Compliance Testing for Random Effects Models With Joint Acceptance Criteria

Crystal D. LINKLETTER

Mathworks
Natick, MA 01760
(*Crystal.Linkletter@mathworks.com*)

Pritam RANJAN

Acadia University
Wolfville, NS, Canada B4P 2R6
(*pritam.ranjan@acadiau.ca*)

C. DEVON LIN

Queen's University
Kingston, ON, Canada K7L 3N6
(*cdlin@mast.queensu.ca*)

Derek R. BINGHAM

Simon Fraser University
Burnaby, BC, Canada V5A 1S6
(*dbingham@stat.sfu.ca*)

William A. BRENNEMAN

The Procter and Gamble Company
Cincinnati, OH 45241
(*brenneman.wa@pg.com*)

Richard A. LOCKHART

Simon Fraser University
Burnaby, BC, Canada V5A 1S6
(*lockhart@sfu.ca*)

Thomas M. LOUGHIN

Simon Fraser University
Burnaby, BC, Canada V5A 1S6
(*tloughin@sfu.ca*)

For consumer protection, many governments perform random inspections on goods sold by weight or volume to ensure consistency between actual and labeled net contents. To pass inspection, random samples must jointly comply with restrictions placed on the individual sampled items and on the sample average. In this article, we consider the current United States National Institute of Standards and Technology joint acceptance criteria. Motivated by a problem from a real manufacturing process, we provide an approximation for the probability of sample acceptance that is applicable for processes with one or more known sources of variation via a random effects model. This approach also allows the assessment of the sampling scheme of the items. We use examples and simulations to assess the quality and accuracy of the approximation and illustrate how the methodology can be used to fine-tune process parameters for a prespecified probability of sample acceptance. Simulations are also used for estimating variance components.

KEY WORDS: Average criterion; Individual criterion; NIST Handbook 133; Regulatory inspections; Variance components.

1. INTRODUCTION

For consumer protection, many governments regulate the production of goods sold by weight or volume to ensure consistency between contents and labeling. Inspections are randomly performed to verify that (a) on average, products being sold have net contents within a reasonable range of the labeled content (e.g., 500 mL of cola) and (b) very few items are defectively underfilled (i.e., have net content significantly less than the labeled content). For regulatory acceptance, a sample of items must jointly satisfy an "average criterion" and an "individual criterion."

From a manufacturing perspective, guaranteeing sample acceptance by overfilling all units is an inefficient and costly solution. It is useful to have a means of exploring the probability of sample acceptance given any set of operating conditions. Thus, it is natural to express the probability of sample acceptance as a function of process parameters (i.e., the process mean and the variance components associated with different manufacturing steps) so that compliance with government regulations can be achieved via more economical parameter fine-tuning.

The joint acceptance criteria vary by country and evolve over time. Under old U.S. guidelines, there were two criteria: one that considers each individual item and another for the average measurement of the sample units. The individual criterion was based on a maximum allowable number of defectively underfilled items, and the average criterion was based solely on the sample mean being greater than the labeled net content. To calculate the acceptance probability, the distribution of the process must be assumed, and to date methodologies assume a normally distributed process with only one source of

TECHNOMETRICS, AUGUST 2012, VOL. 54, NO. 3
DOI: 10.1080/00401706.2012.680394

variability, for example, unit-to-unit variability. With this distributional assumption, Schilling and Dodge (1969) gave tables of exact acceptance probabilities under the old U.S. criterion for small sample sizes using tabulated integral values for the distribution of an extreme deviate from the sample mean (developed by Nair 1948; Grubbs 1950). To accommodate larger sample sizes, Elder and Muse (1982) proposed an approximation that writes the acceptance probability of a sample as the product of the conditional probability of passing the mean criterion, given the individual criterion is met (an approximate truncated normal distribution), and the marginal probability of passing the individual criterion (an exact binomial distribution). This conditioning approach is attractive as it breaks the joint probability into tractable pieces. Alternatively, Vangel (2002) derived a saddlepoint approximation to the bivariate distribution of the sample mean and the sample extremum, addressing the special case where no defective individual items are permitted.

The recent U.S. National Institute of Standard and Technology (NIST) Handbook 133 (2005) now expresses the average criterion as an upper-limit of the confidence interval based on both the sample mean and the sample standard deviation of measured items. While arguably more reasonable, the inclusion of the sample standard deviation in the average criterion presents new challenges. Constantine, Field, and Robinson (2000) suggested two possible ways of handling the additional random entity. First, they replaced the unknown sample standard deviation by a known quantity, such as the true process standard deviation, thereby simplifying the mathematical derivation of the probability of sample acceptance. In this case, the approximation of Elder and Muse (1982) can directly be applied. This ignores the uncertainty in the sample standard deviation. Alternatively, they proposed an approximation that is based on an implicit (but incorrect) assumption that the sample standard deviation and the individual measurements are independent. Rather fortuitously, this oversight has relatively little impact on the probability calculations in regions of the parameter space, which are typically of interest, that is, the regions that correspond to a very high probability of acceptance. The premise of the Constantine, Field, and Robinson (2000) formulation is problematic, and not suitable for extensions to accommodate more realistic assumptions like multiple sources of variations.

In this article, we develop new methodology for determining the probability of sample acceptance as available methodologies are out-of-date in two fundamental ways. First, the probability of acceptance has been successfully derived only under the *old* inspection guidelines, which do not include the sample standard deviation. Second, in reality, many production processes have multiple sources of variation (e.g., lot-to-lot or batch-to-batch) associated with different stages of production. Such variations can have considerable implications for the distribution of the inspection sample measurements and the probability of sample acceptance. In general, we assume these variance components are known from online monitoring or are well measured, as is the case for many companies; however, we discuss uncertainty in the estimation of these parameters in simple cases (see Example 3). The goal of this article is to provide methodology to calculate the probability of sample acceptance given these parameter values, allowing manufacturers to (a) evaluate the probability of acceptance under current operating conditions, (b) choose a process mean that yields a prespecified probability of acceptance, and (c) evaluate the impact of variance reduction programs on the acceptance probability.

2. JOINT ACCEPTANCE CRITERION

During an inspection, n units of a product are sampled from a store shelf and their individual net content (e.g., weight or volume) are measured. Let x_j denote the net content of the jth unit (generally, we use upper case to denote random variables and lower case for measured or computed quantities). To pass the average criterion, the upper limit of a $100(1-\alpha)\%$ confidence interval for the process mean must be at least as large as the labeled net content, V. That is,

$$\bar{x}_n + t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}} \geq V, \tag{1}$$

where $\bar{x}_n$ and s are the sample mean and the sample standard deviation, respectively, and $t_{1-\alpha/2,n-1}$ is the $100(1-\alpha/2)$th percentile of the t distribution with $n-1$ degrees of freedom. We define $b = t_{1-\alpha/2,n-1}/\sqrt{n}$ for convenience, making the criterion $\bar{x}_n + bs \geq V$. The inclusion of "bs" is an evolution from the previous requirement, $\bar{x}_n \geq V$, the criterion addressed by Elder and Muse (1982) and Vangel (2002). From the manufacturer's perspective, the important issues are estimating the probability of passing inspection and also adjusting the process parameters (e.g., mean and, if possible, variability) so that the system passes inspection with a prespecified probability.

