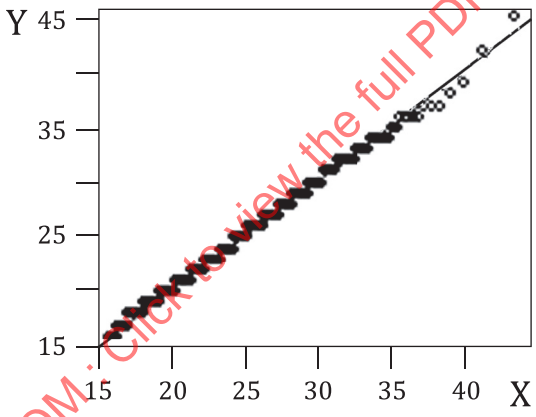
3) Calculate the estimated parameter set ($\hat{\mu},\hat{\sigma},\hat{\xi}$) (refer to Clause A.2) by the extreme samples obtained from steps 1) and 2) using the maximum likelihood estimation (MLE) algorithm.

   Based on the hypothesis that the true probability distribution function (PDF) belongs to the generalized extreme value (GEV) distribution family, apply a maximum likelihood estimation algorithm to obtain the parameter set ($\hat{\mu},\hat{\sigma},\hat{\xi}$).

4) Draw the Q-Q plot and validate the fitness of the estimation.

   A Q-Q plot is a graphical method of statistics that compares two probability distributions by plotting them against each other. First, a set of quantile intervals is selected. A point (*x*, *y*) on the plot corresponds to one (*y* coordinate) of the quantiles of the second distribution plotted against the same quantile of the first distribution (*x* coordinate). Thus, the line is a parametric curve with parameters that connect quantiles. If the two distributions being compared are similar, the point in the Q-Q plot is near the line *y* = *x*. Figure 1 shows an example of a Q-Q plot.

5) Observe the quantile-quantile plot to evaluate the fitness of the estimated model, especially at the extreme value domain (the top-right corner of the plot).



Key

X quantile of rGEV distribution

Y quantile of empirical data

Figure 1 — Example of Q-Q plot of rGEV distribution and empirical data

6) Compare the estimated model and the empirical samples.

Draw the estimated CDF curve by using the GEV model and the parameters obtained in step 3) with 95 % interval of confidence on both sides. Plot the empirical samples on the same plane and compare the fitness of the model, especially the feasibility of the model in the domain where no empirical samples are available.

7) Check the feasibility of the model.

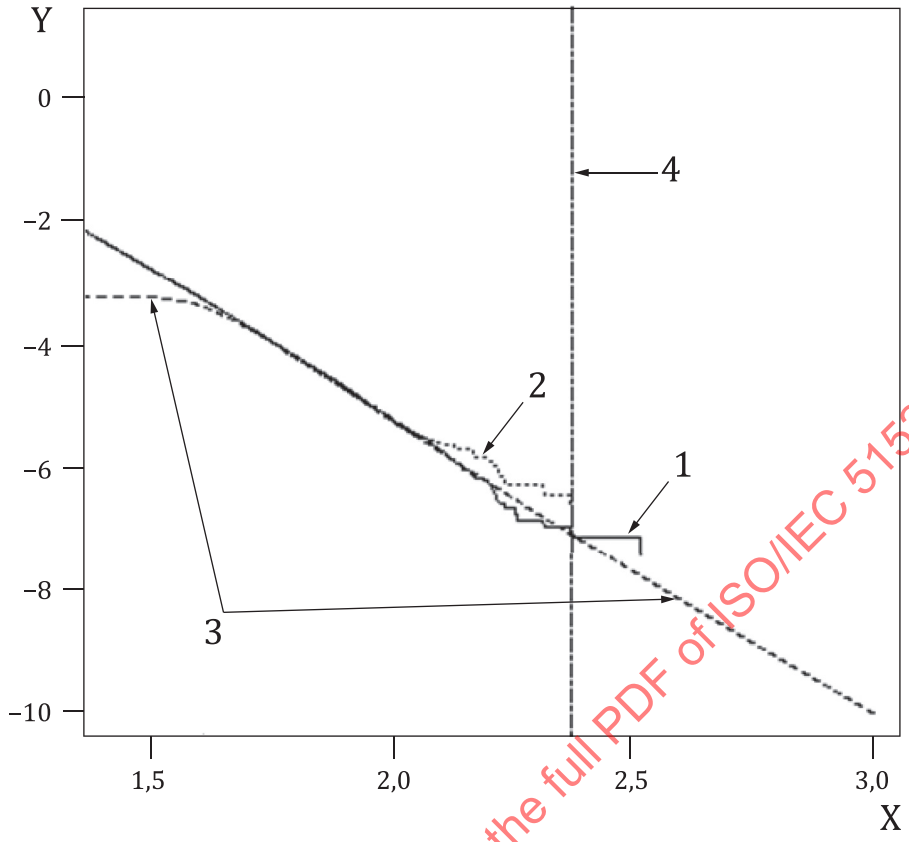
If the fitness of the estimation is good enough (refer to 6.5), go to step 8). Otherwise, go back to step 2) with a different *r* value. If no *r* value gives a good estimation, go back to step 1) and try with a different number of samples in a block, *n*.

8) Obtain the extrapolated FMR from the estimated 1-CDF curve. See Figure 2 and Annex B.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

Choose a point of interest in the estimated 1-CDF curve and report the extrapolated FMR with *m*, *n*, *r* and the interval of confidence at the extrapolated FMR level.



