Bayesian conformity assessment in presence of systematic measurement errors

Bayesian conformity assessment in presence of systematic measurement errors

Carlo Carobbi†§ and Francesca Pennecchi‡

† Universitá degli Studi di Firenze, Department of Information Engineering, 50139 Firenze, Italy

‡ Istituto Nazionale di Ricerca Metrologica (INRiM), 10135 Torino, Italy

Abstract.
Conformity assessment of the distribution of the values of a quantity is investigated by using a Bayesian approach. The effect of systematic, non-negligible measurement errors is taken into account. The analysis is general, in the sense that the probability distribution of the quantity can be of any kind, that is even different from the ubiquitous normal distribution, and the measurement model function, linking the measurand with the observable and non-observable influence quantities, can be non-linear. Further, any joint probability density function can be used to model the available knowledge about the systematic errors. It is demonstrated that the result of the Bayesian analysis here developed reduces to the standard result (obtained through a frequentistic approach) when the systematic measurement errors are negligible. A consolidated frequentistic extension of such standard result, aimed at including the effect of a systematic measurement error, is directly compared with the Bayesian result, whose superiority is demonstrated. Application of the results here obtained to the derivation of the operating characteristic curves used for sampling plans for inspection by variables is also introduced.

Submitted to: *Metrologia*

§ To whom correspondence should be addressed (carlo.carobbi@unifi.it)

1. Introduction

In the framework of the Guide to the Expression of Uncertainty in Measurement (GUM) [1], the measurand is considered as characterized by an essentially unique value. Unfortunately, measurements for industry are often aimed at characterizing a measurand that is better described by a distribution of different possible values. Two distinct situations can be identified. The first situation is the one where a quality characteristic of the products, randomly sampled from a series production, is measured in order to assess the compliance of the series production with specific requirements for that characteristic. Such requirements may be expressed in terms of an upper bound, a lower bound or an interval of acceptable values. The measurand is the distribution of the possible values of the quality characteristic over the whole production, and it has to be inferred from a sample of limited size. The second situation is the one where the measurand is intrinsically unstable in time or inhomogeneous in space. Electromagnetic Compatibility (EMC) measurements, for example, provide several occurences of the latter situation, such as measurements of an unwanted disturbance or noise, or measurements of the electric field intentionally generated inside a reverberation or anechoic chamber for testing purposes. Compliance with a requirement or a specification is frequently to be assessed also in this second situation. Both situations, although related to completely different applications, can be dealt with the same approach in terms of the statistical analysis. In particular, a joint probability density function (pdf) describing the product and measurement dispersions, needs to be formulated when assessing the compliance of the measurand with a certain prescribed limit (for example, an upper bound limit). The main novelty of the present work is to derive, through a deductive reasoning based on Bayesian analysis, the joint pdf of the parameters characterizing the distribution of the measurand, as inferred from a sample of limited size and measured in presence of a dominant (and correlating) systematic effect.

The work naturally flows into the vein of a conspicuous literature dedicated to conformity assessment (see [2] and [3], as recent examples, and the references mentioned therein). Many of the papers dealing with conformity assessment propose criteria and tools for assessing the risk of incorrect decisions of conformity, considering the cost and the impact which may derive from them. A main issue in the field is modelling and treating the measurement uncertainty [4]; it generally appears that there is no single method to integrate the uncertainty into the decision-making process. Moreover, when such uncertainty is of comparable size with respect to the dispersion of the product values, it can be difficult to separate the two terms [5, 6], a fact that may increase the risk of incorrect decisions. Furthemore, very few papers, such as [7], address the problem of systematic effects arising in conformity assessment tests.

In this context, we focused on the determination of an appropriate criterion for the conformity. We considered the case of a distribution for the measurand values sufficiently well spread in comparison with the non-repeatability of the measuring instrumentation, a well satisfied assumption in several testing sectors, such as electrical and mechanical. Nevertheless, we addressed the case of a non-negligible systematic effect affecting and correlating all the measurements, which is a rather frequent occurrence in testing. We also provided some considerations on the applicability of the obtained results to the determination of consumer's and producer's risks by means of the operating characteristic (OC) curves [8], a standard tool developed by the consolidated theory of acceptance sampling.

The structure of the work is as follows: in section 2 the basic notation and assumptions are introduced. The posterior joint pdf for the parameters of the distribution of the quality characteristic, or measurand, is derived in section 3. Also the posterior predictive pdf is derived in the same section since it is deemed to be useful for a simpler, although less informative, approach to conformity assessment, which is the subject of section 4. A general criterion for conformity assessment, which is applicable to any distribution of possible quantity values, is also introduced in section 4, whereas its implementation in a simple but general case is offered in section 5. Section 6 provides the details of the application of the criterion in the case of an upper acceptance limit and compares the results with those obtained through a consolidated frequentistic analysis. The impact of the derived results to the calculation of producer's and consumer's risks is finally investigated in section 7. Conclusions follow in section 8.

