

3GPP TR 34.901 V3.0.0 (2003-06)

Technical Report

**3rd Generation Partnership Project;
Technical Specification Group Terminals;
Test Time Optimisation based on statistical approaches;
Statistical theory applied and evaluation of
statistical significance
(Release 1999)**



The present document has been developed within the 3rd Generation Partnership Project (3GPPTM) and may be further elaborated for the purposes of 3GPP.

The present document has not been subject to any approval process by the 3GPP Organizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPP only. The Organizational Partners accept no liability for any use of this Specification. Specifications and reports for implementation of the 3GPPTM system should be obtained via the 3GPP Organizational Partners' Publications Offices.

Keywords

Testing, UMTS

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

<http://www.3gpp.org>

Copyright Notification

No part may be reproduced except as authorized by written permission.
The copyright and the foregoing restriction extend to reproduction in all media.

© 2003, 3GPP Organizational Partners (ARIB, CWTS, ETSI, T1, TTA, TTC).
All rights reserved.

Contents

1	Scope	5
2	References.....	5
3.	Definitions, symbols and abbreviations	5
3.1	Definitions.....	5
3.2	Symbols	5
3.3	Abbreviations	5
3.4	Equations.....	5
4	Introduction.....	6
5.	Definitions of distribution functions and parameters to be used.....	6
5.1	Properties of the Poisson distribution.....	7
5.2	Equivalence between Poisson distribution and Chi Square distribution	7
5.3	Confidence interval	12
5.4	Application of the confidence interval to decide the outcome of the test	13
5.5	Test time reduction	14
5.6	Calculation of the intersection coordinates	16
5.7	Wrong decision risk F	17
6	F to D conversion in BER BLER tests	17
6.1	Conversion F to D.....	17
6.2	Introduction.....	17
6.3	The simulation procedure	18
6.3.1	Equal pass and fail probability	18
6.3.2	Unequal pass and fail probability	19
6.4	Result of the simulation: M-F array.....	20
7.	Definitions, symbols and abbreviations	20
7.1	Definitions.....	20
7.2	Symbols	21
7.3	Abbreviations	21
7.4	Equations.....	21
8	Introduction.....	21
9	Statistical characteristics of testing processes	22
9.1	Exact probability distribution functions.....	22
9.2	Approximation with continuous distributions.....	23
9.3	Simple test criteria and significance level	24
10	Early Pass/Fail termination of testing.....	24
10.1	Framework of test method	24
10.2	Early Pass/Fail criteria	25
10.3	Estimating the test time	31
10.4	Minimizing the test time	33
10.4.1	Truncation.....	33
10.4.2	Decimation.....	35
10.4.3	Composite criteria	35
11	Accuracy of Continuous Distributions	36
Annex A:	Change history.....	40

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
 - 1 presented to TSG for information;
 - 2 presented to TSG for approval;
 - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

The present document describes the statistical theory and concepts applied in the conformance test of the user equipment (UE) so as to improve test speed.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TS 34.121: "Terminal Conformance Specification; Radio transmission and reception (FDD)".

[2] 3GPP TS 34.122: "Terminal Conformance Specification; Radio transmission and reception (TDD)".

3. Definitions, symbols and abbreviations

Definitions, symbols, abbreviations and equations used in the present document are listed in TR 21.905 [5] and TR 25.990 [6].

3.1 Definitions

For the purposes of the present document, the following additional terms and definitions apply.

3.2 Symbols

For the purposes of the present document, the following symbols apply:

[...] Values included in square bracket must be considered for further studies, because it means that a decision about that value was not taken

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

BER	Bit Error Ratio
BLER	Block Error Ratio
DUT	Device under Test

3.4 Equations

Void.

4 Introduction

This technical report includes two distinct approaches made to determine total test time optimisation. For ease of understanding they are just referred to here as the first and second approach. Furthermore the two approaches differ a little in that they use slightly different terminology.

4.1 First approach

The first approach is found in clauses 5 and 6. It reflects TS 34.121 in that the symbols, abbreviations and equations are consistent with TS 34.121.

4.2 Second approach

The second approach is found in clauses 7 to 11 and does not directly reflect TS 34.121 although it does use the existing theory from TS 34.121. The difference is that it refines the theory and derives further approaches for test time reduction. Some of the symbols, abbreviations and equations have local meaning and these are identified in clause 7.

5. Definitions of distribution functions and parameters to be used

Summary

5.1 Properties of the Poisson Distribution

Description of a statistical experiment by a distribution function and basic characteristics of the distribution.

5.2 Equivalence between Poisson Distribution and Chi Square Distribution

Here it is shown, that both distributions are equal. Just the form is different. On the other hand there are two inverse cumulative operations. One of them is useful for our purpose.

5.3 Confidence interval

Introduction into the notion.

5.4 Application of the confidence interval to decide the outcome of the test

Using the notion of the confidence interval, we calculate the early pass and early fail limit.

5.6 Test time reduction

Using 5.4 the outcome of the test is connected with two qualities, a good one and a worse and variable one. Introducing the bad DUT factor M , the quality of the test is now uniform and test time is further reduced.

5.7 Calculation of the intersection coordinates (maximum number of sample and the normalized test limit)

Calculus for intersection co-ordinates of the early pass and early fail limit.

5.8 Wrong decision risk F

5.1 to 5.6 applies the wrong decision risk for a single test step D . However it is desirable to have a predefined wrong decision risk for the entire test F . The approach to derive F from D this is explained here.

5.1 Properties of the Poisson distribution

NOTE: The following text is expressed in terms of BER=bit error ratio. However it can be used for BLER (Block error ratio) as well. Even for 1- Success Ratio, used in RRM delay tests, the theory can be used.

With a finite number of samples (**ns**), the final bit error ratio **BER** cannot be determined exactly.

Applying a finite **ns**, we measure a number of errors (**ne**).

ne/ns =ber is the preliminary bit error ratio.

In a single test we apply a predefined number of samples **ns** and we measure a number of errors (**ne**). **ne** is connected with a certain differential probability in the Poisson distribution. We don't know the probability and the position in the distribution conducting just one single test.

Repeating this test infinite times, applying repeatedly the same **ns**, we get the complete Poisson distribution. The average number of errors is **NE**. **NE/ns** is the final **BER**.

Poisson Distribution:

$$dpois(ne,NE)=(NE^{ne}/ne!)e^{-NE} \quad (1)$$

TOL:= 10^{-10}
 e.g. : ns := 500 BER:=0.01 NE:=ns·BER ne := 1,2..50

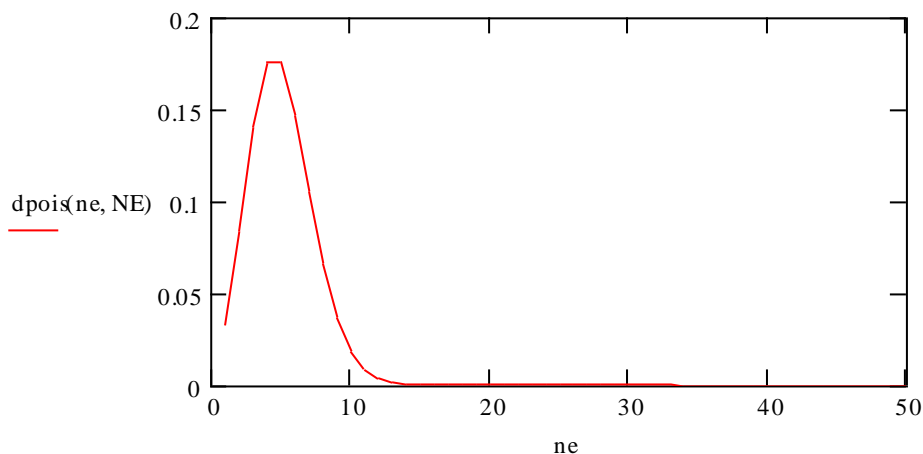


Figure 5-1: Example of Poisson distribution curve

The Poisson distribution has the variable **ne** and is characterised by the parameter **NE**.

Real probabilities to find **ne** between two limits are calculated by integrating between such limits.

Note: The Poisson distribution is an approximation: Independent error occurrence is described by the binomial distribution. If the BER approaches 0 the Poisson distribution approximates the binomial distribution.

5.2 Equivalence between Poisson distribution and Chi Square distribution

The experiment, the Poisson distribution is based on, is having observed a certain number of samples (**ns**), the number of events (**ne**) is counted to calculate the ratio **ne/ns**.

The experiment, the Chi Square distribution is based on, is having observed a certain number of events (**ne**), the number of samples (**ns**) is counted to calculate the ratio **ne/ns**.

Poisson and Chi Square are valid only if $ne \ll ns$

TOL:=10⁻¹⁰ ns := 50 BER:=0.01 NE:=ns·BER ne := 0, 1..50

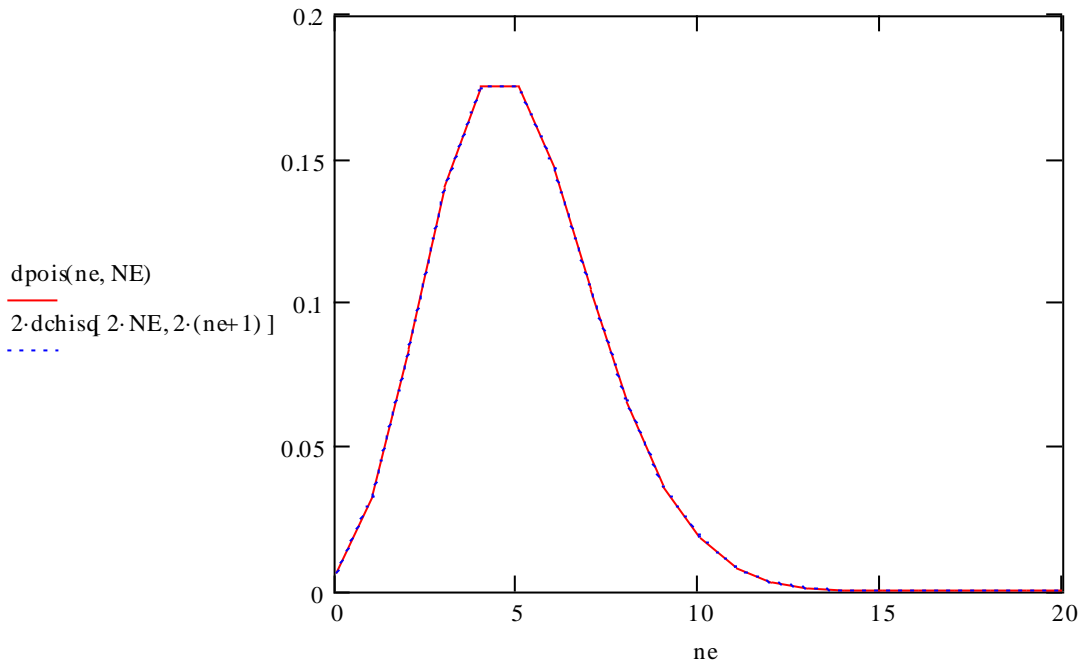


Figure 5-2: Comparison of Chi-Squared and Poisson distribution

The dotted blue function is the Chi-squared distribution, using the parameters of the Poisson distribution. We scaled, offset and changed the interpretation of variable and parameter such that both distributions match. The Poisson distribution is a discrete distribution. Such scaled the Chi Squared distribution interpolates the Poisson distribution exactly for all NE (degenerated for NE=ne=0).

The experiment of the Chi Square distribution is always terminated by an event,

In contrast the experiment of the Poisson distribution almost never is terminated by an event because of $ne/ns \rightarrow 0$. This explains that the Poisson distribution needs one event more, to equal in its form the Chi Square distribution

$$2 \cdot dchisq(2 \cdot NE, 2 \cdot ne) = dpois(ne-1, NE) \text{ describes the experiment, terminated by an error.}$$

$$2 \cdot dchisq(2 \cdot NE, 2 \cdot (ne+1)) = dpois(ne, NE) \text{ describes the experiment, terminated by any sample.}$$

The terminating error may be the artificial error at the beginning of the test, or the last error, causing the fail.

In the next comparison shows dpois versus dchisq.

The first 3D plot shows the Poisson distribution: (Figure 5-1)

Variable: ne	Range 0 to 10	Column in the table	0 to 10- axis in the plot
Parameter: NE	Range 0 to 10	Row in the table	0 to 100 axis in the plot

The second 3D plot shows the Chi Square distribution: (Figure 5-4)

Variable: NE	Range 0 to 10	Column in the table	0 to 100- axis in the plot
Parameter: ne	Range 0 to 10	Row in the table	0 to 10 axis in the plot

Column 0 is degenerated

$$P(ne, NE) := dpois(ne, 0.1NE + 0.000001)$$

$$m := 10 \quad n := 10 \quad O := matrix(m, n, P)$$

Table 5-1: Poisson distribution calculation

	0	1	2	3	4	5	6	7	8	9	10	11
0	1	0.905	0.819	0.741	0.67	0.607	0.549	0.497	0.449	0.407	0.368	0.333
1	$10 \cdot 10^{-7}$	0.09	0.164	0.222	0.268	0.303	0.329	0.348	0.359	0.366	0.368	0.366
2	$5 \cdot 10^{-13}$	$4.524 \cdot 10^{-3}$	0.016	0.033	0.054	0.076	0.099	0.122	0.144	0.165	0.184	0.201
3	0	$1.508 \cdot 10^{-4}$	$1.092 \cdot 10^{-3}$	$3.334 \cdot 10^{-3}$	$7.15 \cdot 10^{-3}$	0.013	0.02	0.028	0.038	0.049	0.061	0.074
4	0	$3.77 \cdot 10^{-6}$	$5.458 \cdot 10^{-5}$	$2.5 \cdot 10^{-4}$	$7.15 \cdot 10^{-4}$	$1.58 \cdot 10^{-3}$	$2.964 \cdot 10^{-3}$	$4.968 \cdot 10^{-3}$	$7.669 \cdot 10^{-3}$	0.011	0.015	0.02
5	0	$7.541 \cdot 10^{-8}$	$2.183 \cdot 10^{-6}$	$1.5 \cdot 10^{-5}$	$5.72 \cdot 10^{-5}$	$1.58 \cdot 10^{-4}$	$3.556 \cdot 10^{-4}$	$6.955 \cdot 10^{-4}$	$1.227 \cdot 10^{-3}$	$2.001 \cdot 10^{-3}$	$3.066 \cdot 10^{-3}$	$4.467 \cdot 10^{-3}$
6	0	$1.257 \cdot 10^{-9}$	$7.278 \cdot 10^{-8}$	$7.501 \cdot 10^{-7}$	$3.813 \cdot 10^{-6}$	$1.316 \cdot 10^{-5}$	$3.556 \cdot 10^{-5}$	$8.114 \cdot 10^{-5}$	$1.636 \cdot 10^{-4}$	$3.001 \cdot 10^{-4}$	$5.109 \cdot 10^{-4}$	$8.19 \cdot 10^{-4}$
7	0	$1.795 \cdot 10^{-11}$	$2.079 \cdot 10^{-9}$	$3.215 \cdot 10^{-8}$	$2.179 \cdot 10^{-7}$	$9.402 \cdot 10^{-7}$	$3.048 \cdot 10^{-6}$	$8.114 \cdot 10^{-6}$	$1.87 \cdot 10^{-5}$	$3.858 \cdot 10^{-5}$	$7.299 \cdot 10^{-5}$	$1.287 \cdot 10^{-4}$
8	0	$2.244 \cdot 10^{-13}$	$5.198 \cdot 10^{-11}$	$1.206 \cdot 10^{-9}$	$1.09 \cdot 10^{-8}$	$5.876 \cdot 10^{-8}$	$2.286 \cdot 10^{-7}$	$7.1 \cdot 10^{-7}$	$1.87 \cdot 10^{-6}$	$4.341 \cdot 10^{-6}$	$9.124 \cdot 10^{-6}$	$1.77 \cdot 10^{-5}$
9	0	$2.494 \cdot 10^{-15}$	$1.155 \cdot 10^{-12}$	$4.018 \cdot 10^{-11}$	$4.842 \cdot 10^{-10}$	$3.265 \cdot 10^{-9}$	$1.524 \cdot 10^{-8}$	$5.522 \cdot 10^{-8}$	$1.662 \cdot 10^{-7}$	$4.341 \cdot 10^{-7}$	$1.014 \cdot 10^{-6}$	$2.163 \cdot 10^{-6}$
10												