To pass the individual criterion, no more than r items can have net content below a defective level, c. NIST Handbook 133 (2005) specifies the defective level as the labeled net content minus a "maximum allowable variation" (MAV). Let Y_c be the number of items in the sample with net content less than c (i.e., $Y_c = \sum_{j=1}^{n} I(X_j \leq c)$). The individual criterion is then

$$y_c \leq r. \tag{2}$$

For most applications, only very small values of r (e.g., $r = 0, 1,$ or 2) are allowed. Overall, the probability that a sample jointly satisfies criteria (1) and (2) (i.e., is accepted) is

$$P(\bar{X}_n + bS \geq V, Y_c \leq r). \tag{3}$$

The evaluation of (3) as a function of process parameters is the main focus of this article.

In assessing compliance to (1) and (2) an inspector does not give any consideration to the mechanism that produced the sample. The sample mean and the sample standard deviation are calculated in the usual way, and the sample is either accepted or not. However, to correctly ascertain the probability of a sample being accepted, it is necessary to understand the production process under different scenarios and thereby the joint distribution of $\bar{X}_n$, S, and Y_c, where $\bar{X}_n = \sum_{k=1}^{n} X_k/n$ and $S^2 = \sum_{k=1}^{n}(X_k - \bar{X}_n)^2/(n-1)$.

For ease of exposition, we initially restrict our attention to processes that have two sources of variation, say a lot-to-lot variance (σ_a^2) and a within-lot unit-to-unit variance (σ_ϵ^2). More general settings are explored later in Sections 3.1 and 3.2. Suppose M different "production lots" are represented in the sample.

In the two variance component case, the process model is

$$X_{ij} = \mu + a_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, M; \sum_{i=1}^{M} n_i = n, \tag{4}$$

where μ is the overall process mean, the lot random effects $\{a_i\}$ are independent $N(0, \sigma_a^2)$ and uncorrelated with independent $N(0, \sigma_\epsilon^2)$ errors $\{\epsilon_{ij}\}$. Thus, the problem of interest is calculating (3) as a function of process mean and variance components, $P(\mu, \sigma_a, \sigma_\epsilon)$.

Clearly from (4), the number of production lots, M, that comprise the sample will play a role in computing the probability of sample acceptance. For example, it is easy to show that under model (4), $\text{var}(\bar{X}_n) = q_1\sigma_a^2 + n^{-1}\sigma_\epsilon^2$, where $q_1 = n^{-2}\sum_{i=1}^{M} n_i^2$. Note that the maximum value of $q_1 = 1$ is attained when $M = 1$ and the minimum value of $q_1 = n^{-1}$ is attained when $M = n$. Meanwhile, $E(S^2) = q_2\sigma_a^2 + \sigma_\epsilon^2$, where $q_2 = n(1 - q_1)/(n - 1)$, and $E(S^2)$ is minimized at $q_2 = 0$ (or, $q_1 = 1$) when $M = 1$ and maximized when $M = n$. Thus, S^2 systematically underestimates the process variability unless each sample is from a different lot, thereby making (1) more difficult to satisfy in most cases. The worst scenario is single-lot sampling, where the bias can be substantial if σ_a^2 is large relative to $n^{-1}\sigma_\epsilon^2$. On the other hand, the lot effects in the sample have an opposite effect on passing the individual criterion: single lot sampling is the easiest case, and $M = n$ the hardest. Assuming that a process is reasonably well tuned, one expects the probability of passing the individual criterion to be fairly high. Therefore, the relative impacts of lot effects will be somewhat smaller on the individual criterion than on the average criterion, and the overall worst case for passing the joint criteria is when all samples are drawn from a single lot. Simulations presented in Section 4 attest to this. We will therefore, use single-lot sampling as a conservative case to provide a lower bound to the probability of sample acceptance under more general sampling scenarios.

Example 1. To illustrate the pitfalls of ignoring variance components when computing the probability of sample acceptance, consider the following example from the Procter and Gamble Company (for proprietary reasons, details on the product cannot be provided). Suppose a product has labeled net content $V = 40$ oz, and so from NIST Handbook 133, the MAV in the product is 1.376 oz. Thus, the critical defective level is $c = 38.624$ oz (calculated as net content minus the MAV). Assume that the number of products on the shelf at the time of inspection is at least n, the number of sampled items, and for the individual criterion, no items are permitted to be defectively underfilled (i.e., $r = 0$). Using the average criterion as stipulated by the NIST Handbook (with percentile $t_{0.975,n-1}$), the probability of sample acceptance is $P(\bar{X}_n + t_{0.975,n-1}S/\sqrt{n} \geq 40, Y_{38.624} = 0)$.

Consider two scenarios: (A) the variance component decomposition is ignored, and suppose the total variability for the process is $\sigma_{\text{total}}^2 = 0.118$; and (B) 30% of the total variability is due to lot-to-lot variation, $\sigma_a^2 = 0.0354$ (here a lot is defined as an 8 hr production period), and the remainder is due to unit-to-unit variation, $\sigma_\epsilon^2 = 0.0826$, and every sample of size n is produced at the same time, that is, all items are from the same lot. To simulate from scenario (A), samples of size n are randomly generated from a $N(\mu, \sigma_{\text{total}}^2)$, while for scenario (B), a random effect a_L is first generated from a $N(0, \sigma_a^2)$ distribution and then a sample of size n is randomly generated from a $N(\mu + a_L, \sigma_\epsilon^2)$ distribution. The probability of sample acceptance is the proportion of generated samples, which jointly satisfy criteria (1) and (2). We use 10 million simulations to reduce the impact of Monte Carlo error to below four decimal places (see Robert and Casella 2005). Moreover, to assess the sensitivity of sample size on the probability of acceptance, we conducted the simulation study for a range of sample sizes, $n = 8$, 10, 12, 15, and 20. These are comparable to sample sizes recommended in the NIST Handbook 133. The results are summarized in Figure 1.

