# Checks by Monitoring
# Quality Assurance Framework

Based on JRC 119733 and TG CbM QA

# ANNEX V
## Statistical background, version 1.0

December 2020

Authors: Dominique Fasbender, Daniele Borio, Natasa Luketic, Wim Devos

## Table of Content

# 1. Release notes (changes/updates from version 1.0)

This is the second draft version of a document consolidating the statistical considerations relevant for the sampling and acceptance decision of the CbM QA.

The changes from the first draft (from August 2020) as follows:

- A simplified methodology for determining the sample size based on tabulated values is now provided in section 2.4
- The quality limit for α errors has been updated to 10%
- A note on sequential processing has been added for step 2
- The quality limit for the abatable errors has been updated to 5%

# 2. Annex V

## 2.1. Rationale

2.1.1. The purpose of sampling in a quality assurance set up is to provide a robust and reliable overall verdict with a sample (and cost) as small as possible.

2.1.2. This requires the sampling methodology to be unbiased (as can be provided by ubiquitous and ever-present Sentinels) but also that a sample is fixed a priori, ring-fencing the costs (and avoiding continuation of inspection until a favourable outcome is achieved)

2.1.3. This annex elaborates the considerations and choices made for the design of the CbM QA inspections and assessment methodology.

## 2.2. Challenge from paired binary inspection results

2.2.1. In the LPIS QA methodology, the industry standard ISO2859-2 tables are used. This standard describes sampling schemes for inspection by attributes, on a pass-fail basis. Tables set the sample size and acceptance criteria given the population size and limiting quality level.

2.2.2. By contrast, CbM QA is not looking to a pass-fail verdict on the inspected items, but we are investigating false verdicts (i.e both false positives and false negatives). As a result, any population of processed items (parcels or FOI's) is composed of two mutually exclusive sub-populations: the positives, parcels where the phenomenon was present, and the negatives, parcels where such phenomenon did not occur. Any CbM algorithm must be able to process both eventualities and thus the results from an inspection yield a paired binary outcome (00, 10, 01 and 11). The binary pair "xy" should be interpreted as "x was decided while y was the real condition".

2.2.3. This particular paired binary income complicates the sampling. The single population for sampling must be subdivided in its two components: the items where something truly happened and those where nothing happened. This is in theory not a problem, but in practice these actual populations are unknown because the CbM system only yields positives and negatives. So, if a single sample would be used and the number of true occurrences differs in the orders of magnitude from items without occurrences, the size of a single sample that caters for both eventualities would be astronomical.

2.2.4. The next chapter describes this interdependency, assuming the limiting qualities for true positives and negatives from the discussion document (DS-CDP-2018-18) and the robustness of verdict underpinning ISO2859-2.

## 2.3. The formula for a single sample

2.3.1. Let us imagine that we know everything in the population. We would know the exact contingency table between a decision 0-1 and the actual truth 0-1:

$$A_{full} = \begin{vmatrix} N_{00} & N_{10} \\ N_{01} & N_{11} \end{vmatrix}$$

where Nij is the number of cases with Decision i and truth j.

2.3.2. The question is "What should be the sample size n, so that the observed contingency table

$$A_{sample} = \begin{vmatrix} n_{00} & n_{10} \\ n_{01} & n_{11} \end{vmatrix}$$

gives "admissible" estimation of $A_{full}$ in percentage ?"

2.3.3. With the same notations, let us define

$$\alpha = \frac{N_{10}}{N_{00} + N_{10}} = \frac{N_{10}}{N_0}$$

and

$$\beta = \frac{N_{01}}{N_{11} + N_{01}} = \frac{N_{01}}{N_1}$$

2.3.4. and their respective estimators based on $A_{sample}$:

$$\hat{\alpha} = \frac{n_{10}}{n_{00} + n_{10}} = \frac{n_{10}}{n_0}$$

and

$$\hat{\beta} = \frac{n_{01}}{n_{11} + n_{01}} = \frac{n_{01}}{n_1}$$

2.3.5. But as stated, we have no a priori idea of the marginal distributions of $n_{00}$, $n_{11}$, $n_{01}$ and $n_{10}$. According to W. Cochran. Sampling techniques, 1977, page 61, in such conditions, the normal approximations to conditional confidence limits on α is