2. Notation and assumptions

We here analyse the case where the measurand is represented by the distribution of the possible values of a quantity Q. Q is measured by using a measuring instrument, or system, affected by a systematic measurement error due to imperfect calibration. The indication of the measuring instrument or system is represented by Q_e, which is the observable quantity. The non-repeatability of the measuring instrumentation or system is assumed to be negligible with respect to the spread of the possible values of Q and the standard uncertainty of the systematic error.

The pdf of Q is given by $f_Q(q; \boldsymbol{p})$, where q is a realization of Q and $\boldsymbol{p}$ is the set of parameters of the pdf. For example if Q is normal then $\boldsymbol{p} = (\mu, \sigma)$ or, equivalently, $Q \sim N(\mu, \sigma^2)$, where μ is the mean and σ is the standard deviation of Q.

The first step of the analysis is to determine, through Bayesian inference, the posterior pdf for $\boldsymbol{p}$ from the available knowledge, which consists of: the pdf of Q (e.g. normal, with *unknown* parameters to be inferred), a sample of n realizations of Q_e, the pdf of the systematic error (e.g. rectangular, with *known* bounds, or normal, with *known* parameters), and the joint prior pdf of $\boldsymbol{p}$, represented by $f_0(\boldsymbol{p})$ (improper or a pdf with *known* parameters). The second step consists in the determination of a conformity assessment criterion for Q, based on the comparison between the distribution of its possible values and an interval of acceptable values I_Q.

Each realization q_i, for $i = 1, 2, ..., n$, of the non-observable quantity Q univocally corresponds to a realization q_{e_i} of the observable quantity Q_e and viceversa. Therefore if $q_i = m(q_{e_i}, \boldsymbol{e})$ and $q_{e_i} = o(q_i, \boldsymbol{e})$, then $m(q_{e_i}, \boldsymbol{e}) = o^{-1}(q_{e_i}, \boldsymbol{e})$, where $\boldsymbol{e} = (e_1, e_2, ..., e_r)$

represents a set of realizations of the systematic error vector $\boldsymbol{E} = (E_1, E_2, ..., E_r)$. Function $m(q_{e_i}, \boldsymbol{e})$ is named *measurand model function*, while its inverse $o(q_i, \boldsymbol{e})$ is named *observation model function*. The pdf of the systematic error is represented by $f_{\boldsymbol{E}}(\boldsymbol{e})$. Note that $\boldsymbol{e}$ is a vector in order to accomodate for general models such as $q_{e_i} = e_1 + e_2 q_i$ or $q_{e_i} = e_1 q_i^2/e_2 + e_3$ ($q_i \geq 0$, in order to work with an invertible function).

The set of the n realizations of Q is $\boldsymbol{q} = (q_1, q_2, ..., q_n)$ and the corresponding set of observations of Q_e is $\boldsymbol{q_e} = (q_{e_1}, q_{e_2}, ..., q_{e_n})$. We have $\boldsymbol{q} = \boldsymbol{m}(\boldsymbol{q_e}, \boldsymbol{e})$ and $\boldsymbol{q_e} = \boldsymbol{o}(\boldsymbol{q}, \boldsymbol{e})$, the interpretation of the symbols being immediate.

Random variables Q_i, of which q_i are the corresponding realizations, are independent each other and each Q_i is independent of $\boldsymbol{E}$.

In the following of the paper, in order to make the notation not too heavy, small symbols will indicate either random variables or realizations of them, the distinction being clear by the role they play within the expression of the relevant conditional pdfs.

3. Derivation of the posterior pdf of $\boldsymbol{p}$

Let us consider the joint pdf of the parameters $\boldsymbol{p}$ given $\boldsymbol{q_e}$; namely, the posterior pdf of $\boldsymbol{p}$. Due to the Bayes rule

$$f(\boldsymbol{p}|\boldsymbol{q_e}) \propto f_0(\boldsymbol{p}) \cdot f(\boldsymbol{q_e}|\boldsymbol{p}) = f_0(\boldsymbol{p}) \cdot \int\limits_{\boldsymbol{D_e}} \int\limits_{\boldsymbol{D_q}} f(\boldsymbol{q_e}, \boldsymbol{q}, \boldsymbol{e}|\boldsymbol{p}) d\boldsymbol{q} d\boldsymbol{e}, \tag{1}$$

where $\boldsymbol{D_e}$ and $\boldsymbol{D_q}$ are the domains of the possible values of $\boldsymbol{E}$ and $\boldsymbol{Q}$, respectively. Further,

$$f(\boldsymbol{q_e}, \boldsymbol{q}, \boldsymbol{e}|\boldsymbol{p}) \propto f(\boldsymbol{q_e}|\boldsymbol{q}, \boldsymbol{e}, \boldsymbol{p}) \cdot f(\boldsymbol{q}, \boldsymbol{e}|\boldsymbol{p}). \tag{2}$$