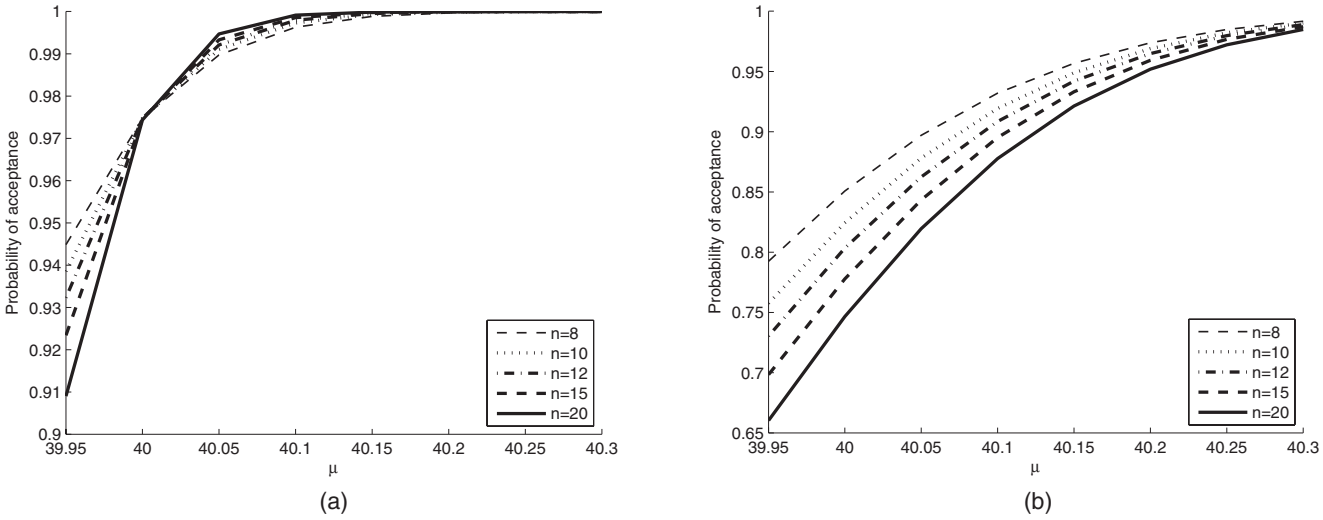

Figure 1. Probability of acceptance as a function of μ with and without variance decomposition. (a) One variance component, (b) two variance components.

Figure 1(a) shows the acceptance probability for different process means under scenario (A) for different sample sizes. There are some immediate observations from the plot. First, as somewhat expected, the acceptance probability is monotonically increasing with the process mean. Second, when $\mu < V$, the probability of passing the inspection appears to be decreasing as the sample size increases for any fixed μ, whereas, the reverse seems to be true when $\mu > V$. For larger sample sizes, the estimated process mean has smaller variance and thus, is known more precisely. As a consequence, the probability of incorrectly passing inspection (when $\mu < V$) or incorrectly failing inspection (when $\mu > V$) is reduced. It turns out that the probability of failing the inspection due to an individual underfill is extremely small under these cases, and thus, the average criterion guides the probability of sample acceptance. That is, such a trend may not always exist.

Figure 1(b) presents the probability of acceptance under scenario (B). The pattern is similar to the region $\mu < V$ in scenario (A). Because of the additional variance component, the probability of acceptance curves in panel (b) are less steep as compared to that in panel (a), and the probability curves tend to 1 as μ increases.

Comparing both panels in Figure 1, it is clear that previously derived methodologies, which ignore variance decomposition, and rely on the assumption that there is only one source of variability, overestimate the probability of sample acceptance. For example, with $n = 12$, if the Procter and Gamble Company equated the process mean to the labeled net content (i.e., set $\mu = 40$ oz) and ignored the variance decomposition, they would erroneously calculate a 97.5% probability of sample acceptance; in reality this probability would be around 80%. To achieve 97.5% probability, they would have to increase the process mean to about 40.25 oz.

For scenario (A), the Monte Carlo estimates of the probabilities of sample acceptance under the mean criterion in (1) and individual criterion in (2) are $P(\bar{X} + t_{0.975,11}S/\sqrt{n} \geq V) = 0.9750$ and $P(Y_{38.624} \leq r) = 0.9997$, respectively. Similarly, the probabilities under scenario (B) are $P(\bar{X} + t_{0.975,11}S/\sqrt{n} \geq V) = 0.8032$ and $P(Y_{38.624} \leq r) = 0.9996$. In both cases, the probability of satisfying the individual criterion is essentially 1, and thus, the probability of satisfying the joint criterion (3) is the same as that of satisfying the mean criterion (1) alone.

3. METHODOLOGY

To outline our approach to estimating the acceptance probability, we begin with the simplest case in which the manufacturing process has two known variance components, a lot-to-lot variance, σ_a^2, and a within-lot unit-to-unit variance, σ_ϵ^2, so the process can be modeled as in (4). We also assume that all n sampled items come from one lot with lot random effect a_L.

Let $\{X_{Lj}, j = 1, \ldots, n\}$ be a random sample such that $X_{Lj}|a_L \sim N(\mu + a_L, \sigma_\epsilon^2)$ for all j, that is, $\bar{X}_n|\ a_L \sim N(\mu + a_L, \sigma_\epsilon^2/n)$ and $\frac{(n-1)S^2}{\sigma_\epsilon^2}|\ a_L \sim \chi^2_{(n-1)}$. Because the lot random effect for a sample is unknown a priori, uncertainty about the random effect can be incorporated by averaging over the lot random effect. Let $f_a(\cdot)$ denote the $N(0, \sigma_a^2)$ density function. Then

$$P(\mu, \sigma_a, \sigma_\epsilon) = P(\bar{X}_n + bS \geq V, Y_c \leq r) = \int_{-\infty}^{\infty} P(\bar{X}_n + bS \geq V, Y_c \leq r|\ a_L) f_a(a_L) da_L. \quad (5)$$

To evaluate the average-individual joint probability in the integrand of (5), we condition in the way *opposite* to Elder and Muse (1982):

$$P(\bar{X}_n + bS \geq V, Y_c \leq r|\ a_L) = P(\bar{X}_n + bS \geq V|\ a_L) P(Y_c \leq r|\ \bar{X}_n + bS \geq V, a_L). \quad (6)$$

We consider the two factors on the righthand side of (6) separately. Leaving the technical details to Appendix A, we find that the first factor can be explicitly computed as

$$P(\bar{X}_n + bS \geq V|\ a_L) = 1 - F_{t':n-1,\delta}(-b\sqrt{n}), \quad (7)$$

where $F_{t':n-1,\delta}(\cdot)$ denotes the cdf of the noncentral t distribution (see e.g., Johnson and Kotz 1970) with $n - 1$ degrees of freedom and noncentrality parameter $\delta = \sqrt{n}(\mu + a_L - V)/\sigma_\epsilon$.