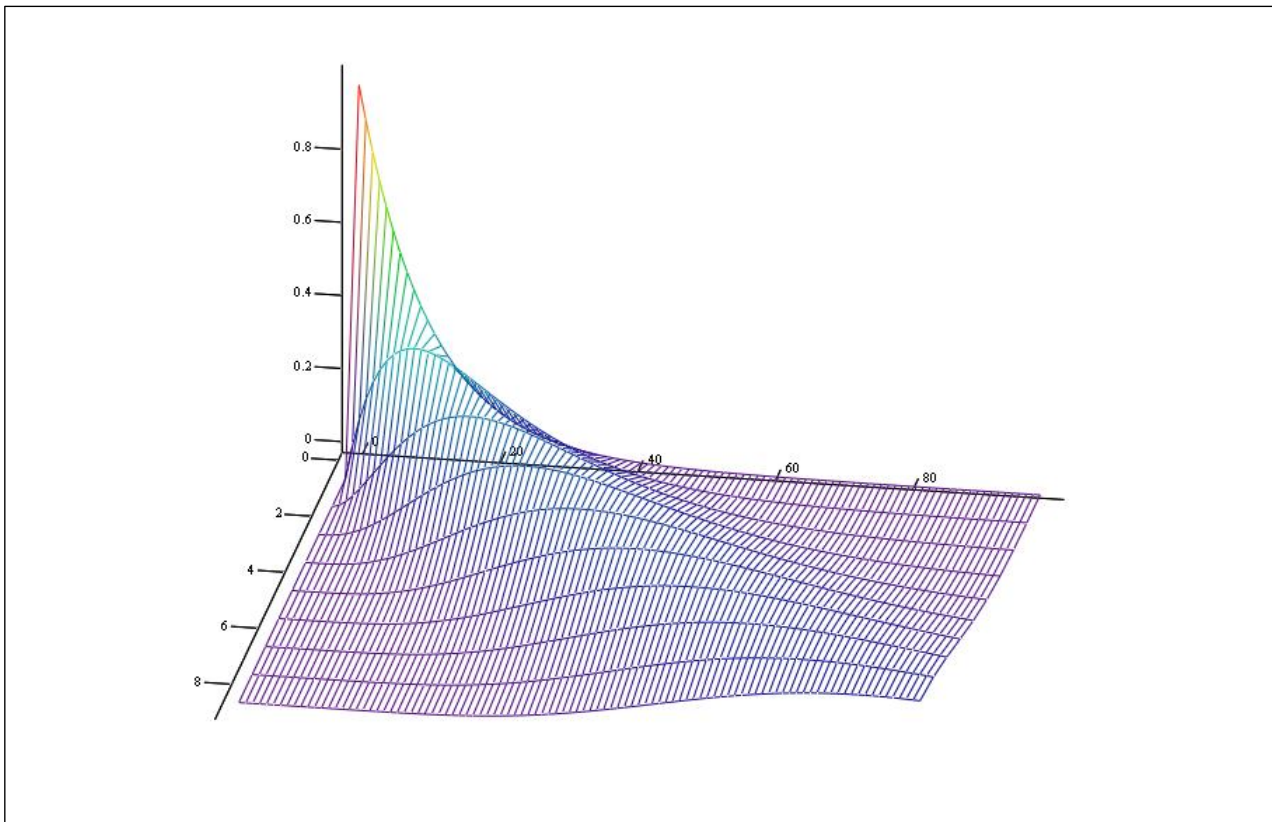


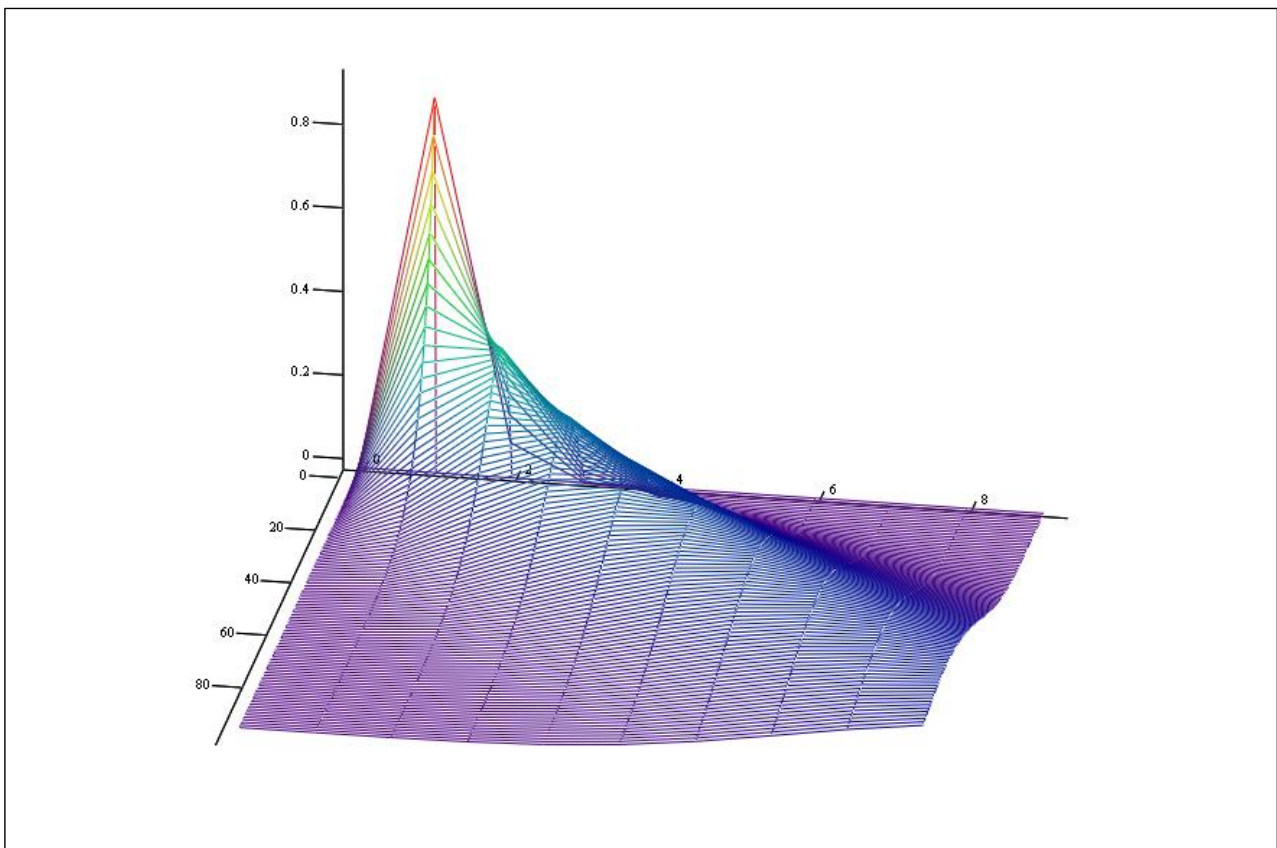
Figure 5-3: 3D plot for Poisson distribution

$C(NE, ne) := 2dchisq[2(0.1NE), 2(ne + .00000)]$

$m := 10 \quad n := 1 \quad (D := \text{matrix}(m, n, C))_{ne}$

Table 5-2: Chi-squared distribution calculation

	0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0	
1	$9.048 \cdot 10^{-6}$	0.905	0.09	$4.524 \cdot 10^{-3}$	$1.508 \cdot 10^{-4}$	$3.77 \cdot 10^{-6}$	$7.54 \cdot 10^{-8}$	$1.257 \cdot 10^{-9}$	$1.795 \cdot 10^{-11}$	$2.244 \cdot 10^{-13}$	
2	$4.094 \cdot 10^{-6}$	0.819	0.164	0.016	$1.092 \cdot 10^{-3}$	$5.458 \cdot 10^{-5}$	$2.183 \cdot 10^{-6}$	$7.278 \cdot 10^{-8}$	$2.079 \cdot 10^{-9}$	$5.198 \cdot 10^{-11}$	
3	$2.469 \cdot 10^{-6}$	0.741	0.222	0.033	$3.334 \cdot 10^{-3}$	$2.5 \cdot 10^{-4}$	$1.5 \cdot 10^{-5}$	$7.501 \cdot 10^{-7}$	$3.215 \cdot 10^{-8}$	$1.205 \cdot 10^{-9}$	
4	$1.676 \cdot 10^{-6}$	0.67	0.268	0.054	$7.15 \cdot 10^{-3}$	$7.15 \cdot 10^{-4}$	$5.72 \cdot 10^{-5}$	$3.813 \cdot 10^{-6}$	$2.179 \cdot 10^{-7}$	$1.09 \cdot 10^{-8}$	
5	$1.213 \cdot 10^{-6}$	0.607	0.303	0.076	0.013	$1.58 \cdot 10^{-3}$	$1.58 \cdot 10^{-4}$	$1.316 \cdot 10^{-5}$	$9.402 \cdot 10^{-7}$	$5.876 \cdot 10^{-8}$	
6	$9.147 \cdot 10^{-7}$	0.549	0.329	0.099	0.02	$2.964 \cdot 10^{-3}$	$3.556 \cdot 10^{-4}$	$3.556 \cdot 10^{-5}$	$3.048 \cdot 10^{-6}$	$2.286 \cdot 10^{-7}$	
7	$7.094 \cdot 10^{-7}$	0.497	0.348	0.122	0.028	$4.968 \cdot 10^{-3}$	$6.955 \cdot 10^{-4}$	$8.114 \cdot 10^{-5}$	$8.114 \cdot 10^{-6}$	$7.1 \cdot 10^{-7}$	
8	$5.617 \cdot 10^{-7}$	0.449	0.359	0.144	0.038	$7.669 \cdot 10^{-3}$	$1.227 \cdot 10^{-3}$	$1.636 \cdot 10^{-4}$	$1.87 \cdot 10^{-5}$	$1.87 \cdot 10^{-6}$	
9	$4.517 \cdot 10^{-7}$	0.407	0.366	0.165	0.049	0.011	$2.001 \cdot 10^{-3}$	$3.001 \cdot 10^{-4}$	$3.858 \cdot 10^{-5}$	$4.341 \cdot 10^{-6}$	
10	$3.679 \cdot 10^{-7}$	0.368	0.368	0.184	0.061	0.015	$3.066 \cdot 10^{-3}$	$5.109 \cdot 10^{-4}$	$7.299 \cdot 10^{-5}$	$9.124 \cdot 10^{-6}$	
11	$3.026 \cdot 10^{-7}$	0.333	0.366	0.201	0.074	0.02	$4.467 \cdot 10^{-3}$	$8.19 \cdot 10^{-4}$	$1.287 \cdot 10^{-4}$	$1.77 \cdot 10^{-5}$	
12	$2.51 \cdot 10^{-7}$	0.301	0.361	0.217	0.087	0.026	$6.246 \cdot 10^{-3}$	$1.249 \cdot 10^{-3}$	$2.141 \cdot 10^{-4}$	$3.212 \cdot 10^{-5}$	
13	$2.096 \cdot 10^{-7}$	0.273	0.354	0.23	0.1	0.032	$8.432 \cdot 10^{-3}$	$1.827 \cdot 10^{-3}$	$3.393 \cdot 10^{-4}$	$5.514 \cdot 10^{-5}$	
14	$1.761 \cdot 10^{-7}$	0.247	0.345	0.242	0.113	0.039	0.011	$2.579 \cdot 10^{-3}$	$5.158 \cdot 10^{-4}$	$9.026 \cdot 10^{-5}$	
15	$1.488 \cdot 10^{-7}$	0.223	0.335	0.251	0.126	0.047	0.014	$3.53 \cdot 10^{-3}$	$7.564 \cdot 10^{-4}$	$1.418 \cdot 10^{-4}$	
16	$1.262 \cdot 10^{-7}$	0.202	0.323	0.258	0.138	0.055	0.018	$4.705 \cdot 10^{-3}$	$1.075 \cdot 10^{-3}$	$2.151 \cdot 10^{-4}$	
17	$1.075 \cdot 10^{-7}$	0.183	0.311	0.264	0.15	0.064	0.022	$6.124 \cdot 10^{-3}$	$1.487 \cdot 10^{-3}$	$3.161 \cdot 10^{-4}$	
18	$9.183 \cdot 10^{-8}$	0.165	0.298	0.268	0.161	0.072	0.026	$7.809 \cdot 10^{-3}$	$2.008 \cdot 10^{-3}$	$4.518 \cdot 10^{-4}$	



D

Figure 5-4: 3D plot for Chi-squared distribution

Observation:

- 1) The rows in Poisson distribution correspond the columns in the scaled Chi-squared distribution and vice versa.
- 2) Poisson distribution at $ne=0$ versus NE is the exponential distribution
Chi-squared distribution at $ne=0$ (degree of freedom=1) versus NE is also the exponential distribution

see the next plot:

NE := 0.01, 0.02.. 10 ne := 0, 1.. 10

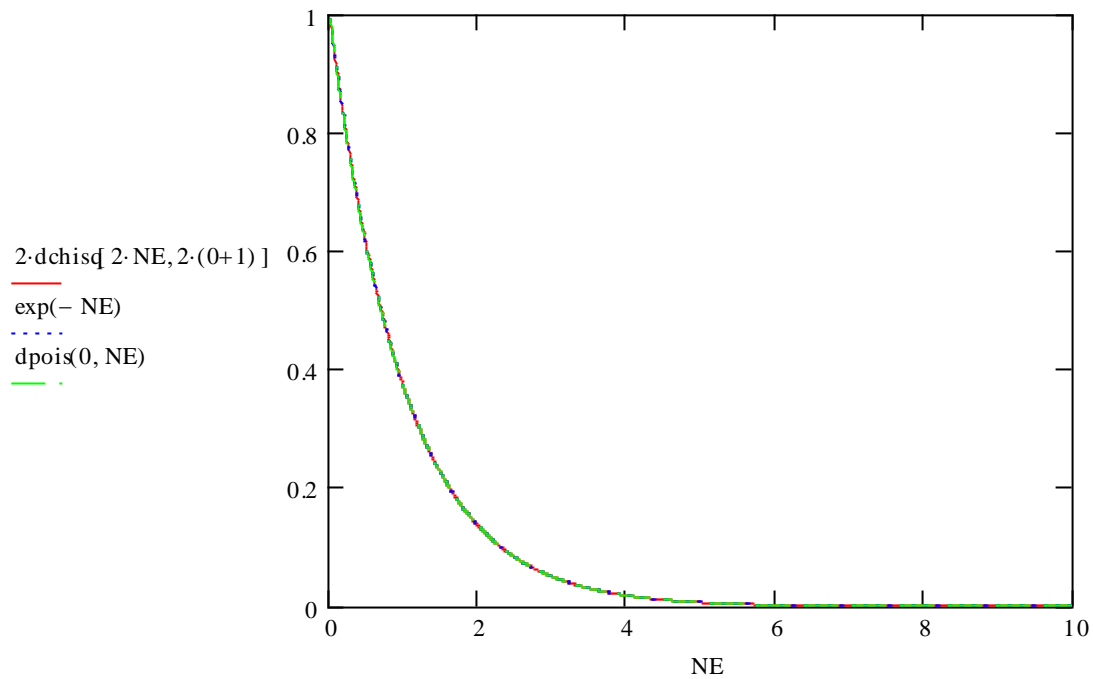


Figure 5-5: Comparison between Poisson and Chi-squared distribution

Inverse Cumulative Operation:

We have seen: Chi Square and Poisson both describe the same array: ne versus NE.