Now consider that if $\boldsymbol{q}$ and $\boldsymbol{e}$ are given then $\boldsymbol{q_e}$ is deterministically obtained through the observation model function, independently on $\boldsymbol{p}$, i.e.,

$$f(\boldsymbol{q_e}|\boldsymbol{q}, \boldsymbol{e}, \boldsymbol{p}) = f(\boldsymbol{q_e}|\boldsymbol{q}, \boldsymbol{e}) = \prod_{i=1}^{n} \delta[q_{e_i} - o(q_i, \boldsymbol{e})]. \tag{3}$$

Since $\boldsymbol{q}$ and $\boldsymbol{e}$ are independent each other and $\boldsymbol{e}$ does not depend on $\boldsymbol{p}$ then we have

$$f(\boldsymbol{q}, \boldsymbol{e}|\boldsymbol{p}) = f_Q(\boldsymbol{q}; \boldsymbol{p}) \cdot f_{\boldsymbol{E}}(\boldsymbol{e}). \tag{4}$$

Since (q_i)s are independent from each other, we have

$$f_Q(\boldsymbol{q}; \boldsymbol{p}) = \prod_{i=1}^{n} f_Q(q_i; \boldsymbol{p}). \tag{5}$$

If we now substitute (5), (4) and (3) into (2) and then into (1) we find

$$f(\boldsymbol{p}|\boldsymbol{q_e}) \propto f_0(\boldsymbol{p}) \int\limits_{\boldsymbol{D_e}} \int\limits_{\boldsymbol{D_q}} \prod_{i=1}^{n} \delta[q_{e_i} - o(q_i, \boldsymbol{e})] \cdot f_Q(q_i; \boldsymbol{p}) \cdot f_{\boldsymbol{E}}(\boldsymbol{e}) d\boldsymbol{q} d\boldsymbol{e}. \tag{6}$$

Solving the integral with respect to $\boldsymbol{q}$ in the right-hand term of (6) we obtain (see [9], p. 71)

$$f(\boldsymbol{p}|\boldsymbol{q_e}) \propto f_0(\boldsymbol{p}) \cdot \int\limits_{\boldsymbol{D_e}} \prod_{i=1}^{n} f_Q[m(q_{e_i}, \boldsymbol{e}); \boldsymbol{p}] \cdot \left| \frac{\partial}{\partial q_{ei}} m(q_{ei}, \boldsymbol{e}) \right| \cdot f_{\boldsymbol{E}}(\boldsymbol{e}) d\boldsymbol{e}. \tag{7}$$

Finally, (7) can be written in the more usual and compact form

$$f(\boldsymbol{p}|\boldsymbol{q_e}) \propto l(\boldsymbol{p}|\boldsymbol{q_e}) \cdot f_0(\boldsymbol{p}), \tag{8}$$

where $l(\boldsymbol{p}|\boldsymbol{q_e})$ is the likelyhood and

$$l(\boldsymbol{p}|\boldsymbol{q_e}) \propto \int\limits_{\boldsymbol{D_e}} \prod_{i=1}^{n} f_Q[m(q_{e_i}, \boldsymbol{e}); \boldsymbol{p}] \cdot \left| \frac{\partial}{\partial q_{ei}} m(q_{ei}, \boldsymbol{e}) \right| \cdot f_{\boldsymbol{E}}(\boldsymbol{e}) d\boldsymbol{e}. \tag{9}$$

3.1. Posterior predictive pdf of Q

The posterior predictive pdf of Q is obtained as follows

$$f_Q(q|\boldsymbol{q_e}) = \int\limits_{\boldsymbol{D_p}} f_Q(q; \boldsymbol{p}) \cdot f(\boldsymbol{p}|\boldsymbol{q_e}) d\boldsymbol{p}. \tag{10}$$

$f_Q(q|\boldsymbol{q_e})$ represents the average pdf of Q over the possible values of $\boldsymbol{p}$ as inferred through $f(\boldsymbol{p}|\boldsymbol{q_e})$ and it is useful for predicting the portion of the Q values lying, in the average, in a given interval.

4. Conformity assessment criterion for Q

The conformity assessment criterion can be stated as follows: at least the portion α_1 of the distribution of the possible values of Q shall be within a given interval (or semi-interval) I_Q with probability not less than α_2. The implementation of this criterion requires the use of the joint posterior pdf $f(\boldsymbol{p}|\boldsymbol{q_e})$. Let $\boldsymbol{D_p}(I_Q, \alpha_1)$ be the region of the $\boldsymbol{p}$ parameter space such that, for any value of $\boldsymbol{p}$ belonging to that region, at least the portion α_1 of the distribution of the possible values of Q is within I_Q, that is, $P(Q \in I_Q|\boldsymbol{p}) \geq \alpha_1$. In the Bayesian framework, parameters $\boldsymbol{p}$ are random variables whose posterior pdf $f(\boldsymbol{p}|\boldsymbol{q_e})$ can say which is the probability for the parameters of lying within the interval $\boldsymbol{D_p}(I_Q, \alpha_1)$. Then, the conformity assessment criterion is mathematically expressed as

$$\int\limits_{\boldsymbol{D_p}(I_Q, \alpha_1)} f(\boldsymbol{p}|\boldsymbol{q_e}) d\boldsymbol{p} \geq \alpha_2. \tag{11}$$

Another criterion for conformity assessment, based on the posterior predictive pdf (10), is stated as follows: the portion of the distribution of the possible values of Q

that is expected to be within a given interval (or semi-interval) I_Q shall be greater than α. Such criterion is mathematically expressed as

$$\int_{I_Q} f_Q(q|\boldsymbol{q_e})dq \geq \alpha. \tag{12}$$

This alternative criterion requires a more simple calculation with respect to (11) but it is less informative because it applies in the average and to a single item randomly selected from the production.

