Improving Biometric and Forensic Technology: The Future of Research Datasets
January 26-27, 2015

Hari Iyer

Statistical Engineering Division

Information Technology Laboratory

hari@nist.gov

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Questions Considered

ISO/IEC 19795-1:
Sufficient samples shall be collected per test subject so that the total number of attempts exceeds that required by the Rule of 3 or Rule of 30 as appropriate

- What is the **RULE OF 3** and how is it applied when determining sample sizes?

- What is the **RULE OF 30** and how is it applied when determining sample sizes?

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

- A new matching algorithm is used to perform 1000 non-mated comparisons and **no false matches** were found. What can we say about the **false match rate** based on this data?

- Suppose we have a matching algorithm that **we believe has a false match rate of 0.1 % or less**. If this is really the case, what is the **minimum number of tests** that must be run to even have a chance of statistically demonstrating this at a 95% confidence level ?

Rule of THREE is useful for answering such questions

NIST
**National Institute of
Standards and Technology**
U.S. Department of Commerce

MOTIVATION

- Wish to estimate the proportion **p** of prints in a large database that have a particular interesting feature present, e.g. a Trifurcation

- Rather than go through the entire database, we wish to examine a random subset of prints. How many prints should be examined to assure us (with 90% confidence) that the true proportion **p** is within 30% of the sample proportion p_{est} ?

Rule of THIRTY can be useful for answering such questions.

Selected Publications on the Rule of 3

Thomas A. Louis (1981),
Confidence Intervals for a Binomial Parameter after Observing No Successes
The American Statistician, Vol. 35, No. 3, p. 154

Hanley, J. A., and Lippman-Hand, A. (1983),
"If Nothing Goes Wrong, Is Everything Allright? Interpreting Zero Numerators,"
Journal of the American Medical Association, 249(13), 1743-1745.

B. D. Jovanovic and P. S. Levy (1997),
A Look at the Rule of Three
The American Statistician, Vol. 51, No. 2, pp. 137-139

Gerald van Belle and Steven P. Millard (1998),
STRUTS: Statistical Rules of Thumb
Technical Report, University of Washington.

A. J. Mansfield and J. L. Wayman (2002)
Best Practices in Testing and Reporting Performance of Biometric Devices (2002)
Version 2.01, NPL Report CMSC 14/02

NIST
**National Institute of
Standards and Technology**
U.S. Department of Commerce

Rule of 2.3, Rule of 3, Rule of 4.6, Rule of 6.9

90%	95%	99%	99.9%

Hanley, J. A., and Lippman-Hand, A. (1983),
"If Nothing Goes Wrong, Is Everything Allright? Interpreting Zero Numerators,"
Journal of the American Medical Association, 249(13), 1743-1745.

To be 95% confident that our interval estimate of the long-run risk is correct, a simple rule (of unknown origin) can be applied.⁴ This "rule of three" states that if none of n patients shows the event about which we are concerned, we can be 95% confident that the chance of this event is at most three in n (ie, $3/n$). In other words, the upper 95% confidence limit of a $0/n$ rate is approximately $3/n$. (This approximation is remarkably good: when n is larger than 30, the rule of three agrees with the exact calculation to the nearest percentage point; below 30, it slightly overestimates the risk, but then the

For a 99% confidence interval. the corresponding shortcut is a "rule of 4.6" (ln $0.01 = -4.6051...$), while for 99.9% confidence one uses a "rule of 6.9."

Thomas B. Newman, MD, MPH (1995)
If Almost Nothing Goes Wrong, Is Almost Everything All Right? Interpretating Small Numerators. --- **JAMA**

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

EXAMPLE

Rule of 3:

Suppose p = probability of an event occurring in a single trial
Conduct N (iid) trials. NO EVENTS OCCUR.
Then we can be 95% confident that the value of p does not exceed 3/N.

Application: No FALSE MATCHES are found In 1000 independent non-mated searches.
We can be 95% confident that the false match rate p does not exceed 3/1000 = 0.3 %

NIST
**National Institute of
Standards and Technology**
U.S. Department of Commerce

EXAMPLE

- Suppose we believe that a new algorithm has a **false match rate of at most 0.1%.** What is the minimum number of comparisons needed to establish this statistically?

- If we conduct N searches, the *most favorable outcome* we could have is NO FALSE MATCHES.

- Even in this most favorable case we could only demonstrate (at 95% confidence level) that p is at most 3/N.