The figures above show, that NE and ne in both functions are not commutative.

Hence there are two inverse operations (a) and (b):

$$D = \int_0^{ne} dpois(ni, NE) dni = 2 * \int_0^{ne} dchisq[2NE, 2 \cdot (ni + 1)] dni \tag{a}$$

with D=wrong decision probability or confidence level (input).

ni is the integration variable

ne is the measured value.(input, discrete) It is the integration limit

NE (real) is tuned such that the integral is consistent.

It returns an NE as a function of the two parameters D and ne. qchisq(D,ne)

$$D = \int_0^{NE} dpois(ne, NI) dNI = 2 * \int_0^{NE} dchisq[2NI, 2 \cdot (ne + 1)] dNI \tag{b}$$

NI is the integration variable

NE (real) is the integration limit

ne (discrete) is tuned such that the integral is consistent.

It returns ne as a function of the two parameters D and NE: qpois(D,NE)

Our target requires a). This is usually called the Inverse Cumulative Chi Square function.

(b) is the solution for another target. This is usually called the Inverse Cumulative Poisson function.

(a) returns a greater **NE** than (b) returns with respect to **ne**. (easily visible in the figures)

The difference (a)-(b) is small. This is also visible from the figures: **ne** and **NE** are close to commutative.

(a) returns a continuous **NE**, (b) returns a discrete **ne**.

5.3 Confidence interval

In a single test we apply **ns** samples and measure **ne** errors. The result can be member of different distributions each characterized by another parameter **NE**. We ask for two of them:

1) The worst possible distribution **NE_{high}**, containing our measured **ne** with [D= 0.0085%] probability in the sense

$$0.000085 = \int_0^{ne} \text{dpois}(ni, NE_{high}) \text{dni} \quad (2)$$

ni is the integration variable

ne is the measured value

NE is the variable to tune in order to make the integral consistent.

The result of the inverse cumulative operation is **NE_{high}**

2) The best possible distributions **NE_{low}**, containing our measured **ne** with [D=0.0085%] probability in the sense

$$0.000085 = \int_{ne}^{\infty} \text{dpois}(ni, NE_{low}) \text{dni} \quad (3)$$

The result of the inverse cumulative operation is **NE_{low}**

To illustrate the meaning of the range between **NE_{low}** and **NE_{high}**:

In the case our measured value **ne** is a rather untypical result (just [0.0085%] probability) nevertheless the final result **NE** can still be found in this range, called confidence interval.

The probabilities **D** in (1) and (2) can be independent like in GSM, but we want to have them dependent and equal.

The inverse cumulative Chi Squared distribution gives the wanted results:

Inputs: number of errors **ne**, measured in this test.

Probabilities **D** and the complementary probability 1- **D**

Output: **NE**, the parameter describing the average of the distribution.

E.G.: **ne** := 15 **D** := .00008.

$$NE_{low} := 0.5 \text{qchisq}(D, 2ne) \quad (4)$$

$$NE_{low} = 4.56$$

$$NE_{high} := 0.5 \text{qchisq}(1 - D, 2ne) \quad (5)$$

$$NE_{high} = 34.085$$

ni := 1, 2.. 50

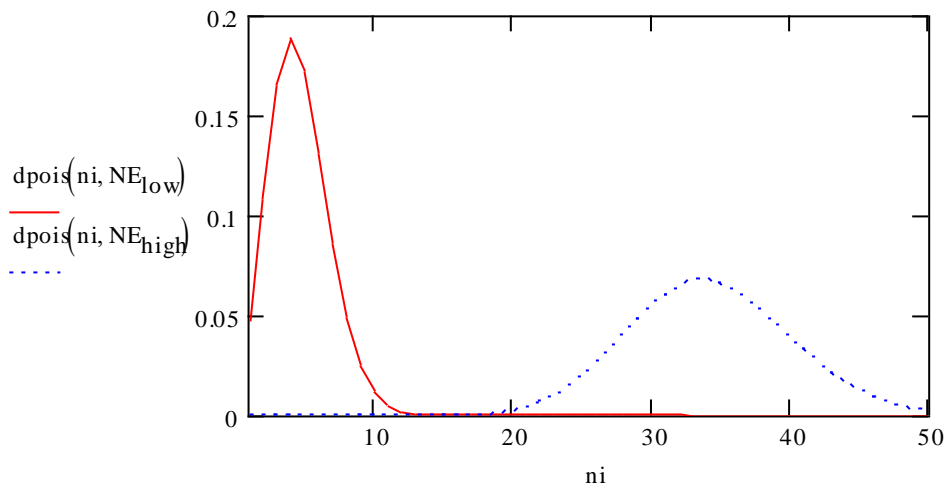


Figure 5-6: Confidence Interval

Same as the width of the distributions the confidence interval increases proportional to \sqrt{NE} , that means, it increases absolutely, but decreases relatively to the measured number of errors.

5.4 Application of the confidence interval to decide the outcome of the test

If we find the entire confidence range, calculated from a single result **ne**, on the good side of the specified limit we can state: With high probability **1-D**, the final **NE** is better than the limit.

If we find the entire confidence interval, calculated from a single result **ne**, on the bad side of the specified limit we can state: With high probability **1-D**, the final **NE** is worse than the limit.

With each new test we update our preliminary data for **ns**, **ne** and **ber**. For each new sample we calculate the **confidence interval** and check it against the test limit.

Once we find the entire **confidence interval** on the good side of the specified limit we allow an early pass.

Once we find the entire **confidence interval** on the bad side of the specified limit we allow an early fail.

If we find the confidence interval on both sides of the specified limit, it is evident neither to pass nor to fail the DUT early.

Transcription of the above text into formulas:

The current number of samples **ns** is calculated from the preliminary **ber** and the preliminary **ne**

$$ber = ne/ns \tag{6}$$

$$BER_{lim} = NE_{limit} / ns \tag{7}$$

for abbreviation in the formula: $ber_{norm} = ber/BER_{limit} = ne/NE_{limit}$ (normalised ber)

Early pass stipulates:

$$NE_{high} < NE_{limit} \tag{8}$$

Early fail stipulates:

$$NE_{low} > NE_{limit} \tag{9}$$

$D := 0.00008$ $ne := 1, 2.. 1000$

The early fail and the early pass limit are displayed in Figure 5-7:

early pass limit

$$bernompas(ne, D) := \frac{2ne}{qchisq(1 - D, 2 \cdot ne)} \tag{10}$$

early fail limit

$$bernorfai(ne, D) := \frac{2ne}{qchisq(D, 2 \cdot ne)} \tag{11}$$

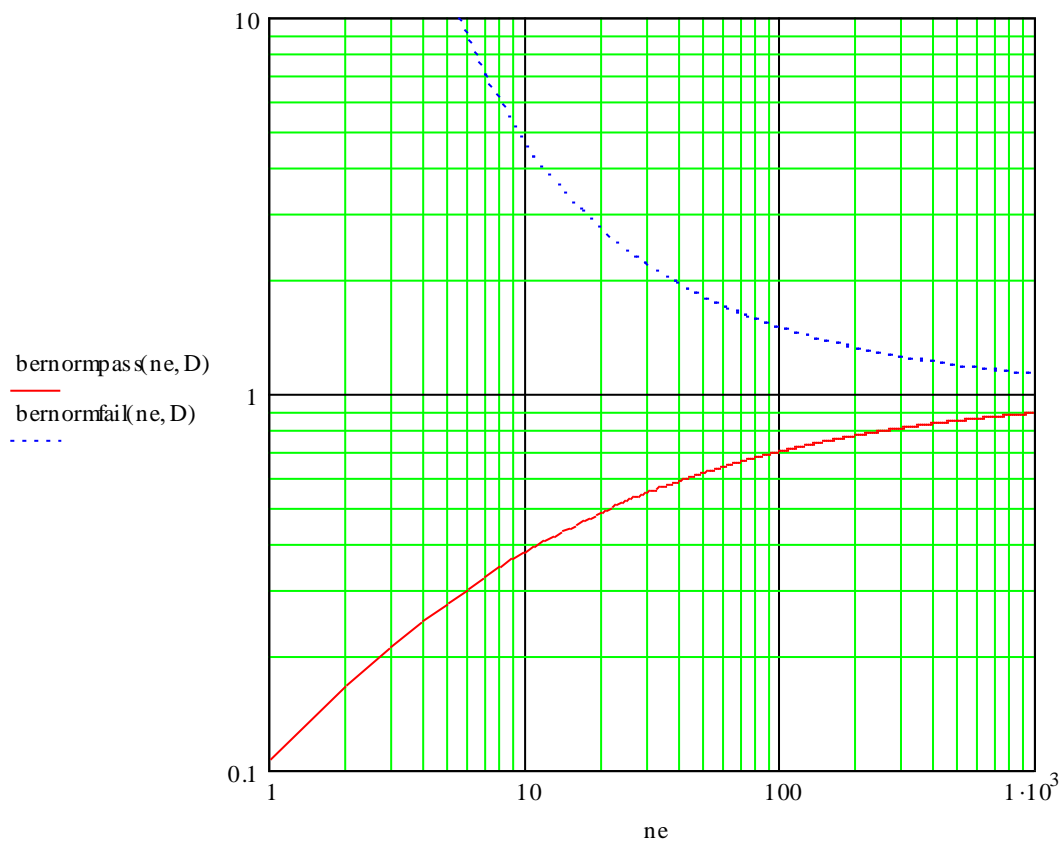


Figure 5-7: Early pass and early fail curves

5.5 Test time reduction

Whichever **ne** we propose as a final stop condition e.g. **ne** =200, the test can leave the area between the early pass and the early fail limit through the open end of the right side of Figure 5-7. This situation needs an arbitrary pass or fail decision. E.g. pass, if the test hits the vertical 200-error line. This situation has the following drawback: The test has two different qualities. A good one, when the test hits an early pass or early fail limit, and a worse and variable one, when the test hits the vertical 200 error line; variable, depending on the height, it crosses the line. The quality of the test in terms of wrong decision risk is variable in the range D up to as bad as 50%. We can replace the situation against a better trade-off:

Instead a test with different qualities against one limit,

we design a test with a fixed uniform quality against two limits,
(gaining further test time reduction).

We maintain the definition of the early fail limit:

(a) We fail a DUT and accept the probability of $D=0.0085\%$ that it is actually better than the limit.

We propose a meaningful redefinition of the early pass limit:

(b) We pass a DUT and accept the probability of $D=0.0085\%$ that it is actually worse than M times the limit ($M>1$), ($M = \text{Bad DUT factor}$)

This produces the following consequences:

- (1) The early pass limit is shifted upwards by the factor of M
- (2) The early fail and the early pass limit intersect.
- (3) The intersection coordinates are:

the normalized test limit
and the maximum number of events

Transcription of the above text into formulas:

$$\text{berlim}_{\text{bad_pass}} := \text{early pass limit against the bad DUT limit} \tag{12}$$

$$\text{berlim}_{\text{fail}} := \text{early fail limit against the specified limit} \tag{13}$$

$$M := 1.5 \quad D := 0.00008 \quad n_e := 1, 2, \dots, 1000$$

$$\text{berlim}_{\text{bad_pass}}(n_e, D) := 2 \cdot \frac{n_e}{\text{qchisq}(1 - D, 2 \cdot n_e)} \cdot M$$

$$\text{berlim}_{\text{fail}}(n_e, D) := 2 \cdot \frac{n_e}{\text{qchisq}(D, 2 \cdot n_e)}$$

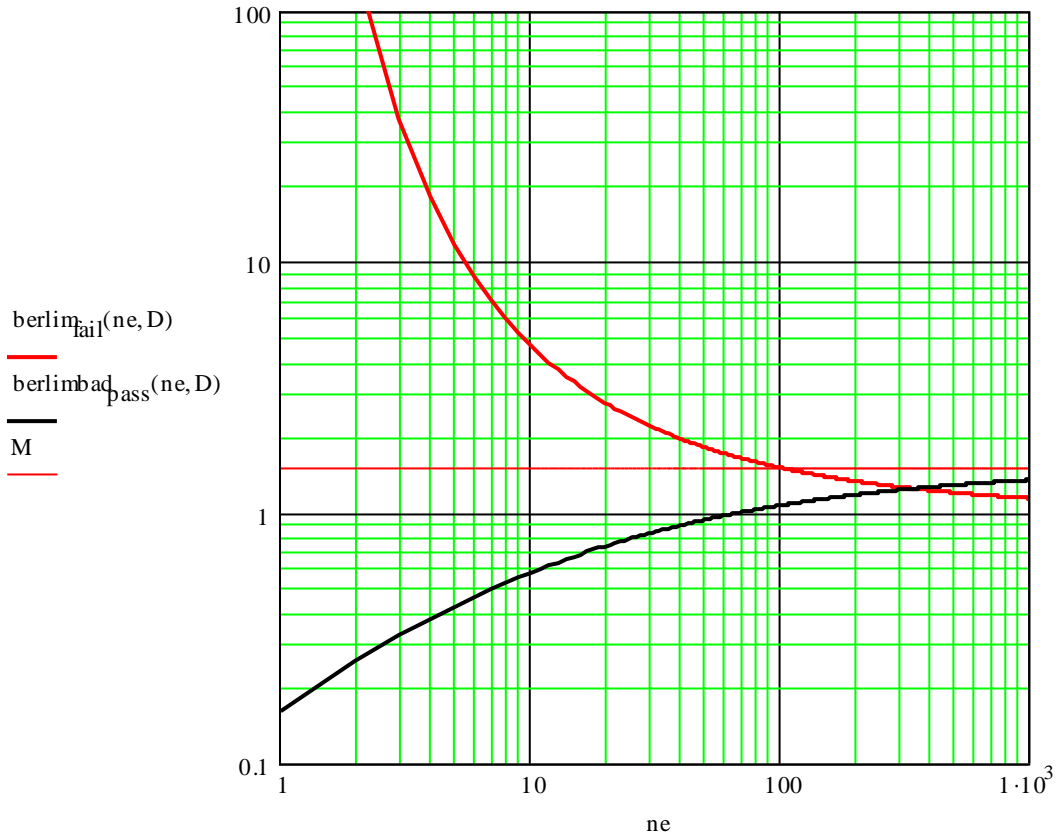


Figure 5-8: Early pass and early fail curves with multiplication factor M

5.6 Calculation of the intersection coordinates

$M := 1.5$
 $D := 0.00008$
 $ne := 1, 2, \dots, 1000$

$$berlimbad_{pass}(ne, D) := 2 \cdot \frac{ne}{qchisq(1 - D, 2 \cdot ne)} \cdot M$$

$$berlim_{fail}(ne, D) := 2 \cdot \frac{ne}{qchisq(D, 2 \cdot ne)}$$

$$Diff(k) := berlim_{fail}(k, D) - berlimbad_{pass}(k, D)$$

initial guess of target number of events

$$k := 200$$

root finds the zero of the function

$$k_{1M} := \text{root}(Diff(k), k)$$

target number of events

$$k_{1M} = 344.942$$

normalized test limit

$$berlim_{fail}(k_{1M}, D) = 1.234$$

5.7 Wrong decision risk F

Provided a single BER trajectory with final BER on the limit hits the early fail limit. If a fail is decided at this instant of the test, the wrong decision risk is as small as D. For each member of a large population of DUTs a wrong decision can happen, with probability D, accumulating to an amount $F > D$ for the entire population.