To approximate the second factor in (6), we first define $W_{Lj} = I[X_{Lj} \leq c]$, $j = 1, \ldots, n$, where $W_{Lj} = 1$ if unit j is defectively underfilled, and 0 otherwise. The conditional probability $p_W = P(W_{Lj} = 1|\ \bar{X}_n + bS \geq V, a_L) = P(X_{Lj} \leq c|\ \bar{X}_n + bS \geq V, a_L)$ does not depend on j, and can be computed by noting that

$$p_W = \frac{P(X_{Lj} \leq c, \bar{X}_n + bS \geq V|\ a_L)}{P(\bar{X}_n + bS \geq V|\ a_L)}. \quad (8)$$

The denominator in (8) is simply the noncentral t probability specified in (7). To tackle the numerator, define $f_{\bar{X}_n|a_L}(\cdot)$ and $f_{S|a_L}(\cdot)$ to be the conditional sampling distributions of $\bar{X}_n$ and S specified above. Using results in Appendix B, the numerator in (8) can be written as

$$P(X_{Lj} \leq c, \bar{X}_n + bS \geq V|\ a_L) = \int_0^\infty \int_{V-bs}^\infty F_B(c_0) f_{\bar{X}_n|a_L}(\bar{x}_n) f_{S|a_L}(s) d\bar{x}_n ds, \quad (9)$$

where $F_B(\cdot)$ is the cumulative distribution function of Beta($\frac{n-2}{2}, \frac{n-2}{2}$) and

$$c_0 = \frac{1}{2}\left[1 + \frac{\sqrt{n}(c - \bar{x}_n)}{(n-1)s}\right].$$

The double integral in (9) can be evaluated using numerical integration techniques. We used the Gaussian quadrature routine `dblquad` in Matlab with error tolerance for the integral approximation set to be 10^{-6}. The second factor in (6) can now be written as

$$P(Y_c \leq r|\ \bar{X}_n + bS \geq V, a_L) = P(W \leq r|\ \bar{X}_n + bS \geq V, a_L),$$

where $W = \sum_{j=1}^n W_{Lj}$. It is important to note that because of the conditioning constraint that the average criterion is satisfied, the indicators W_{Lj} are conditionally *dependent*. In our case—though not independent the indicators are identically distributed— $W_{Lj}|\ (\bar{X}_n + bS \geq V, a_L) \sim \text{Bin}(1, p_W)$ for every j. Approximations for the distribution of the sum of

dependent indicators have been discussed extensively (Bahadur 1960; Altham 1978; Soon 1996). After exploring several of these options for the distribution of $W|\ (\bar{X}_n + bS \geq V, a_L)$, we chose the $_{\mathcal{D}}$approximation from Soon (1996), $W|\ (\bar{X}_n + bS \geq V, a_L) \approx \text{Bin}(n, p_W)$, for obtaining

$$P(Y_c \leq r|\ \bar{X}_n + bS \geq V, a_L) \approx \sum_{q=0}^{r} \binom{n}{q} p_W^q (1 - p_W)^{n-q}. \tag{10}$$

For more details on the approximate distribution of $W|\ (\bar{X}_n + bS \geq V, a_L)$ see Remark 2 at the end of this section. In summary, the two components of (6) are evaluated using (7) and (10). Plugging these into (5), a reasonable approximation for the probability of acceptance is

$$P(\mu, \sigma_a, \sigma_\epsilon) \approx \int_{-\infty}^{\infty} [1 - F_{t':n-1,\delta}(-b\sqrt{n})] \times \sum_{q=0}^{r} \binom{n}{q} p_W^q (1 - p_W)^{n-q} f_{a_L}(a_L) da_L. \tag{11}$$

This approximation is valid for a sample from the two variance component model given in (4) and the conservative sampling assumption that all items in the sample were produced close together in time. We now consider two possible generalizations.

3.1 Multiple Variance Components

If we maintain the sampling assumption that all items in the sample were produced at the same time, it is straightforward to generalize to process models with more (or less) than two variance components. First, if there is *only* unit-to-unit variability in the process (as assumed by Elder and Muse (1982) and others), averaging over the extra source of variability in (5) is not needed. In this simpler case, the probability of sample acceptance reduces to

$$P(\mu, \sigma_\epsilon) \approx [1 - F_{t':n-1,\delta_0}(-b\sqrt{n})] \sum_{q=0}^{r} \binom{n}{q} p_W^q (1 - p_W)^{n-q}. \tag{12}$$

The computational difference to note here is the absence of the outer most integral and the reduced noncentrality parameter of the noncentral t distribution in (12), $\delta_0 = \sqrt{n}(\mu - V)/\sigma_\epsilon$, also used in the evaluation of p_W in (8) for this case.

Alternatively, consider a production process with K different stages (e.g., personnel shifts, batches, plant locations, lots, etc.). The more general version of model (4) with K variance components—plus unit-to-unit variability—is

$$X_{i_1 i_2 \ldots i_K j} = \mu + \sum_{k=1}^{K} a_{i_k} + \epsilon_{i_1 i_2 \ldots i_K j},$$

where all the means associated with the different stages, a_{i_k}, $k = 1, \ldots, K$, are independently distributed as $N(0, \sigma_{a_k}^2)$, independent of $\epsilon_{i_1 i_2 \ldots i_K j} \overset{\text{iid}}{\sim} N(0, \sigma_\epsilon^2)$. Though on the surface more complicated, results readily follow the given assumption that all sampled items were produced at the same time. Suppose that $(i_1, \ldots, i_K) = (L_1, \ldots, L_K)$ for all $j = 1, \ldots, n$, with lot random effects $a_{i_k} = a_{L_k}$ given for $k = 1, \ldots, K$, that is, the production stages are the same for all items in the sample. Define $a_L = \sum_{k=1}^{K} a_{L_k}$ and $\sigma_a^2 = \sum_{k=1}^{K} \sigma_{a_k}^2$. Substituting these redefined parameters, the sampling distributions of $\bar{X}_n|a_L$ and $S|a_L$ are the same as in the two variance component case, and the probability of sample acceptance can still be written as (11).

3.2 Mixed Sampling Schemes

We now consider relaxing the assumption that $M = 1$ in (4). For notational convenience, we use the two variance component model (4) to illustrate how more general sampling plans can be incorporated. More variance components can be considered similarly.