Key

X score
Y log(extrapolated FMR)
1 empirical data (all)
2 empirical data (estimation)
3 rGEV estimation
4 max estimation sample

Figure 2 — Comparison of the empirical 1-CDF and the extrapolated false match rate

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

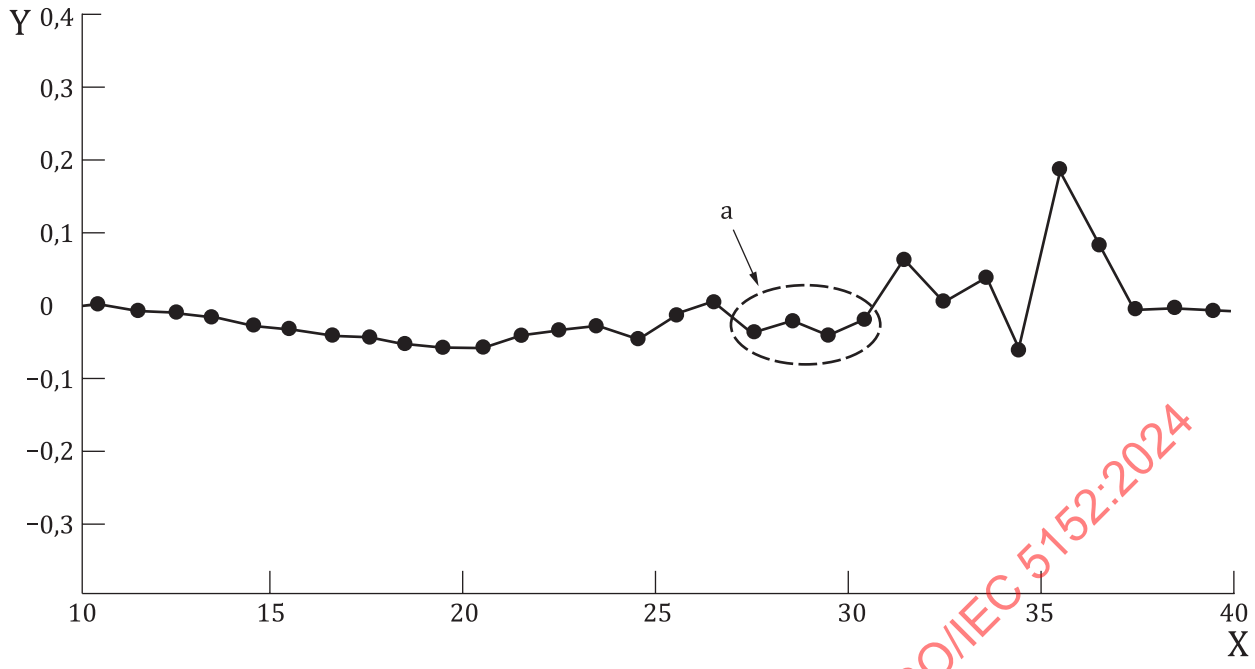
6.4 Generalized Pareto distribution

The GP estimation is based on the GP model as described in Clause A.3. The validation processes are as follows.

1) Determine μ, the location parameter for the GP.

Upon applying the GP model, it is necessary to find an appropriate threshold value μ to extract the extreme scores from the entire score set. The appropriate μ can be obtained by observing the stability of the scale parameter σ and the shape parameter ξ. Figure 3 shows the optimum shape parameters ξ obtained by MLE for corresponding threshold values, plotted in *y* and *x* axis, respectively. The shape parameter ξ is regarded as stable in the circled range and the appropriate threshold μ can be selected from the values close to the lower end of the range. It is also possible to use the scale parameter σ vs threshold graph to find the optimum μ in the same manner.

© ISO/IEC 2024 – All rights reserved



Key

X threshold

Y shape parameter

[a] The values of shape parameter ξ are stable.

Figure 3 — Example of plot of the shape parameter ξ

2) Estimate parameters (σ, ξ) of GP.

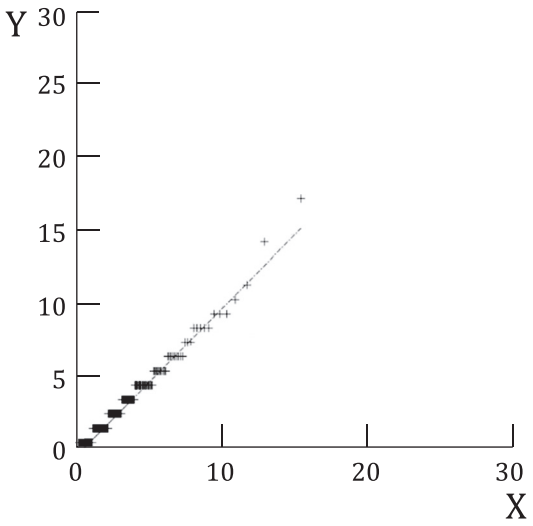
The GP distribution is fitted to the data exceeding the threshold μ, which is selected in the step 1). The scale parameter σ and the shape parameter ξ are then estimated by using the maximum likelihood estimation algorithm.

3) Diagnosis of GP model.

To diagnose whether parameters of an estimated GP are appropriate, Q-Q plots are commonly used for extreme value theory. A diagnostic example using the Q-Q plot is shown in the Figure 4. Figure 4 shows an example of the Q-Q plot between the empirical data and GP distribution.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved



Key

X quantile of GP distribution

Y quantile of empirical data

Figure 4 — Example of Q-Q plot of GP distribution and empirical data

4) Determine parameters of GP model.

If there is a problem with the diagnosis using Q-Q plot, the threshold μ is reselected. For example, in some cases where there is a problem with the Q-Q plot, there can be a clear separation relative to $y = x$.

5) Obtain CDF of comparison scores.

The CDF of the non-mated comparison score is obtained as follows. A value less than the threshold μ is a distribution obtained from the empirical values, and a value more than the threshold μ is a GP. The GP is extrapolated to the extent that there are no actual measurements. Each is rescaled based on the ratio of the number of scores in the measured value, and the CDF is combined to obtain the CDF of the non-mated comparison scores.

Calculate the PDF using these parameters and extrapolate a range with no score.

6) Obtain the extrapolated FMR from the estimated 1-CDF curve. See Annex B.