D is the wrong decision risk based on the statistical totality of samples with BER on the limit.

F is the wrong decision risk based on the statistical totality of DUTs with BER on the limit.

(The same holds for a bad DUT, hitting the early pass limit.)

We call D the wrong decision risk at a single test step and F the wrong decision risk for the entire test. For a real test it is desirable to define in advance the wrong decision risk F of the entire test. An exact theory is not available for this problem. It is proposed to derive D from F by the following simulation:

A large population of DUTs with BER on the limit (limit-DUT) is simulated and decided against the early pass and early fail bound, with a free D-parameter in the early pass and fail limit. The simulation will show, that a certain fraction F ($D < F < 1$) falsely fails.

The complementary simulation is:

A large population of DUTs with $M \cdot \text{BER}$ (bad DUT) is simulated and decided against the early pass and early fail bound, with a free D-parameter in the early pass and fail limit. The simulation will show, that a certain fraction F ($D < F < 1$) falsely passes.

Both false decision fractions are approximately equal and represent the wrong decision probability F for the entire test. D is tuned such that F corresponds to the predefined wrong decision probability.

6 F to D conversion in BER BLER tests

6.1 Conversion F to D

Annex F.6.1 in TS 34.121 gives a statistical approach for BER BLER tests. It gives early pass and early fail conditions. The formulas for this condition contain the parameter D, the wrong decision probability for a single test step. However it is desirable to have a wrong decision probability for the entire test F. This contribution explains the way, how to derive F from D and gives results for a set of parameters.

6.2 Introduction

Provided a single BER trajectory with final BER on the limit hits the early fail limit. If a fail is decided at this instant of the test, the wrong decision risk is as small as D. For each member of a large population of DUTs a wrong decision can happen, with probability D, accumulating to an amount $F > D$ for the entire population.

D is the wrong decision risk based on the statistical totality of samples with BER on the limit.

F is the wrong decision risk based on the statistical totality of DUTs with BER on the limit.

(The same holds for a bad DUT, hitting the early pass limit.)

We call D the wrong decision risk at a single test step and F the wrong decision risk for the entire test. For a real test it is desirable to define in advance the wrong decision risk F of the entire test. An exact theory is not available for this problem. It is proposed to derive D from F by the following simulation:

A large population of DUTs with BER on the limit (limit-DUT) is simulated and decided against the early pass and early fail bound, with a free D-parameter in the early pass and fail limit. The simulation will show, that a certain fraction F ($D < F < 1$) falsely fails.

The complementary simulation is:

A large population of DUTs with $M \cdot \text{BER}$ (bad DUT) is simulated and decided against the early pass and early fail bound, with a free D-parameter in the early pass and fail limit. The simulation will show, that a certain fraction F ($D < F < 1$) falsely passes.

Both false decision fractions are approximately equal and represent the wrong decision probability F for the entire test. D is tuned such that F corresponds to the predefined wrong decision probability.

6.3 The simulation procedure

6.3.1 Equal pass and fail probability

A population of DUTs on the limit is established. Quantity 10 000

Preselected BER 1%

An early fail and an early pass limit is established. With D (wrong decision risk) and M (bad DUT factor)

With target n_e and test limit as a side result.

D is tuned in the inner loop

M and $F_{\text{predefined}}$ are varied in the outer loop

During the simulation

A member of the population leaves the statistical totality if

An error happens and the early fail limit is hit (false fail)

An error happens and the early pass limit is hit or crossed (correct pass)

The fraction false fails / 10 000 = F is recorded.

Inner loop: In repeated trials D is tuned, such that $F \leq F_{\text{predefined}}$ (conservative approach).

Having decided for a specific D the simulation is repeated again 10 times and F_{min} , F_{max} , and F_{mean} are recorded.

The complementary simulation is done with a population of bad DUTs

(same quantity, same M , same $F_{\text{predefined}}$, same D)

Observation 1: the false pass fraction is slightly lower than the false fail fraction.

Hence the result is even more conservative for the false pass.

Outer Loop: M is varied from 1.1 to 1.5 in steps of 0.1

$F_{\text{predefined}}$ is varied from 0.2 %, 0.5 %, 1%, 2% to 5%.

Observation 2: For lower wrong decision risks F the false decisions in 10 000 DUT are less.

Hence the variance of F in the 10 repetitions relatively increases.

For lower wrong decision risks F the simulation time increases.

Hence the compensation of the increasing variance of F by more repetitions is limited by the simulation time, or vice versa: the simulation results for F converge to a final value, investing infinite effort for simulations.

For practical and security reasons the D s for lower F s are decided more conservative than the equivalent ones for higher F s.

6.3.2 Unequal pass and fail probability

For statistical test, frequently repeated, a single false fail can fail the composite test. This can be combated by a decreased false fail risk. This costs test time.

A single false pass does not have this effect on the composite test. Hence it is not necessary to consume increased test time due to decreased false pass risk for the pass probability. Hence unequal pass and fail probabilities are treated for very low false fail risk.

A population of DUTs on the limit is established. Quantity 10 000

Preselected BER 1%

In the complementary simulation

a population of bad DUTs is established (same parameters)

Common for both simulations:

An early fail limit is established with D_{fail} and an early pass limit is established with D_{pass} (D wrong decision risk $D_{\text{fail}} < D_{\text{pass}}$) and M (bad DUT factor)

With target n_e as a side result.

D_{fail} and D_{pass} are tuned independently in the inner loop

M is varied in the outer loop

During the simulation

A member of the population leaves the statistical totality if

An error happens and the early fail limit is hit (false fail)

For the complementary simulation: (correct fail)

An error happens and the early pass limit is hit or crossed (correct pass)

For the complementary simulation: (false pass)

The fraction false fails / 10 000 = F_{fail} and the fraction false pass/10 000 = F_{pass} are recorded.

Inner loop: In repeated trials D_{fail} and D_{pass} are tuned independently, such that F_{fail} and $F_{\text{pass}} \leq F_{\text{predefined}}$ (conservative approach).

Having decided for a specific D_{fail} and D_{pass} the simulation is repeated again 10 times and $F_{\text{min fail}}$, $F_{\text{max fail}}$, and $F_{\text{mean fail}}$ and $F_{\text{min pass}}$, $F_{\text{max pass}}$, and $F_{\text{mean pass}}$ are recorded.

Observation 3: The D_{pass} must be slightly lower than the equivalent D in the case for equal probabilities (a).

Due to lower D_{fail} the target number of errors increases (e.g. 345 \rightarrow 403), accumulating more single step wrong decisions. This is compensated by a lower D_{pass} .

6.4 Result of the simulation: M-F array

Table 6-1: M-F table

Format	M=1.1	M=1.2	M=1.3	M=1.4	M=1.5
Pass:0.2%	False fails%	False pass%			0.0161 0.129
Fail:	min / max%	min / max%			0 / 0.05 0.08 / 0.21
F=0.02%	Dfail	Dpass			0.0008 0.008
		target-ne			403
Pass:0.2%					
Fail:					
F=0.05%					
Pass:0.2%					
Fail:					
F=0.1%					
	False fails%	False pass%	0.151	0.154	0.169 0.151 0.159 0.129
Pass&Fail	min / max%	min / max%	0.12 / 0.19	0.1 / 0.2	0.12 / 0.22 0.1 / 0.22 0.1 / 0.22 0.1 / 0.16
F=0.2%	D	target-ne	0.0065	852	0.0075 509 0.0085 345
					0.396 0.331 0.417 0.332
Pass&Fail					0.34 / 0.5 0.27 / 0.37 0.3 / 0.62 0.28 / 0.48
F=0.5%					0.02 444 0.025 296
					0.901 0.704
Pass&Fail					0.74/1.05 0.59/0.78
F=1%					0.06 256
					1.737 1.284
Pass&Fail					1.52/2.04 1.1/1.47
F=2%					0.14 218
					4.993 3.971
Pass&Fail					4.4/5.41 3.49/4.41
F=5%					0.6 154

7. Definitions, symbols and abbreviations

Definitions, symbols, abbreviations and equations used in the present document are listed in TR 21.905 [5] and TR 25.990 [6].

7.1 Definitions

For the purposes of the present document, the following additional terms and definitions apply.

Wrong Decision Probability: Chance of incorrect judgement based on the given test results.

Significance Level: Chance of incorrect judgement based on the given test results. This term is exchangeable with Wrong Decision Probability.

Individual Significance Level: The significance level of a simple test or a single decision. Denoted by "D".

Total Significance Level: The significance level of a test system as a whole (a set of simple tests.) Denoted by "F".

Confidence Coefficient (Level): In this report, this is defined by 1 – Significance Level.

Specified Error Ratio: General term that is referred to BER/BLER or other error ratios that are specified in the test specifications (TS34.121 and TS34.122.)

Average Error Ratio: The error ratio specific to a DUT whose value can be determined by averaging the infinite number of measurement data of error ratio.

Early Pass/Fail Criteria: a set of simple tests, each of which consists of a pair of thresholds of the measured error ratio.

7.2 Symbols

For the purposes of the present document, the following symbols apply:

[...]	Values included in square bracket must be considered for further studies, because it means that a decision about that value was not taken
C	Confidence Coefficient (Confidence Level)
D	Individual Significance Level
D_p	Individual Significance Level of an early pass criterion
D_f	Individual Significance Level of an early fail criterion
F	Total Significance Level
F_p	Total Significance Level of a set of early pass criteria
F_f	Total Significance Level of a set of early pass criteria
M	Bad DUT factor
R_0	Specified error ratio
R	Measured error ratio (rate) (calculated from the measured error count and time duration or sample number)
R_{th}	Threshold level against measured error ratio
R_{th-p}	Threshold of an early pass test
R_{th-f}	Threshold of an early fail test
r	Average error ratio (rate) of DUT

7.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

BER	Bit Error Ratio
BLER	Block Error Ratio
DUT	Device under Test
pdf	Probability Distribution Function
CDF	Cumulative Distribution Function

7.4 Equations

Chi-square distribution: of a degree of freedom n

$$\chi_{pdf}^2(x:n) = \frac{x^{n/2-1} e^{-x/2}}{\Gamma(n/2) 2^{n/2}}$$

where $\Gamma()$ is gamma function.

Exponential distribution:

$$Exp(x:r) = \frac{1}{r} e^{-x/r}$$

8 Introduction

BER/BLER measurements are inherently statistical processes, and a statistical approach should be introduced in order to make a BER/BLER test method objective and accurate.

In the GSM standard specifications, this has been achieved by introducing "wrong decision probability (significance level)" and "bad DUT factor." That is, the standard requires the test method to give a well-defined (predictable) "wrong decision probabilities", F_f/F_p or confidence coefficients $C_f (= 1 - F_f)/C_p (= 1 - F_p)$. Here, it should be noted that the significance levels are defined against the population of UEs with two error rates (0.01 and 0.015).

These parameters, the significance level and bad DUT factor, should be determined *a priori*, since there seems to be no good reason to pick a set of values rather than others. So, it would be a reasonable decision to employ the same value with GSM standard: $F = 0.002$ (0.2 %) and $M = 1.5$ for BER and BLER measurements since we now have a long enough experience to prove the validity of these parameters.

The GSM standard recommends that BER/BLER should be calculated after 200 errors were observed and the resultant BER/BLER should be compared against a threshold (around $1.24 R_0$). This test method gives the significance level presented above.

However, this method is not optimal from the viewpoint of test time. Apparently, if the DUT population consists of very good UEs (UE with a very low BER), it would take a long time to observe 200 errors. In practice, such a situation is unacceptable, so some supplementary criterion should be used. For example, if a UE doesn't report any bit/block errors for the certain time period, it should pass the test. This supplementary criterion may save many of the potentially time-consuming cases, but still not optimised for the test time.

In this report, a more sophisticated method will be developed to optimise the test time by introducing the early pass/fail criteria, whose basic idea was inspired by I-95 standard.

9 Statistical characteristics of testing processes

9.1 Exact probability distribution functions

BER/BLER measurements treat phenomena that are characterized as:

- 1) The experiment consists of a sequence of independent trials.
- 2) Each trial has two possible outcomes, E (Error) or N (Non Error).
- 3) The probability of E ($r = f(E)$) is constant from one trial to another.

That is, the BER/BLER measurement can be reduced to observation of Bernoulli sequences. Many other test objects, such as RRM delay measurements results, can also be reduced to a process that is characterized by these requirements. In these case, the outcomes are either "Pass or Successful (the measured delay was within the limit value, for example)" or "Fail (the delay exceeded the limit)."

Our objective is to find a way to tell whether r is larger than $R_0 * M$ or smaller than R_0 with a certain significance level from the result of a certain number of trials.

This report proposes to use a set of many simple tests, instead of a single simple test. Here, a test is a comparison of resultant error rate against a predefined criterion (a threshold value). Practically speaking, there would be two ways to do this comparison:

- a) To do the comparison for every trial, or
- b) To do the comparison each time an error is observed.

For method a) above, the probability of giving an error number, m , is given by a binomial distribution,

$$Bi(m : r, s) = \binom{s}{m} r^m (1-r)^{s-m} \quad (m = 0, 1, 2, \dots, s).$$

Where r : average error rate, m : the number of errors, s : the number of samples. The resultant error ratio, R , is m/s , and its probability can be derived from the equation above.

On the other hand, for method b), the probability that m -th error is observed at the s -th sample follows a negative binomial distribution,

$$nBi(s : r, m) = \binom{s-1}{m-1} r^m (1-r)^{s-m} \quad (s = m, m+1, m+2, \dots).$$

The error rate R is again given by m/s .

9.2 Approximation with continuous distributions

In this section, the notion r for average error ratio is used for average error rate too.

The distributions introduced above are exact, but sometimes they are very time-consuming to simulate the process on a computer. So, it is convenient to use appropriate continuous distribution functions that approximate the discrete functions.

If r is very low (that is, $m/s \ll 1$), the phenomenon described in the previous section can be reduced to the one that satisfies following requirements:

- 1) Every error can be described by the time at which it occurred.
- 2) Each error occurs independently
- 3) The average error rate $r (= m/T)$ is constant throughout the testing.

It is known that the time interval between the consecutive errors, t , should follow an exponential distribution.