Suppose that n_1 units in the sample were produced at the same time and the other n_2 units were also produced close together (but at a different time), $n_1 + n_2 = n$. Denote the data vector in this case as $\mathbf{X} = (X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2})'$, where the first subscript indicates whether an observation came from lot 1 or lot 2. Let $\boldsymbol{\theta} = (a_1, a_2)'$ be the random effects for the two lots, $\boldsymbol{\eta} = (n_1, n_2)'$ be the sampling scheme, and let

$$\mathbf{H} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} \end{bmatrix},$$

where $\mathbf{1}_{n_i}$ and $\mathbf{0}_{n_i}$ are $n_i \times 1$ vectors of 1s and 0s, respectively, for $i = 1, 2$. Given the lot random effects $\boldsymbol{\theta}$ and the sampling scheme $\boldsymbol{\eta}$, we can write $\mathbf{X}|(\boldsymbol{\theta}, \boldsymbol{\eta}) \sim N_n(\mu\mathbf{1}_n + \mathbf{H}\boldsymbol{\theta}, \sigma_\epsilon^2 \mathbf{I}_n)$. This implies the sampling distributions $\bar{X}_n|(\boldsymbol{\theta}, \boldsymbol{\eta}) \sim N(\mu + \bar{a}, \sigma_\epsilon^2/n)$ and $\frac{(n-1)S^2}{\sigma_\epsilon^2}|(\boldsymbol{\theta}, \boldsymbol{\eta}) \sim \chi_{n-1}^2(\lambda)$, where

$$\bar{a} = \frac{n_1 a_1 + n_2 a_2}{n}$$

and $\chi_{n-1}^2(\lambda)$ distribution is a noncentral χ^2 distribution with noncentrality parameter

$$\lambda = \frac{n_1(a_1 - \bar{a})^2 + n_2(a_2 - \bar{a})^2}{\sigma_\epsilon^2}.$$

Let $f_a(\boldsymbol{\theta})$ be the bivariate normal density, $N_2(\mathbf{0}_2, \sigma_a^2 \mathbf{I}_2)$, then the probability of sample acceptance can be computed by averaging over the random effects similar to (5), that is,

$$P(\mu, \sigma_a, \sigma_\epsilon) = \int \int_{\Theta} P(\bar{X}_n + bS \geq V, Y_c \leq r|\ \boldsymbol{\theta}, \boldsymbol{\eta}) f_a(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{13}$$

where $\Theta = \mathbb{R}^2$. The integrand in (13) can be evaluated as in (6), but when there is a mixture of two lots in the sample, the first factor on the righthand side of (6) is now given by the doubly noncentral t distribution with $n - 1$ degrees of freedom and noncentrality parameters $\delta = \sqrt{n}(\mu + \bar{a} - V)/\sigma_\epsilon$ and λ, defined above. To evaluate the second term in (6) with mixed lots, we define two sets of indicators $W_{ij} = I[X_{ij} \leq c]$, $j = 1, \ldots, n_i$, $i = 1, 2$, where $W_{ij} = 1$, if unit j that was produced in lot i is defective, and 0 otherwise. Conditional on the average criterion being satisfied, the probability that unit j in lot i is defective is

$$\begin{aligned} p_{W_i} &= P(W_{ij} = 1|\ \bar{X}_n + bS \geq V, \boldsymbol{\theta}, \boldsymbol{\eta}) \\ &= P(X_{ij} \leq c|\bar{X}_n + bS \geq V, \boldsymbol{\theta}, \boldsymbol{\eta}),\ i = 1, 2;\ j = 1, \ldots, n_i. \end{aligned}$$

Derivations of p_{W_1} and p_{W_2} are similar to the one lot case (see Appendix C). Using these indicators,

$$P(Y_c \le r|\bar{X}_n + bS \ge V, \boldsymbol{\theta}, \boldsymbol{\eta}) = P(W' \le r|\bar{X}_n + bS \ge V, \boldsymbol{\theta}, \boldsymbol{\eta}),$$

where $W' = \sum_{j=1}^{n_1} W_{1j} + \sum_{j'=1}^{n_2} W_{2j'}$. As before, these indicators are conditionally dependent, but now the distribution of $W_{ij}|(\bar{X}_n + bS \ge V, \boldsymbol{\theta}, \boldsymbol{\eta})$ are also not identical between lots (i.e., for different values of i), but rather only within lots. An alternative approximation from Soon (1996) can be used to deduce that $W'|(\bar{X}_n + bS \ge V, \boldsymbol{\theta}, \boldsymbol{\eta}) \overset{\mathcal{D}}{\approx} \text{Bin}(n', p'_W)$, where

$$n' = \left\lfloor \frac{(n_1 p_{W_1} + n_2 p_{W_2})^2}{n_1 p_{W_1}^2 + n_2 p_{W_2}^2} + \frac{1}{2} \right\rfloor,$$

and

$$p'_W = \frac{n_1 p_{W_1} + n_2 p_{W_2}}{n'}.$$

The notation $\lfloor x \rfloor$ denotes the integer part of x.

In summary, when n_1 items in the sample are from one lot and the remaining n_2 items are from a second lot, the probability of sample acceptance becomes

$$P(\mu, \sigma_a, \sigma_\epsilon) \approx \int_\Theta [1 - F_{t'':n-1,\delta,\lambda}(-b\sqrt{n})] \times \sum_{q=0}^{r} \binom{n'}{q} (p'_W)^q (1 - p'_W)^{n'-q} f_a(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (14)$$

where $F_{t'':n-1,\delta,\lambda}$ is the cumulative distribution function of a doubly noncentral t distribution.

Remarks:

(1) In addition to incorporating beliefs about how many lots are represented, prior distributions can also be placed on $\boldsymbol{\eta}$, the number of items from each lot. For example, for $\boldsymbol{\eta} = (n_1, n_2)'$, $\boldsymbol{\eta} \sim$ multinomial$(n; 1/2, 1/2)$ would specify an uninformative prior on the sample size distribution. To incorporate this into (13),

$$P(\mu, \sigma_a, \sigma_\epsilon) = \sum_{\boldsymbol{\eta}} \int_\Theta P(\bar{X}_n + bS \ge V, Y_c \le r|\, \boldsymbol{\theta}, \boldsymbol{\eta}) f_a(\boldsymbol{\theta}) d\boldsymbol{\theta}\, P(\boldsymbol{\eta}),$$

where $P(\boldsymbol{\eta})$ denotes the joint probability mass function of $\boldsymbol{\eta}$. Note that the multinomial prior places most of its weight on a mixed sample, but does also give small weight to a one-lot-only sample. Subsequently, the procedure for evaluating the conditional joint probability in the integrand will depend on if it is a mixed sample or a one-lot-only sample.

