Estimators for Fault Tolerance Coverage Evaluation

David Powell
dpowell@laas.fr

Eliane Martins
eliane@dcc.unicamp.br

Jean Arlat
arlat@laas.fr

Yves Crouzet
crouzet@laas.fr

LAAS-CNRS
7 avenue du Colonel Roche
31077 Toulouse - France

Research Report 92471

Revized March 30, 1995

Abstract. This paper addresses the problem of estimating the coverage of a fault tolerance mechanism through statistical processing of observations collected in fault injection experiments. A formal definition of coverage is given in terms of the fault and system activity sets that characterize the input space. Two categories of sampling techniques are considered for coverage estimation: sampling in the whole space and sampling in a space partitioned into classes. The estimators for each technique are compared by means of hypothetical examples. Techniques for early estimations of coverage are then studied. These techniques allow unbiased estimations of coverage to be made before all classes of the sampling space have been tested. Then, the "no-reply" problem that hampers most practical fault-injection experiments is discussed and an a posteriori stratification technique is proposed that allows the scope of incomplete tests to be widened by accounting for available structural information about the target system.

Index Terms. Coverage, fault injection, fault tolerance, estimation, sampling, variance reduction

1. Introduction

A central problem in the validation of fault-tolerant systems is the evaluation of the efficiency of their fault tolerance mechanisms. One parameter used to quantify this efficiency is the *coverage factor* defined as the probability of system recovery given that a fault exists [1]. The sensitivity of dependability measures (such as reliability and availability) to small variations in the coverage factor is well known [1, 2]. Consequently, it is very important to determine coverage as accurately as possible. This paper addresses the problem of obtaining accurate and useful estimations of coverage through the statistical processing of observations collected in *fault injection* experiments.

This work was partially supported by the CEC in the framework of the ESPRIT Basic Research Project n° 6362, Predictably Dependable Computing Systems.

The authors are with the Laboratoire d'Analyse et Architecture des Systèmes (LAAS-CNRS), Toulouse, France. E. Martins is now with DCC - IMECC - UNICAMP, Campinas, S.Paulo, Brazil.

Fault injection can take various forms according to the level of abstraction used to represent the system (from empirical models to physical prototypes) and the level of application of the faults [3]. We consider in this paper the *physical fault injection* approach, where physical faults (stuck-at-0, stuck-at-1, etc.) are directly applied to the pins of the integrated circuits (ICs) that compose a prototype of the system.

A fault injection experiment is performed on a physical system by corrupting the digital signal values of an IC contained in the system via a fault injection module. This fault injection module places the desired fault condition over one or more IC pins at the appropriate point in time. The behavior of the system is observed to determine whether or not the injected fault has been properly handled by the system's fault tolerance mechanisms. This may be done in various ways, for example: by monitoring specific hardware signals in the system, by assessing the truth of application-level predicates that define "correct operation" or by comparing the system outputs with the fault free case (via a duplicate system or by reprocessing the same inputs with the fault injection module deactivated).

One of the difficulties of the approach is the selection of the faults to be injected, since it is not always possible to physically inject all faults that could occur during the system operational life. The determination of these possible faults is another difficulty: the complexity of current VLSI chips and the need to account for temporary faults, which represent the majority of the faults that actually occur in computer systems [4], makes exhaustive testing intractable. As a consequence, coverage evaluation is a problem of statistical estimation, where inferences about a population are based on sample observation.

Since the effect of a fault is dependent on system activity at the moment of its occurrence, we consider a sampling space consisting of the combination of the set of faults and the set of system "activities" at the moment of fault occurrence. As recommended in [5], a *weighted coverage* is calculated: weights are assigned to each point in the sampling space based on their relative probability of occurrence.

The paper is organized as follows. Section 2 gives the basic system definitions and formalizes the notion of a coverage factor. Section 3 presents and compares two categories of sampling techniques for coverage factor estimation: sampling in the whole input space and sampling in a space partitioned into *classes*. Several estimators are analyzed and compared by applying them to data relative to three hypothetical systems. Section 4 proposes further sampling techniques that allow unbiased estimations of coverage to be made before all classes of the sampling space have been tested. Section 5 is devoted to the "no-reply" problem that hampers most practical fault-injection experiments. This problem occurs if some parts of the target system cannot be tested by the fault injection tool due to inaccessibility or parasitic mutations. An *a posteriori* stratification technique is proposed that allows the scope of incomplete tests to be widened by accounting for available structural information about the target system. Finally, Section 6 concludes the paper.

2. Definitions

In this section, we give a formal definition of the coverage *factor* of a fault tolerance mechanism and relate it to the often-used notion of the coverage *proportion.*

2.1. Coverage Factor

We consider a fault tolerance mechanism or fault-tolerant system subjected to faults f in a given fault space F. Let Y be a variable characterizing the coverage of a particular fault, such that $Y = 1$ if the mechanism correctly handles the fault (0 otherwise). What exactly constitutes "correct handling" of a fault depends on the considered fault tolerance mechanism or designer viewpoint, as pointed out in the original definition of coverage [1]. For our purposes here, it suffices to say that the boolean variable Y can be defined in terms of any observable predicate on the system state, such that $Y = 1$ corresponds to the case where the fault is deemed to have been "correctly handled", e.g., the fault has been activated as an error, an error has been detected or the fault has been successfully tolerated, or any boolean function of such or similar predicates [6][1].

The effect of a given fault is dependent on system activity at the moment of, and following, the occurrence of the fault. This system activity can be modelled as a trajectory in the system state space starting from the system state at the moment of fault occurrence and evolving in function of the sequence of system inputs (including the passage of time). Here, we call such a trajectory a system *activity* and let *A* represent the set of all possible activities in the considered operational profile of the system.

The input space of a fault tolerance mechanism can thus be considered in two parts: (a) the *activity set A* due to the system's functional inputs, and (b) the *fault set F* which the mechanism should handle. The complete input space of a fault tolerance mechanism is then defined as the Cartesian product $G = F \times A$. The "output" of the mechanism from the viewpoint of fault tolerance is the predicate *Y*.

For a given fault/activity space *G*, coverage is defined as the *cumulative distribution* of the time interval between the occurrence of a fault and its correct handling by a fault tolerance mechanism [3, 7]. In this paper, we focus on the *asymptotic* value of this distribution, which is called the *coverage factor,* defined formally as:

$$c(G) = \Pr\left\{Y = 1 \middle| G\right\} \tag{1}$$

i.e., the conditional probability of correct fault handling, given the occurrence of a fault/activity pair $g \in G$.

Y is a random variable that can take the values 0 or 1 for each element of the fault/activity space *G*, the coverage factor $c(G)$ can be viewed as $E\{Y|G\}$, the expected value of *Y* for the population *G*. In terms of each fault/activity pair $g \in G$, let $y(g) = 1$ if $Y = 1$ when the system is submitted to g (0 otherwise), and let $p(g|G)$ be the *relative probability of occurrence* of *g*. Expression (1) can then be rewritten as:

$$c(G) = \sum_{g \in G} y(g)\mathrm{p}(g|G) \tag{2}$$

It should be stressed that the distribution $p(g|G)$ is an inherent part of the very *definition* of coverage as a conditional probability parameter to be used, for example, in models for predicting system dependability. In the total absence of knowledge about $p(g|G)$, one can do no better than to assume a uniform distribution. In this case, the coverage factor becomes identical to the coverage *proportion*, which is the subject of the next paragraph.

2.2. Coverage Proportion

Using the same notation as before, the efficiency of a fault tolerance mechanism can alternatively be characterized by the *proportion* of fault/activity pairs in the population *G* which are correctly handled, i.e., the arithmetic mean of the $y(g)$ in *G*:

$$\tilde{y} = \frac{1}{|G|} \sum_{g \in G} y(g) \tag{3}$$

where $|G|$ represents the cardinal of *G*.

[1] The exact definition of *Y* is important when deciding for example the impact of faults that are activated as errors but are not detected despite the fact that no system failure is observed. Such masked faults may be considered to have been "correctly handled" from the application viewpoint (no failure occurred), but certainly not from the error detection viewpoint.

The relationship between $c(G)$ and $\tilde{y}$ can be clarified by introducing a variable *P*, representing the fault/activity occurrence process in G, i.e., P has the values $p(g|G)$. The *covariance* between *Y* and *P*, S_{YP} is, by definition:

$$S_{YP} = \frac{1}{|G|}\sum_{g\in G}\left(y(g) - \tilde{y}\right)\left(p(g|G) - \tilde{p}\right)$$

where: $$\tilde{p} = \frac{1}{|G|}\sum_{g\in G} p(g|G) = \frac{1}{|G|} \qquad (4)$$

From (2-4), it is easy to show that:

$$\tilde{y} = c(G) - |G|S_{YP} \qquad (5)$$

From (5), it can be seen that the coverage proportion $\tilde{y}$ is less (resp. greater) than the coverage factor $c(G)$ as defined in (2) if S_{YP} is positive (resp. negative). A positive correlation ($S_{YP} > 0$) between *Y* and *P* could be expected, for example, if a system designer took account of the relative occurrence probabilities, $p(g|G)$, to implement mechanisms that “cover” the most probable fault/activity pairs [8].

3. Coverage Estimation

The most accurate way to determine $c(G)$ as defined in (2) would be to submit the system to all $g \in G$ and to observe all values of $y(g)$. However, such exhaustive testing is only possible under very restrictive hypotheses (for example, the consideration of permanent stuck-at faults only, or when testing only a small part of the system (e.g., see the test sequence carried out on the computerized interlocking system presented in [3]). For this reason, coverage evaluation is in practice carried out by submitting the system to a subset of fault/activity occurrences $G^* \subset G$. The selection of the sub-set G^* can be done either (i) deterministically, or (ii) by random sampling in the space *G*.

In the first case, the experiments allow one to state, for example, that “*x* percent of the faults injected of type *y* for a workload of type *z* are covered”. Such experiments can be considered as “benchmark” tests that are useful for *comparing* alternative systems or design solutions (e.g., see [9]). However, it is not possible from these experiments to infer anything about the coverage of the system with respect to the complete fault/activity space *G*.

In the second case, if the random selection of $G^* \subset G$ is fair, i.e., there is a non-zero probability that $\forall g \in G$: $g \subset G^*$, then it is possible to make statistical inferences about the complete space *G* based on the observations of the results with respect to G^*. Consequently, conditioned on the assumption that the considered fault/activity space *G* is characteristic of the faults and activities that may occur during the tested system’s operational life, the estimated value of the coverage $c(G)$ can be used in evaluations for predicting measures of the system’s dependability (e.g., see [6]).

This paper focuses on the second approach identified above, i.e., the statistical estimation of the coverage $c(G)$. The important questions that must be addressed are: (i) how to estimate the value of $c(G)$ and to obtain inferences about the error committed in the estimation; (ii) how to select samples and (iii) how to obtain a sufficiently accurate estimation in a reasonable time.

In this section, we successively consider two approaches for estimating coverage factors based on techniques that carry out the sampling (a) directly in the complete space *G*, and (b) in sub-spaces defined by a partition of *G*.