Again, there are two possibilities to determine the measured error rate R and test it against the criteria:

- a) To calculate R after the predetermined time period, or
- b) To calculate R when m -th error is observed.

In case a), the probability that m errors are observed within the certain time period, T follows the Poisson distribution.

$$P_o(m) = \frac{\lambda^m e^{-\lambda}}{m!}$$

Where λ : average error count in the time duration T , so $\lambda = rT$. So, the function can be represented as:

$$P_o(m : r, T) = \frac{(rT)^m e^{-rT}}{m!}$$

The measured error rate, R , is given by m/T .

For the test method b), the time to m -th error, T , follows an m-Erlang distribution.

$$mEr(T : r, m) = \frac{T(rT)^{m-1} e^{-rT}}{(m-1)!} = 2T \chi_{pdf}^2(2rT, 2m)$$

Where $\chi_{pdf}^2(x, n)$ is a chi-square distribution function of a degree of freedom n . m-Erlang distribution's CDF is given by

$$mEr_{CDF}(T : r, m) = 1 - \sum_{k=0}^{m-1} \frac{(rT)^k e^{-rT}}{k!} = \chi_{CDF}^2(2rT, 2m)$$

This function can be transformed to a function of $R(=m/T)$, $mEr_{CDF}(R)$.

$$mEr_{CDF}(R : r, m) = 1 - \chi_{CDF}^2(2(r/R)m, 2m)$$

Apparently, the CDF is a function of (R/r) and this means the distribution does not depend on the absolute value of average or measured error rate. This would make a handling very easy.

Almost all the RF and RRM measurements that require the statistical approach are of the discrete nature, and then the continuous pdf is accurate only for smaller r values. It's hard to tell the exact condition under which the approximation should be good, but it can be said that we cannot use the continuous pdf when $r = 0.1$. This subject will be discussed in the chapters below.

9.3 Simple test criteria and significance level

For BER/BLER testing, our test can be reduced to a comparison of the measured error rate, R against a threshold. It would follow the procedure like:

```
if  $R < R_{th-p}$  then the DUT passes
else if  $R > R_{th-f}$  then the DUT fails
else no decision should be made
```

Where R_{th-p} and R_{th-f} are the threshold for pass and fail criteria, and in general $R_{th-p} \leq R_{th-f}$.

The significance level for the first test, F_p , is defined as the probability that a bad DUT ($r > R_0 M$) should pass. Similarly, F_f is defined as the probability that a good DUT ($r < R_0$) should fail in this test.

The severest condition from the viewpoint of the significant level is that $r = R_0 M$ and $r = R_0$ for bad DUT and good DUT population respectively. So, it would be reasonable to define the significance levels for such populations.

Now that it is presumed that r of the DUT population is unique (R_0 or $M \cdot R_0$), we can calculate the probability distribution of R according to the functions introduced in the previous chapter. For a pdf, $f(R: r, m)$, the threshold can be determined so that:

$$\sum_{x=0}^{R_{th-p}} f(x: R_0 M, m) = D_p \quad (1)$$

$$\sum_{x=R_{th-f}}^{\infty} f(x: R_0, m) = D_f \quad (2)$$

Where, it should be noted that x takes only discrete values that are given by m/s .

For the continuous pdf's, the summation should be replaced with integration, or using their CDF ($F(R: r, m)$)

$$F(R_{th-p}: R_0 M, m) = D_p \quad (1')$$

$$F(R_{th-f}: R_0, m) = 1 - D_f \quad (2')$$

When we choose an appropriate combination of M , m , and $D (= D_p = D_f)$, we can set the R_{th} so that the outcome of the test is either of Pass or Fail. For GSM standard, $M = 1.5$, $D = 0.2\%$, and then $m = 200$ and $R_{th} = 1.24 R_0$. This can be derived from both of m-Erlang distribution ($mER_{CDF}(R:m,r)$ above) and the negative binomial distribution.

10 Early Pass/Fail termination of testing

10.1 Framework of test method

If a UE reported 50 block errors for the first 100 frames, we would intuitively know that the UE should be bad (its average BLER should be higher than 0.01), even though the number of the observed errors is much smaller than 200. We are quite sure about that, since it should be very rare for a UE with an average BLER of 0.01 to report 50 block errors out of 100 frames. This can be shown using the binomial distribution, which gives its probability as

$$Bi(50: r = 0.01, s = 100) = 6.1 \times 10^{-72}$$

The probability that 50th error should occur at 100th sample is given by negative binomial distribution:

$$nBi(100: r = 0.01, m = 50) = 3.1 \times 10^{-72}$$

We can safely say that both cases should be very unlikely. That is, r cannot be 0.01 and should be much larger.

This suggests that, for very good or very poor UEs, the test sequence can be terminated at earlier stages, and as a result, the test time can be cut short. This is the origin of the idea of "early pass/fail criteria."

Such the test method will consist of many simple tests and its pseudo script would look like:

```

m = 0
for each s: # repeat for sample forever
  if an error is reported:
    m = m + 1
    R = m/s
    if R > Rth_f(m): # fail test
      UE failed
      break # stop procedure
    else if R < Rth_p(m):
      UE passed
      break
    else
      continue

```

This method is straightforward. Each time an error reported, the error rate should be calculated and then compared with thresholds that depend on m . However, there is still a room for improvement from viewpoint of test time reduction. The problem is apparent if we imagine the case in which no errors occur at all. The test would not be completed within a finite time period.

Assume that the no error samples last long enough after the m -th error. If the duration is long enough and the error rate calculated from imaginary $(m+1)$ th error is lower than $Rth_p(m+1)$, the UE can be considered to have passed the test. This idea would be implemented like:

```

m = 0
for each s: # repeat for sample forever
  if an error is reported:
    m = m + 1
    R = m/s
    if R > Rth_f(m): #fail test
      UE failed
      break
  else: # if not error
    if s > Tdp(m): #pass test
      UE passed
      break
    else:
      continue

```

Where $Tdp(m)$ is a predefined time duration (in sample number) which gives $R < Rth_p$ for $(m+1)$ th error.

To make the test method complete, a table like the example below should be prepared:

m	R _{th_p}	T _{dp}	R _{th_f}
0	---	417	---
1	0.0024	540	---
2	0.0034	750	0.031
3	0.0040	800	0.025
...

Where, "---" means "any decision should not be made at that error count." For example, while no error is reported, any decisions should not be made based on the error rate, and only when the time duration in terms of the number of samples exceeds 417, DUT is decided to pass.

10.2 Early Pass/Fail criteria

As described in the previous section, once the framework of the test procedure is established, a test can be defined by presenting the table of R_{th_p} , T_{dp} , and R_{th_f} . These values will be referred to as "Early Criteria."

The requirements for such criteria should be:

- 1) F should be 0.2 % for BER/BLER measurement, and 5% for RRM delay measurement and others.
- 2) The test shall be terminated in a finite time period (in terms of the number of samples.)

Presuming these requirements, we will try to minimize the test time.

There seems to be a broad freedom in choosing a set of early criteria. However, it would be a good start point to choose these criteria so that they provide a constant D at each error count m . When the error-count based and discrete distribution approach is presumed, the pdf to be used is a negative binomial distribution, $nBi(s: r, m) = nBi(m/R: r, m)$. Then, equation (1) and (2) will be:

$$\sum_{s=S_p}^{\infty} nBi(s: R_0 M, m) \leq D_p \quad R_{th-p} = m / S_p \quad (3)$$

$$\sum_{s=m}^{S_f} nBi(s: R_0, m) \leq D_f \quad R_{th-f} = m / S_f \quad (4)$$

Using these equations, R_{th-f} and R_{th-p} can be determined and from the latter, T_{dp} can also be derived. The problem is how to determine D . The target is clear; we have to get an F of 0.2%, but the population keeps changing in the procedure (some DUT hit either of the criteria and will be removed from the population), so it seems almost impossible to determine D to give a predefined F by an analytical method.

So an experimental method should be used to determine D values, which goes like:

- 1) Pick a D value
- 2) Calculate R_{th-p} and R_{th-f}
- 3) With a simulation with random generator of Bernoulli distribution of $r = R_0$ and $r = R_0 M$, and the criteria above, determine F value.
- 4) If the resultant F value is not close enough to the target F value (total significance level), start over with step 1) with a slightly different D value.

In the experiments that give results shown in this report bisection method was used to perform this iteration efficiently, and a binomial distribution generator was used as Bernoulli generator. The simulations have been done for a population of 100,000 devices.

The resultant criteria for typical parameters are given in the following tables.

Table 10-1: Early Pass/Fail criteria for RRM delay measurements ($R_0=0.1$, $F=0.05$)

m	Rth-p	Tdp	Rth-f
0	---	32	---
1	0.03125000	46	---
2	0.04347826	57	---
3	0.05263158	68	0.75000000
4	0.05882353	79	0.50000000
5	0.06329114	89	0.41666667
6	0.06741573	98	0.35294118
7	0.07142857	108	0.30434783
8	0.07407407	117	0.28571429
9	0.07692308	126	0.26470588
10	0.07936508	136	0.25000000
11	0.08088235	144	0.23404255
12	0.08333333	153	0.22641509
13	0.08496732	162	0.21666667
14	0.08641975	171	0.21212121
15	0.08771930	179	0.20547945
16	0.08938547	188	0.20000000
17	0.09042553	197	0.19540230
18	0.09137056	205	0.18947368
19	0.09268293	213	0.18627451
20	0.09389671	222	0.18348624
21	0.09459459	230	0.18103448
22	0.09565217	238	0.17741935
23	0.09663866	247	0.17557252
24	0.09716599	255	0.17266187
25	0.09803922	263	0.17123288
26	0.09885932	271	0.16883117
27	0.09963100	280	0.16666667
28	0.10000000	288	0.16568047
29	0.10069444	296	0.16384181
30	0.10135135	304	0.16216216
31	0.10197368	312	0.16062176
32	0.10256410	320	0.15920398
33	0.10312500	328	0.15789474
34	0.10365854	336	0.15668203
35	0.10416667	344	0.15555556
36	0.10465116	352	0.15450644
37	0.10511364	360	0.15352697
38	0.10555556	368	0.15261044
39	0.10597826	376	0.15175097
40	0.10638298	383	0.15094340
41	0.10704961	391	0.15018315
42	0.10741688	399	0.14946619
43	0.10776942	407	0.14878893
44	0.10810811	415	0.14814815
45	0.10843373	423	0.14705882
46	0.10874704	430	0.14649682
47	0.10930233	438	0.14596273
48	0.10958904	446	0.14545455
49	0.10986547	454	0.14454277
50	0.11013216	462	0.14409222
51	0.11038961	469	0.14366197
52	0.11087420	477	0.14285714
53	0.11111111	484	0.14247312
54	0.11134021	492	0.14210526
55	0.11178862	500	0.14138817
56	0.11200000	508	0.14105793
57	0.11220472	516	0.14039409
58	0.11240310	523	0.14009662
59	0.11281071	531	0.13981043
60	0.11299435	539	0.13921114
61	0.11317254	546	0.13895216
62	0.11355311	554	0.13839286
63	0.11371841	561	0.13815789
64	0.11408200	569	0.13763441

65	0.11423550	577	0.13742072
66	0.11438475	584	0.13692946
67	0.11472603	592	0.13673469
68	0.11486486	600	0.13627255
69	0.11500000	607	0.13609467
70	0.11532125	615	0.13565891
71	0.11544715	622	0.13523810
72	0.11575563	630	0.13508443
73	0.11587302	637	0.13468635
74	0.11616954	645	0.13454545
75	0.11627907	653	0.13416816
76	0.11638591	660	0.13380282
77	0.11666667	668	0.13368056
78	0.11676647	675	0.13333333
79	0.11703704	683	0.13299663
80	0.11713031	690	0.13289037
81	0.11739130	698	0.13256956
82	0.11747851	705	0.13225806
83	0.11773050	713	0.13216561
84	0.11781206	720	0.13186813
85	0.11805556	728	0.13157895
86	0.11813187	735	0.13149847
87	0.11836735	743	0.13122172
88	0.11843876	750	0.13095238
89	0.11866667	758	0.13088235
90	0.11873351	765	0.13062409
91	0.11895425	773	0.13037249
92	0.11901682	780	0.13012730
93	0.11923077	788	0.13006993
94	0.11928934	795	0.12983425
95	0.11949686	803	0.12960437
96	0.11955168	810	0.12938005
97	0.11975309	818	0.12916112
98	0.11980440	825	0.12911726
99	0.12000000	832	0.12890625
100	0.12019231	840	0.12870013
101	0.12023810	847	0.12849873
102	0.12042503	855	0.12830189
103	0.12046784	862	0.12826899
104	0.12064965	870	0.12807882
105	0.12068966	877	0.12789281
106	0.12086659	884	0.12771084
107	0.12104072	892	0.12753278
108	0.12107623	899	0.12735849
109	0.12124583	907	0.12733645
110	0.12127894	914	0.12716763
111	0.12144420	922	0.12700229
112	0.12147505	929	0.12684032
113	0.12163617	936	0.12668161
114	0.12179487	944	0.12652608
115	0.12182203	951	0.12637363
116	0.12197687	958	0.12622416
117	0.12212944	966	0.12621359
118	0.12215321	973	0.12606838
119	0.12230216	981	0.12592593
120	0.12232416	988	0.12578616
121	0.12246964	995	0.12564901
122	0.12261307	1003	0.12551440
123	0.12263210	1010	0.12538226
124	0.12277228	1017	0.12525253
125	0.12291052	1025	0.12512513
126	0.12292683	1032	0.12500000
127	0.12306202	1039	0.12487709
128	0.12319538	1047	0.12475634
129	0.12320917	1054	0.12475822
130	0.12333966	1062	0.12464046
131	0.12335217	1069	0.12452471
132	0.12347989	1076	0.12441093

133	0.12360595	1084	0.12429907
134	0.12361624	1091	0.12418906
135	0.12373969	1098	0.12408088
136	0.12386157	1106	0.12397448
137	0.12386980	1113	0.12386980

The resultant thresholds (criteria) for BER, BLER and RRM Delay measurements are shown in Figure 10-1, Figure 10-2 and Figure 10-3 respectively.