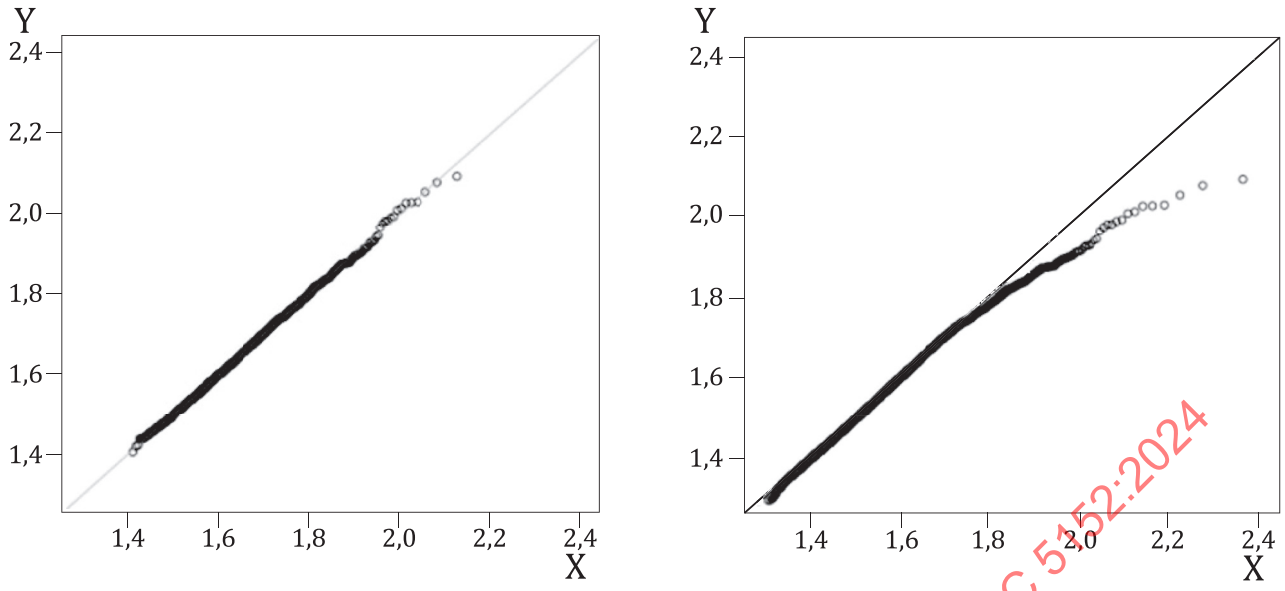
Choose a point of interest in the graph drawn in the step 5) and report the extrapolated FMR with μ, σ and ξ. Similarly, report an upper 95 % confidence interval of extrapolated FMR.

6.5 Evaluation of the fitness of the model

The fitness of the model shall be evaluated by using a Q-Q plot. The Q-Q plot is a quantile-quantile comparison between two distributions, i.e. the estimated data versus the empirical data in this context, and, therefore, the plotted values always increase monotonically. If the two distributions are identical, the Q-Q plot follows the 45 degrees line $y = x$. Since the fluctuations of the top few scores are essentially large, the plots for those scores tend to deviate from the line $y = x$. Therefore, it is important to observe the deviation between the estimation model and the empirical data carefully, especially at the top-right corner of the Q-Q plot. Figure 5 illustrates two Q-Q plot examples, one for a model with good fitness and the other for bad fitness. The horizontal and the vertical axes are for the estimation result and the empirical scores, respectively. If the Q-Q plot is almost on the line $y = x$ throughout the score range [Figure 5 a)], the fitness of the estimation model is regarded as adequate, and the estimation results can be adopted. On the other hand, if the Q-Q plot is not on the line $y = x$ as illustrated on Figure 5 b), the fitness of the model is not good enough and thus the estimation results shall not be adopted. If the Q-Q plot indicates the fitness is poor, the parameters of the estimation models should be reviewed, or otherwise, the evaluator should reconfirm that the dataset is free from errors such as wrong labels.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved



a) Good example **b) Bad example**

Key

X model

Y empirical

Figure 5 — Q-Q plot comparison: a good example and a bad example

6.6 Selection of rGEV and GP

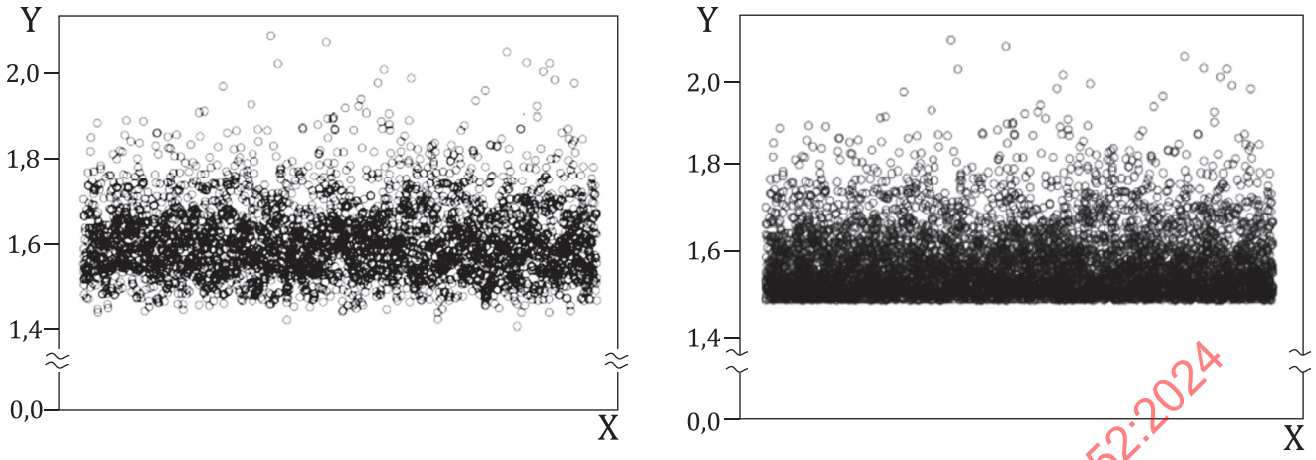
6.6.1 Differences between the two methodologies

It is known that both rGEV and GP estimation methodologies perform well in the extreme value estimation applications in general. Apart from the mathematical models, the major difference between the two estimation methods is the score sampling policy for estimation. The rGEV model utilizes the top r scores of each block and does not have any explicit lower limit for the scores to be used. The number of samples is determined by the parameters r and m. The GP model, however, utilizes all samples above the chosen threshold value μ and thus the number of samples depends on the dataset.