For each technique, the coverage estimator, the variance of the coverage estimator and the estimator of this variance are defined.

3.1. Sampling in a Non-Partitioned Space

We first consider the theory of coverage factor estimation by sampling in a non-partitioned sampling space and then discuss practical aspects of the implementation of such an approach.

3.1.1. Representative Sampling

This sampling technique consists in sampling (with replacement) a group of n fault/activity pairs g in G. To each element of G is assigned a *sampling probability*, $t(g|G)$, such that $\forall g \in G,\ t(g|G) > 0$ and $\sum_{g \in G} t(g|G) = 1$.

To obtain the estimation of $c(G)$, we recall that $c(G)$ is the expected value of the variable Y in the space G. We are therefore faced with a problem of estimation of a population mean. It is shown in the appendix that an unbiased estimator $\hat{c}'(G)$ of this mean and its variance $V\{\hat{c}'(G)\}$ are given by:

$$\hat{c}'(G) = \frac{1}{n}\sum_{i=1}^{n} y(g_i)\frac{p(g_i|G)}{t(g_i|G)} \tag{6}$$

$$V\{\hat{c}'(G)\} = \frac{1}{n}\left(\sum_{g \in G}\left[y(g)\frac{p^2(g|G)}{t(g|G)}\right] - c^2(G)\right) \tag{7}$$

If the sampling distribution is chosen such that $\forall g \in G,\ t(g|G) = p(g|G)$, then (6) and (7) may be rewritten as:

$$\hat{c}(G) = \frac{d}{n} \text{ with } d = \sum_{i=1}^{n} y(g_i) \tag{8}$$

$$V\{\hat{c}(G)\} = \frac{c(G) - c^2(G)}{n} \tag{9}$$

The variance $V\{\hat{c}(G)\}$ can be estimated by:

$$\hat{V}\{\hat{c}(G)\} = \frac{\hat{c}(G) - \hat{c}^2(G)}{n-1} \tag{10}$$

The reader will recognize the well-known formulas for estimating a binomial *proportion*. However, this sampling technique does indeed give an unbiased estimation of the coverage *factor* $c(G)$. The sampling experiments are Bernoulli trials with outcome $Y = 1$ with probability π and $Y = 0$ with probability $1 - \pi$ where $\pi = \sum_{g \in G} y(g)\mathrm{t}(g|G)$. If the sampling distribution were uniform (i.e., $\forall g \in G,\ t(g|G) = 1/|G|$) then we would have $\pi = \tilde{y}$, the coverage *proportion* (cf. (3)). However, by setting the sampling distribution equal to the fault/activity occurrence distribution, we obtain $\pi = c(G)$, the coverage *factor* (cf. (2)). A sample obtained in this way will henceforth be termed a *representative sample*.

3.1.2. Practical Implementation

By definition, a fault/activity pair $g \in G$ corresponds to an activity $a \in A$ and a fault $f \in F$. The representative sampling technique described above therefore requires the random selection of *n* pairs $\langle a, f\rangle$ such that *a* and *f* are selected independently with probabilities $t(a|A) = p(a|A)$ and $t(f|F) = p(f|F)$, where $p(a|A)$ represents the distribution of activities over the considered activity space *A* at the instant the fault occurs and $p(f|F)$ is the distribution of fault occurrences over the considered fault space *F*, such that $p(g|G) = p(a|A) \times p(f|F)$.

3.1.2.1. The activity distribution. The condition $t(a|A) = p(a|A)$ can be satisfied without having to explicitly define the activity distribution $p(a|A)$. The activity space *A* depends on the considered target system and its functional input profile. A real fault can occur at any random point in time, so the probability $p(a|A)$ of it occurring in coincidence with a particular activity *a* (or system state trajectory, cf. Section 2.1) is dependent only on the frequency at which that activity recurs, given the system's functional input profile. Therefore, to ensure that an injected fault "chooses" *a* with a probability $t(a|A) = p(a|A)$, it suffices to simulate this independence between system activity and the instant of occurrence of a real fault. This can be achieved as follows:

a) the target system is reset (to remove the effects of previous experiments) and is then activated with the considered functional input profile;

b) selection of a fault $f \in F$ according to the fault occurrence distribution $p(f|F)$ (see below);

c) the selected fault *f* is injected at some random delay after initiating the system activation.

3.1.2.2. The fault distribution. The definition of $p(f|F)$ can be conceptually simplified by characterizing the set of faults *F* according to different attributes whose distributions can be defined independently. One possible set of attributes suitable for IC-based systems, inspired from those presented in [10], consists of: (a) the *location* of the fault in the target system (i.e., the affected IC), (b) the *multiplicity* of the fault (number of IC pins affected, noted mx), (c) the affected *pins* of the faulted IC, (d) the fault *value model* (stuck-at-0, stuck-at-1, etc.) for each affected IC pin, and (e) the fault *timing model* (transient, intermittent, permanent, etc.) of each affected IC pin. The fault occurrence probability $p(f|F)$ can then be expressed as the product of a set of conditional probabilities defining the distributions of the attributes of the given fault in each attribute category:

$$\begin{aligned} p(f|F) = {} & p(ic|set\_of\_ICs) \\ & \times p(mx|\{1..MX\}) \\ & \times p(pin\_set|set\_of\_pin\_sets(ic, mx)) \\ & \times p(value\_model|set\_of\_fault\_value\_models(mx)) \\ & \times p(timing\_model|set\_of\_fault\_timing\_models) \end{aligned}$$

where *MX* defines the maximum fault multiplicity.

The fault occurrence distribution $p(f|F)$ should be as representative as possible of the faults that will affect the system during its operational life. Data concerning the attribute distributions is of course extremely hard to find so one must resort to assumptions about these distributions based on data from real systems (preferably), from experiments or from fault simulations, or failing all else, on (worst-case) engineering judgement. In the absence of evidence to the contrary, uniform distributions can be adopted. For example, the distribution of the location of the fault, $p(ic|set\_of\_ICs)$, might be determined based on IC failure rate data. As another example, in a system in which error detection is based on a watchdog timer, engineering judgement might dictate that *MX*=1 represents a worst-case scenario. By way of illustration, the distributions used in the experiments described in [3, 11] are presented in Table 1 (although with the hindsight of experience, an alternative set of distributions might have been chosen).

It should again be stressed that the very definition of a coverage factor as a conditional probability (cf. Section 2.1) *must* involve the distribution $p(g|G)$ and thereby $p(a|A)$ and $p(f|F)$. *Any* statement about system coverage should be conditioned by the assumptions that are made about these distributions. The sensitivity of coverage estimations to changes in these distributions is an interesting area for future research and will not be discussed further in this paper.

3.1.2.3. Selection and injection of the sample fault set. Given a definition of F in terms of fault attributes, a representative sample of n faults, noted F_n, can thus be characterized by a set of n vectors where the elements of each vector define the fault attributes that are selected randomly according to the different attribute distributions:

$$F_n = \begin{cases} f(1) = \{location_1,\ multiplicity_1,\ pins_1,\ value\_model_1,\ timing\_model_1\} \\ f(2) = \{location_2,\ multiplicity_2,\ pins_2,\ value\_model_2,\ timing\_model_2\} \\ \qquad \cdots \\ f(n) = \{location_n,\ multiplicity_n,\ pins_n,\ value\_model_n,\ timing\_model_n\} \end{cases}$$

A practical fault injection tool, such as the *MESSALINE* injector developed at LAAS [3], can automatically carry out several fault injection experiments at a single location (IC) of the target system. However, moving the injector probe from one location to another requires manual intervention so it is more practical to sort the set of selected faults by the attribute *location* before carrying out the experiments. Furthermore, the selection of the other attributes can be carried out dynamically for each location to avoid having to store all the *n* vectors of F_n.

In a target system with N_c ICs, a practical representative fault injection campaign thus consists of N_c subsets of experiments. A random number of experiments are carried out on each IC of the target system according to the number of occurrences of the given *location* value in F_n. These subsets of the set of possible fault locations effectively partition the fault space *F* and thus the fault/activity space *G* into N_c disjoint subsets. We will now consider other techniques for estimating the coverage factor that rely on such a partitioning.

3.2. Sampling in a Partitioned Space

For the sampling techniques that follow, the sampling space *G* is considered as partitioned into *classes*. Each class will be referenced by a subscript, according to the following convention: subscripts in Greek letters are used to refer to classes in the sampling space *G* and subscripts in Latin letters are used to refer to classes in the sample (the reason for this convention will only become apparent in the presentation of the 2-stage sampling techniques described in Section 4.2).

By definition of a partition, the classes form *M* disjoint subsets:

$$G = \bigcup_{\alpha=1}^{M} G_\alpha \text{ such that } \forall \alpha,\beta,\ \alpha \neq \beta,\ G_\alpha \cap G_\beta = \emptyset$$

We can rewrite the coverage factor definition (2) as follows:

$$c(G) = \sum_{\alpha=1}^{M} \sum_{g \in G_\alpha} y(g) p(g|G) = \sum_{\alpha=1}^{M} \sum_{g \in G_\alpha} y(g) p(g|G_\alpha) p(G_\alpha|G) = \sum_{\alpha=1}^{M} p(G_\alpha|G) \sum_{g \in G_\alpha} y(g) p(g|G_\alpha)$$

which can be written: $$c(G) = \sum_{\alpha=1}^{M} p(G_\alpha|G) c(G_\alpha)$$

where $c(G_\alpha)$ is the coverage factor for fault/activity class G_α:

$$c(G_\alpha) = \sum_{g \in G_\alpha} y(g) p(g|G_\alpha)$$

We will now consider two sampling techniques based on the above definitions of a partitioned sampling space.

3.2.1. A Naive Estimator

The first sampling technique that can be considered in a partitioned sampling space is to take an equal number of *representative samples* $n_i = n/M$ in each class $G_i, \forall i \in [1, M]$, to count the number of successfully covered faults for each class, d_i, and to apply an estimator derived directly from that given in (8) for sampling in a non-partitioned space:

$$\hat{c}_{na}(G) = \frac{1}{n} \sum_{i=1}^{M} d_i = \frac{d}{n}$$

The variance of this estimator is given by:

$$V\{\hat{c}_{na}(G)\} = \frac{1}{nM} \sum_{\alpha=1}^{M} \left(c(G_\alpha) - c^2(G_\alpha) \right)$$

We call this a *naive* estimator since it is *biased* if the fault occurrences in each class are not equally probable — it can be easily shown that:

$$E\{\hat{c}_{na}(G)\} = \tilde{c}(G_\alpha) = \frac{1}{M} \sum_{\alpha=1}^{M} c(G_\alpha)$$

Reasoning in the same way as in Section 2, the covariance S_{CP} between the coverage $c(G_\alpha)$ and the fault/activity occurrence probability $p(G_\alpha|G)$ of each class is given by:

$$S_{CP} = \frac{1}{M} \sum_{\alpha=1}^{M} \left(c(G_\alpha) - \tilde{c}(G_\alpha) \right) \left(p(G_\alpha|G) - \frac{1}{M} \right)$$

from which it can be shown that:

$$\tilde{c}(G_\alpha) = c(G) - M\, S_{CP}$$

The estimator $\hat{c}_{na}(G)$ can therefore provide pessimistic or optimistic estimations of the system coverage depending on whether the covariance S_{CP} is positive or negative. This will be illustrated by the examples presented in Section 3.4.