$$\alpha \in \left[ \hat{\alpha} - \left( t \sqrt{\left(1 - \frac{n_0}{N_0}\right) \frac{\hat{\alpha}\,(1 - \hat{\alpha}\,)}{n_0 - 1}} + \frac{1}{2n_0} \right); \hat{\alpha} + \left( t \sqrt{\left(1 - \frac{n_0}{N_0}\right) \frac{\hat{\alpha}\,(1 - \hat{\alpha}\,)}{n_0 - 1}} + \frac{1}{2n_0} \right) \right]$$

where t is the quantile of the standard normal distribution (defining the confidence level of the interval, e.g. at 97.5%), $N_0 = N_{00}+N_{10}$ and $n_0 = n_{00}+n_{10}$. Using this approximation, we can build a statistical test that the true ratio is smaller than a given quality target (e.g. 5%).

2.3.6. In planning the sample size n for a given precision of $\hat{\alpha}$, we would like to have the minimum value of n so that the precision of $\hat{\alpha}$ is smaller than a target threshold (e.g. 1%). The problem again is that we are sampling n observations but the test is based on $n_0$ (i.e. the number of true "0"s, absences of event in the sample), which is unknown. If we assume further that:

$$\frac{n_0}{N_0} \approx \frac{n}{N}$$

and

$$n_0 \approx xn$$

where x is the expected percentage of true "0"s in the population, then it is possible to find the minimum value of n that satisfies the condition on the expected precision, p, with the formula:

$$t\sqrt{\left(1-\frac{n}{N}\right)\frac{\hat{\alpha}\,(1-\hat{\alpha}\,)}{nx-1}}+\frac{1}{2nx} \le p$$

2.3.7. When planning of n for testing that $\hat{\alpha}$ is equal or smaller than 5%, we would like to have the minimum value of n so that the hypothesis "$H_0$: α is equal or less than 5%" is rejected with a large probability (e.g. 90%) if the actual value of α is equal to $\alpha_1$ (e.g. 10%). With the same conditions on $n_0$ and $N_0$ as above, the minimum value of n that satisfies the condition on the detection of a wrong hypothesis $H_0$ is found with

$$\left[n(\alpha_0-\alpha_1)+\frac{1}{2x}\right]^2 > n^2(t_{0.10}-t_{0.95})^2\left(1-\frac{n}{N}\right)\frac{\alpha_1(1-\alpha_1)}{xn-1}$$

where $t_{0.1}$ and $t_{0.95}$ are the 10% and 95% quantiles of a standard Gaussian distribution.

2.3.8. The same result holds true for β where $\hat{\alpha}$ is replaced by $\hat{\beta}$, $n_0$ by $n_1=n_{11}+n_{01}$ and $N_0$ by $N_1 = N_{11} + N_{01}$. Conditions on both α and β are tested: the most stringent is used to determine n as demonstrated in the examples provided in the next section.

2.3.9. Both formulas are implemented on an excel worksheet and available for download on WikiCAP. Simulating the various parameters shows that the impact of x is predominant; if x is close to 0 or 1 then n will have to be

larger. It follows that if an algorithm is much targeted and specific (i.e. very successful and operating on a high probability of occurrence), a single sample size n should be very large to accommodate those few cases that were not in scope. The population size N is also important but its impact on the size of n is not linear. In fact, large or very large populations make little difference; the sample size levels off after a certain population size.

## 2.4. Setting a single sample size

2.4.1. Let's use the spreadsheet to simulate a condition where the chance of occurring is roughly 50%, i.e. x = .5, α=0.05 and β=0.1 and N=400000. We notice that the highest sample need is 365. If we increate N to 4M, the number stays the same.

| INPUTS | | | | OUTPUTS | | | |
|---|---|---|---|---|---|---|---|
| Population parameters | N (population size) | 400000 | | alpha | | beta | |
| | x (expected percentage of true "1") | 50% | | n for test | 365 | n for test | 175 |
| alpha | | | | n for precision | 185 | n for precision | 315 |
| Statistical test parameters | alpha0 | 0% | | | | | |
| | alpha1 | 5% | | | | | |
| | level of confidence of test | 95.00% | | | | | |
| Precision parameter | p (target precision of alpha) | 5% | | | | | |
| 2 | | | | | | | |
| Statistical test parameters | beta0 | 0% | | | | | |
| | beta1 | 10% | | | | | |
| | level of confidence of test | 95.00% | | | | | |
| Precision parameter | p (target precision of beta) | 5% | | | | | |