Figure 6 a) and Figure 6 b) visualize the samples used by rGEV and GP, respectively.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved



a) rGEV samples **b) GP samples**

Key

X index

Y score

Figure 6 — Samples used for estimation

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

6.6.2 Features of the two methodologies

Since the rGEV estimation is based on the order statistics, it is more robust against the gradual change of the trend. For example, some biometric modalities tend to produce higher similarity scores as users familiarise themselves with the biometric system. rGEV estimation is based on the extreme scores selected by the order in each block rather than thresholding. This means that the number of scores used for estimation remains the same even if the score distributions become different in the future visits. On the other hand, GP depends on all scores above a fixed threshold and the number of scores used for estimation will be different if the population distribution is different. If the dataset is a concatenation of subsets that are created at different times, e.g. 1st visit + 2nd visit + 3rd visit, then the rGEV model is more suitable as it will keep up with the extreme values of each subset. Note that this feature will be eliminated if the order of samples is randomized in such dataset.

Except for the above-mentioned influence originating from the sampling policy, the two methodologies do not have any significant difference in the context of extrapolated FMR estimation. Evaluators shall use either or both rGEV and GP estimations and report the results together with other information defined in the Clause 8.

7 Performance metrics

The extrapolated FMR shall be prepared in the following manner.

a) Plot the extrapolated FMR on the graph with the likelihood score along the x-axis in the linear scale and the extrapolated FMR along the y-axis in the logarithmic scale.

b) Plot the 95 % confidence interval lines on both sides of the extrapolated FMR plot.

© ISO/IEC 2024 – All rights reserved

c) The 95 % confidence interval shall be obtained by using a Monte Carlo approach. Prepare at least 100 models by applying randomly generated model parameters to the chosen extreme value model or models, i.e. μ,σ and ξ for rGEV and σ and ξ for GP models. The random numbers shall be generated by algorithms designed to reflect the distribution of the profile likelihood of each parameter. Plot the upper and lower 2,5 % percentile points on both side of each extrapolated FMR.

NOTE 1 Reporting the worst value in the confidence interval is in line with the idea of the rule of 3.

d) If extrapolated FMR is to be reported in a single metrics manner, it shall be described in either of the following formats.

1) Report the upper 2,5 % percentile value of the extrapolated FMR at the threshold score of evaluator's choice.

Example:

Extrapolated FMR = $2{,}4 \times 10^{-6}$ or less with 95 % confidence level

2) Report the extrapolated FMR at the threshold score of evaluator's choice with the 95 % confidence interval on both sides.

Example:

Extrapolated FMR = $2{,}1 \times 10^{-6} \pm 0{,}3 \times 10^{-6}$ at 95 % c.i.

Extrapolated FMR = $2{,}1 \times 10^{-6}$ [$1{,}9 \times 10^{-6}$, $2{,}2 \times 10^{-6}$] at 95 % c.i.

e) If confidence levels other than 95 % is selected, the confidence level shall be clearly stated.

NOTE 2 One of such algorithms can be implemented by using revdbayes package of R.

8 Record keeping

Records regarding the methods used to derive performance measures shall be retained, including but not limited to statistical models, parameters, diagnostic diagrams. Enough information shall be kept to enable the evaluation to be repeated under conditions as close as possible to the original.

9 Reporting estimation results

9.1 Reporting one-to-one comparison performance

The performance metrics in Table 1 shall be reported.

Table 1 — One-to-one comparison performance metrics

Metric	Threshold	Reporting	Details of report
Extrapolated FMR	Corresponding threshold	Mandatory	See Clause 7.
FNMR	Corresponding threshold	Optional	See ISO/IEC 19795-1

NOTE FNMR reporting is optional because this document is concerned with application of extreme value techniques to FMR.

9.2 Reporting estimation results

The estimation results in Table 2 shall be reported.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

Table 2 — Estimation results

Estimation details	Reporting	Details of report
Statistical model	Mandatory	Including the statistical model (rGEV, GP etc.) used for estimations.
Parameters	Mandatory	Including the parameters of the statistical model. In the case of rGEV: $\hat{\mu},\hat{\sigma},\hat{\xi}$, block number n In the case of GP: threshold μ, scale parameter σ, shape parameter ξ.
Diagnostic graphs	Mandatory	Include Q-Q plots and other relevant graphs to show the validity of the estimation results. Examples of such graphs are shown in Figure 7.