3.2.2. Stratified Sampling

In a stratified sampling, a number of samples n_α is predetermined[2] for each class or *stratum* $G_\alpha, \forall \alpha \in [1, M]$. For each class, a *representative* sample of size $n_i = n_\alpha$ is taken and the class coverage factor is estimated using (8) applied to the class instead of the complete sampling space:

$$\hat{c}(G_i) = \frac{d_i}{n_i} \tag{11}$$

where d_i is the number of covered faults in class G_i. The system coverage factor is then estimated by:

$$\hat{c}_{st}(G) = \sum_{i=1}^{M} p(G_i|G)\hat{c}(G_i) \tag{12}$$

The variance of this estimator is:

$$V\{\hat{c}_{st}(G)\} = \sum_{\alpha=1}^{M} p^2(G_\alpha|G)V\{\hat{c}(G_\alpha)\} \tag{13}$$

where the variance of the estimator $\hat{c}(G_\alpha)$ is given by (9) applied to the class G_α:

$$V\{\hat{c}(G_\alpha)\} = \frac{1}{n_\alpha}\left(c(G_\alpha) - c^2(G_\alpha)\right) \tag{14}$$

Similarly, the variance $V\{\hat{c}_{st}(G)\}$ can be estimated by:

$$\hat{V}\{\hat{c}_{st}(G)\} = \sum_{i=1}^{M} p^2(G_i|G)\hat{V}\{\hat{c}(G_i)\} \tag{15}$$

with, from (10):

$$\hat{V}\{\hat{c}(G_i)\} = \frac{\hat{c}(G_i) - \hat{c}^2(G_i)}{n_i - 1} \tag{16}$$

From (13) and (14) it can be seen that the variance of the estimator of the system coverage depends on the allocation of the sample size in each class, n_α. After the sample size *n* is chosen, there are many ways to divide *n* into the individual classes. Hence our objective is to use an allocation that minimizes $V\{\hat{c}_{st}(G)\}$ under the constraint $n = \sum_{\alpha=1}^{M} n_\alpha$. By using the Lagrange multiplier method [12], it can be shown that $V\{\hat{c}_{st}(G)\}$ is minimal for a given total sample size *n* if the n_α are fixed such that:

$$n_\alpha = p(G_\alpha|G)n\left(\frac{\sqrt{c(G_\alpha) - c^2(G_\alpha)}}{\sum_{\alpha=1}^{M} p(G_\alpha|G)\sqrt{c(G_\alpha) - c^2(G_\alpha)}}\right) \tag{17}$$

[2] The term "*a priori* stratification" is used to underline this fact and to distinguish this approach from "*a posteriori* stratification" considered in Section 5.1.

This means that the best value for the n_α depends upon two parameters, the relative probability of fault/activity occurrences in each class, $p(G_\alpha|G)$, and coverage variability within each class — larger sample sizes should be assigned to classes presenting higher $p(G_\alpha|G)$ and greater dispersion of the coverage values. However, such an optimal sample size allocation requires prior knowledge of $\sqrt{c(G_\alpha)-c^2(G_\alpha)}$ for each class. Since these are not known before the experiments are carried out, we can do no better than to suppose that they are constant for all classes, in which case, (17) may be rewritten as:

$$n_\alpha = p(G_\alpha|G)\,n \tag{18}$$

A sample size allocation as given by (18) will be called a *stratified sample with representative allocation* and the corresponding estimator will be denoted $\hat{c}_{stR}(G)$. By substituting the value of n_α in (11) and using (12) we obtain the following estimator for the system coverage:

$$\hat{c}_{stR}(G) = \sum_{i=1}^{M} p(G_i|G)\frac{d_i}{n\,p(G_i|G)} = \frac{1}{n}\sum_{i=1}^{M} d_i = \frac{d}{n}$$

which is analogous to the estimator of a representative sample in the whole space presented in (8). However, since a *pre-determined* number of samples is taken in each class, the variance of this estimator is different to that given in (9) — by substituting n_α in (14) and using (13), we obtain (after a little algebraic manipulation):

$$V\{\hat{c}_{stR}\} = \frac{1}{n}c(G) - \frac{1}{n}\sum_{\alpha=1}^{M} p(G_\alpha|G)\,c^2(G_\alpha)$$

Another possible sample allocation can be defined by taking the same number of samples in each class, i.e., $\forall\alpha,\ n_\alpha = n/M$. The estimator, noted $\hat{c}_{stH}(G)$, and corresponding variance for such a *homogeneous allocation* are obtained in a similar way to above:

$$\hat{c}_{stH}(G) = \frac{M}{n}\sum_{i=1}^{M} p(G_i|G)\,d_i \tag{19}$$

$$V\{\hat{c}_{stH}(G)\} = \frac{M}{n}\sum_{\alpha=1}^{M} p^2(G_\alpha|G)\left(c(G_\alpha)-c^2(G_\alpha)\right)$$

3.3. Confidence Intervals

The distributions of the various estimators defined in the previous section are quite complex and difficult to calculate. Therefore, as is common in many statistical analyses, we shall suppose, based on the central limit theorem applied to large sample sizes, that all the estimators are approximately normally distributed around their expected values, such that:

$$P\left\{\left|\hat{X}-E\{\hat{X}\}\right| \quad \sqrt{V\{\hat{X}\}} \le z_\gamma\right\} = \gamma$$

where $\hat{X}$ is an estimator, $E\{\hat{X}\}$ is its expected value (which is equal to the parameter being estimated if $\hat{X}$ is unbiased), $V\{\hat{X}\}$ is the corresponding variance, z_γ is the 100γth standard normal percentile. Under these conditions, a two-sided $100\gamma\%$ confidence interval for $E\{\hat{X}\}$ can be defined by:

$$\hat{X} - K_\gamma \sqrt{\hat{V}\{\hat{X}\}} < E\{\hat{X}\} < \hat{X} + K_\gamma \sqrt{\hat{V}\{\hat{X}\}} \tag{20}$$

where $\hat{V}\{\hat{X}\}$ is an estimate of $V\{\hat{X}\}$. and K_γ is the $[100(1+\gamma)/2]$th standard normal percentile (one-sided confidence limits are obtained by using z_γ in place of K_γ).

Similarly, the accuracy and precision of the various estimation techniques can be compared in terms of the expected range of variation, noted $h(\hat{X})$, with $100\gamma\%$ confidence, of the corresponding estimators, i.e., for estimator $\hat{X}$:

$$E\{\hat{X}\} - K_\gamma \sqrt{V\{\hat{X}\}} < h(\hat{X}) < E\{\hat{X}\} + K_\gamma \sqrt{V\{\hat{X}\}} \tag{21}$$

3.4. Comparative Examples

In this Section, the various estimators defined in sections 3.1 and 3.2 are compared by way of three hypothetical systems whose characteristics are defined in Figures 1, 3 and 5. Each system is partitioned into $M = 50$ classes; the distribution of the coverage, $c(G_\alpha)$, and the relative fault/activity occurrence probability $p(G_\alpha|G)$ of each class are presented, as well as the values of the system coverage, $c(G)$, and the mean coverage per class, $\tilde{c}(G_\alpha)$.

For each example, the value of $c(G)$ is given along with that of the “normalized” covariance (the *correlation factor*) between $p(G_\alpha|G)$ and $c(G_\alpha)$ defined as:

$$\rho_{CP} = \frac{S_{CP}}{\sqrt{V(C).V(P)}} \times 100\%$$

with: $V(C) = \frac{1}{M}\sum_{\alpha=1}^{M}\left(c(G_\alpha) - \bar{c}(G)\right)^2$ and $V(P) = \frac{1}{M}\sum_{\alpha=1}^{M}\left(P(G_\alpha|G) - \frac{1}{M}\right)^2$

Figures 2, 4 and 6 compare the various estimators in terms of the expected range of variation of coverage estimates as defined by (21) with $K_\gamma = 2.58$ (99% confidence). The figures give the bounds of the system coverage estimation as a function of n, the sample size.

From Figure 1, it can be noted that the main characteristics of system A are: a relative homogeneity among the classes with respect to the coverage, low variability of the relative probabilities of the classes and a slight (negative) correlation ρ_{CP} with the consequence that $c(G)$ and $\tilde{c}(G_\alpha)$ are close to one another, with $\tilde{c}(G_\alpha)$ slightly greater than $c(G)$. For this system, the gain in precision provided by stratification is negligible — Figure 2 shows that the theoretical confidence intervals are almost the same.

For system B (Figure 3) there is a greater variability for the coverage and the relative occurrence probability in each class than for system A. Furthermore, the correlation factor ρ_{CP} is positive and greater than 40%. Consequently, $c(G)$ is quite different from $\tilde{c}(G_\alpha)$. The estimator $\hat{c}_{na}(G)$, which converges to $\tilde{c}(G_\alpha)$, provides a very pessimistic value of the system coverage (Figure 4). Concerning the other estimators, it can be noticed that stratified sampling with representative allocation, $\hat{c}_{stR}(G)$, provides better precision in the estimations, especially for small n.

System C (Figure 5) has a very high coverage with quite a high variability over the classes and a large negative correlation with respect to the fault/activity occurrence probabilities. As a consequence, $\tilde{c}(G_\alpha)$ is an optimistic evaluation for the system coverage — the same is therefore true for the estimations provided by $\hat{c}_{na}(G)$ (Figure 6). It can also be noted that the stratification with representative allocation is equivalent to the representative sampling in a non partitioned population. However the gain in precision over a stratification with homogeneous allocation is appreciable.

In summary, these three examples illustrate that:

1) a naive estimation of coverage by the estimator $\hat{c}_{na}(G)$ can be very pessimistic (system B) or optimistic (system C) depending on the sign of the covariance S_{CP};

2) representative stratification allows an appreciable gain in precision when the classes have very different values of $c(G_\alpha)$ (system B).

Concerning the last point, it can be shown [12] that stratification with a representative allocation of samples is in fact never worse than an equivalent sample in the whole space. However, as illustrated by system A (Figure 2), the gain in precision is negligible when the classes are homogeneous.

4. Early Estimations

When validating a prototype of a fault-tolerant system, it is of real practical interest to obtain estimations of coverage as soon as possible to provide rapid feedback to the design process. However, fault injection experiments can be very time-consuming. In this section, we consider two techniques that enable unbiased estimations of coverage to be obtained as soon as possible. We first consider what early conclusions can be drawn when a stratified sampling approach has been followed but not yet completed. Second, we introduce another method, called *two-stage* sampling. A comparison of these techniques is then given.