(2) The conditional dependence of W_{ij}'s makes the computation of $P(Y_c \le r|\bar{X}_n + bS \ge V, *)$ challenging under the sampling scenarios described above. For instance, in the one variance component case, $\text{corr}(W_i, W_j|\bar{X}_n + bS \ge V) = (p_{12} - p_1^2)/(p_1 - p_1^2)$, for $i \ne j$, where $p_1 = P(X_i \le c|\bar{X} + bS \ge V)$ and $p_{12} = P(X_i \le c, X_j \le c|\bar{X} + bS \ge V)$. By using the definition of conditional probability, it is easy to show that $P(X_i \le c, X_j \le c|\bar{X} + bS \ge V) = P(X_i \le c|X_j \le c, \bar{X} + bS \ge V)P(X_j \le c|\bar{X} + bS \ge V)$, and $P(X_i \le c|\ X_j \le c, \bar{X} + bS \ge V) \le P(X_i \le c|\bar{X} + bS \ge V)$. Consequently, $p_{12} \le p_1^2$ and hence, $\text{corr}(W_i, W_j|\bar{X}_n + bS \ge V) \le 0$. However, the magnitude of the correlation coefficient depends on several quantities such as the mean (μ) and variance component (σ_ϵ^2) in the distribution of X_{Lj}, and parameters in the acceptance criteria, such as the sample size (n), the individual cutoff (c) and the average cutoff (V). If there are more variance components and/or multiple lots in the sample, even more parameters contribute to the dependence structure.

For the setup in Example 4, we investigated the accuracy of the binomial approximation (proposed by Soon 1996) used for the distribution of $W|(\bar{X}_n + bS \ge V)$. It turns out that the discrepancy measured in terms of Kolmogorov–Smirnov distance between the simulated distribution of $W|(\bar{X}_n + bS \ge V)$ based on Monte Carlo simulation with 10^6 random samples, and the binomial approximation, by Soon (1996), is essentially zero in the regions of interest (see Figure 6). The discrepancy turned out to be of the order of 10^{-3} in regions where probability of passing is quite small (less than 0.7).

4. EXAMPLES

Example 2. Recall Example 1 that was outlined in Section 2 and helped motivate this work. This production process follows the two variance component model in (4). The initial question posed was as follows: given a production process which has two known sources of variability, what production mean should be targeted to achieve a prespecified probability of passing?

Due to the presence of the lot-to-lot variability component, the answer to this question depends on how many lots were included in the sample. Under the conservative assumption that all sampled items come from the same production lot, the probability of sample acceptance can be calculated using (11). Alternatively, under the more favorable sampling scenario that the inspection sample consists of a mixture of two production lots, the probability of sample acceptance is given by (14). For comparison, (12) gives the probability of sample acceptance when the variance components are ignored.

Using the specified regulatory constants ($V = 40$ oz, $c = 38.624$ oz, $r = 0$, $n = 12$) and variance components ($\sigma_a^2 = 0.0354$, $\sigma_\epsilon^2 = 0.0826$, $\sigma_{\text{total}}^2 = \sigma_a^2 + \sigma_\epsilon^2 = 0.118$), Figure 2 shows the probability of sample acceptance as calculated by (11), (14), and (12) for a sequence of production targets, μ, and sample size mixtures (n_1, n_2). In Figure 2, the two curves corresponding to the one-lot sampling scenario with and without variance decomposition (which were previously calculated and plotted in Figure 1 via simulation) are highlighted in bold. As can be seen from Figure 2, in the two variance component case the calculated probability of sample acceptance is always lowest for a given target under the one-lot sampling scheme, emphasizing the conservative nature of this scenario.

In the situation where the manufacturer has little flexibility in tuning process variability, the curves in Figure 2 can be used to determine how much overfilling is required to achieve a prespecified probability of sample acceptance under the fixed process variability settings and a variety of sampling scenarios. In this example, if the Procter and Gamble Company aims for a 95%

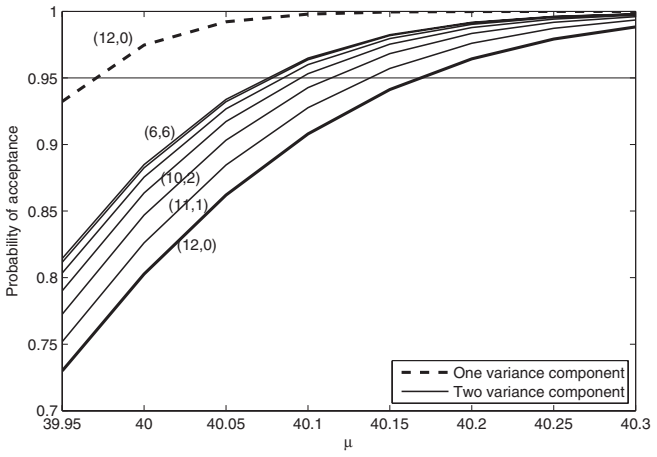

Figure 2. Probability of sample acceptance as a function of μ with and without variance decomposition. All total sample sizes are $n = 12$; lot size mixtures given as (n_1, n_2).

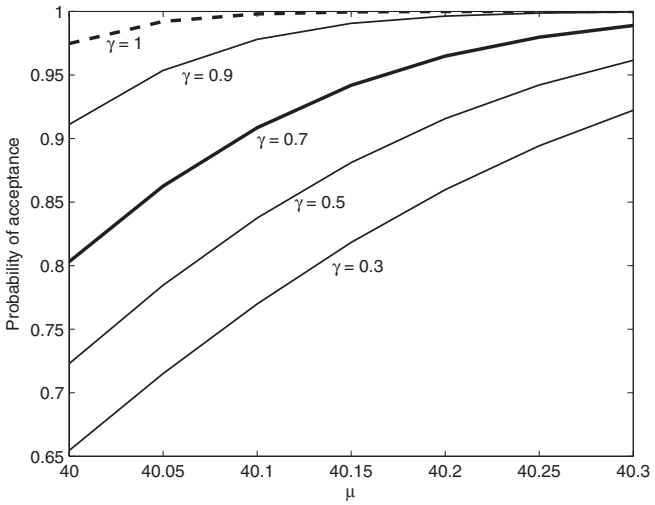

Figure 3. Probability of sample acceptance as a function of μ for a variety of variance decompositions, γ. Calculations based on a one-lot sampling scheme with $n = 12$ and fixed σ^2_{total}.

probability of sample acceptance, they should conservatively set their production target to $\mu = 40.17$. In contrast, if it is known that the sample will consist of an equal mixture of two lots, the same probability of sample acceptance could be achieved by only overfilling to $\mu = 40.07$. Erroneously tuning the process to the dotted line (calculated without the variance decomposition), the target corresponding to 95% sample acceptance is $\mu = 39.97$, effectively resulting in an acceptance probability of only 76.1% under the conservative sampling scheme.