- So we would like 3/N to be smaller or equal to 0.001.

- Thus the **minimum number** N of (iid) searches needed to demonstrate our claim with 95% confidence is obtained by solving **3/N = 0.001, i.e., N = 3,000**.

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Example

Rule of 6.9 (99.9% confidence level)
Suppose p = probability of an event occurring in a single trial
Suppose, in N (iid) trials, NO EVENTS OCCUR.
Then we can be 99.9% confident that the value of p does not exceed 6.9/N.

Application:

No FALSE MATCHES are found In 1000 independent non-mated searches. We can be 99.9% confident that the false match rate p does not exceed 6.9/1000 = 0.69 %

Application:

Suppose we believe that a new algorithm has a false match rate of at most 0.1%. The minimum number N of iid searches needed to demonstrate this claim with 99.9% confidence is obtained by solving the equation   6.9/N = 0.001,   i.e., N = 6900.

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Exact Formulae

- Confidence level = C %

- Error Rate *alpha* = 1 – C/100   (so *alpha* = 0.05 for C = 95)

- Suppose N (iid) tests resulted in ZERO false matches.

- Then, with C% confidence we can say
  the true false match rate

$$\textbf{p is no greater than } \textbf{\textit{1-alpha}}^{\textbf{1/N}}$$

EXAMPLE:

- Suppose, in N = 10 (iid) trials, NO FALSE MATCHES OCCUR.

- C = 95%,   *alpha* = 1- 95/100 = 0.05.

- Then we can be 95% confident that the actual false match rate does not exceed
$$1\text{-}0.05^{1/10} = 0.26 = 26\%.$$

- The rule of 3 would give 3/10 = 30%

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Exact Formulae

- For 99% confidence, C = 99, alpha = 1- 99/100 = 0.01.

- Suppose, in N = 10 identical, independent trials, NO FALSE MATCHES OCCUR.

- Then we can be 99% confident that the actual false match rate does not exceed

$$1-0.01^{1/10} = 0.37 = 37\%.$$

- But the rule of 4.6 would give 4.6/10 = 0.46 = 46%

  NOTE:

  -ln(alpha) = 2.3,   3,   4.6,   6.9
  for C = 90,   95,   99,   99.9

- **The Rule of 3 works satisfactorily for N greater than or equal to 30.**

Rule of 30

Doddington et. al. (2000), Speech Communication (31), 225-254

The NIST speaker recognition evaluation – Overview, methodology, systems, results, perspective

George R. Doddington [a,b], Mark A. Przybocki [b], Alvin F. Martin [b,*], Douglas A. Reynolds [c]

2.5.2.2. *The rule of 30.* In determining the required size of a corpus, a helpful rule is what might be called *"the rule of 30"*. This comes directly from the binomial distribution, assuming independent trials. Here is the rule:

To be 90% confident that the true error rate is within ±30% of the observed error rate, there must be at least 30 errors.

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

How Not To Interpret "Rule of 30"

> *2.5.2.2. The rule of 30.* In determining the required size of a corpus, a helpful rule is what might be called *"the rule of 30"*. This comes directly from the binomial distribution, assuming independent trials. Here is the rule:
>
> To be 90% confident that the true error rate is within ±30% of the observed error rate, there must be at least 30 errors.

- This rule seems to say that you should keep on testing until you get at least 30 errors.
- If one does that, the sample size N becomes random and one does not know the required sample size up front.

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Other References to the RULE OF 30

A. J. Mansfield and J. L. Wayman (2002)
Best Practices in Testing and Reporting Performance of Biometric Devices (2002)
Version 2.01, NPL Report CMSC 14/02

James Wayman, Anil Jain, Davide Maltoni and Dario Maio (Eds) (2005)
Biometric Systems: Technology, Design and Performance Evaluation
Chapter 9, Large-Scale Identification System Design
Herve Jarosz and Jean-Christophe Fondeur

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

How to Use the Rule of 30

- Suppose the performance goal for a new algorithm is a 1% false non-match rate.

- Expected number of false non-matches = N x 0.01

- We would like this number to be equal to 30.

- So solve N x 0.01 = 30.   Get N = 3,000.

- Suppose the performance goal for a new algorithm is a 0.1% false match rate.

- So solve N x 0.001 = 30.   Get N = 30,000.