2.4.2. Alternatively, if we simulate a condition where the chance of occurring is skewed, i.e. 99% to one side, i.e. x = .99, α=0.05 and β=0.1 and N=400000. We notice that the highest sample needed becomes 15.000. If we increate N to 4M, the number stays the same. This high number comes from the need to have a representative sample for β, but as the occurrences of negatives are rare and unknown, enough observations on them are needed.

| INPUTS | | | | OUTPUTS | | | |
|---|---|---|---|---|---|---|---|
| Population parameters | N (population size) | 400000 | | alpha | | beta | |
| | x (expected percentage of true "1") | 99% | | n for test | 185 | n for test | 8600 |
| alpha | | | | n for precision | 95 | n for precision | 15000 |
| Statistical test parameters | alpha0 | 0% | | | | | |
| | alpha1 | 5% | | | | | |
| | level of confidence of test | 95.00% | | | | | |
| Precision parameter | p (target precision of alpha) | 5% | | | | | |
| 2 | | | | | | | |
| Statistical test parameters | beta0 | 0% | | | | | |
| | beta1 | 10% | | | | | |
| | level of confidence of test | 95.00% | | | | | |
| Precision parameter | p (target precision of beta) | 5% | | | | | |

2.4.3. These findings raise some questions:

- Which is the worst case of the two above situations? We argue the case where x is near 0.5.

- Is there a need to be statistically robust for cases that are rare (with x close to 0 or 1) because the system has eliminated those cases by targeting? We argue there is not.

2.4.4. Indeed, whichever positive cases or negative cases that constitute the smaller share of N are less important than the complementary larger share. The smaller share represents fewer cases and hence has a lower impact. Therefore it should not be driving the sample sizes. The more numerous cases of the larger share should. And in the worst case of equal halves, the maximum sample size need that can occur for any half with the proposed α and β is 365.

2.4.5. This does not mean we abandon the smaller share (whether positives or negatives) altogether. We continue to inspect and statistically analyse its cases that happen to turn up during the inspection of sample. However, we are no longer pursuing the original robustness of verdict (consumer risk) for that smaller share.

2.4.6. To make up practical solutions and avoid complexity and mistakes, we propose to have a single sample of size n as sum of $n_1$ and $n_0$. We established above that the highest value of $n_1$ or $n_0$ in any given share of the population is 365; and we simply ignore its complementary share.

2.4.7. A sample size equal to 365 is obtained for large population sizes, N. While this sample size should be adopted for most cases, smaller sample sizes may be considered if small populations are under analysis. While, in principle, sample sizes can be calculated using the formulas provided above, the procedure for their evaluation is quite complex and the use of tabulated values has been preferred. Simplifications have been adopted to reduce the size and complexity of the table providing the sample sizes as a function of N.

2.4.8. The following principles have been adopted for the generation of the sample size table:

a. The sample should not exceed a practical size. For instance, 365 is the value obtained using the approach detailed above for very large N and the strictest testing parameters.

b. Oversampling provides more reliable tests. Thus, approximation by excess is acceptable where it matters most (while preserving the limiting condition of 365 elements). Oversampling could also allow for on-the-fly follow-up analysis

d. The sample sizes, n, should assume only a limited number of values: 125, 200, 315 and 365, in line with practices of ISO2859