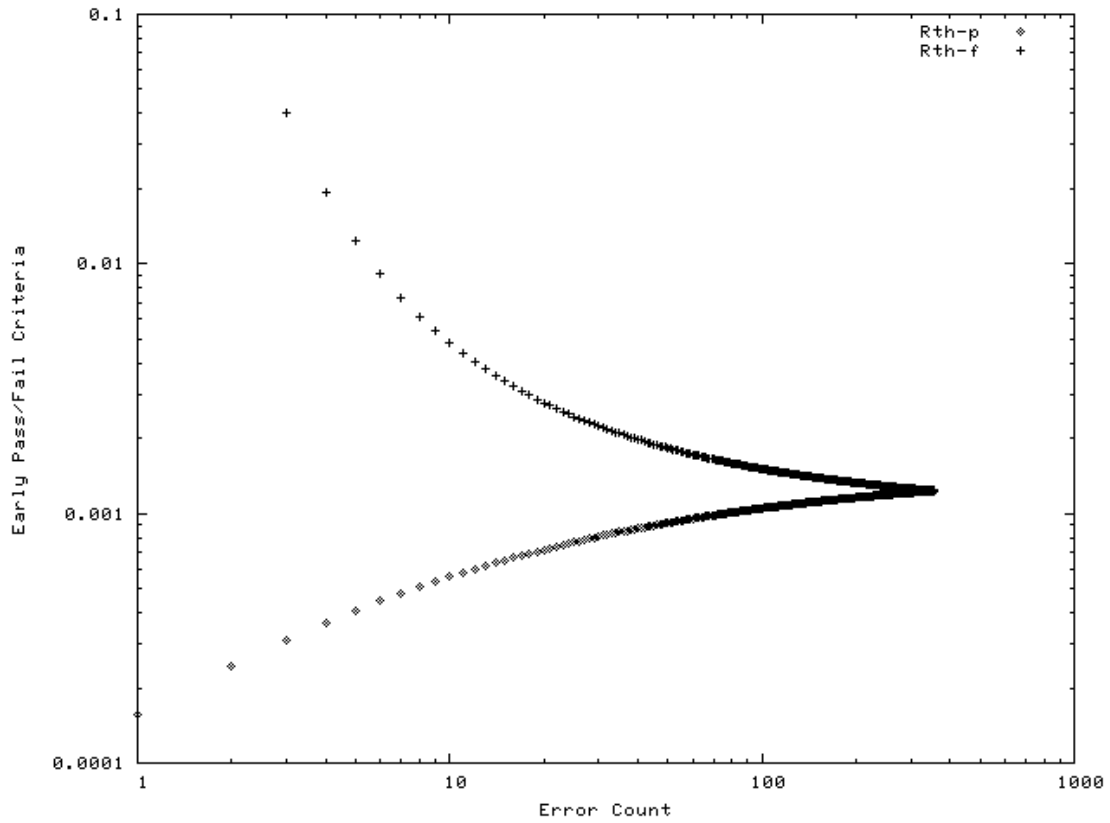


Figure 10-1: Early Pass/Fail Criteria for BER Measurement ($R_0 = 0.001$, $F = 0.002$)

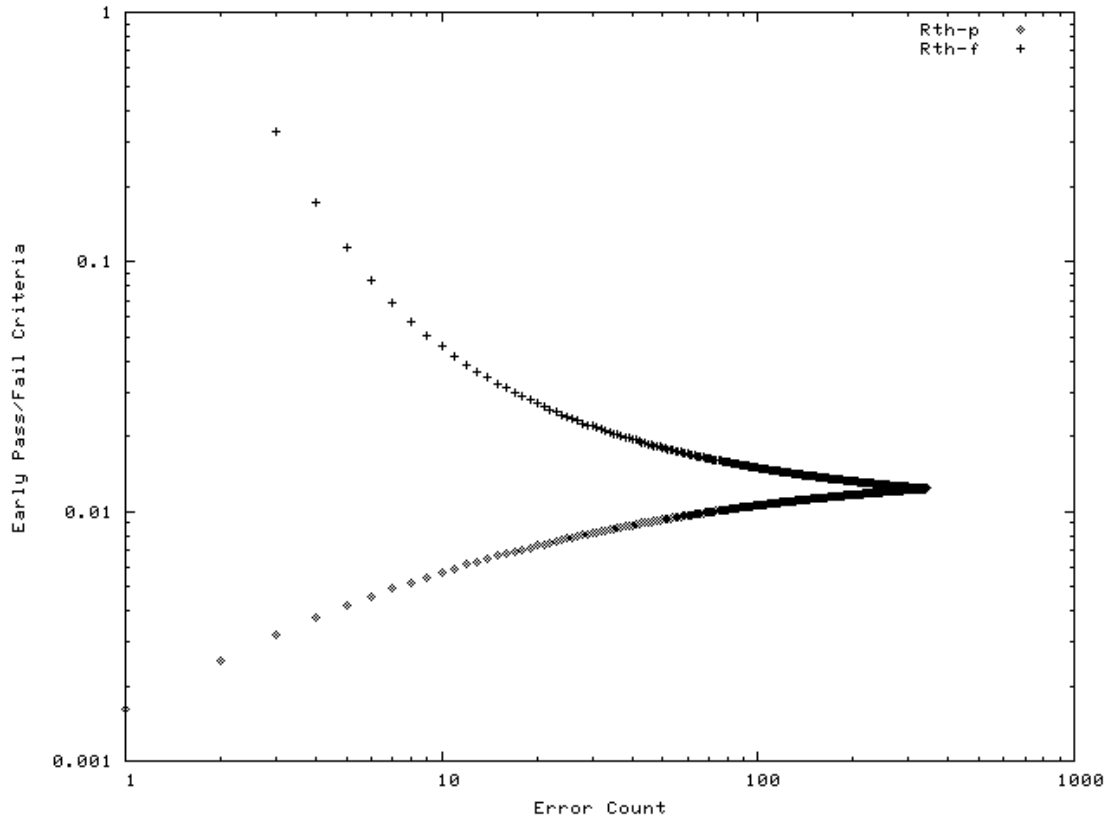


Figure 10-2: Early Pass/Fail Criteria for BLER Measurement ($R_0=0.01, F=0.002$)

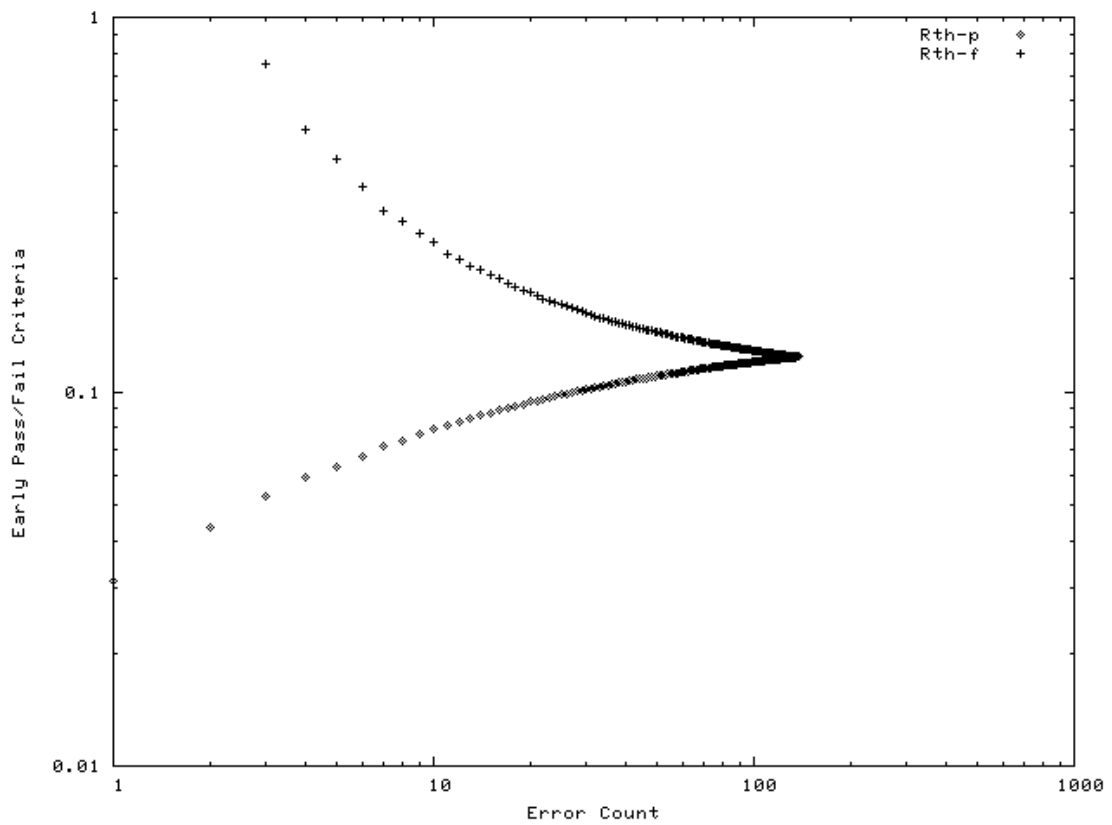


Figure 10-3: Early Pass/Fail for RRM Delay Measurement ($R_0=0.1, F=0.05$)

10.3 Estimating the test time

The final goal of this report is to reduce the total test time, so each set of criteria should be rated by estimating its test time. This task is fairly simple, and can be done using almost the same simulation program used in the previous section. The sample number at which the measurement is terminated is recorded and averaged to give the test time of the criteria. The estimation is repeated for various average error rates, r .

Some results are shown in Figure 10-4 through Figure 10-6.

<Figure will be inserted here>

Figure 10-4: Test Time for BLER Measurement (R0 = 0.001, F = 0.002)

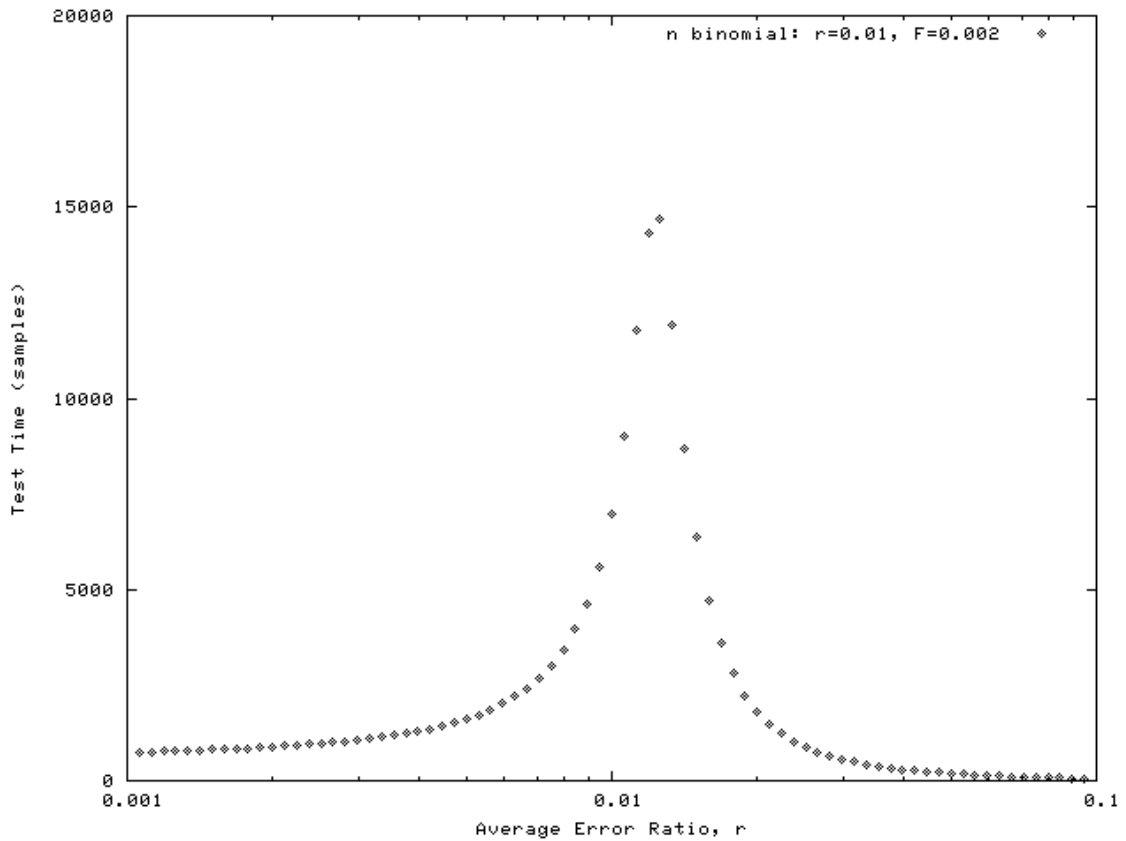


Figure 10-5: Test Time for BLER Measurement (R0 = 0.01, F = 0.002)

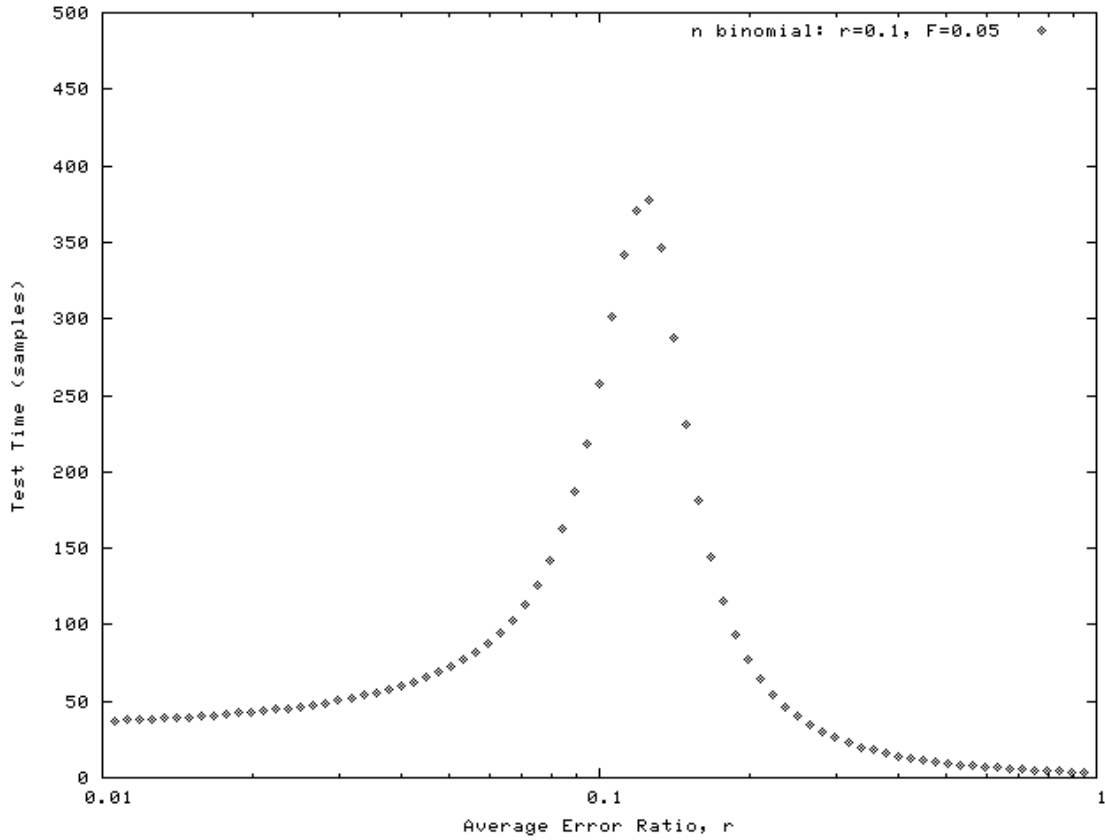


Figure 10-6: Test Time for RRM Delay Measurement (R0 = 0.1, F = 0.05)

In Figure 10-7, the BLER test time (Figure 10-5) is compared with a conventional test method, in which no early criteria are not incorporated. Apparently, the required test time is the time to see 200 errors, and the test time is simply given by $200/r$. It should be noted that the averaged test time values almost coincide at $r = 0.0124$, and the early criteria greatly reduce the test time in the areas of $r \ll 0.0124$ and of $r \gg 0.0124$.