4.1. Incomplete Stratified Sampling

When a stratified sampling technique has been followed but not yet completed, the sampling space G can be decomposed in two parts:

- the subset G' containing the m classes that have been tested, with $m \in [1, M]$, with a corresponding probability of fault/activity occurrences:

$$p(G'|G) = \sum_{\alpha=1}^{m} p(G_\alpha|G)$$

- the subset $\overline{G'} = G - G'$, containing the classes not yet covered by the tests, with a corresponding probability of fault/activity occurrences:

$$p(\overline{G'}|G) = 1 - p(G'|G)$$

The system coverage $c(G)$ can then be expressed as:

$$c(G) = p(G'|G)c(G') + p(\overline{G'}|G)c(\overline{G'}) \tag{22}$$

Letting $p(G_i|G') = p(G_i|G)/p(G'|G)$, the coverage relative to the part tested, $c(G')$, can be estimated according to (12) for the sub-population G' by:

$$\hat{c}_{st}(G') = \sum_{i=1}^{m} p(G_i|G').\hat{c}(G_i)$$

Since classes are selected deterministically, we cannot make any statistical inference about $c\left(\overline{G'}\right)$ the coverage for the sub-population not yet covered by the tests. The only thing we can say about $c\left(\overline{G'}\right)$ is that it is some value between zero and one! This means that we can only estimate *bounds* on the *system* coverage $c(G)$ taking into account the lack of knowledge about $c\left(\overline{G'}\right)$. If, for a given confidence level, we have: $\hat{c}_{st}(G') - \delta_{G'} \le c(G') \le \hat{c}_{st}(G') + \delta_{G'}$, then, applying this relation to (22), we can write with the same confidence level:

$$p(G'|G).\left[\hat{c}_{st}(G') - \delta_{G'}\right] \le c(G) \le p(G'|G).\left[\hat{c}_{st}(G') + \delta_{G'}\right] + p\left(\overline{G'}|G\right) \qquad (23)$$

From relation (23), it can be concluded that in incomplete stratified sampling, the error in extending the estimation to the complete system space is minimized if $p(G'|G)$ is maximized for a given *m*. This occurs if classes are tested by decreasing order of $p(G_\alpha|G)$.

4.2. Two-Stage Sampling

Another way to obtain unbiased early estimations of system coverage is to use *two-stage* sampling. In this technique, the sampling space is again divided into *M* classes but the sampling process is carried out in two steps:

1) selection of *random* sample of *m* classes among M;
2) selection of a representative random sample of *predetermined* size n_α, $\alpha \in [1, M]$ in each of the *m* classes.

Three different two-stage sampling procedures are considered. They differ by the sampling technique used in the first stage and the type of estimator used:

- sampling without replacement, with an equal probability of choosing each class — a *linear* and a *quotient* estimator are considered, noted $\hat{c}_{2epL}(G)$ and $\hat{c}_{2epQ}(G)$;
- sampling with replacement, with different probabilities A_α of choosing each class G_α — the corresponding estimator is noted $\hat{c}_{2dp}(G)$.

For the sake of simplicity, the theory behind these two-stage sampling techniques is not detailed here — we use results presented in [13] by adapting the notation. The expressions for the different estimators, their variances and variance estimators are presented in Table 2.

In Table 2, the estimators $\hat{c}(G_i)$, the variances $V\{\hat{c}(G_\alpha)\}$ and the variance estimators $\hat{V}\{\hat{c}(G_i)\}$ are given by equations (11), (14) and (16) respectively. Note that the estimators $\hat{c}_{epL}(G)$ and $\hat{c}_{epQ}(G)$ both become equal to the stratified sampling estimator $\hat{c}_{st}(G)$ (cf. (12)) when $m = M$. This is so because the first stage sampling is carried out *without* replacement. Also, the first term of the variances of these two estimators disappears when $m = M$ and the second term becomes equal to that given by (13).

4.3. Comparative Examples

The same three hypothetical systems that were presented in Figures 1, 3 and 5 are used to compare the estimators obtained from an *incomplete stratified sampling* (with classes selected by decreasing probability) and *two-stage sampling*. In all methods, a representative sample size in each class is fixed in advance: $n_\alpha = p(G_\alpha|G)n$. In addition, the 1st-stage class sampling probabilities for estimator $\hat{c}_{2dp}(G)$ are given by: $A_\alpha = p(G_\alpha|G)$.

Figures 7, 8 and 9 compare the various two-stage estimators in terms of the expected range of variation of coverage estimates as defined by (21) with $K_\gamma = 2.58$ (99% confidence). The figures give the bounds of the system coverage estimation as a function of m, the number of classes tested. For incomplete (representative) stratified sampling, the expected value of the partial estimate $E\{\hat{c}_{stR}(G')\}$ is shown, together with the lower and upper limits of the overall coverage, noted $\underline{\hat{c}_{stR}(G')}$ and $\overline{\hat{c}_{stR}(G')}$, obtained from the two extremes of relation (23) with $\hat{c}_{st}(G') = \hat{c}_{stR}(G')$ and $\delta_{G'} = K_\gamma \sqrt{V\{\hat{c}_{stR}(G')\}}$.

For all three systems, the estimator $\hat{c}_{2epL}(G)$ gives the worst results, since the estimations obtained are subject to great variability.

The incomplete stratified sample shows better precision for system B only, especially for $m \geq 13$. This is because class 13 — i.e., the class with the 13th highest value of $p(G_\alpha|G)$, (cf. Figure 3) has a very low coverage value and its inclusion in the sample has a significant effect on the global coverage estimate.

The estimator $\hat{c}_{2dp}(G)$ displays a good precision for systems A and C, but not for system B. This is due to the fact that system B has a greater variability of the coverage values than the other two. Moreover, the estimations provided do not converge to $\bar{c}(G)$ when $m \to M$, because the first stage samples are selected with replacement.

Finally, the estimator that provides quite good results for the three systems considered is $\hat{c}_{2epQ}(G)$. This estimator provides relatively good precision for small values of m. Only for system B is it necessary to wait till $m \approx 30$. This is due to the heterogeneity of the class coverage values of system B, which means that more classes must be tested to reduce the variance of the estimator. In conclusion, this estimator provides the best overall early estimations — moreover, the estimator becomes equivalent to stratified sampling when $m = M$.

5. The No-Reply Problem

One source of estimation errors that is quite common in opinion polls is the "no-reply" problem that occurs when it is not possible to obtain measures from some elements in the sample [14]. In physical fault injection, a similar problem occurs because of non-significant experiments, which can occur for at least two reasons.

1) Some injected faults may not activated[3].

2) Some experiments foreseen in the sample set may not be feasible due to physical problems such as injection probes not adapted to certain circuits or parasitic mutations (e.g., capacitive loading effects) that prevent the target system from working even though no faults are explicitly injected.

In case (1) above, it is sometimes possible to carry out further experiments to increase the effective sample size and restore the level of precision. In case (2), however, it is not usually possible to carry out further experiments. Whatever sampling technique was adopted at the outset, as soon as a circuit is chosen that cannot be sampled, then — from a circuit-only viewpoint — it will be impossible to sample the whole system and the only way to make an

[3] Faults can be observed as being activated only if the fault injector is equipped with a current-sensing device at the level of the injection probe. If this is not the case, then experiments with non-activated faults cannot be discarded, with pessimistic or optimistic consequences on the coverage estimation, depending on the observed fault-handling predicate Y.

inference about the overall coverage is to extend a partial estimate to the bounds expressed by relation (23) for incomplete stratified sampling.

In this section, we study an "*a posteriori* stratification" technique that enables better estimations to be obtained if structural information about the target system is available.

5.1. A Posteriori Stratification

A posteriori stratification is based on the definition of "strata" or classes *after* having sampled the fault/activity space. We consider the case where an attempt is made to test the target system using the homogeneous (*a priori*) stratification technique (estimator $\hat{c}_{stH}(G)$, cf. (19)). The classes are defined here by the integrated circuits that compose the target system.

When a fault is injected directly onto a single pin of an IC (i.e., a fault of multiplicity 1, cf. Section 3.1.2.2), for instance, by forcing a particular voltage pattern on it, the fault that is injected simulates a fault that could have occurred in that IC or, indeed, in any IC that has a pin connected to the same equipotential line (wire). In the general case of faults of multiplicity $mx \geq 1$, the same can be said for the *group* of affected equipotentials[4].

This suggests a different way of "counting" the fault injection experiments — instead of counting the experiments for each circuit, they can be counted by groups of equipotentials. This can be done if the detailed wiring diagram of the target system is available. If the target system has Q equipotentials and faults of multiplicity $mx \leq MX$ are taken into account during the fault injection experiments, the sampling space can be partitioned according to the $\Theta = \sum_{mx=1}^{MX} \theta(mx)$ equipotential groups that can be affected, where $\theta(1) = Q$ and $\theta(mx) \leq \binom{Q}{mx}$ for $mx \in [2, MX]$.

The estimator, noted $\hat{c}''_{st}(G)$, that applies to *a posteriori* stratification is the same as that for *a priori* stratification with a representative sample in each class, i.e., from (11) and (12), and letting G''_i represent a class in the sampling space partitioned according to equipotential groups:

$$\hat{c}''_{st}(G) = \sum_{i=1}^{\Theta} p\left(G''_i \middle| G\right) \frac{d_i}{n_i}$$

where n_i is the number of injected faults that affect equipotential group G''_i, d_i the number of covered faults and $p(G''_i|G)$ is the probability of fault/activity occurrence affecting that equipotential group. The probabilities $p(G''_i|G)$ are calculated from the distributions of the various fault attributes as defined in Section 3.1.2.2, together with the connectivity information derived from a description of the target system, which enables the probability of faults at the IC pin level to mapped onto the corresponding equipotential group [15].

The variance of this estimator cannot be determined in advance however, since the number of faults affecting each equipotential group is no longer fixed but is in fact itself a random variable. Nevertheless, once a set of experiments has been carried out, the number of faults n_α, $\alpha = 1..\Theta$ affecting each equipotential group is known, so the variance of the estimator $\hat{c}''_{st}(G)$, given the actual sample distribution $\left[n_1, n_2, .., n_\Theta\right]$, can be calculated in exactly the same way as for *a priori* stratification:

[4] Note that for *mx*>1, the group of equipotentials could have less than *mx* members if the faulted circuit has more than one pin on the same equipotential.

$$V\left\{\hat{c}''_{st}(G)\middle|\langle n_1,n_2,..,n_\Theta\rangle\right\} = \sum_{\alpha=1}^{\Theta} \frac{p^2(G''_\alpha|G)}{n_\alpha}\left(c(G''_\alpha)-c^2(G''_\alpha)\right) \tag{24}$$

Now, if the original sampling distribution was chosen such that the probability of an injected fault/activity occurring in equipotential group G''_α is equal to $p(G''_\alpha|G)$ and, given that all equipotential groups are sampled at least once, then the random variables $n_\alpha,\ \alpha=1..\Theta$ are distributed according to a positive binomial distribution with parameters *n* (the total sample size) and $p(G''_\alpha|G)$. Therefore, we can write [16, page 73]:

$$E\left\{\frac{1}{n_\alpha}\right\} \approx \frac{1}{np(G_\alpha|G)-\left(1-p(G_\alpha|G)\right)}$$

Taking expectations of expression (24), and approximating for large *n*, we can thus show that the expected estimator variance is given by:

$$E\left\{V\{\hat{c}''_{st}(G)\}\right\} \approx \sum_{\alpha=1}^{\Theta} \frac{p(G''_\alpha|G)}{n}\left(c(G''_\alpha)-c^2(G''_\alpha)\right)$$

Irrespectively of the distribution of injected fault/activity occurrences across the equipotential groups, the estimator variance can be estimated in exactly the same way as for *a priori* stratification, i.e., by adapting (15) and (16) to the new class definition.