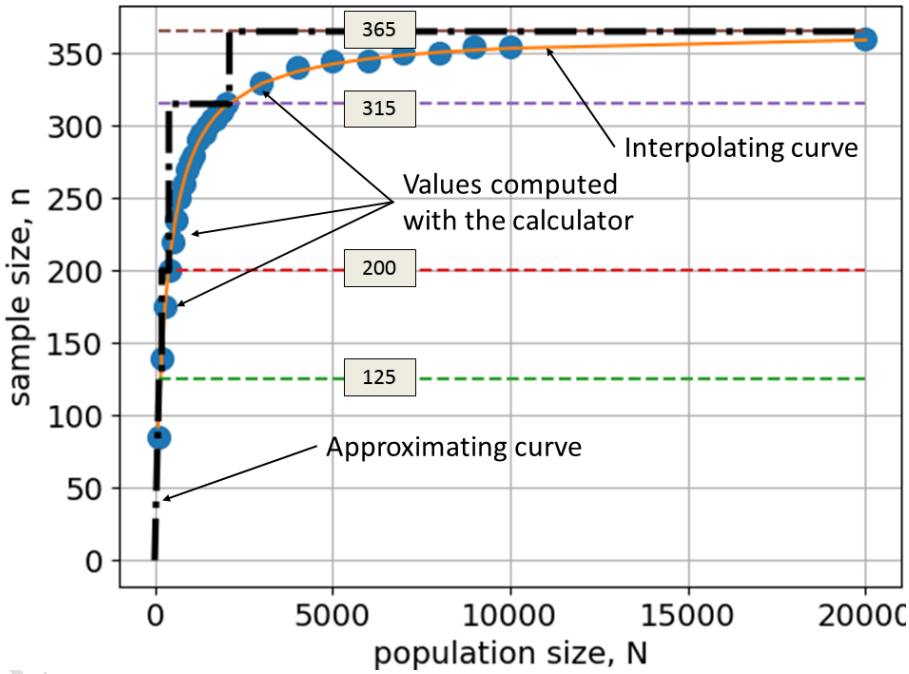
2.4.9. To derive tabulated values for the sample sizes, the following approach has been adopted:

1. Values of n have been computed using the formulas provided above

2. For population sizes (N) lower than 125, the full population should be inspected (n = N)

3. For N > 125 a step-wise approximation has been adopted. This approximation assumes only four values (125, 200, 315 and 365) and respects the condition, N ≥ n. Upper-bounding and oversampling have been adopted were possible.

2.4.10. The result obtained using this approach is shown in the figure below.




2.4.11. Tabulated values of n as a function of N are provided below:

| N | [1-124] | [125-199] | [200-399] | [400-2100] | > 2100 |
|---|---|---|---|---|---|
| n | N | 125 | 200 | 315 | 365 |

2.4.12. The table has been obtained by assuming a quality limit equal to 5% for α and a quality limit equal to 10% for β. While the quality limits can evolve to take into account feedback from MS and from experimentation, the values used for the derivation of the sample size table are considered a good compromise between computational and analysis requirement and

robustness of the tests. In this respect, the sample size specification is not expected to change.

2.4.13. This approach prevents direct application of the ISO2859-2 tables as they determine a priori the sample size $n_1$ or $n_0$ given their populations $N_1$ and $N_0$. By contrast, this approach sets a constant n for unknown $N_1$ and $N_0$. Fortunately, the standard does provide the probability distributions that allow us to create equivalent tables for the accidental $n_1$ and $n_0$ that result from our fixed sample n. This solution also provides a clear ring-fence around the inspection activities.

## 2.5. Calculating acceptance tables for $n_1$ and $n_0$

2.5.1. As ISO2859-2 does not provide us with a directly applicable solution, we go back to its original logic and construct a CbM QA solution for our situation having a fixed sample size of n with a set of unknown positives and negatives.

2.5.2. The sample of size n is partitioned in two subpopulations of size $n_0$ (negatives) and $n_1$ (positives), respectively. Items are adjudicated to their respective partition during the inspection process and a separate test is performed on each subpopulation, which defines a different error type.

2.5.3. The test performed on the subpopulation of negatives/non-existing events (size $n_0$) answers the question "Is the number of type I errors (α, false positives) lower than 10%?". The test has the form:

$$n_{10} \leq AC_0$$

where $AC_0$ is the acceptance number and $n_{10}$ is the number of false positives determined during the inspection process. Note that $^{n_{10}}/_{n_0}$ is an estimate of the probability of false positives.