9.3 Reporting form

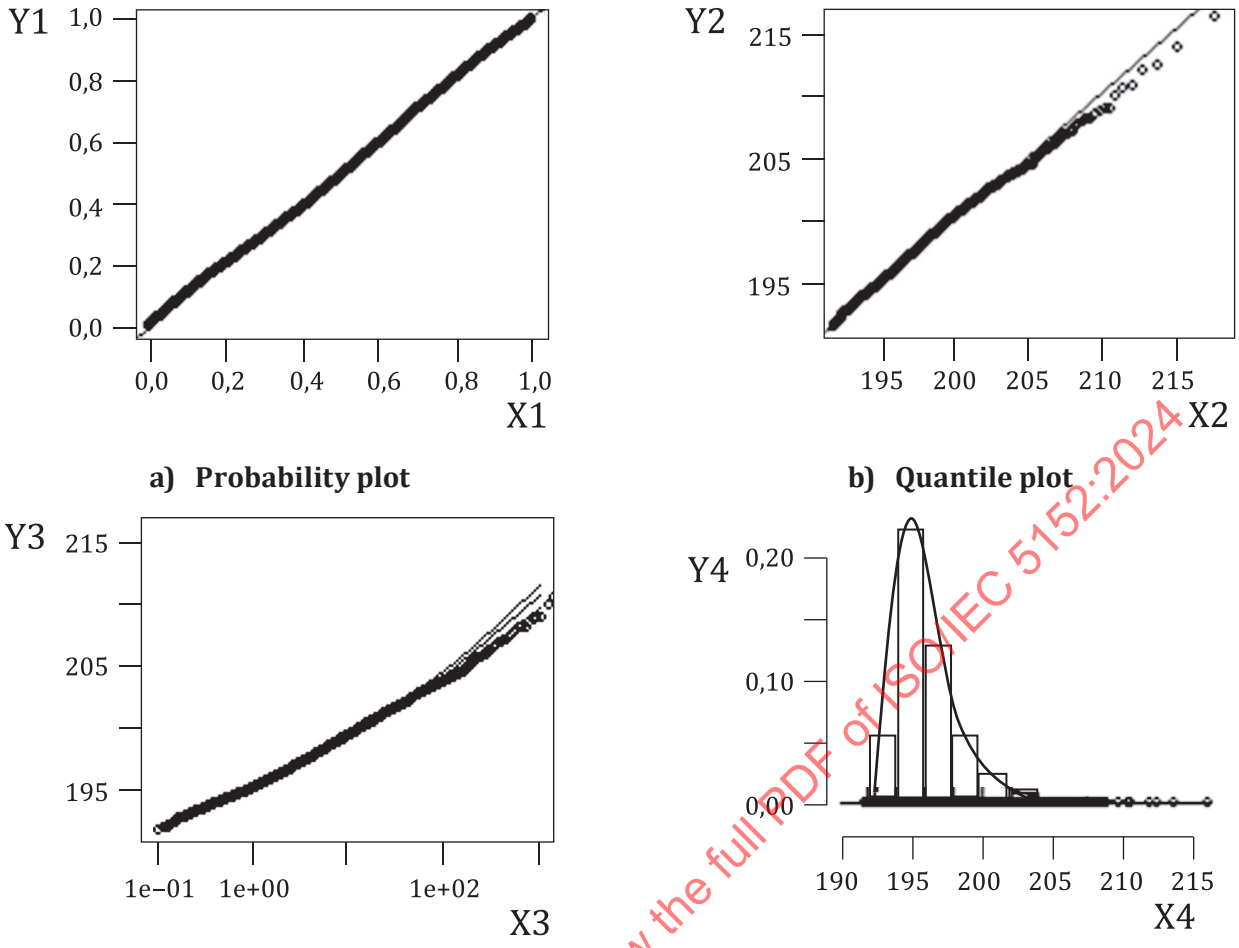
The evaluation results shall be reported according to the reporting form in Table 3.

Table 3 — Reporting form

Value	Threshold	Reporting
Extrapolated FMR	Corresponding threshold	Mandatory
Empirical FMR	Corresponding threshold	Mandatory
Empirical FNMR	Corresponding threshold	Optional
Maximum observed non-mated score	-	Optional
Maximum score with 30 + false match cases	-	Optional

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved



a) Probability plot

b) Quantile plot

c) Return level plot

d) Density plot

Key

X1 empirical
X2 model
X3 return period
X4 z
Y1 model
Y2 empirical
Y3 return level
Y4 f(z)

Figure 7 — Example of diagnostic graphs

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

Annex A
(informative)

Extreme value theory

A.1 Fundamental premises

The extreme value theory is based on the following three fundamental premises.

a) The probability distribution function (PDF) F(x) of the target population is non-degenerate distribution.

In the context of extrapolated FMR estimation, biometric systems that always return a constant score are regarded as degenerated and the score distribution of such systems is called a degenerate distribution. Extreme value theory requires the target population to have a non-degenerate distribution and thus extrapolated FMR can be calculated only for non-degenerate biometric systems.

Degenerate distribution in an n-dimensional Euclidean space is any probability distribution having support on some (linear) manifold of dimension smaller than n. Otherwise the distribution is called non-degenerate.[3]

An example of univariate degenerate distribution is shown in Figure A.1.

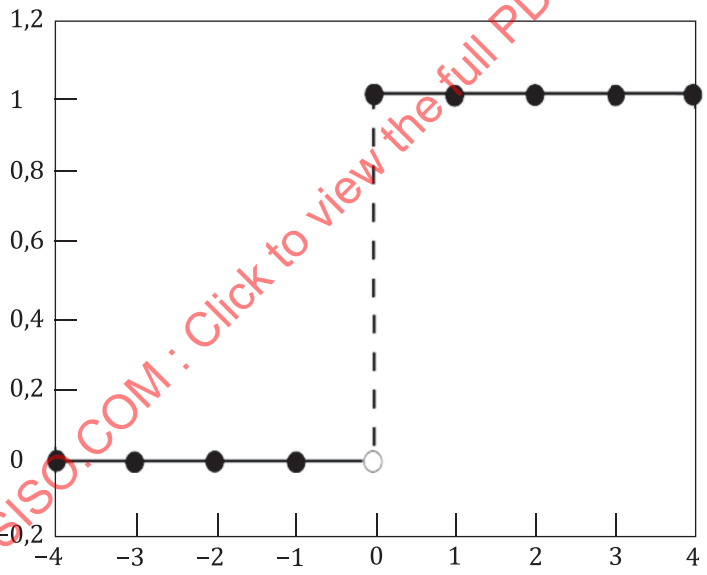


Figure A.1 — An example of univariate degenerate distribution

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

b) The PDF F(x) belongs to the domain of attraction of an extreme value distribution G(z), i.e. F ∈ D(G).