It should be noted that the number of equipotential groups to be considered is usually much greater than the number of integrated circuits — this means that the number of injected faults that affect a given equipotential group can be quite small. This can lead to an overall *decrease* in precision due to the higher variance of the estimations of the equipotential group coverage. This will be illustrated by the example given in the next section.

5.2. Comparative Example

The gain in precision obtained using *a posteriori* stratification based on equipotentials is again illustrated by means of a hypothetical example. Since structural information about the target system is necessary to map the fault counts and the fault occurrence probabilities from circuit pins to equipotential groups, structural information from a real target system is considered. The considered system is composed of a card with 111 integrated circuits and 597 equipotential lines. The considered fault-injector (*MESSALINE*) only has adequate test-probes for 64 of the 111 integrated circuits. The “no-reply” problem therefore occurs for the 47 non-testable circuits. Values for the relative probabilities of fault occurrences at the circuit level were calculated based on IC failure rate data. Coverage values were randomly assigned to each equipotential (and thereby to the target system ICs) on a purely hypothetical basis — they do not in any way represent the real coverage values of the target system and serve only as an illustration of the proposed technique. The theoretical coverage of this hypothetical system is 0.99.

The results presented correspond to a fault injection campaign in which the initial plan was to inject 150 faults on each IC of the target system (homogeneous *a priori* circuit-level stratification). For simplicity, we only consider faults of multiplicity $mx=1$. The comparison is based on the upper and lower limits of the overall system coverage obtained from an incomplete stratified sample as expressed by relation (23). The probability of fault/activity occurrences in the tested part of the system, $p(G'|G)$, is obtained by summing the fault/activity probabilities corresponding to classes that have been tested either from the circuit-level *a priori* stratification viewpoint or from the equipotential-level *a posteriori* stratification viewpoint. The estimation of the coverage of the tested part of the system $\hat{c}_{st}(G')$, and the associated two-sided

99% confidence half-interval $\delta_{G'}$, are obtained from the *a priori* and *a posteriori* estimators $\hat{c}_{stH}(G')$ and $\hat{c}''_{st}(G')$, and their corresponding variances (in the latter case, the *expected* variance).

Figure 10 shows the results obtained when the circuits are tested in an order corresponding to decreasing fault/activity occurrence probabilities, discounting the non-testable circuits. The curves are plotted in function of the number of tested circuits m.

It can be seen that as soon as 7 circuits have been tested, the *a posteriori* stratification by equipotential gives better results than the estimation with *a priori* circuit-level stratification. Also, the final result with *a priori* circuit-level stratification is very poor: the 64 testable circuits have a total relative fault/activity occurrence probability of only 58% — the consequent uncertainty on the overall coverage means that it can only be bounded by $0.57 < c(G) < 0.996$. When the additional information concerning the target system structure is taken into account, the *a posteriori* stratification technique gives an appreciable improvement since now the 99%-confidence interval on overall coverage becomes $0.968 < c(G) < 0.994$.

6. Discussion and Conclusions

In this paper, various sampling methods have been presented that can be applied when estimating coverage based on physical fault injection. A formal definition of coverage has been given in terms of the relative probabilities of points in the complete input space of a fault-tolerant system that includes both system activities and fault occurrences. The considered sampling techniques have been compared from the viewpoints of the precision that they procure in the estimation of overall system coverage and the testing effort required to obtain sufficiently precise results.

When all the circuits in the complete system can be tested, *a priori* circuit-level stratification with a representative sample allocation enables an unbiased estimation of system coverage that is never worse than representative sampling in the complete space and is sometimes appreciably better.

Two-stage sampling techniques allow unbiased early estimations of overall coverage to be obtained that are usually better than the bounds on coverage that can be deduced from incomplete stratified sampling. Of the three two-stage sampling techniques considered, a quotient estimator based on first-stage sampling with equal probabilities (without replacement) gives the best results. Moreover, the technique is equivalent to *a priori* stratification when all classes have been tested.

The “no reply” problem that unfortunately affects most practical fault-injection experiments means however that little can be gained from two-stage circuit-level sampling and one must often resort to the poor bounds obtained from incomplete stratified sampling.

By taking into account available structural information about the target system, we were able to consider an alternative stratification technique based on equipotential groups. This stratification technique was used in an *a posteriori* fashion to improve the results obtained after a set of experiments initially carried out with circuit-level stratification in mind.

We are currently carrying a set of fault injection experiments on a real system in which equipotential groups are used as a basis for *a priori* stratification with the aim of avoiding, rather than correcting, the “no-reply” problems posed by inadequate injection probes and parasitic mutations.

Further research on coverage estimation techniques is being considered in two directions. First, we are investigating estimation techniques that allow stratification to be used to estimate coverage confidence limits for systems with very high coverage, i.e., so high that very few (or even zero) fault-tolerance deficiencies are observed during fault injection experiments. Under such extreme conditions, the normal approximation for confidence interval calculation is no longer valid and other techniques must be developed. Second, since the definition of a coverage factor as a conditional probability must involve the distributions of fault and system activity occurrences, it is essential to assess the impact on the estimations of innaccuracies in our assumptions about these distributions. It would thus be of very real interest to study the sensitivity of the results to variations in the distributions. Note however, that if new evidence about the distributions is obtained after carrying out the fault injection experiments then *a posteriori* stratification can often be used to correct the estimations.

Appendix : *Estimator and corresponding variance for generalized global sampling*

Theorem A1: Given n fault/activity pairs obtained by sampling the population G with replacement and with a sampling probability $t(g|G)$ assigned to each element of G, such that $\forall g \in G,\ t(g|G) > 0$ and $\sum_{g \in G} t(g|G) = 1$, an unbiassed estimator of $c(G) = \sum_{g \in G} y(g)\mathrm{p}(g|G)$ is given by:

$$\hat{c}'(G) = \frac{1}{n}\sum_{i=1}^{n} y(g_i)\frac{p(g_i|G)}{t(g_i|G)}$$

Proof: We show that the mathematical expectation of $\hat{c}'(G)$ is equal to $c(G)$:

$$E\{\hat{c}'(G)\} = E\left\{\frac{1}{n}\sum_{i=1}^{n} y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\} = \frac{1}{n}\sum_{i=1}^{n} E\left\{y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\} \qquad \text{(A.1)}$$

Since the sampling is carried out with replacement, then, from the definition of mathematical expectation and from (2) (cf. Section 2.1), we have:

$$E\left\{y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\} = \sum_{g \in G} t(g|G)\left(y(g)\frac{p(g|G)}{t(g|G)}\right) = \sum_{g \in G} y(G)p(g|G) = c(G) \qquad \text{(A.2)}$$

Substituting $E\left\{y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\} = c(G)$ into (A.1), we obtain:

$$E\{\hat{c}'(G)\} = \frac{1}{n}\sum_{i=1}^{n} c(G) = c(G)$$

❑

Theorem A2: The variance of the estimator $\hat{c}'(G)$ defined in theorem A1 is given by:

$$V\{\hat{c}'(G)\} = \frac{1}{n}\left(\sum_{g \in G}\left[y(g)\frac{p^2(g|G)}{t(g|G)}\right] - c^2(G)\right)$$

Proof: The experiments are independent so we can write:

$$V\{\hat{c}'(G)\} = V\left\{\frac{1}{n}\sum_{i=1}^{n} y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\} = \frac{1}{n^2}\sum_{i=1}^{n} V\left\{y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\} \tag{A.3}$$

From the definition of variance and using (A.2):

$$\begin{aligned} V\left\{y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\} &= E\left\{\left(y(g_i)\frac{p(g_i|G)}{t(g_i|G)} - E\left\{y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\}\right)^2\right\} \\ &= E\left\{\left(y(g_i)\frac{p(g_i|G)}{t(g_i|G)} - c(G)\right)^2\right\} \end{aligned}$$

From the definition of mathematical expectation:

$$\begin{aligned} V\left\{y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\} &= \sum_{g\in G} t(g|G)\left(y(g)\frac{p(g|G)}{t(g|G)} - c(G)\right)^2 \\ &= \sum_{g\in G}\left(y^2(g)\frac{p^2(g|G)}{t(g|G)} - 2y(g)p(g|G)c(G) + t(g|G)c^2(G)\right) \\ &= \sum_{g\in G}\left(y^2(g)\frac{p^2(g|G)}{t(g|G)}\right) - 2c(G)\sum_{g\in G}\big(y(g)p(g|G)\big) + c^2(G)\sum_{g\in G} t(g|G) \end{aligned}$$

Since $y(g)$ may take only the values 0 and 1, and by definition $\sum_{g\in G} y(G)p(g|G) = c(G)$ and $\sum_{g\in G} t(g|G) = 1$, we obtain:

$$V\left\{y(g_i)\frac{p(g_i|G)}{t(g_i|G)}\right\} = \sum_{g\in G}\left(y(g)\frac{p^2(g|G)}{t(g|G)}\right) - c^2(G)$$

whence, from (A.3), we finally obtain:

$$V\{\hat{c}'(G)\} = \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{g\in G}\left[y(g)\frac{p^2(g|G)}{t(g|G)}\right] - c^2(G)\right) = \frac{1}{n}\left(\sum_{g\in G}\left[y(g)\frac{p^2(g|G)}{t(g|G)}\right] - c^2(G)\right)$$

❑

Acknowledgement

The authors would like to thank the anonymous referees for their very detailed and constructive comments on an earlier version of this paper.

References

[1] W. G. Bouricius, W. C. Carter and P. R. Schneider, “Reliability Modeling Techniques for Self-Repairing Computer Systems”, in *Proc. 24th National Conference,* 1969, pp. 295-309 (ACM).

[2] T. F. Arnold, “The Concept of Coverage and its Effect on the Reliability Model of Repairable Systems”, *IEEE Trans. Computers*, vol. C-22, pp. 251-254, March 1973.

[3] J. Arlat, M. Aguera, L. Amat, Y. Crouzet, J.-C. Fabre, J.-C. Laprie, E. Martins and D. Powell, “Fault Injection for Dependability Validation — A Methodology and Some

Applications”, *IEEE Trans. Software Engineering*, vol. 16, pp. 166-182, February 1990.

[4] D. P. Siewiorek and R. S. Swarz, *Reliable Computer Systems — Design and Evaluation*, Burlington, MA, USA: Digital Press, 1992.