Alternatively, if tuning of the variance components is possible, (11) can similarly be used to assess variance reduction strategies (restricting attention to the one-lot sampling scenario). Here we consider two possibilities: (a) the total variability, $\sigma^2_{\text{total}} = 0.118$, is held constant but the proportion of variability due to noise, which we denote $\gamma = \sigma^2_\epsilon / \sigma^2_{\text{total}}$, can be varied; and (b) the proportion of variability due to noise, γ, is fixed to 70% but the total variability can be changed. Figure 3 shows the acceptance probability as a function of the target mean for a variety of variance decompositions. As γ increases (i.e., the lot-to-lot variability decreases), less overfilling is required to achieve the same probability of acceptance. Figure 4 shows contours of the probability of acceptance surface for a number of combinations of μ and σ^2_{total}. For this example, there seems to be a limited benefit in reducing the total variability (with the proportion of lot-to-lot variability, γ, held fixed at 0.7). While less total variability does decrease the target required for a specified probability of acceptance, the narrow gaps between contour lines suggests that the probability of acceptance decreases quickly with small changes in μ when the total variability is reduced. In practice, costs of overfilling can be weighed against costs of reducing variability to achieve an optimum strategy.

In some instances the variances of the random effects can be estimated with very little error from a wealth of data acquired by online monitoring and inspection. This may not be the case in some applications, and the variance components (e.g., unit-to-unit, lot-to-lot, etc.) have to be estimated from experiments. In the next example, we present a simulation study to investigate the impact of (a) the estimation error in the variance component, (b) misspecification of the number of variance components and (c) the percentage of lot-to-lot variance in computing the joint probability of acceptance.

Example 3. Consider the setup of Examples 1 and 2. The parameters were set to $\mu = 40$, MAV $= 1.376$, $c = \mu - \text{MAV} = 38.624$, $r = 0$, $\sigma^2_{\text{total}} = (\text{MAV}/k)^2$ for $k = 1, 2.5, 4$ and $\sigma^2_\epsilon = \gamma\sigma^2_{\text{total}}$ for $0 \le \gamma \le 1$. The choices of k span the range of capability that most processes will have in relation to the MAV. Note that a small value of k implies large variation in the data. Consequently, as k increases, the probability of passing the individual criterion ($Y_c = 0$) increases. In this example, $k = 1$ implies $\sigma^2_{\text{total}} \approx 1.89$ and the six-sigma range for the data is $(35.87, 44.13)$. That is, the individual criterion ($Y_{38.624} = 0$) would be difficult to satisfy. The results for $k = 1$ is not very appealing from a company's perspective, however, they show how the estimation affects the performance by using the two models in an extreme case.

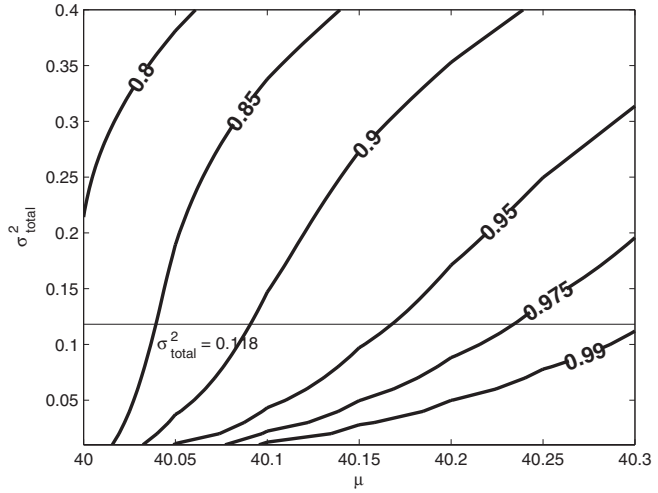

Figure 4. Contours of the probability of sample acceptance as a function of μ and σ^2_{total}. Calculations based on a one-lot sampling scheme with $n = 12$ and fixed $\gamma = 0.7$.

In this example, we use simulation results to illustrate the impact of the three issues (a)–(c) mentioned above in estimating the probability of passing the joint acceptance criterion. The variance components σ_ϵ^2 and σ_a^2 are estimated using the restricted maximum likelihood (REML) approach from randomly generated populations with m lots of 10 units per lot. The simulation results in Tables 1 and 2 are summarized for different combinations of $m = 10, 30, 50$, $k = 1, 2.5, 4$, and $\gamma = 0.75, 0.9$. We used $n = 12$ as this is the most common type of sampling for NIST. Since the probability estimation under two variance components model is computationally time consuming, the simulation results summarized here are based on roughly 50 realizations.

Truth—One Variance Component. The goal of this simulation is to find out how the misspecification in the number of variance components and the uncertainty in the estimates themselves affect the calculated probabilities. Assume that the true process is in perfect control and there is a single variance component (unit-to-unit). That is, all observations in the random population (of m lots with 10 units per lot) are iid $N(\mu, \sigma_{\text{total}}^2)$, where $\sigma_{\text{total}}^2 = \sigma_\epsilon^2$. But we include a lot-to-lot variance component in our analysis, using REML to compute the estimates. Table 1 summarizes the mean and standard error of several realizations of (a) P_{true}—the probability of passing, computed using the true parameters $\sigma_\epsilon = \sigma_{\text{total}} = \text{MAV}/k$ and $\sigma_a = 0$ (b) $P_{\text{est}}(1$ var)—the probability of passing, computed using the estimate of σ_ϵ^2 obtained from REML (under the assumption of $\sigma_a = 0$) and (c) $P_{\text{est}}(2$ var)—the probability of passing computed using the estimate of σ_ϵ^2 and σ_a^2 from REML (under the assumption of two variance components with $n_1 = 0, 3, 6$ and $n_2 = n - n_1$). The numbers in the parentheses are the standard errors.

From Table 1, it is clear that $P_{\text{est}}(1$ var) provides the best approximation for the truth in all the cases considered here. Similar to Figure 2, the estimated probabilities for $k \geq 2.5$ slightly increase with the mixing of the samples from the two lots, but the probability estimates for $k = 1$ show somewhat opposite trend as it becomes more difficult to satisfy the individual criterion ($Y_c = 0$) due to large variation in the data. This is because the individual criterion becomes dominant when k is small, whereas for large k the mean criterion is the dominant one. Moreover, Table 1 also shows that the standard error of the estimated probabilities is approximately directly proportional to σ_{total}^2 and inversely proportional to the number of lots (M).

Truth—Two Variance Components. Assume that the true underlying population has two variance components, lot-to-lot variance σ_a^2 and within lot (unit-to-unit) variance σ_ϵ^2. The observations for the ith lot of the random population are generated from $N(\mu + a_i, \sigma_\epsilon^2)$, where $a_i \sim N(0, \sigma_a^2)$ is the lot random effect for the ith lot and $\sigma_{\text{total}}^2 = \sigma_\epsilon^2 + \sigma_a^2$. The objective of this simulation is to show how our method will more accurately provide the correct probability (i.e., under the two variance components model) over the single variance model by showing $P_{\text{true}}(2$ var) and $P_{\text{est}}(2$ var) along with $P_{\text{est}}(1$ var). Similar to the one variance component case, Table 2 summarizes the mean and the standard errors. Unlike the one variance component case, $P_{\text{true}}(2$ var) values are computed for $n_1 = 0, 3, 6$ and $n_2 = n - n_1$. In addition to different combinations of the number of lots (M) and σ_{total}^2 (or equivalently k), the simulation results are also classified by the percentage of lot-to-lot variance component $\gamma = \sigma_\epsilon^2/\sigma_{\text{total}}^2 = 0.9$ and 0.75.