In the context of extrapolated FMR estimation, if the right-hand tail of the likelihood score histogram asymptotically approaches to zero, the probability distribution function the histogram follows is practically considered to belong to the domain of attraction. Figure A.2 is an example of PDF that belongs to the domain of attraction.

© ISO/IEC 2024 – All rights reserved

Let $X_1, X_2, \ldots X_n$ be independent identically distributed random variables. Let P be probability distribution. Let $F(x) = P(X_i \leq x)$, $i = 1,2,\ldots$ be its identical distribution and Z_n be extreme statistic defined as in Formula (A.1).

$$Z_n := \max\{X_1, X_2, \ldots X_n\} = \max_{1 \leq i \leq n} X_i \tag{A.1}$$

Suppose that a random variable Z follows a non-degenerative distribution G and that there exist constants $a_n > 0$ and $b_n \in R$ such that Formula (A.2) and Formula (A.3) are satisfied.

$$\frac{Z_n - b_n}{a_n} \xrightarrow{d} Z, n \to \infty \tag{A.2}$$

and

$$P\left(\frac{Z_n - b_n}{a_n} \leq x\right) \to P(Z \leq x) = G(x) \tag{A.3}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

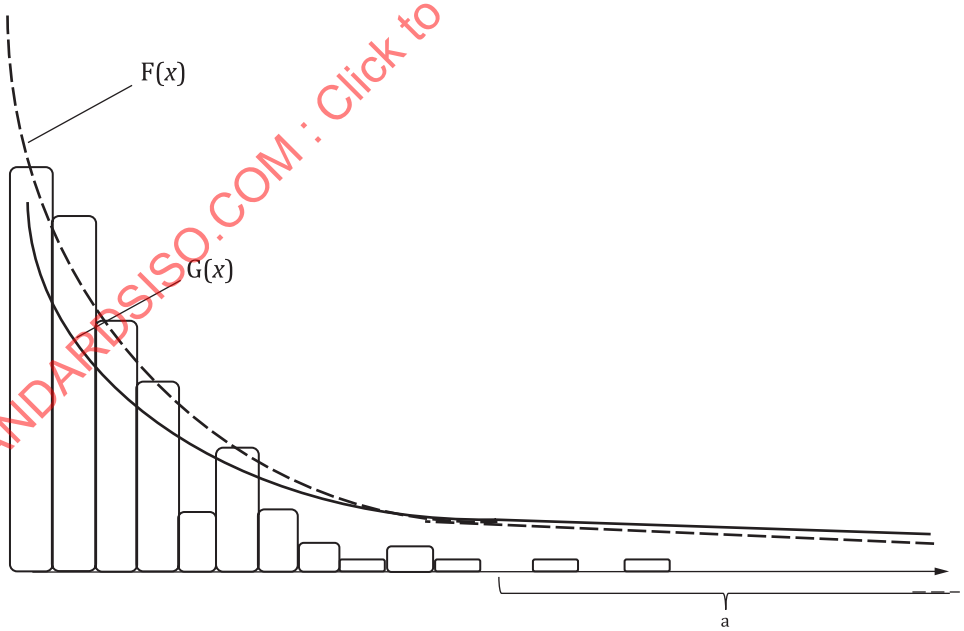
Here, $G(x)$ is called an extreme value distribution.

Since X_i and Z have distribution functions F and G, respectively, F is said to be attracted to G. Also, F is said to belong to the domain of attraction of G, which is denoted as $F \in D(G)$.

c) Each attempt is regarded as independent and identically distributed (i.i.d.).

Thus i.i.d. random variables are independent variables all having the same distribution. The most common situation involving i.i.d. random variables arises when a random sample of observations is taken from a single population[3].

If the score distribution of the target biometric system does not satisfy the above fundamental premises, the methodology defined in this document shall not be used.



a F converges to G in the extreme value domain.

Figure A.2 — An example PDF that belongs to the domain of attraction

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

A.2 Generalized extreme value distribution

A.2.1 Preparation

Generalized extreme value distribution, also known as GEV, is an extreme value distribution model widely used to estimate the maximum value expected in a certain period of time in the future. The statistical estimation of FMR is made by using the top largest r values in a unit period (or a "block") as extreme values. This statistic model is called rGEV model.

Suppose each of non-mated attempts is independent and identically distributed (i.i.d.), and let x_1, x_2, ... x_n denote the scores for attempt #1 to #n. Then, x_i for positive integer $i(1 \le i \le n)$ can be considered as a probabilistic variable and, therefore, the identical distribution F(x) is defined according to Formula (A.4):

$$\mathrm{F}(x) = \mathrm{P}(x_i \le z), i = 1,2,\ldots,n \tag{A.4}$$

where P() is the probabilistic distribution function (PDF) of the non-mated score distribution.

Sort the non-mated score series x_i in the ascending order and prepare the order statistics series $x_{(i:n)}$ using Formula (A.5).

$$x_{(1:n)} \le x_{(2:n)} \le \ldots \le x_{(n:n)} \tag{A.5}$$

The extreme value set, or extreme statistics, Z_n is called a block maxima set, which is defined as Formula (A.6):

$$Z_n := z_{(n:n)} = \max\{x_1, x_2, \ldots, x_n\} = \max_{1 \le i \le n} x_i \tag{A.6}$$

Also, the order statistics consisting of the top largest r values Z_{nr} is defined as Formula (A.7):

$$Z_{nr} := \left\{ z_{(n:n)}, z_{(n-1:n)}, z_{(n-2:n)}, \ldots, z_{(n-r+1:n)} \right\} \tag{A.7}$$

Note that the block maxima Z_n is a special case of Z_{nr} when r = 1, namely, Z_{n1}.

A.2.2 The rGEV Model

Suppose n_0 non-mated scores are available for the FMR estimation. Choose positive integers m and n such that $mn = n_0$, where m is the number of blocks and n is the number of samples in each block.