[5] A. Avizienis and D. Rennels, “Fault-Tolerance Experiments with the JPL-STAR Computer”, in *Proc. 6th Ann. IEEE Computer Society Conference*, San Francisco, CA, USA, 1972, pp. 321-324.

[6] J. Arlat, A. Costes, Y. Crouzet, J.-C. Laprie and D. Powell, “Fault Injection and Dependability Evaluation of Fault-Tolerant Systems”, *IEEE Trans. Computers*, vol. 42, pp. 913-923, August 1993.

[7] J. B. Dugan and K. S. Trivedi, “Coverage Modeling for Dependability Analysis of Fault-Tolerant Systems”, *IEEE Trans. Computers*, vol. 38, pp. 775-787, June 1989.

[8] B. Bjurman, G. M. Jenkins, C. J. Masreliez and J. E. Templeman, “Airborne Advanced Reconfigurable Computer System”, Report no. CR-145024, NASA, 1976.

[9] D. P. Siewiorek, J. J. Hudak, B.-H. Suh and Z. Segall, “Development of a Benchmark to Measure System Robustness”, in *Proc. 23rd Int. Conf. on Fault-Tolerant Computing (FTCS-23),* Toulouse, France, 1993, pp. 88-97 (IEEE Computer Society Press).

[10] J. Arlat, *Dependability Validation by Fault Injection: Method, Implementation, Application,* State Doctoral Dissertation, INPT, France, 1990 (in French).

[11] J. Arlat, M. Aguera, Y. Crouzet, J. Fabre, E. Martins and D. Powell, “Experimental Evaluation of the Fault Tolerance of an Atomic Multicast Protocol”, *IEEE Trans. Reliability,* vol. 39, pp. 455-467, October 1990.

[12] B. Grais, *Statistical Methods*, Paris: Dunod, 1991 (in French).

[13] J. Desbadie, *Theory and Practice of Sample Surveys*, Paris: Dunod, 1966 (in French).

[14] W. G. Cochran, *Sampling Techniques*, New York: John Wiley & Sons, 1977.

[15] E. Martins, *Validation of Distributed Systems by Fault Injection*, Doctoral Dissertation, ENSAE, Toulouse, France, 1992 (in French).

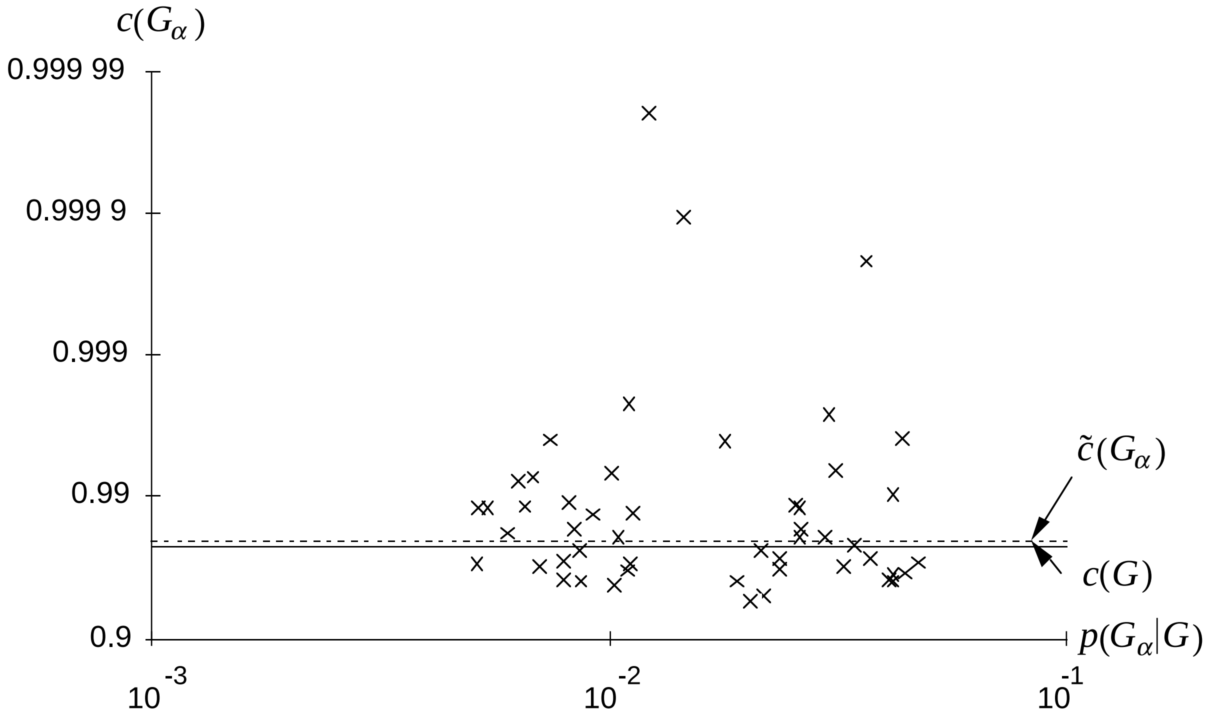
[16] N. L. Johnson and S. Kotz, *Distributions in Statistics — Discrete Distributions*, New York: John Wiley & Sons, 1969.

Table 1 — Fault attribute distributions adopted in [3,12]

Attribute	**Distribution**
Location	$p(ic \mid set\_of\_ICs) = [cardinal\{set\_of\_ICs\}]^{-1}$ (the uniform distribution)
Multiplicity	$MX = 3$ and $p(1 \mid \{1..3\}) = 50\%$, $p(2 \mid \{1..3\}) = 30\%$, $p(3 \mid \{1..3\}) = 20\%$
Pins	$p(pin\_set \mid set\_of\_pin\_sets(ic, mx)) = \left[\binom{N_{ic}}{mx}\right]^{-1}$ (uniform distribution over all combinations of mx pins out of the total of N_{ic} pins for the selected circuit ic)
Value model	Only stuck-at-0 and -1 faults were considered. All 0-1 combinations of *mx* pins were considered equally probable fault values — e.g., for *mx*=2, $\forall value\_model$: $p(value\_model \mid \{00, 01, 10, 11\}) = 0.25$
Timing model	Only intermittent faults were considered in these experiments. Such faults were modelled by asserting the prescribed fault pattern on the selected pins for a given *duration*and with a given repetition *period*. The *period* was logarithmically distributed over the interval $[10\mu s, 30ms]$ and the *duration* was uniformly distribution over the interval $[2\mu s, \min(duration/2; 1ms)]$.

Table 2 — Expressions for two-stage sampling techniques

1st-stage sampling with equal probabilities (without replacement) — linear estimator
$\hat{c}_{2epL}(G) = \frac{M}{m}\sum_{i=1}^{m} p(G_i \mid G)\ \hat{c}(G_i)$
$V\{\hat{c}_{2epL}(G)\} = \frac{M}{m}\ \left(\frac{M-m}{M-1}\right)\sum_{\alpha=1}^{M}\left[p(G_\alpha \mid G)\ c(G_\alpha) - \frac{c(G)}{M}\right]^2 + \frac{M}{m}\sum_{\alpha=1}^{M} p^2(G_\alpha \mid G)\ V\{\hat{c}(G_\alpha)\}$
$\hat{V}\{\hat{c}_{2epL}(G)\} = \frac{M}{m}\ \left(\frac{M-m}{m-1}\right)\sum_{i=1}^{m}\left[p(G_i \mid G)\ \hat{c}(G_i) - \frac{\hat{c}_{2epL}(G)}{M}\right]^2 + \frac{M}{m}\sum_{i=1}^{m} p^2(G_i \mid G)\ \hat{V}\{\hat{c}(G_i)\}$
1st-stage sampling with equal probabilities (without replacement) — quotient estimator
$\hat{c}_{2epQ}(G) = \sum_{i=1}^{m} p(G_i \mid G)\ \hat{c}(G_i) \Big/ \sum_{i=1}^{m} p(G_i \mid G)$
$V\{\hat{c}_{2epQ}(G)\} = \frac{M}{m}\left(\frac{M-m}{M-1}\right)\sum_{\alpha=1}^{M} p^2(G_\alpha \mid G)\ \left[c(G_\alpha) - c(G)\right]^2 + \frac{M}{m}\sum_{\alpha=1}^{M} p^2(G_\alpha \mid G)\ V\{\hat{c}(G_\alpha)\}$
$\hat{V}\{\hat{c}_{2epQ}(G)\} = \frac{M}{m}\left(\frac{M-m}{m-1}\right)\sum_{i=1}^{m} p^2(G_i \mid G)\ \left[\hat{c}(G_i) - \hat{c}_{2epQ}(G)\right]^2 + \frac{M}{m}\sum_{i=1}^{m} p^2(G_i \mid G)\ \hat{V}\{\hat{c}(G_i)\}$
1st-stage sampling with different probabilities (with replacement)
$\hat{c}_{2dp}(G) = \frac{1}{m}\sum_{i=1}^{m}\frac{p(G_i \mid G)}{A_i}\ \hat{c}(G_i)$
$V\{\hat{c}_{2dp}(G)\} = \frac{1}{m}\sum_{\alpha=1}^{M} A_\alpha\ \left[\frac{p(G_\alpha \mid G)}{A_\alpha}\ c(G_\alpha) - c(G)\right]^2 + \frac{1}{m}\sum_{\alpha=1}^{M}\frac{p^2(G_\alpha \mid G)}{A_\alpha}\ V\{\hat{c}(G_\alpha)\}$
$\hat{V}\{\hat{c}_{2dp}(G)\} = \frac{1}{m}\left(\frac{1}{m-1}\right)\sum_{i=1}^{m}\left[\frac{p(G_i \mid G)}{A_i}\ \hat{c}(G_i) - \hat{c}_{2dp}(G)\right]^2$