- IID trials is a key assumption. When the trials are based on reusing enrolees the rule cannot be mathematically justified.
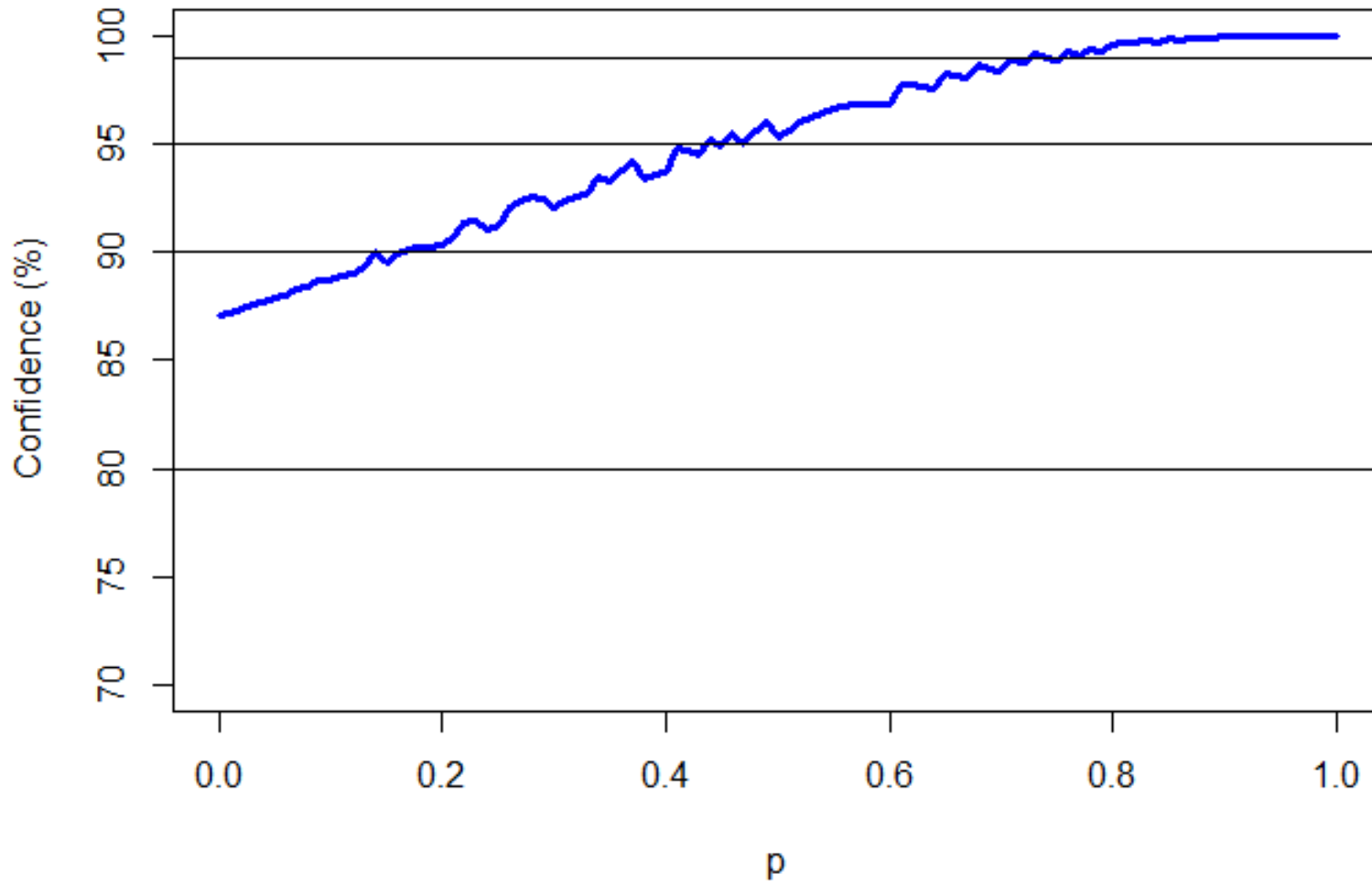
NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Extensions of the Rule of 30

Percent Relative Error

Confidence

	30%	20%	10%	5%	1%
80%	18	43	166	660	16500
90%	**30**	73	276	1090	27100
95%	50	103	386	1550	38500
99%	92	187	680	2650	66000

The numbers in the table are approximate solutions to a nonlinear equation. Exact values will require some additional computations but this is quite easily accomplished.

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Performance Graph for "RULE OF 30"

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

THANKS