Here, consider a non-mated sample set Z, which consists of the top r scores from each block, namely, as in Formula (A.8):

$$Z = \{Z_{nr}^i, 1 \le i \le m, r \ll n\} = \{Z_{nr}^1, Z_{nr}^2, \ldots, Z_{nr}^m\} \tag{A.8}$$

where Z_{nr}^i is a non-mated score set Z_{nr} created from the i-th block.

The block size m shall be chosen to be large enough so that some extreme values are included in each block. The r value shall be chosen to be much smaller number than n as the top r similarity scores in the block are regarded as extreme values. Both m and r (and consequently n), shall be carefully selected by observing the diagnostic graphs described in 6.3. Figure A.3 shows an example of blocks for rGEV estimation.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

The rGEV model is fitted to the scores $z \in Z$ extreme values and the model parameters are typically estimated by using maximum likelihood estimation. Generalized extreme value distribution GEV(μ,σ,ξ): = G(z) is mathematically described as in Formula (A.9):

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\} = G_\xi\left(\frac{z-\mu}{\sigma}\right) \tag{A.9}$$

where G_ξ is defined by Formula (A.10)

$$G_\xi(z) = \exp\left[-(1+\xi z)_+^{-1/\xi}\right] \tag{A.10}$$

and $(a)_+$ is defined by Formula (A.11)

$$(a)_+ = \max\{a, 0\} \tag{A.11}$$

The parameters are defined as follows:

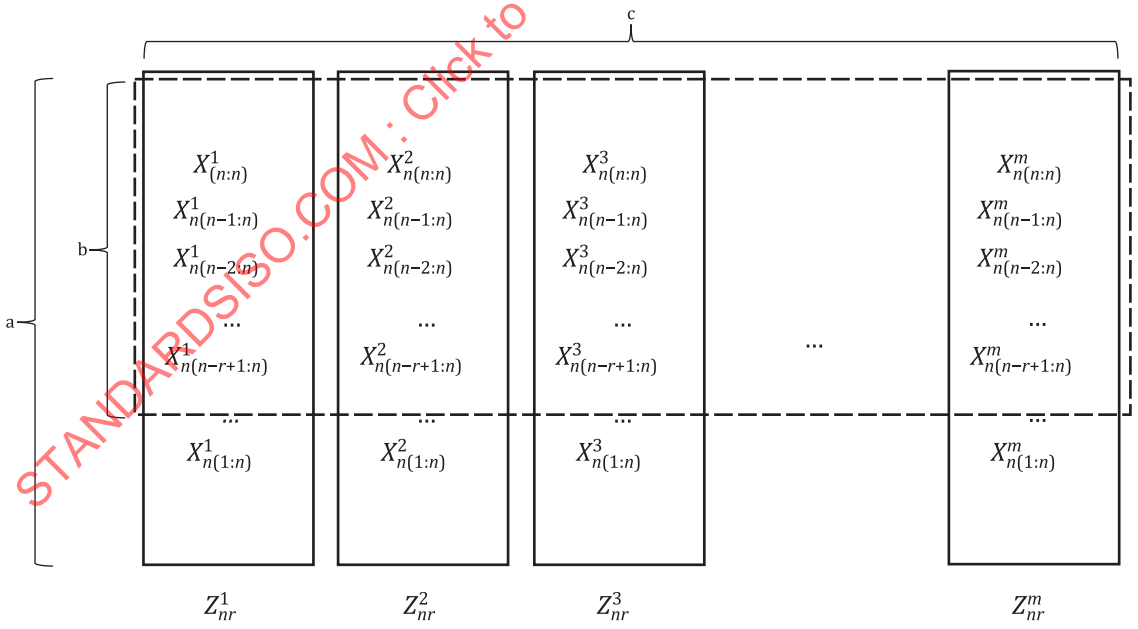
μ = location

σ = scale

ξ = shape

These parameters determine the estimated PDF G(z), which is used to obtain the extrapolated FMR in the extreme value domain. In other words, finding G(z) that best describes the real PDF in the extreme value domain resolves itself into an estimation of the optimized parameters μ, σ and ξ for given n and r.

In the context of extrapolation-based FMR estimation, biometric scores that are extraordinarily large for non-mated pairs are regarded as extreme values and the above-mentioned "block" is regarded as a unit number of non-mated attempts, which are mutually exclusive subsets sampled from the entire set of non-mated scores.



a The size of each block: n.

b The set consists of the top r scores of each block: Z.

c The number of blocks: m.

Figure A.3 — Blocks for rGEV estimation

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

A.3 Generalized Pareto distribution

The appearance of large data depends on the right tail of the population distribution. Therefore, the peak over threshold method is used to estimate the right tail of the population distribution using data exceeding a sufficiently large threshold. A statistical model and analysis method for threshold excess data are described.

The threshold excess data $\{x_1, x_2, \ldots, x_n\}$ are measured values of random variables which follow the generalized Pareto distribution independently and identically. The generalized pareto distribution[4] has cumulative distribution function as indicated by Formula (A.12):

$$F(x) = \begin{cases} 1 - \left(1 + \xi \frac{x-\mu}{\sigma}\right)_+^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right), & \xi = 0 \end{cases} \quad \text{(A.12)}$$

where:

σ is the scale parameter for generalized Pareto distribution;

ξ is the shape parameter for generalized Pareto distribution;

μ is the location parameter for generalized Pareto distribution;

These parameters can be obtained by using the maximum likelihood method. Generally, maximum likelihood estimates are easily obtained by using software R[5].

Here, in order to perform data analysis with high accuracy in the generalized Pareto distribution, it is necessary to select the optimum threshold θ. The smaller the threshold, the more data exceeds the threshold, but the fit of the generalized Pareto distribution is poor. Therefore, the variance of the estimation becomes small, but the bias becomes large. On the other hand, the larger the threshold, the smaller the data, but the better the fit of the generalized Pareto distribution. The deviation of the estimate becomes smaller, but the variance becomes larger. In this method, there are two types of errors: errors due to the fact that the generalized Pareto distribution is an approximate distribution to be applied to the threshold excess data, and errors due to estimation. The threshold is selected taking these factors into consideration.