5. Application to additive model and normally distributed Q

The convenience of the assumption of normality stems from the fact that it physically models several practical applications. Further, it permits to obtain results that reduce to standardized ones (obtained through classical, non-Bayesian, derivations) for the conformity assessment criterion, when the systematic measurement error is negligible. The additive observation model has wide applicability, minimizes the mathematical burden of the derivations and results are of immediate interpretation. It should be borne in mind however that (8) and (9) are general in that they apply to any non-linear (but monothonic) model functions and to any (non-normal) distributions for the measurand.

The addictive observation model is

$$q_{e_i} = o(q_i, \boldsymbol{e}) = q_i + e. \tag{13}$$

Note that the systematic error is of a single kind ($r = 1$). Then from (9)

$$l(\boldsymbol{p}|\boldsymbol{q_e}) \propto \int_{-\infty}^{\infty} \prod_{i=1}^{n} f_Q(q_{e_i} - e; \boldsymbol{p}) \cdot f_E(e)de. \tag{14}$$

Further, assume that Q is normally distributed with parameters $\boldsymbol{p} = (\mu, \sigma)$, i.e.

$$f_Q(q; \boldsymbol{p}) = f_Q(q; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(q-\mu)^2}{2\sigma^2}\right]. \tag{15}$$

Substituting (15) into (14) and after manipulation we have

$$l(\mu, \sigma|\boldsymbol{q_e}) \propto \int_{-\infty}^{\infty} \frac{1}{\sigma^n} \exp\left\{-\frac{(n-1)s^2 + n\left[\overline{q_e} - (\mu + e)\right]}{2\sigma^2}\right\} \cdot f_E(e)de, \tag{16}$$

where $\overline{q_e} = \frac{1}{n}\sum_{i=1}^{n} q_{e_i}$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(q_{e_i} - \overline{q_e})^2$. Now, substituting (16) into (8) and assuming the improper prior $f_0(\mu, \sigma) \propto \frac{1}{\sigma}$ we obtain

$$f(\mu, \sigma|\boldsymbol{q_e}) \propto \int_{-\infty}^{\infty} \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{(n-1)s^2 + n\left[\overline{q_e} - (\mu + e)\right]}{2\sigma^2}\right\} \cdot f_E(e)de. \tag{17}$$

For what concerns the posterior predictive pdf of Q, substituting (17) into (10), we obtain, after manipulation

$$f_Q\left(q|\boldsymbol{q_e}\right) = \int\limits_{\boldsymbol{D_e}} t_{n-1}(q+e;\overline{q_e}, s\sqrt{1+1/n}) \cdot f_E(e)de, \tag{18}$$

where $t_{n-1}(q+e;\overline{q_e}, s\sqrt{1+1/n})$ is a Student's t pdf with $n-1$ degrees of freedom, of argument $q+e$, shifted by $\overline{q_e}$ and scaled by $s\sqrt{1+\frac{1}{n}}$. Therefore the posterior predictive pdf (18) is simply the pdf of the difference between a scaled and shifted Student's t random variable and the random variable describing the error. Hence, the expected value of (18) is

$$E_Q = \overline{q_e} \tag{19}$$

and the variance is

$$Var_Q = \frac{n-1}{n-3}\left(1+\frac{1}{n}\right)s^2 + u_e^2. \tag{20}$$

By comparing (19) and (20) with (B.6) and (B.9) in [4] it may be acknowledged that the posterior predictive pdf (18) is a valid Bayesian alternative to $g_0(\eta)$, as derived in Annex B of [4], for the prior pdf of the measurand based on a measured sample of items.

Two cases are of particular interest for the pdf of the residual systematic measurement error $f_E(e)$: normal pdf and rectangular pdf. Expressions of the posterior joint pdf of μ and σ in the two cases are given below. Note that, in those cases, the evaluation of the posterior predictive pdf (19) is easily affordable by using commercial or freely available software tools for numerical analysis. Also the application of the posterior predictive pdf to conformity assessment by using (12) is straightforward and therefore left to the interested reader.

Since known systematic effects are assumed to be corrected for, then the expected value of the residual error is zero. Further the standard uncertainty of the residual error is assumed to be known and given by u_e. The generalization to the case where the systematic effect is known but uncorrected is immediate.