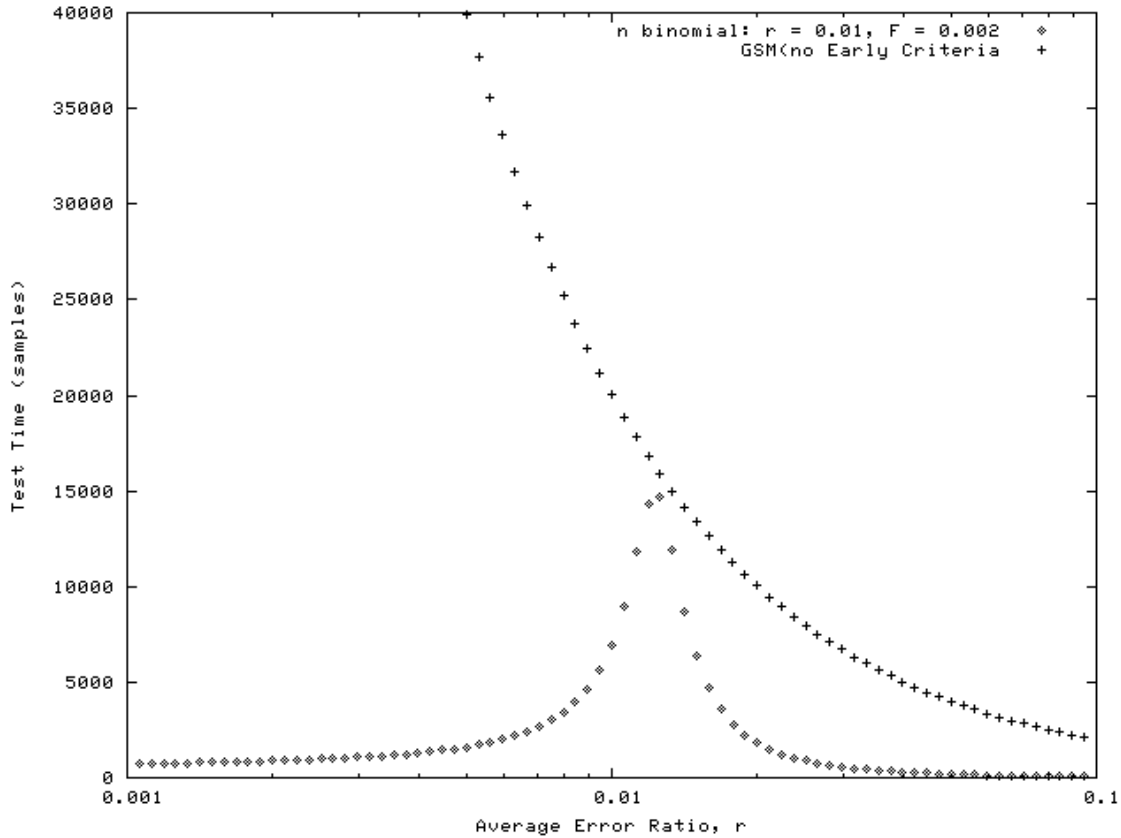


Figure 10-7: Reduction of the Test Time by Employing Early Criteria

10.4 Minimizing the test time

The criteria set in Section 10.2 ("constant D criteria") were chosen in an arbitrary way, and there is no guarantee that that should give the shortest test time. So, to seek the optimized method, some other criteria that are basically a variation of the "constant D criteria" will be proposed and their test time will be estimated in the following subsections.

10.4.1 Truncation

As a matter of fact, while the simple test is terminated at $m = 200$ ($F = 0.2\%$), the cross point of the criteria is located at $m = 374$ for the "constant D criteria." Moreover, only few DUTs survive till later stages. So, it seems to be a good idea truncating the procedure at a little earlier stage.

This means that R_{th_p} and R_{th_f} should be to $1.24 R_0$ at an m value between 200 and 374 in the criteria table. The change is simple, but has an impact on F , so we have to re-evaluate D .

The results are shown in Figure 10-8, which shows no significant improvement in the test time, while the (rare) maximum test time of 1000 samples were reduced to 700 as shown in Figure 10-9. In the figure, the red solid line shows the average test time as a reference.

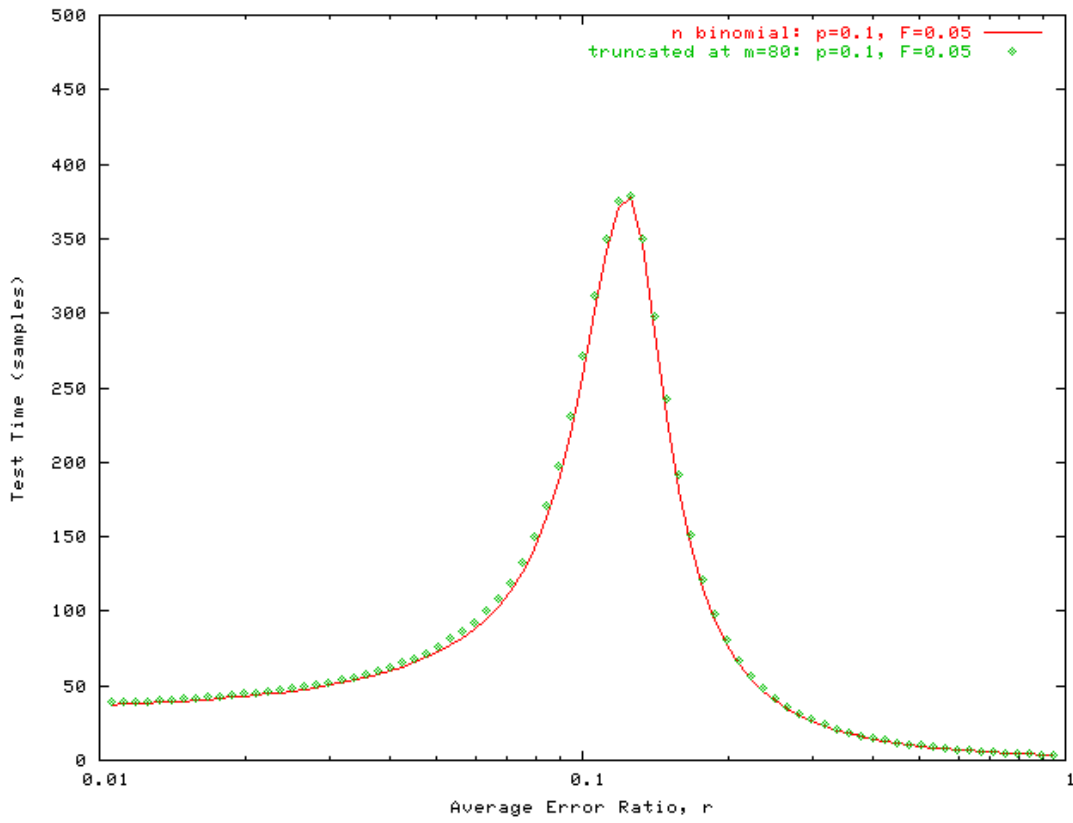


Figure 10-8: Test Time by Truncated Early Criteria ($R_0 = 0.1, F = 0.05$)

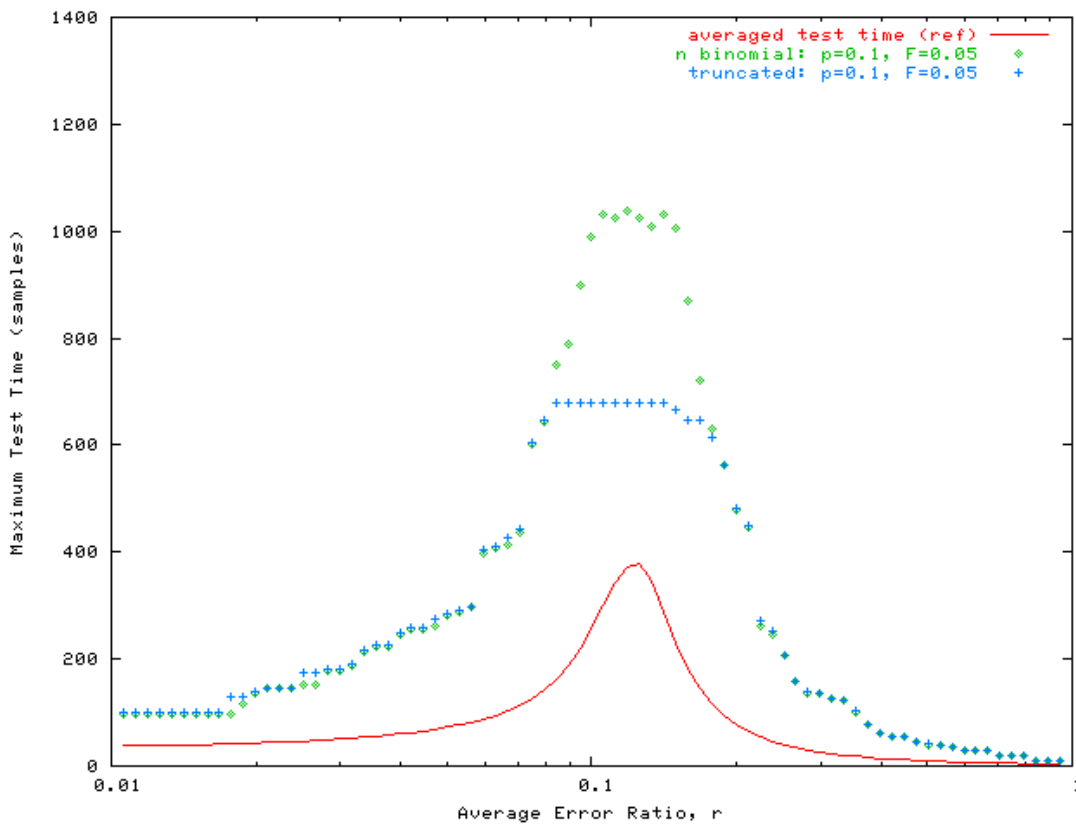


Figure 10-9: Maxim Test Time by Truncated Early Criteria ($R_0 = 0.1, F = 0.05$)

10.4.2 Decimation

If the simple tests (comparisons) were done at a smaller number of m , that would result in a smaller F value. In other words, a larger D value can be employed for a certain F value, this can mean a reduced test time since the criteria curve with a higher D value will be located at more left hand side. The real expected advantages are that such method can be presented by a much shorter table and that it may reduce the requirement for calculation.

The resultant test time values for the criteria with every m , and one with every 10 m points are shown in Figure 10-10. Apparently, the decimation didn't improve the test time.

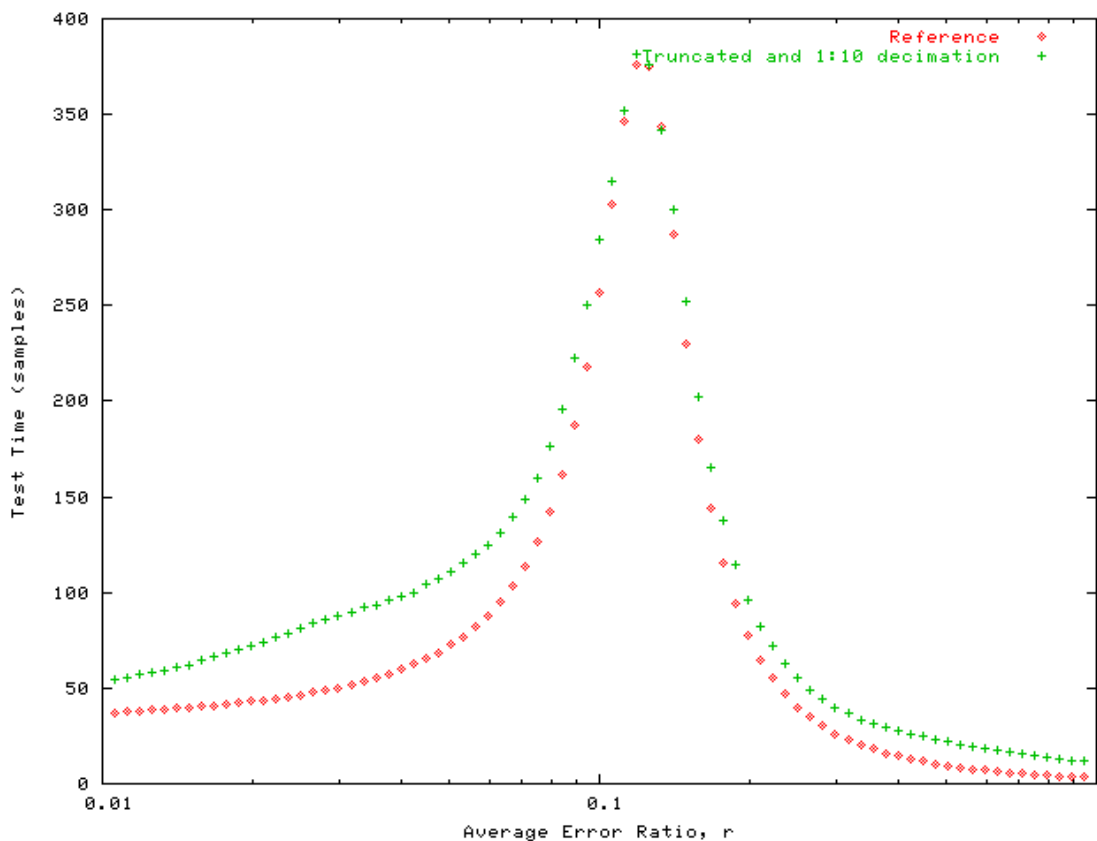


Figure 10-10: Test time of 1:10 Decimated Criteria ($R_0 = 0.1, F = 0.05$)

10.4.3 Composite criteria

The results of the simulation stated above show that almost all the devices are picked out at relatively early stages, and only few reach later stages. This observation would lead to the idea that the test time could be reduced if stricter criteria at the later stages. Suppose that each threshold in a criterion, $R_{th-fix}(m) / R_{th-var}(m)$, have individual significance levels, D_{fix} / D_{var} , respectively. A new criteria, $R_{th-com}(m)$, can be built by merging these two criteria like,

$$R_{th-com}(m) = \frac{m-1}{m_{max}-1} R_{th-fix}(m) + \frac{m_{max}-m}{m_{max}-1} R_{th-var}(m) \quad (m = 1, 2, \dots, m_{max})$$

That is, the composite threshold, R_{th-com} , is a weighted average of R_{th-fix} and R_{th-var} . Where m_{max} corresponds to the cross point of pass/fail criteria of R_{th-fix} , and $D_{fix} > D_{var}$.

The experimental method to find D value witch meets the requirement for F was introduced in Section 10.2, and it can be easily modified to find D_{var} for a fixed D_{fix} . The resultant D values are listed in Table 10-2.

Table 10-2: Combinations of D values to give $F = 0.05$ ($R_0 = 0.1$)

D_{fix}	D_{var}	m_{max}	Comment
---	0.005125	143	Reference
0.01	0.00425	118	
0.02	0.00250	92	
0.03	0.001125	77	

Using these criteria, the test time was evaluated, and the results are shown in Figure 10-11. In comparison with the reference ("Constant D" criteria), the composite criteria give a slightly shorter test time in the worst case ($r = 1.24 R_0$), but these give a longer test time for $r < R_0$.