$c(G) = 0.9780,\ \rho_{CP} = -13.5\%$

Figure 1 — Characteristics of system A

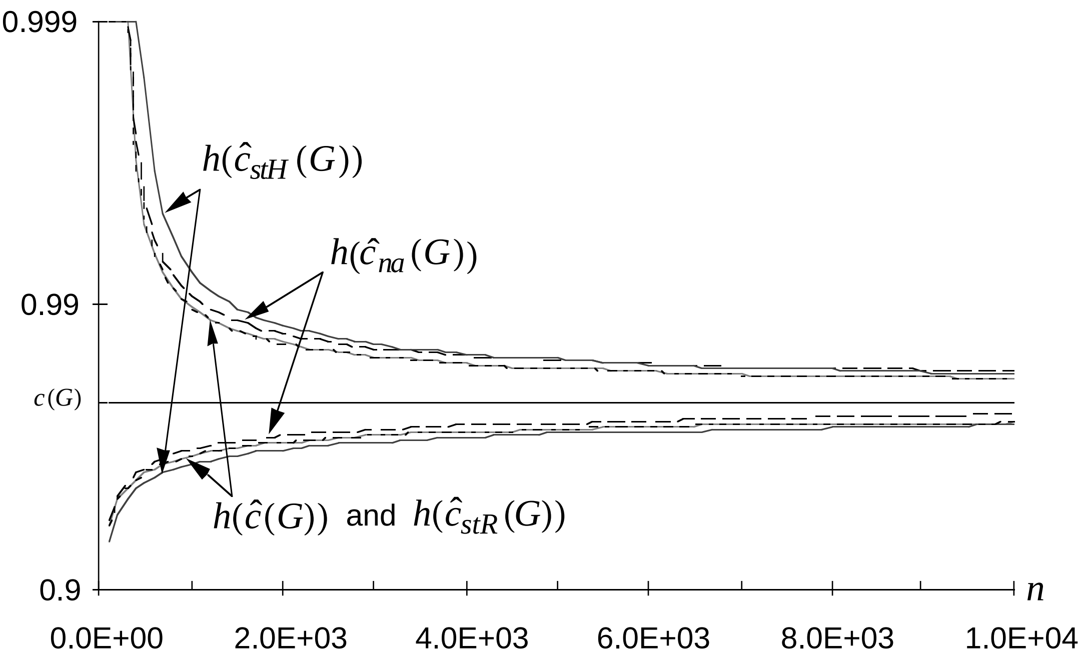
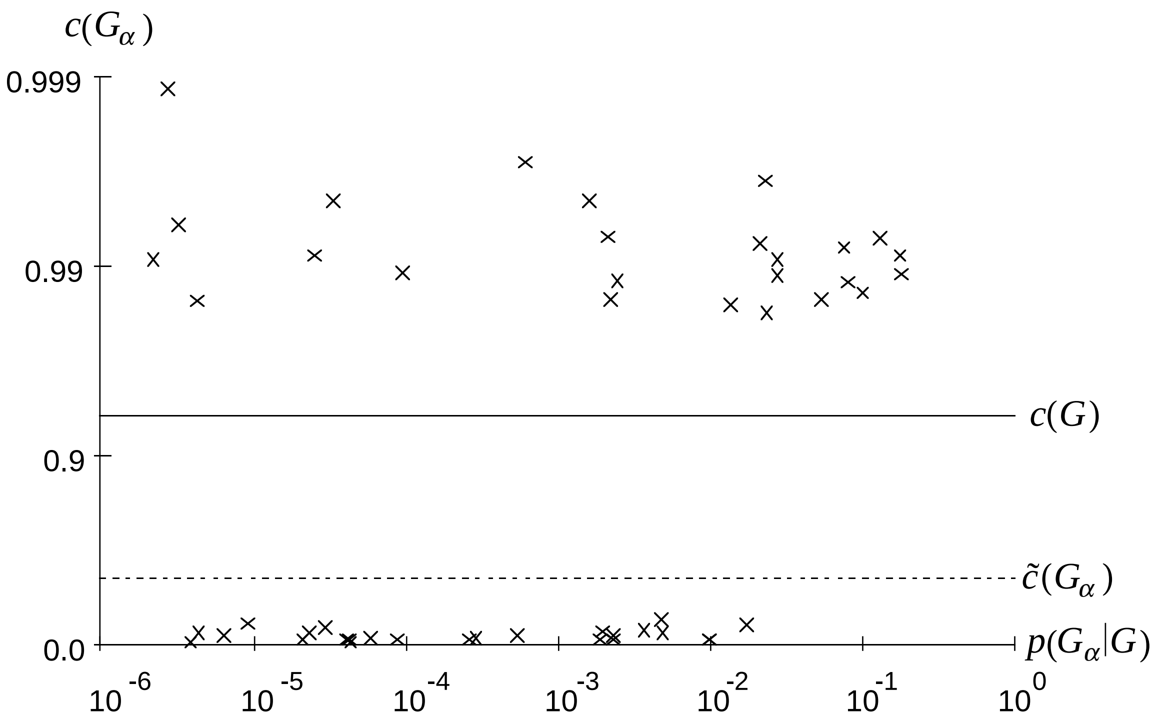


Figure 2 — Expected ranges of variation of coverage estimates for system A (99% confidence)



$c(G) = 0.9379,\ \rho_{CP} = +41.5\%$

Figure 3 — Characteristics of system B

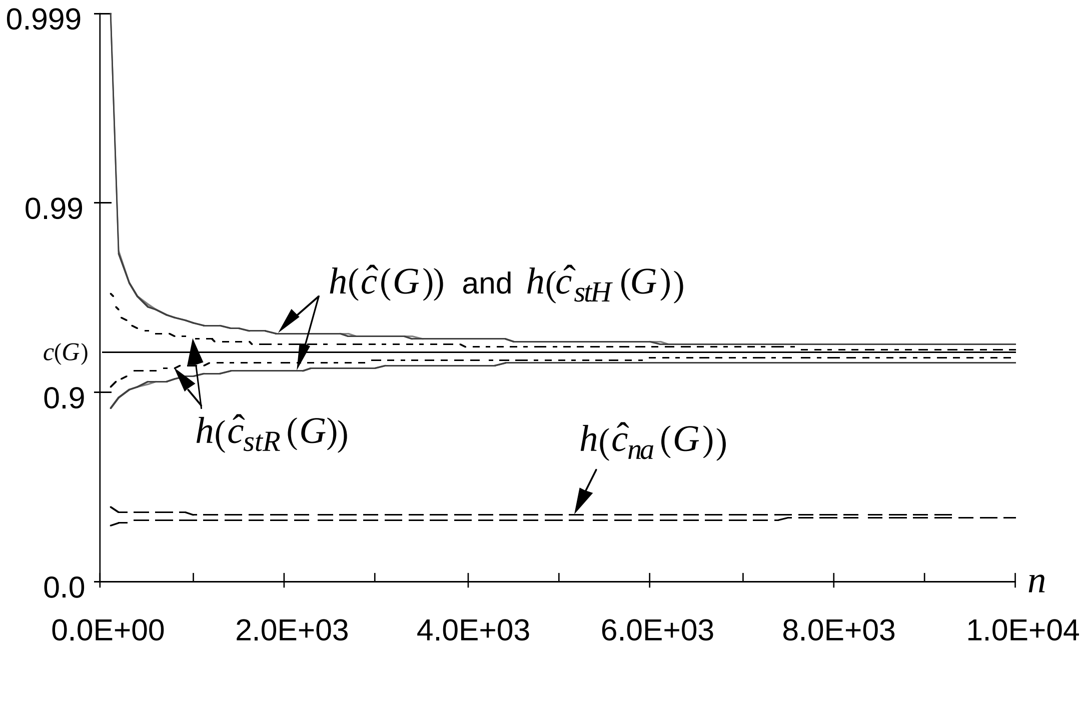
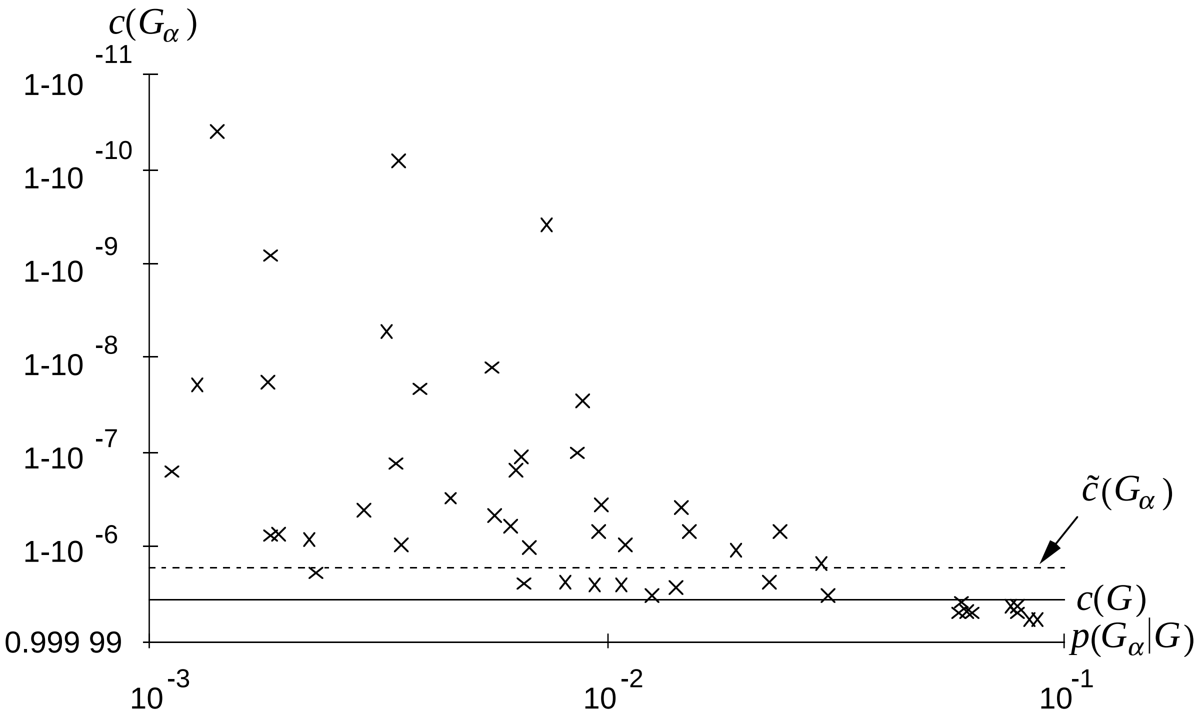


Figure 4 — Expected ranges of variation of coverage estimates for system B (99% confidence)



$c(G) = 0.9999963,\ \rho_{CP} = -89.1\%$

Figure 5 — Characteristics of system C

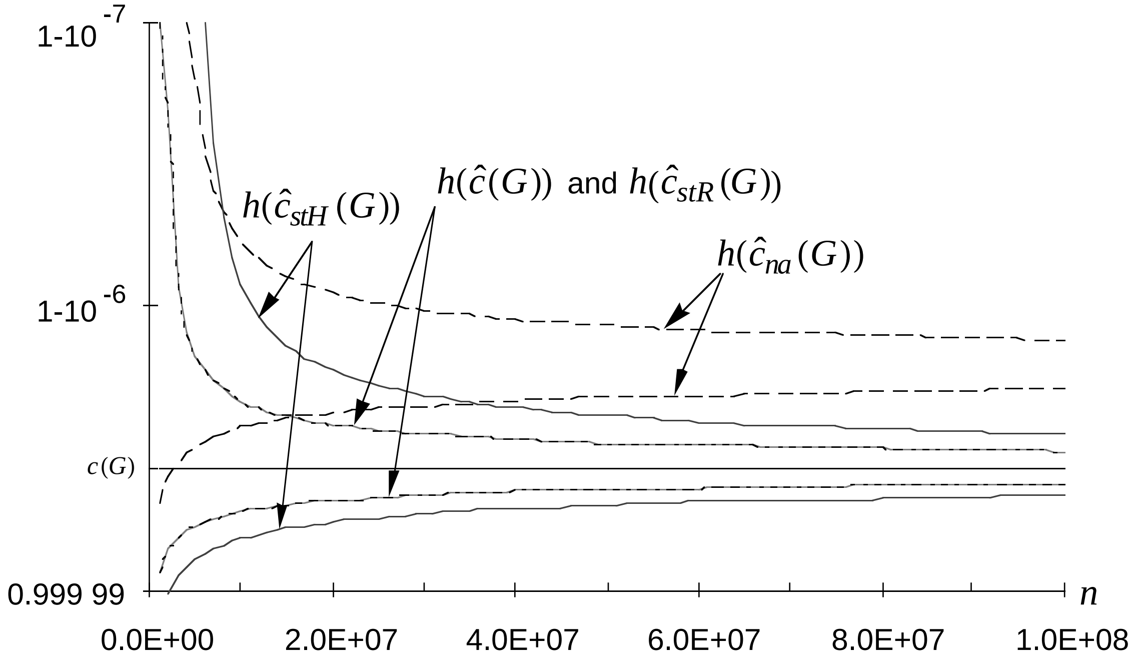


Figure 6 — Expected ranges of variation of coverage estimates for system C (99% confidence)