5.1. Normal pdf for the residual systematic error

We have

$$f_E(e) = \frac{1}{\sqrt{2\pi}u_e}\exp\left(-\frac{e^2}{2u_e^2}\right). \tag{21}$$

Substituting (21) into (17) and integrating we find, after normalization,

$$f(\mu,\sigma|\boldsymbol{q_e}) = \frac{\frac{2}{s}\left(\frac{s}{\sigma}\right)^n\left(\frac{n-1}{2}\right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\cdot\sqrt{2\pi}\sqrt{\sigma^2/n+u_e^2}} \cdot \exp\left[-\frac{1}{2}\frac{(n-1)s^2}{\sigma^2}\right] \cdot \exp\left[-\frac{1}{2}\frac{(\overline{q_e}-\mu)^2}{\sigma^2/n+u_e^2}\right]. \tag{22}$$

This same result was obtained in [10] by using an entirely different derivation from that described in section 3, tailored to the specific assumptions of an additive model and of normality and independence of Q and E, hence resorting to a multivariate normal likelihood for the observations, whose covariance matrix took into account u_e^2 as the covariance term between the observations. In the same paper, the expression for the marginal posterior pdfs for μ and σ are also provided and discussed.

5.2. Rectangular pdf for the residual systematic error

The pdf of the residual systematic measurement error is

$$f_E(e) = \begin{cases} \frac{1}{2T} & \text{if } -T < e < T \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

where $T > 0$.
Substituting (23) into (17) and integrating we find, after normalization,

$$\begin{aligned} f\left(\mu, \sigma | \boldsymbol{q_e}\right) &= \frac{1}{2Ts} \frac{\left(\frac{n-1}{2}\right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{s}{\sigma}\right)^n \exp\left[-\frac{(n-1)\,s^2}{2\sigma^2}\right] \\ &\cdot \left[\operatorname{erf}\left(\sqrt{n}\frac{T-\mu+\overline{q_e}}{\sqrt{2}\sigma}\right) + \operatorname{erf}\left(\sqrt{n}\frac{T+\mu-\overline{q_e}}{\sqrt{2}\sigma}\right)\right]. \end{aligned} \tag{24}$$

6. Comparison with an upper limit L

Given the values of μ and σ, the fraction α_1 of the production satisfing the (upper) limit prescription value L is

$$P(Q < L|\mu, \sigma) = P\left(\frac{Q-\mu}{\sigma} < \frac{L-\mu}{\sigma}|\mu, \sigma\right) = \alpha_1. \tag{25}$$

Hence, whenever $(L-\mu)/\sigma \geq z_{\alpha_1}$, where z_{α_1} is the α_1^{th} quantile of a standard normal distribution, probability (25) is larger or equal to α_1.
However, according to the Bayesian view, μ and σ are random variables, the state of knowledge about which is encoded by their joint pdf (17). Hence, the requirement is that the probability of $(L-\mu)/\sigma$ being larger than z_{α_1} should be at least equal to α_2. Therefore the scope is to determine the limit value L so that

$$P\left(\frac{L-\mu}{\sigma} > z_{\alpha_1}|\boldsymbol{q_e}\right) = P\left(\mu + z_{\alpha_1}\sigma < L|\boldsymbol{q_e}\right) \geq \alpha_2. \tag{26}$$

Let L_{α_2} be the limit value such that $P\left(\mu + z_{\alpha_1}\sigma < L_{\alpha_2}|\boldsymbol{q_e}\right) = \alpha_2$, then L_{α_2} satisfies the following equation:

$$\int_0^\infty \int_{-\infty}^{L_{\alpha_2}-z_{\alpha_1}\sigma} f(\mu, \sigma|\boldsymbol{q_e}) d\mu d\sigma = \alpha_2. \tag{27}$$

Therefore, for any given probabilty values α_1 and α_2, the corresponding limit value L_{α_2} was numerically determined by calculating integral (27) of expression (22) or (24), relation (27) holding, consequently, for any $L \geq L_{\alpha_2}$. It can be shown that

$$L_{\alpha_2} = k_{\alpha_2} s + \overline{q_e}, \tag{28}$$

where k_{α_2} is a function of n and s/u_e. Tables 1 (normal residual systematic measurement error with zero mean and standard deviation equal to u_e) and 2 (rectangular residual systematic measurement error with $T = \sqrt{3}u_e$, hence having the same standard deviation as the normal error) show some k_{α_2} values calculated for several values of n and s/u_e, when $\alpha_1 = \alpha_2 = 0.8$. The resulting conformity criterion, relying on the whole available information about the series production, is then a relationship between probabilities α_1 and α_2 and limit L: given two of the three ingredients, the third one can be always obtained.

Table 1. Values of k_{α_2} obtained for several values of n and s/u_e, when $\alpha_1 = \alpha_2 = 0.8$ and the pdf of the residual systematic measurement error is normal. The values of k_{α_2} reported in the table are accurate to within ± 0.01.