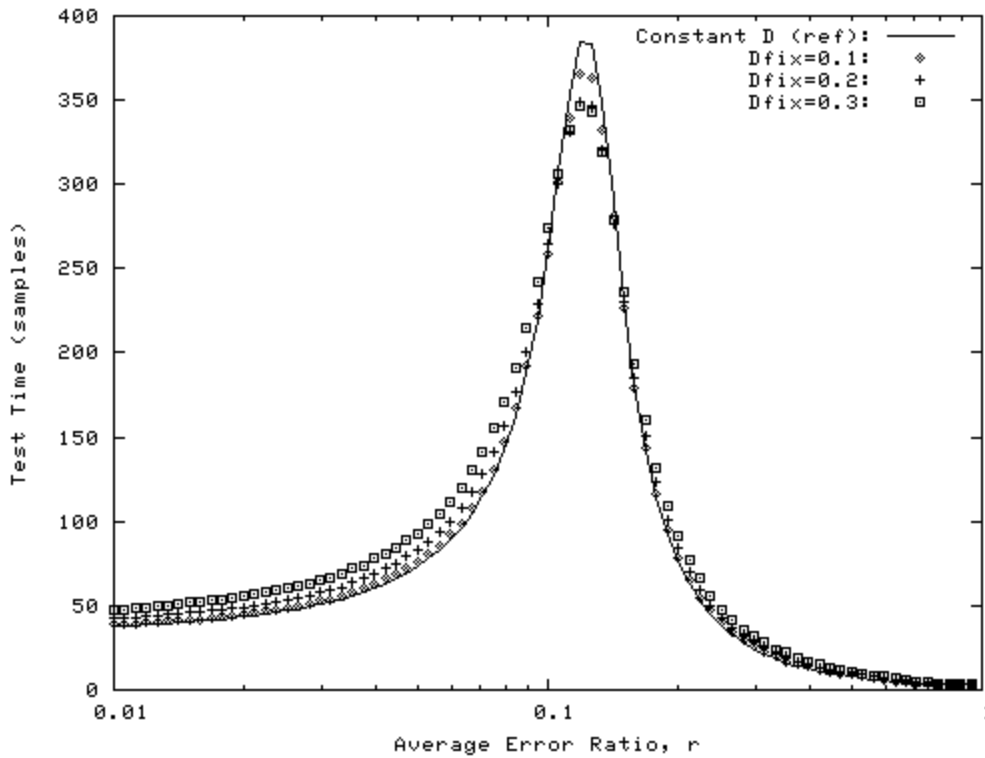


Figure 10-11: Test time for composite test criteria

11 Accuracy of Continuous Distributions

As stated in clause 9.2, the negative binomial distributions can be approximated by m-Erlang distribution. In this Appendix, we will discuss about how good this approximation is.

For continuous distributions, Equation (1') and (2') determine early Pass/Fail thresholds, and these can be transformed to more specific forms shown below by presuming m-Erlang distribution.

$$mER_{CDF}(R_{th-p} : r, m) = 1 - \chi_{CDF}^2(2(r/R_{th-p})m, 2m) = D_p$$

$$mER_{CDF}(R_{th-f} : r, m) = 1 - \chi^2_{CDF}(2(r/R_{th-f})m, 2m) = D_f$$

There is no explicit way to know R_{th} from D , but $\chi^2_{CDF}()$ is a continuous function, so Newton method can be employed to find R_{th} .

The Early Pass/Fail criteria obtained this way are compared with those from negative binomial distributions in Figure 11-1 through Figure 11-3.

For the first two cases (Figure 11-1 and Figure 11-2), the two approaches show a good agreement. However, in Figure 11-3, there is a significant difference between them. This difference results in a significant difference in the test time as shown in Figure 11-4. The difference is largest for a marginal UEs ($r \sim 0.124$).

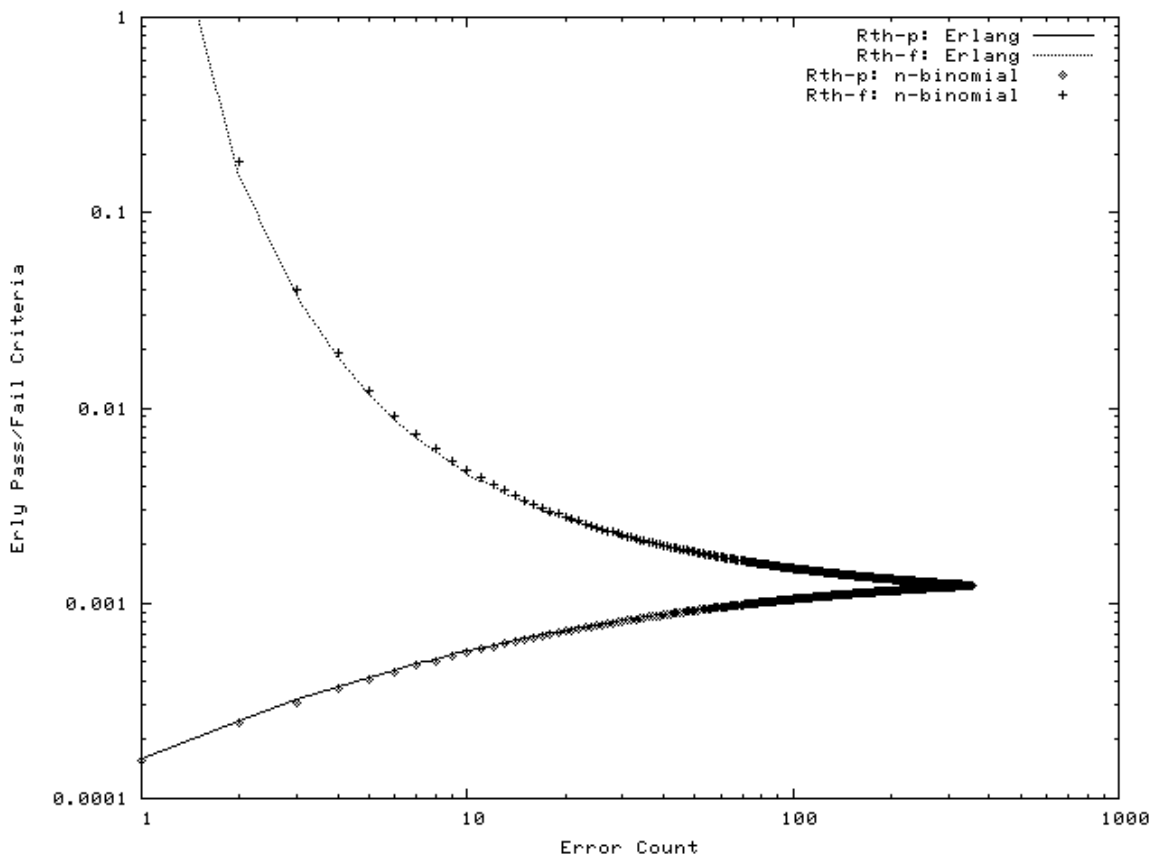


Figure 11-1: Comparison of Early Pass/Fail Criteria for BER Measurement (R0 = 0.001, F = 0.002)

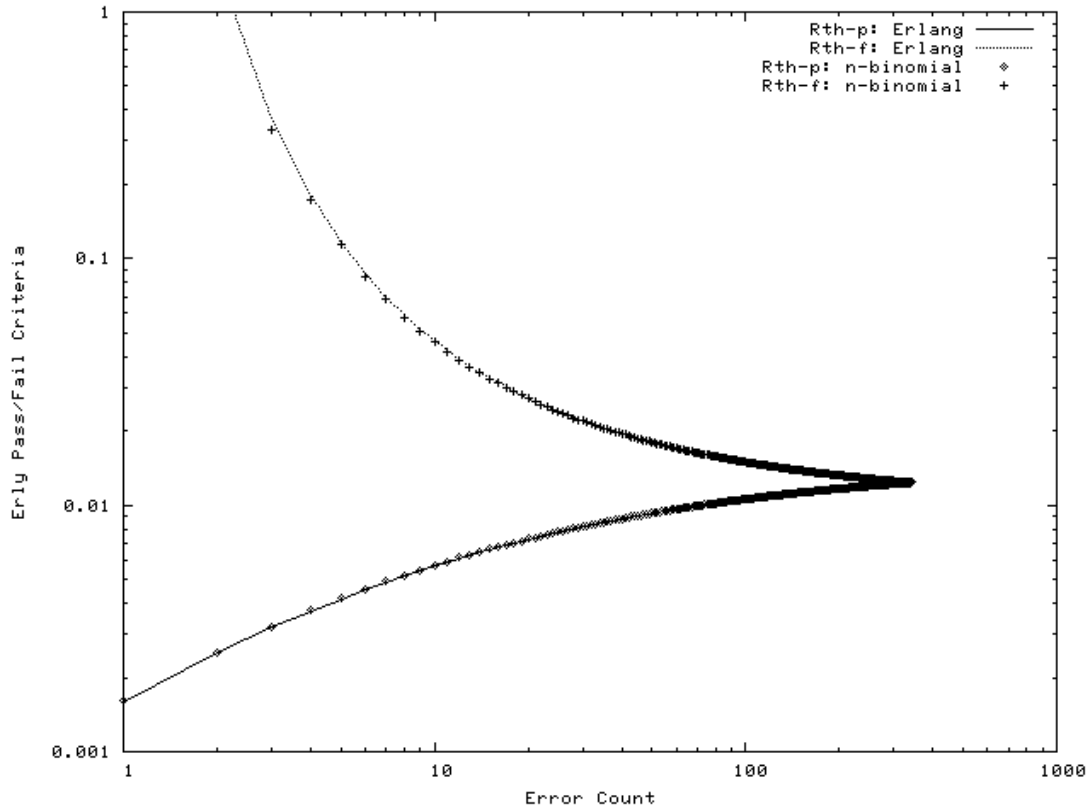


Figure 11-2: Comparison of Early Pass/Fail Criteria for BLER Measurement ($R_0 = 0.01, F = 0.01$)

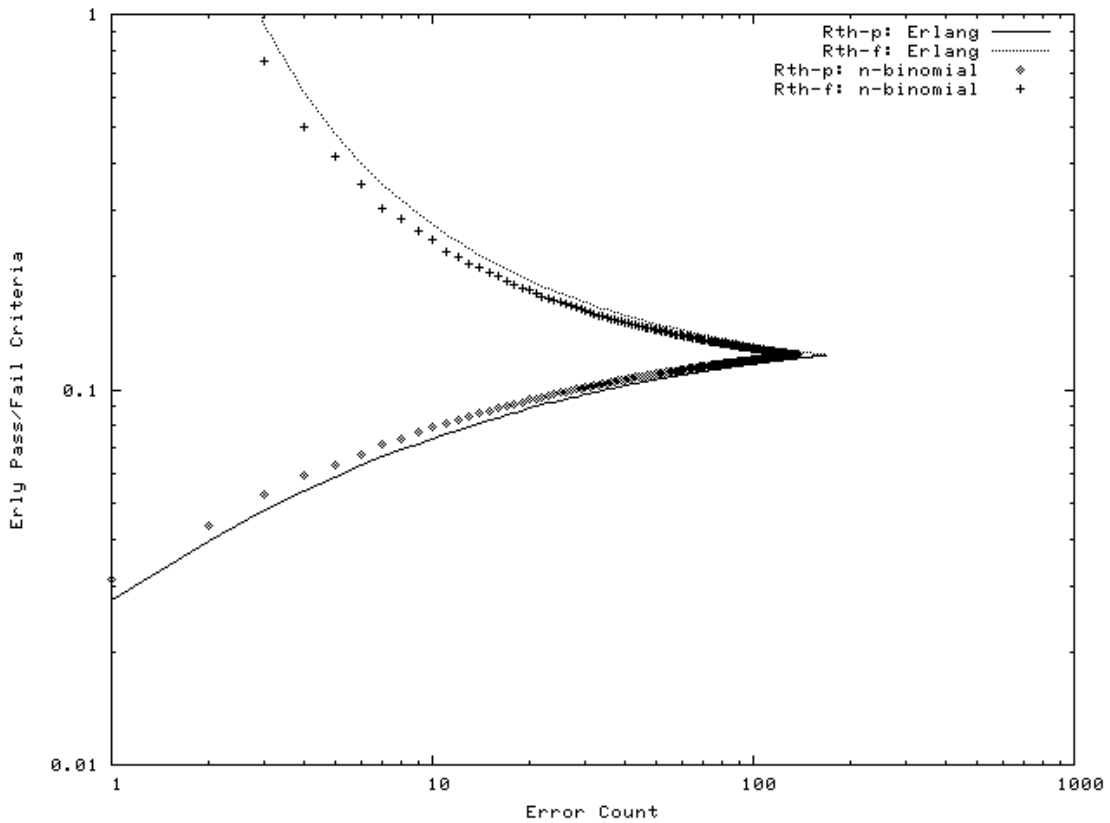


Figure 11-3: Comparison of Early Pass/Fail Criteria for RRM Delay Measurement ($R_0 = 0.1, F = 0.05$)

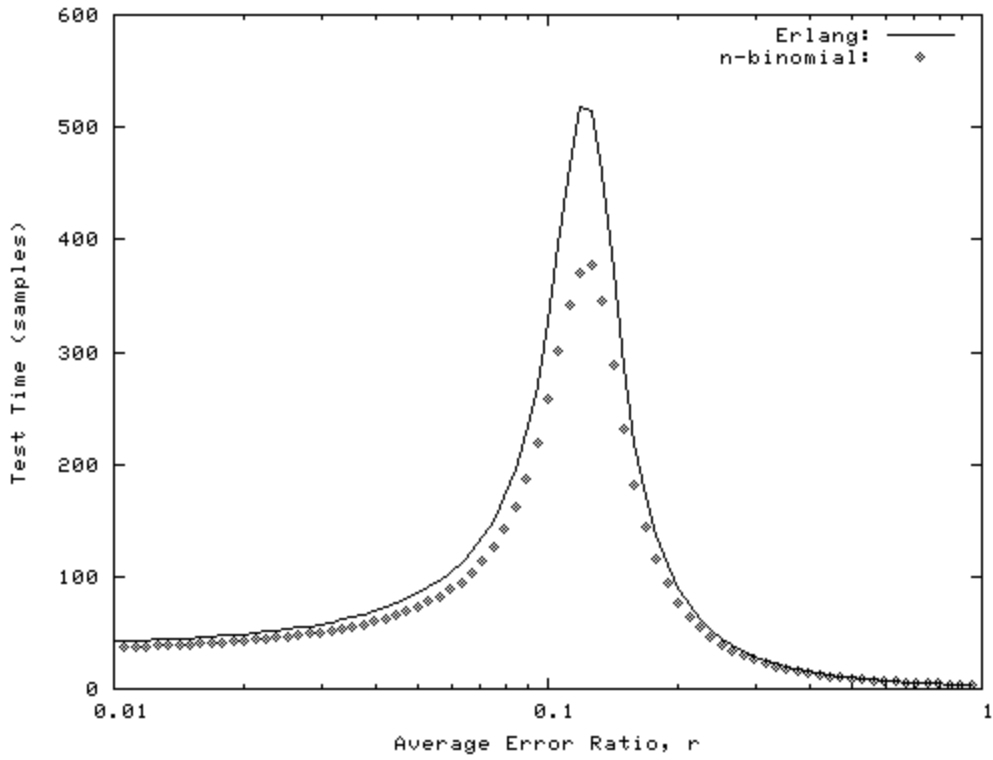


Figure 11-4: Comparison of Test Time (R0 = 0.1, F = 0.05)

Annex A: Change history

T Meeting	Doc-1st-Level	CR	Rev	Subject	Cat	Version -Current	Version -New	Doc-2nd-Level
TP-20				Approval of the report		2.0.1	3.0.0	