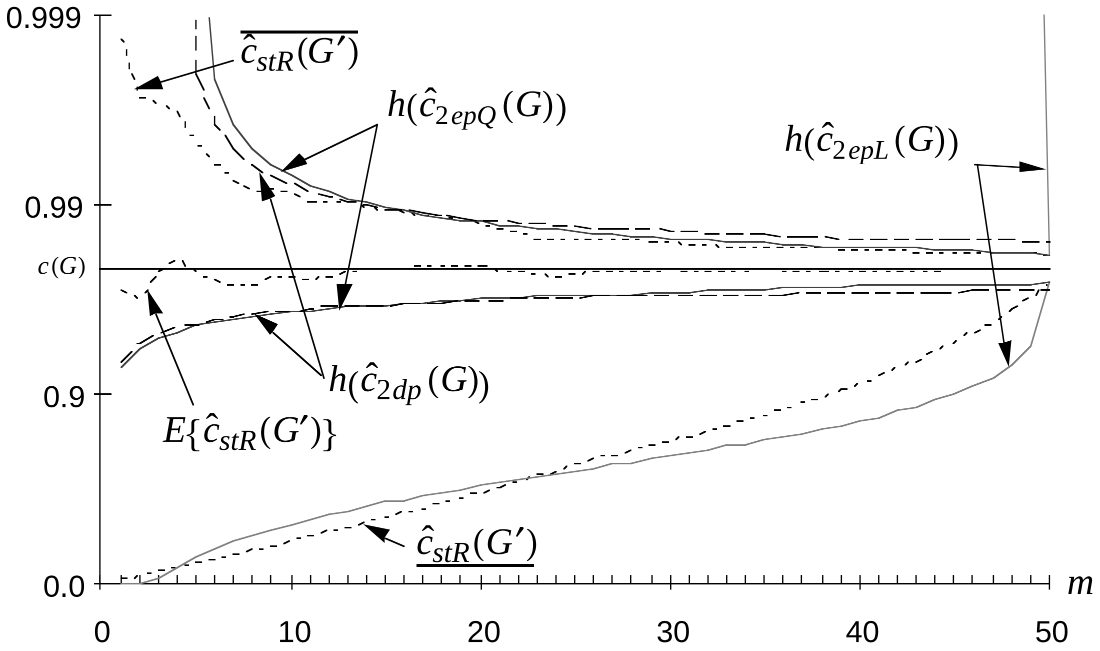


Figure 7 — Expected ranges of variation of coverage estimates for system A (99% confidence)

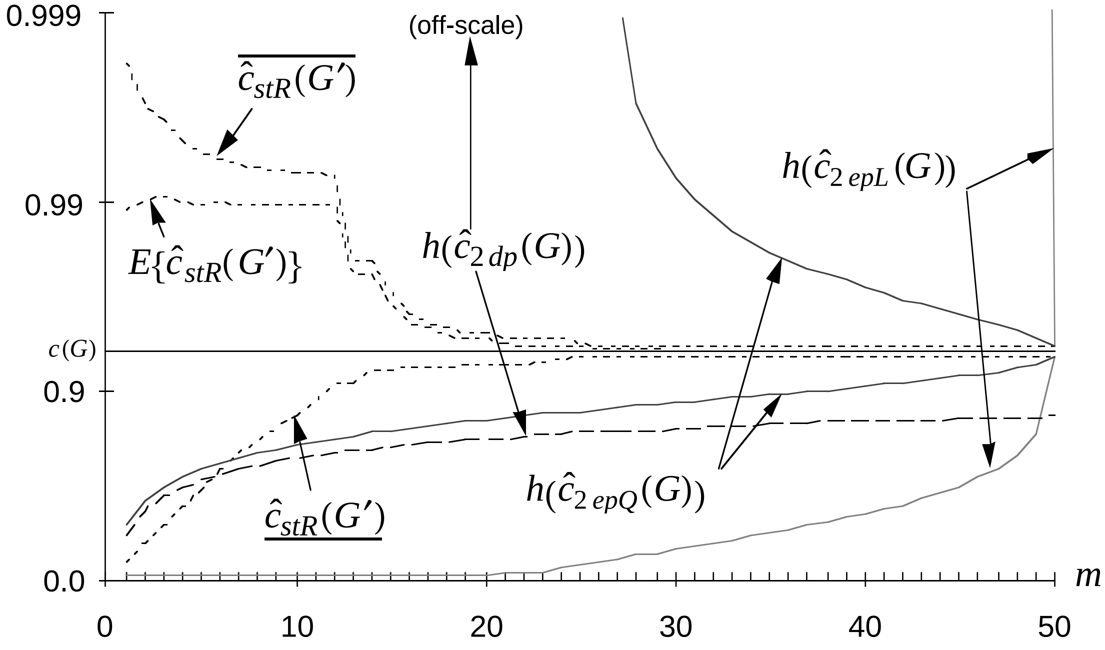


Figure 8 — Expected ranges of variation of coverage estimates for system B (99% confidence)

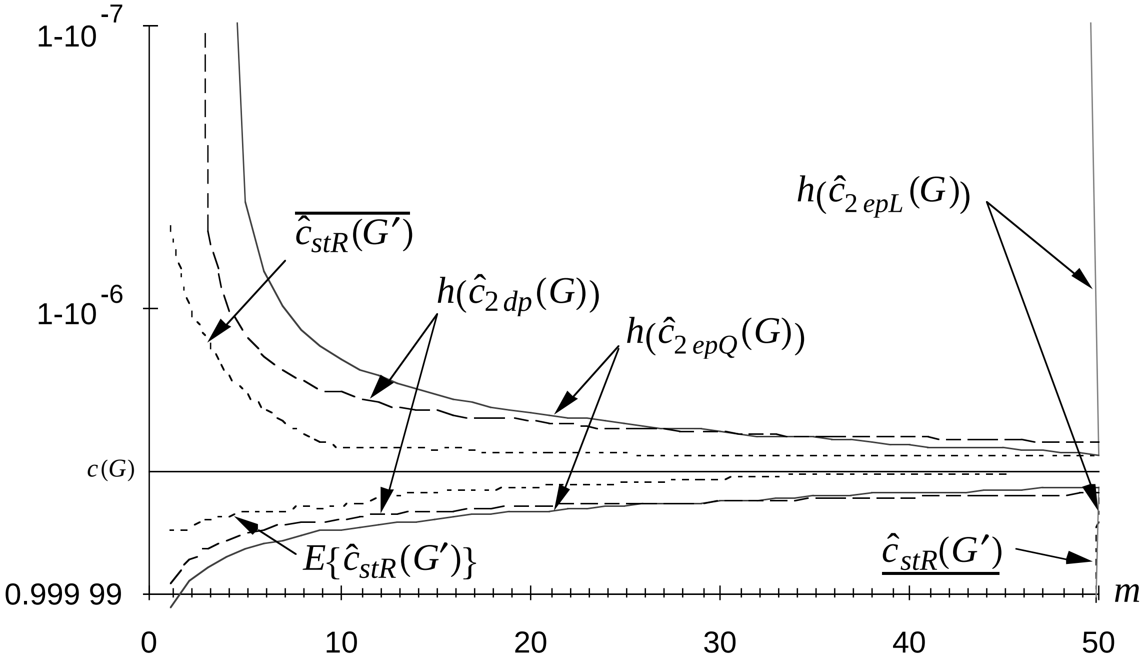


Figure 9 — Expected ranges of variation of coverage estimates for system C (99% confidence)

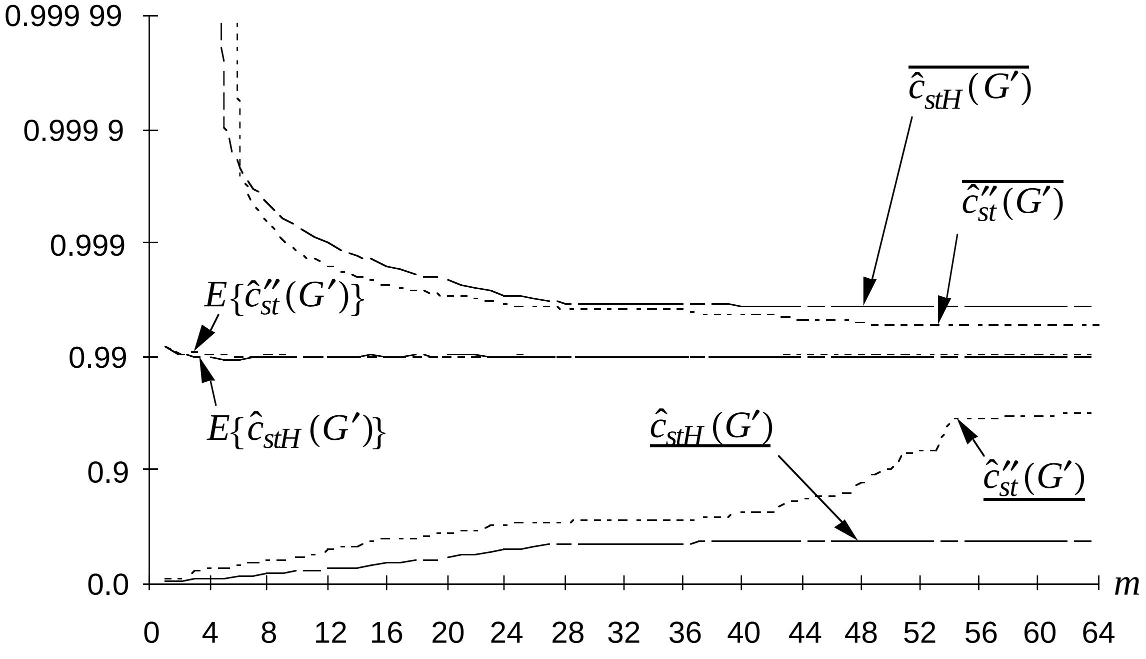


Figure 10 — Reduction of “no-reply” problem with *a posteriori* stratification