2.5.4. The test performed on the subpopulation of positives/existing events (size $n_1$) answers the question "Is the number of type II errors (β, false negatives) lower than 10%?". The test has the form:

$$n_{01} \leq AC_1$$

where $AC_1$ is the acceptance number and $n_{01}$ is the number of false negatives determined during the inspection process. Note that $^{n_{01}}/_{n_1}$ is an estimate of the probability of false negatives.

2.5.5. Acceptance numbers $AC_0$ and $AC_1$ are determined using the logic of the ISO2859-2 standard. In particular, for the false positive case, $n_{10}$ is a hyper-geometric random variable:

$$n_{10} \sim HyperGeometric(N_0, K_0, n_0)$$

where $N_0=N_{10}+N_{00}$ is the total number of negatives/not-existing events in the full population, N. $K_0$ is the proportion of false positives that the CbM would detect over the set of events that the QA would not confirm. The fact that $n_{10}$ is hyper-geometrically distributed derives from the random sampling scheme used to select the n elements considered for inspection. The approach is therefore equivalent to randomly sample $n_0$ elements from $N_0$. In principle, $N_0$ is unknown. However, it can be safely assumed that $N_0$ is significantly larger than $n_0$. Also $K_0$ is unknown, however it will be determined according to the principle discussed below.

2.5.6. $AC_0$ is obtained by fixing the probability of declaring the CbM decision process erroneously compliant. In particular, the probability of erroneously passing the QA test is given by:

$$P(n_{10} \leq AC_0 | L \geq L_q)$$

where L is the actual type I error probability of the CbM decision process and $L_q$ is the maximum acceptable type I error probability (here we want to verify if the CbM decision process has a type I error probability within the limit). $L_q$ corresponds to the quality limit in the ISO 2859-2 test.
Thus, $AC_0$ is found by solving

$$P(n_{10} \leq AC_0 | L \geq L_q) \leq \beta_0$$

where $\beta_0$ is the probability of erroneously accepting the test as compliant. In the ISO 2859-2 standard, $\beta_0$, is denoted as consumer's risk (CR) and is set to 0.1 (10%). The same CR as that used in the ISO 2859-2 is adopted here. Probability $P(n_{10} \leq AC_0 | L \geq L_q)$ is maximized when $L = L_q$. This condition allows one to determine $K_0$ as $K_0=L_qN_0$

2.5.7. For a large $N_0$, the distribution of a hyper-geometric random variable can be approximated with that of a Binomial random variable. Thus, $P(n_{10} \leq AC_0 | L = L_q)$ can be approximated by the cumulative probability density function (cdf) of a Binomial random variable:

$$P(n_{10} \leq AC_0 | L = L_q) = F(AC_0, n_0, L_q) = \sum_{i=0}^{AC_0} \binom{n_0}{i} L_q^i (1 - L_q)^{n_0 - i}$$

In the Binomial approximation, the dependency on $N_0$ and $K_0$ disappears.

2.5.8. Thus, $AC_0$ is obtained by inverting the last equations using the Binomial inverse cdf. In Microsoft Excel, this function is implemented by the "BINOM.INV" function. In particular, for $L_q$ = 10% and $\beta_0$ = 0.1 (Consumer's risk specified by the ISO standard), $AC_0$ is found as:

$$AC_0 = BINOM.INV(n_0, 0.10, 0.1)\text{-}1$$

2.5.9. The "-1" term introduced in the previous equation, is required since the 'BINOM.INV' function "returns the smallest value for which the cumulative binomial distribution is greater than or equal to a criterion value". Since we are in interested in finding the greatest value for which the cumulative binomial distribution is lower than or equal to the criterion value, the additional "-1" is needed.

2.5.10. $AC_1$ can be derived using the same logic adopted for $AC_0$. In particular, $AC_1$ is found by solving

$$P\left(n_{01} \leq AC_1 \middle| L = L_q\right) \leq \beta_1$$

where $\beta_1$ is the CR set to 0.1. In this case, the quality limit, $L_q$, is set to 0.1.