$\boldsymbol{k_{\alpha_2}}$ values		$\boldsymbol{n}$											
		2	3	4	5	6	7	8	9	10	20	50	100
$\boldsymbol{s/u_e}$	∞	3.42	2.02	1.67	1.51	1.42	1.35	1.30	1.27	1.24	1.10	0.99	0.95
	10	3.43	2.02	1.68	1.52	1.43	1.36	1.31	1.28	1.25	1.11	1.02	0.98
	3	3.46	2.07	1.75	1.60	1.51	1.45	1.41	1.38	1.35	1.24	1.17	1.15
	2	3.49	2.15	1.83	1.69	1.61	1.55	1.51	1.48	1.46	1.36	1.30	1.28
	1	3.71	2.47	2.18	2.04	1.97	1.91	1.88	1.85	1.83	1.75	1.71	1.69
	0.5	4.50	3.28	2.97	2.84	2.76	2.71	2.68	2.66	2.64	2.58	2.54	2.53
	0.3	5.72	4.40	4.07	3.93	3.86	3.81	3.78	3.76	3.74	3.69	3.66	3.66
	0.2	7.28	5.79	5.45	5.31	5.25	5.20	5.18	5.16	5.14	5.09	5.07	5.07
	0.15	8.81	7.19	6.84	6.70	6.64	6.60	6.58	6.56	6.54	6.49	6.47	6.47
	0.1	11.8	9.99	9.63	9.50	9.44	9.40	9.38	9.36	9.34	9.30	9.27	9.27

6.1. Frequentist approach

The theory for acceptance sampling by variables, adopted by the ISO 3951 standard series, is developed by using a frequentist approach and it is based on the assumptions of normal distribution for Q and negligible residual systematic measurement error. In the last edition of the standard ISO 3951-2 [8] an informative annex has been introduced for the purpose of taking into accout a non-negligible *random* measurement error. The case of a non-negligible residual *systematic* measurement error is analysed in [11] (a reference which appears among those listed in the bibliography of [8]). It is worth comparing the frequentist result in [11] with the Bayesian one obtained here and reported in table 1, both valid for zero mean, normal systematic measurement error. From section 3 of [11] we have

$$k_{\alpha_2} = \frac{t_{\alpha_2}}{\sqrt{n_*}}, \tag{29}$$

Table 2. Values of k_{α_2} obtained for several values of n and s/u_e, when $\alpha_1 = \alpha_2 = 0.8$ and the pdf of the residual systematic measurement error is rectangular. The values of k_{α_2} reported in the table are accurate to within ± 0.01.

$\boldsymbol{k_{\alpha_2}}$ values		$\boldsymbol{n}$											
		2	3	4	5	6	7	8	9	10	20	50	100
$\boldsymbol{s/u_e}$	∞	3.42	2.02	1.67	1.51	1.42	1.35	1.30	1.27	1.24	1.10	0.99	0.95
	10	3.42	2.02	1.68	1.52	1.43	1.36	1.31	1.28	1.25	1.11	1.02	0.98
	3	3.45	2.07	1.75	1.60	1.51	1.46	1.41	1.38	1.36	1.26	1.21	1.20
	2	3.48	2.14	1.84	1.70	1.62	1.57	1.53	1.50	1.48	1.41	1.38	1.37
	1	3.68	2.48	2.23	2.12	2.06	2.02	1.99	1.98	1.96	1.92	1.89	1.89
	0.5	4.40	3.45	3.23	3.13	3.08	3.05	3.03	3.01	3.00	2.96	2.93	2.93
	0.3	5.81	4.87	4.62	4.52	4.47	4.43	4.41	4.40	4.39	4.34	4.32	4.31
	0.2	7.73	6.63	6.36	6.25	6.20	6.17	6.14	6.13	6.12	6.07	6.05	6.05
	0.15	9.62	8.37	8.09	7.98	7.93	7.90	7.88	7.86	7.85	7.80	7.78	7.78
	0.1	13.3	11.9	11.6	11.4	11.4	11.4	11.3	11.3	11.3	11.3	11.2	11.2

where

$$n_* = \frac{n}{1 + n(u_e/\sigma)^2} \tag{30}$$

and t_{α_2} is the α_2^{th} quantile of the non-central Student's t pdf with $n-1$ degrees of freedom and non-centrality parameter $\delta = z_{\alpha_1}\sqrt{n_*}$. Table 3 is similar to table 1 but its entries are calculated by using (29) and (30) in place of (22), (27) and (28). When

Table 3. Values of k_{α_2} obtained for several values of n and σ/u_e, when $\alpha_1 = \alpha_2 = 0.8$ and according to the frequentist analysis outlined in [11]. The pdf of the residual systematic measurement error is normal. The values of k_{α_2} reported in the table are accurate to within ± 0.01.