Some threshold selection methods are based on the properties of the following generalized Pareto distributions. Let x denote data exceeding the value u and let x_{max} denote the maximum value of the data.

For $\mu < x_{max}$, a generalized Pareto distribution is fitted to the data $\{x_\mu - \mu\}_{i=1}^{n_\mu}$ exceeding μ to obtain maximum likelihood estimates σ and ξ. By changing the value μ, $\left(\mu, \hat{\sigma}^*\right)$ and $\left(\mu, \hat{\xi}\right)$ are plotted. Here $\hat{\sigma}^* = \hat{\sigma}_\mu - \hat{\xi}\mu$ is an estimate of the correction factor. When the estimated values $\hat{\sigma}^*$ and $\hat{\xi}$ are both assumed to be constant to the right of a certain value in these two plots, the minimum of these values μ is determined as the threshold.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

Annex B
(informative)

Examples applied to multiple modality datasets to demonstrate the validity of the methodology

B.1 Overview

This annex provides examples applied to biometric datasets[6] representing different modalities. The purpose of this annex is to provide basic information for developing an understanding of how to use this estimation methodology with extreme value theory. This annex only shows application examples in comparison score data of various modalities, and does not show characteristics of each modality.

In ISO/IEC 19795-1, accuracy metrics such as FMR or FNMR are calculated based on either the rule of 30 or the rule of 3. The rule of 30 requires the test size to be sufficiently large so the number of errors at the desired threshold exceeds 30. The rule of 3 gives the minimum test size needed for the desired FMR or FNMR and requires a perfect match in the test. These accuracy metrics are statistical estimations based on all available comparison scores with pre-determined confidence levels.

On the other hand, extrapolated FMR is aimed to estimate the false match rate even if the available test size is too small to satisfy the above-mentioned rules. The extrapolated FMR is a statistical estimation based only on the comparison scores beyond a certain threshold value, or those within the top r scores per block. The confidence level of the estimation is to be reported together with the corresponding confidence intervals.

B.2 Datasets and test protocol

Table B.1. shows the details of the non-mated comparison score data used in the application of the estimation methodology. As a protocol, the test data was created by randomly extracting a small number of data (10 % or 20 %) from all score data. rGEV and GP were applied to these test data, and the fitness of rGEV and GP were evaluated by the Q-Q plot. If there were no problems in the fitness, the extrapolated FMR was calculated. The extrapolated FMR was compared to the FMR measured from all score data. When the approximation using rGEV or GP was not appropriate for 10 % of the test data, the evaluation was made by increasing the number of scores to 20 %.

Table B.1 — Score dataset

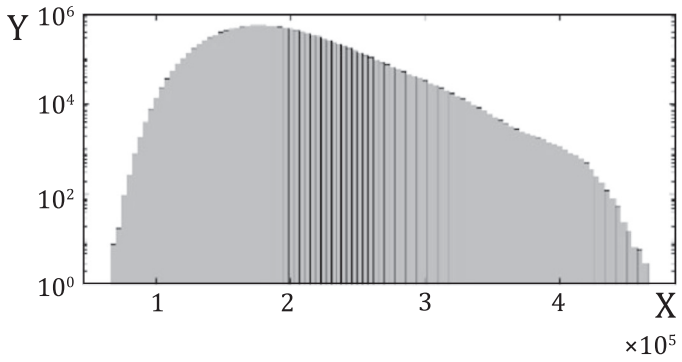
Biometrics	Comparison score dataset	Number of comparisons
Face	Idiap BIOSCOTE 2014, Face Recognition Grand Challenge (FRGC) v2.0[7]	4 million
Gait	Osaka University, Gait Energy Image (GEI)[8][9]	13,73 million

B.3 Application examples

Figure B.1 through B.5 shows the rGEV and GP results for FRGC. Figure B.1 shows a histogram of the comparison score of the FRGC. Twenty percent of the comparison scores (approximately 800 000 scores) were randomly extracted from all score data and applied to the rGEV and GP. Figure B.2 shows the Q-Q plot when approximated by the rGEV, while Figure B.3 shows the comparison between extrapolated FMR estimated with rGEV and measured FMR. Figure B.4 and B.5 show similar results for GP. Table B.2 shows rGEV and GP parameters for FRGC. The Q-Q plots are located near $y = x$, and the right tail of the comparison score distribution can be well approximated by the rGEV and GP. With the FRGC score, the FMR for all scores can be estimated from 20 % of the scores.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved

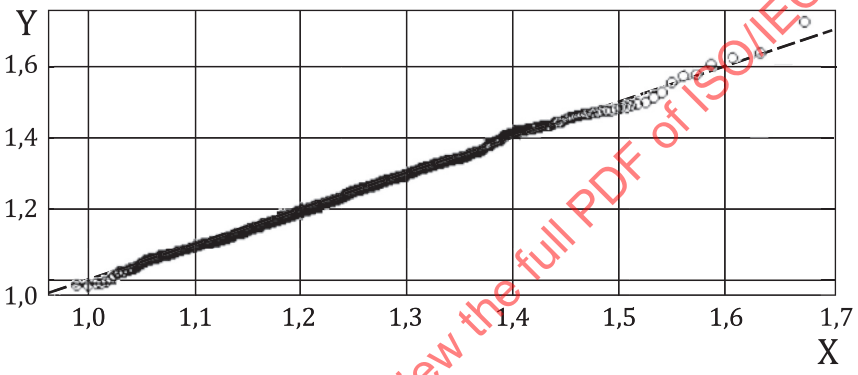


Key

X score

Y frequency

Figure B.1 — FRGC example: histogram of non-mated comparison scores



Key

X rGEV

Y measured score quantiles

Figure B.2 — FRGC example: Q-Q plot of rGEV

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 5152:2024

© ISO/IEC 2024 – All rights reserved