2.5.11. The distribution of $n_{01}$ can be approximated with that of a Binomial random variable and

$$P\left(n_{01} \leq AC_1 \middle| L = L_q\right) = F\left(AC_1, n_1, L_q\right) = \sum_{i=0}^{AC_1} \binom{n_1}{i} L_q^i \left(1 - L_q\right)^{n_1 - i}$$

2.5.12. Finally, $AC_1$ is found as:

$$AC_1 = BINOM.INV(n_1, 0.1, 0.1) - 1.$$

2.5.13. These formulas are embedded in a worksheet, available on WikiCAP, which allows calculating the acceptance numbers for any size $n_1$ or $n_0$ at the pre-set α and β error expectation.

2.5.14. Given that the sample sizes $n_1$ and $n_0$ are dynamic (unknown from start but revealed after inspection) the probability of acceptance is inverted, i.e. once a number of non-compliant items and the sample size have been revealed, one can establish whether the lot is accepted.

2.5.15. To do this, we produced dedicated tables that are indexed on acceptance numbers (the maximum allowed non-compliant items) to identify the maximum sample size that can be statistically tolerated to contain the applicable error rate. These tables thus provide a simple look-up facility and are included in the technical guidance of CBM QA step 1.

## 2.6. Rationale and definitions for Step 2

2.6.1. The second step of the CbM QA translates detection results to a compliance decision. In step 1, a lot level analysis is performed where, for each lot, a sample of size n is inspected. In step 2, a system level verdict is performed by parcel-level eligibility assessment. Since compliance errors from different

lots (G1, G2, T1-T4, C1) are jointly analysed, a superscript 'i' is introduced to denote quantities related to the ith lot.

2.6.2. For the ith lot (G1, G2, T1-T4, C1), a sample denoted here as $\Sigma_i$ was inspected. Among this sample, $n_{10}^i$ elements were identified as false positives and $n_{01}^i$ as false negatives. Each error type is then further divided in "abatable" and "end-stage" errors.

2.6.3. A false compliance outputs on each agricultural parcel is "abatable" if the farmer is expected to come forward and demand correction, i.e. the applicant has an interest or incentive to have the error reversed because the current state is disadvantageous for him/her.

2.6.4. A compliance error is "end-stage" if the farmer has no incentive to contest. End-stage errors lead to undue payments and to direct financial losses, because the applicant has no interest or incentive to have the error reversed.

2.6.5. The type of compliance error (abatable vs. end-stage) is determined by the unique combination of land use/ cover and type of detection error (false positive/negative). The individual type of compliance error is identified when the samples for the different lots are inspected.

2.6.6. Using these definitions, abatable and end-stage compliance errors can be identified in each sample, $\Sigma_i$

2.6.7. CbM QA step 2 aims at analysing the overall CbM performance in terms of eligibility conditions. In this respect, several lots, considered for different tests, may be jointly analysed. For the joint analysis, the total (system level) sample is constructed as

$$\Sigma = \bigcup_i \Sigma_i$$

$\Sigma$ is the union of the original samples, $\Sigma_i$, and contains a list of unique FOIs. The union is used to avoid double-counting a FOI used in more than one sample.

2.6.8. $\Sigma$ has size $n^s$:

$$n_s = \#(\Sigma)$$

2.6.9. For each FOI in $\Sigma$, it is possible to determine if it is an abatable or end-stage error. If the FOI in $\Sigma$ was from a single sample, $\Sigma_i$, then it is an abatable/end-stage error for $\Sigma$ if it was an abatable/end-stage error for $\Sigma_i$. If the FOI in $\Sigma$ was present in more than one lot, then it is an abatable/end-stage error for $\Sigma$ if it was an abatable/end-stage error for at least one of the sample where it was originally present.

2.6.10. For systems, where sequential processing of the final parcel/FOI eligibility is applied, abatable and end-stage errors can be waivered. A waiver can be applied only in case when, for a single FOI, the QA confirms the correctness of the final eligibility decision generated by the CbM, within a given scenario. When a series of step 1 detection processes are concatenated only the final decision should be taken into account for determining the number of abatable/end-stage errors. An example of sequential processing is discussed in the technical guidance.

2.6.11. The numbers of abatable/end-stage errors in Σ are denoted as $n_a^s$ and $n_e^s$, respectively

2.6.12. The ratios $\frac{n_a^s}{n^s}$ and $\frac{n_e^s}{n^s}$ determine the percentage of abatable/end-stage errors in the overall sample and are estimates of the percentage of abatable/end-stage compliance errors at the CbM system level. The goal of CbM QA step 2 is to verify that the percentage of abatable/end-stage errors in the full system is below pre-defined quality limits ($L_q$).