$\boldsymbol{k_{\alpha_2}}$ values		$\boldsymbol{n}$											
		2	3	4	5	6	7	8	9	10	20	50	100
$\boldsymbol{\sigma/u_e}$	∞	3.42	2.02	1.67	1.51	1.42	1.35	1.30	1.27	1.24	1.10	0.99	0.95
	10	3.42	2.02	1.68	1.52	1.43	1.36	1.31	1.28	1.25	1.11	1.02	0.98
	3	3.47	2.09	1.75	1.60	1.51	1.45	1.41	1.37	1.35	1.24	1.17	1.15
	2	3.53	2.17	1.84	1.69	1.61	1.55	1.51	1.48	1.46	1.36	1.30	1.28
	1	3.88	2.53	2.21	2.07	1.98	1.93	1.89	1.87	1.85	1.76	1.71	1.70
	0.5	4.95	3.46	3.10	2.94	2.85	2.79	2.75	2.72	2.70	2.61	2.56	2.54
	0.3	6.63	4.82	4.37	4.17	4.06	3.98	3.93	3.89	3.87	3.75	3.69	3.67
	0.2	8.84	6.55	5.98	5.72	5.58	5.48	5.42	5.37	5.33	5.18	5.10	5.07
	0.15	11.1	8.31	7.60	7.28	7.10	6.99	6.91	6.85	6.80	6.61	6.51	6.48
	0.1	15.6	11.8	10.9	10.4	10.2	10.0	9.89	9.81	9.74	9.48	9.34	9.30

$u_e = 0$, the obtained k_{α_2} values in tables 1 and 3 are identical each other and to those actually prescribed by the "80 %/80 % rule" as implemented so far, see [12]. For $u_e \neq 0$, the values obtained through the frequentist analysis are larger than those obtained through the Bayesian one. This suggests that the supplemental information provided by the pdf of the residual systematic error is more efficiently taken into account through the Bayesian analysis, especially in those cases where it is expected that the frequentist analysis may not work well (i.e. when n is small and u_e is large). Finally note from (30) that, in order to calculate n_*, the true value of the standard deviation σ should, in principle, be known, which is not actually possible. Thus, in practice, one has to assume that $\sigma \sim s$ in order to select the relevant entry of table 3, a typical inconvenient arising from the frequentist approach to inference.

7. Evaluation of producer's and consumer's risks: the OC curve

In statistical quality control it is of interest to determine the acceptability of a lot of items on the basis of the fraction of nonconforming items in the lot [8]. The decision about accepting or rejecting the lot relies upon measurements of a quality characteristic performed on a random sample of items from the lot. Assume that the consumer is expected to reject a lot whose fraction of nonconforming items is $1 - \alpha_1$ or greater. Further, let $1 - \alpha_2$ be the risk that the consumer accepts a lot whose fraction of nonconforming items is greater than $1 - \alpha_1$. The consumer accepts the lot if, see (28), $L > \overline{q_e} + k_{\alpha_2} s$ and the lot is not conforming if $L - \mu < z_{\alpha_1}\sigma$. Therefore the risk of the consumer (risk of acceptance of a nonconforming lot) is

$$P\left(\overline{q_e} + k_{\alpha_2} s - \mu < L - \mu < z_{\alpha_1}\sigma \middle| \boldsymbol{q_e}\right) = 1 - \alpha_2, \tag{31}$$

then, from (27) and (31),

$$1 - \alpha_2 = 1 - \int_0^\infty \int_{-\infty}^{k_{\alpha_2} s + \overline{q_e} - z_{\alpha_1}\sigma} f(\mu, \sigma | \boldsymbol{q_e}) d\mu d\sigma, \tag{32}$$

The operating characteristic (OC) curve is a diagram showing the probability of acceptance of a lot as a function of the fraction of nonconforming items in the lot [8]. If f is the fraction of nonconforming items and z_f is the $(1 - f)^{\text{th}}$ quantile of the standard normal distribution, then the probability of acceptance of a lot whose fraction of nonconforming items is f is, from (32),

$$OC(f) = 1 - \int_0^\infty \int_{-\infty}^{ks + \overline{q_e} - z_f\sigma} f(\mu, \sigma | \boldsymbol{q_e}) d\mu d\sigma, \tag{33}$$

where $k = k_{\alpha_2}$ is the *acceptability constant*, according to the terminology in [8]. The OC curves shown in figure 1 with continuous and dashed lines are obtained assuming a sample size $n = 6$ and $\alpha_1 = \alpha_2 = 0.8$. The residual systematic measurement error is taken as normal. The continuous line corresponds to the case $u_e = 0$ and the dashed line corresponds to the case $u_e = 1/2\ s$. The chosen values of α_1 and α_2 are those stipulated by the "80 %/80 % rule" and the value of k is taken from table 1 ($k = 1.42$ if $u_e = 0$ and $k = 1.61$ if $u_e = 1/2\ s$). Note that expression (22) or (24) is substituted in (33) when assuming normal or rectangular residual systematic error, respectively. For the normal case, for example, figure 1 shows that the effect of the systematic error is significant since, in order to obtain a producer's risk of 5 % (i.e. an acceptance probability of 95 %), the maximum acceptable fraction of nonconforming items reduces from $9 \cdot 10^{-3}$ ($u_e = 0$) to $14 \cdot 10^{-4}$ ($u_e = 1/2\ s$).