## 2.7. Acceptance numbers for $n_a^s$ and $n_e^s$

2.7.1. The tests performed in step 2 use the combined set of items sampled and inspected during step 1. The numbers of compliance errors, $n_a^s$ and $n_e^s$, are determined during the sample inspection process.

2.7.2. The test performed on the abatable compliance errors answers the question "Is the number of abatable errors in the full lot lower than 5%?". The test has the form:

$$n_a^s \leq AC_a$$

where $AC_a$ is the acceptance number and $n_a^s$ is the number of abatable errors determined during the inspection process.

2.7.3. The test performed on the end-stage compliance errors answers the question "Is the number of end-stage errors in the full population lower than 5%?". The test has the form:

$$n_e^s \leq AC_e$$

where $AC_e$ is the acceptance number and $n_e^s$ is the number of end-stage compliance errors determined during the inspection process.

2.7.4. Acceptance numbers, $AC_a$ and $AC_e$, are determined using the same logic adopted in step 1 and in line with the ISO 2859-2 standard. In particular, acceptance numbers are obtained by controlling the Consumer's Risk (CR) and solving the following equations:

$$P\left(n_a^s \leq AC_a \middle| L \geq L_q = 0.05\right) \leq \beta_a$$
$$P\left(n_e^s \leq AC_e \middle| L \geq L_q = 0.05\right) \leq \beta_e$$

The terms in the left sides of the previous equations correspond to the probabilities that the numbers of abatable/end-stage errors are lower than the acceptance number even if the quality limit is above 5%. The acceptance numbers selected in order to have such probabilities lower than the CRs, $\beta_a$ and $\beta_e$. The CRs are both set to 0.1, which is the value adopted by the ISO 2859-2 standard.

2.7.5. Let us consider the case of abatable errors. The problem can be modelled with the well-known urn model: among a large population of $N^s$ elements, a sample of size $n^s$ is randomly sampled. $N^s$ is significantly larger than $n^s$ such that the effect of drawing without replacement can be neglected. This hypothesis was already used in step 1 where hyper-geometric probabilities were approximated with a binomial model. For each element considered in the sample extracted for inspection, a decision is taken and the following random variable is defined:

$$A_h = \begin{cases} 1 & abatable\ error\ committed \quad p \\ 0 & otherwise \quad 1-p \end{cases}$$

$A_h$ is a Bernoulli random variable and indicates if an abatable error was committed when analysing the h-th sample element. p is the probability that an abatable error is committed and is at this point still unknown. Since $N^s$ is significantly larger than $n^s$, it is possible to assume that all the elements inspected lead to independent and identically distributed (i.i.d.) random variables and that all $A_h$'s are characterized by the same p.

2.7.6. The total number of abatable compliance errors is then given by

$$n_a^s = \sum_{h=1}^{n^s} A_h \sim Binomial(n^s, p)$$

and follows a Binomial distribution with $n^s$, the number of trials, and p the success probability. While p is unknown the goal of the test is to verify that p is lower than the quality limit $L_q$ = 0.05. Thus, the worst case scenario, p = $L_q$ should be considered.

2.7.7. Finally, the first probability in 2.7.4 can be computed using the cumulative density function (cdf) of a Binomial distribution. $AC_e$ is found as

$$AC_a = BINOM.INV(n^s, 0.05, 0.1) - 1$$

where "BINOM.INV" is the inverse binomial cumulative distribution as provided by Microsoft Excel software and already used in step 1. The '-1' term takes into account the fact that BINOM.INV "returns the smallest value for which the cumulative binomial distribution is greater than or equal to a criterion value".

2.7.8. The acceptance number for the end-stage compliance errors is found using the same approach and in particular

$$AC_e = BINOM.INV(n^s, 0.05, 0.1) - 1.$$

2.7.9. The two formulas were used to produce a dedicated table that is indexed by the combined sample size, $n^s$. This table provide a simple look-up facility and is included in the technical guidance of CbM QA step 2.