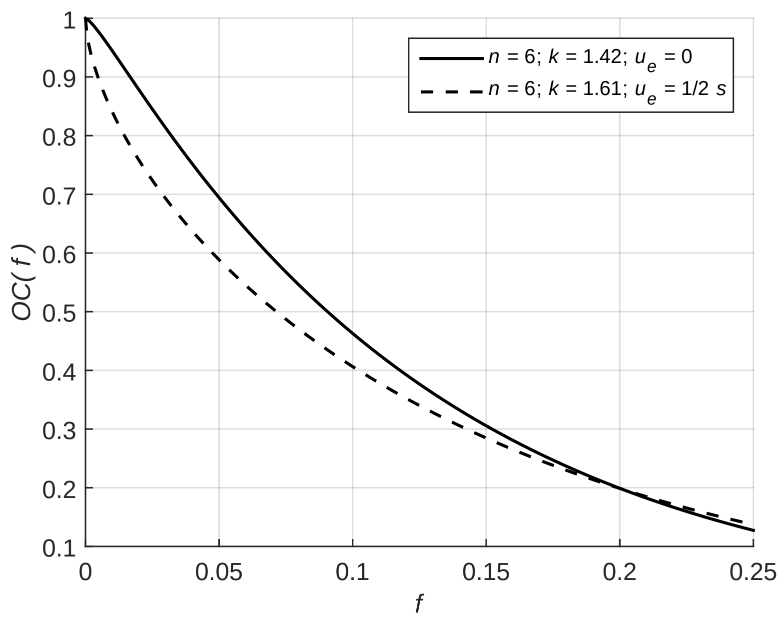

Figure 1. OC curves corresponding to a sample size $n = 6$, minimum fraction of rejectable items equal to 0.2, consumer's risk equal to 0.2 and normal residual systematic measurement error with $u_e = 0$ (continuous line) and $u_e = 1/2\ s$ (dashed line).

8. Conclusions

Bayesian analysis provides a powerful and consistent theoretical framework within which the measurement uncertainty associated with random variability is propagated and mixed up with the uncertainty associated with unknown, but sizeable, systematic effects. Random variability may arise from both the fluctuation of the observed quantity and/or the non repeatability of the measurement system. Results are derived here that generalize the classical (and standard) analysis of acceptance sampling by variables to the case where systematic measurement errors are not negligible with respect to the spread of the quality characteristic (the measurand) over the produced items. Assessment of compliance of a series production is performed and producer's and consumer's risks, as a function of the fraction of nonconfoming items in a lot, are calculated taking measurement uncertainty into account. It is demonstrated that the standard results are obtained in the limiting case where measurement uncertainty tends to zero. The frequentist approach to the inclusion of systematic measurement errors is proved to provide more conservative results than the Bayesian one. Further investigation will be devoted to extending the analysis in order to include the non repatability of the measuring system (which superimposes to the random variability of the production process) and vague knowledge of the standard uncertanty associated with the systematic effects.

References

[1] JCGM 100:2008, *Evaluation of measurement data - Guide to the expression of uncertainty in measurement*, http://www.bipm.org/en/publications/guides/gum.html.

[2] L R Pendrill, *Using measurement uncertainty in decision-making and conformity assessment*, Metrologia 51, S206-S218, 2014.

[3] L R Pendrill, H Karlsson, N Fischer, S Demeyer, and A Allard, *A guide to decision-making and conformity assessment*, Deliverable 3.3.1, EMRP project NEW04 - Novel mathematical and statistical approaches to uncertainty evaluation, May 2015. Available at: http://www.ptb.de/emrp/new04-publications.html.

[4] JCGM 106:2012, *Evaluation of Measurement Data - The role of measurement uncertainty in Conformity Assessment*, http://www.bipm.org/en/publications/guides/gum.html.

[5] R Kacker, N F Zhang and C Hagwood, *Real-time control of a measurement process*, Metrologia 33, 433-445, 1996.

[6] G B Rossi and F Crenna, *A probabilistic approach to measurement-based decisions*, Measurement, 39, 101-119, 2006.

[7] A B Forbes, *Measurement uncertainty and optimized conformance assessment*, Measurement 39, 808-814, 2006.

[8] ISO 3951-2:2013, "Sampling procedures for inspection by variables - Part 2: General specification for single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection of independent quality characteristics."

[9] Ignacio Lira, *Evaluating the Measurement Uncertainty, Fundamentals and Practical Guidance*, Series in Measurement Science and Technology, IoP Publishing Ltd, Bristol and Philadelphia, 2002.

[10] Carlo Carobbi and Francesca Pennecchi, *Bayesian Analysis of Repeated Measurements Affected by a Systematic Error and its Application to Conformity Assessment*, XXI IMEKO World Congress Measurement in Research and Industry, August 30 - September 4, 2015, Prague, Czech Republic.

[11] Henrik Melgaard and Poul Thyregod, *Acceptance sampling by variables under measurement uncertainty*, pp. 47-57. In Lenz, H.J., and Wilrich, P.-Th. [Editors]. Frontiers in Statistical Quality Control 6, Physica-Verlag, Heidelberg, New York, 2001.

[12] CISPR TR 16-4-3:2007, "Specification for radio disturbance and immunity measuring apparatus and methods Part 4-3: Uncertainties, statistics and limit modelling Statistical considerations in the determination of EMC compliance of mass-produced products."