Similar to the one variance component model case, Table 2 shows that the proposed approach provides a good approximation of the truth under the correct assumption of the number of variance components and $[n_1, n_2]$ combination. Similar to Table 1, the simulation results for $k = 1$ exhibit the reverse trend

Table 1. Truth: the data come from one variance component scenario. The entries, which are in the format: mean (standard error) of the probabilities, show the distribution of the estimated probability of passing the joint acceptance criterion under different assumptions

				$P_{\text{est}}(2$ var)		
		P_{true}	$P_{\text{est}}(1$ var)	[0, 12]	[3, 9]	[6, 6]
# lots = 10	$k = 1$	0.1294	0.1352	0.1481	0.1350	0.1337
			(0.0010)	(0.0012)	(0.0103)	(0.0102)
	$k = 2.5$	0.9061	0.9042	0.8939	0.8905	0.8923
			(0.0010)	(0.0011)	(0.0102)	(0.0099)
	$k = 4$	0.9746	0.9743	0.9614	0.9656	0.9675
			(2.7×10^{-5})	(0.0006)	(0.0030)	(0.0023)
# lots = 30	$k = 1$	0.1294	0.1317	0.1391	0.1387	0.1375
			(0.0005)	(0.0006)	(0.0041)	(0.0038)
	$k = 2.5$	0.9061	0.9047	0.8992	0.8950	0.8960
			(0.0005)	(0.0006)	(0.0054)	(0.0054)
	$k = 4$	0.9746	0.9745	0.9679	0.9709	0.9716
			(1.2×10^{-5})	(0.0003)	(0.0015)	(0.0012)
# lots = 50	$k = 1$	0.1294	0.1314	0.1368	0.1366	0.1357
			(0.0004)	(0.0005)	(0.0033)	(0.0032)
	$k = 2.5$	0.9061	0.9049	0.9004	0.9023	0.9036
			(0.0004)	(0.0005)	(0.0028)	(0.0028)
	$k = 4$	0.9746	0.9745	0.9692	0.9735	0.9731
			(9.1×10^{-6})	(0.0002)	(0.0003)	(0.0006)

Table 2. Truth: the data come from the two variance components scenario. The entries, which are in the format: mean (standard error) of the probabilities, show the distribution of the estimated probability of passing the joint acceptance criterion under different assumptions

			P_{true}(2 var)				P_{est}(2 var)		
			[0, 12]	[3, 9]	[6, 6]	P_{est}(1 var)	[0, 12]	[3, 9]	[6, 6]
# lots = 10	$k = 1$	$\gamma = 0.90$	0.1948	0.1686	0.1611	0.1333	0.1994	0.1675	0.1604
						(0.0037)	(0.0051)	(0.0081)	(0.0075)
		$\gamma = 0.75$	0.2858	0.2263	0.2115	0.1359	0.2849	0.2115	0.2035
						(0.0044)	(0.0059)	(0.0081)	(0.0128)
	$k = 2.5$	$\gamma = 0.90$	0.8599	0.8766	0.8522	0.9072	0.8625	0.8673	0.8711
						(0.0030)	(0.0044)	(0.0074)	(0.0072)
		$\gamma = 0.75$	0.8027	0.8391	0.8509	0.9035	0.8149	0.8460	0.8464
						(0.0047)	(0.0054)	(0.0181)	(0.0157)
	$k = 4$	$\gamma = 0.90$	0.9110	0.9371	0.9454	0.9742	0.9120	0.9401	0.9463
						(0.0001)	(0.0047)	(0.0048)	(0.0038)
		$\gamma = 0.75$	0.8269	0.8812	0.8997	0.9740	0.8259	0.9060	0.8842
						(0.0001)	(0.0056)	(0.0104)	(0.0125)
# lots = 30	$k = 1$	$\gamma = 0.90$	0.1948	0.1686	0.1611	0.1322	0.1946	0.1647	0.1569
						(0.0018)	(0.0033)	(0.0042)	(0.0048)
		$\gamma = 0.75$	0.2858	0.2263	0.2115	0.1365	0.2853	0.2287	0.2138
						(0.0024)	(0.0034)	(0.0061)	(0.0048)
	$k = 2.5$	$\gamma = 0.90$	0.8599	0.8766	0.8822	0.9058	0.8592	0.8722	0.8763
						(0.0021)	(0.0029)	(0.0039)	(0.0047)
		$\gamma = 0.75$	0.8027	0.8391	0.8509	0.9076	0.8071	0.8385	0.8502
						(0.0023)	(0.0031)	(0.0075)	(0.0056)
	$k = 4$	$\gamma = 0.90$	0.9110	0.9371	0.9454	0.9746	0.9220	0.9441	0.9509
						(3.8×10^{-5})	(0.0029)	(0.0025)	(0.0024)
		$\gamma = 0.75$	0.8269	0.8812	0.8997	0.9745	0.8313	0.8895	0.9089
						(4.5×10^{-5})	(0.0030)	(0.0055)	(0.0041)
# lots = 50	$k = 1$	$\gamma = 0.90$	0.1948	0.1686	0.1611	0.1289	0.1925	0.1693	0.1615
						(0.0013)	(0.0022)	(0.0026)	(0.0027)
		$\gamma = 0.75$	0.2858	0.2263	0.2115	0.1287	0.2835	0.2245	0.2092
						(0.0017)	(0.0022)	(0.0034)	(0.0026)
	$k = 2.5$	$\gamma = 0.90$	0.8599	0.8766	0.8821	0.9044	0.8604	0.8741	0.8785
						(0.0016)	(0.0024)	(0.0035)	(0.0036)
		$\gamma = 0.75$	0.8027	0.8391	0.8509	0.9057	0.8040	0.8371	0.8478
						(0.0019)	(0.0024)	(0.0045)	(0.0032)
	$k = 4$	$\gamma = 0.90$	0.9110	0.9371	0.9454	0.9746	0.9143	0.9388	0.9460
						(2.5×10^{-5})	(0.0018)	(0.0017)	(0.0013)
		$\gamma = 0.75$	0.8269	0.8812	0.8997	0.9745	0.8334	0.8838	0.9040
						(3.4×10^{-5})	(0.0025)	(0.0037)	(0.0021)

as compared to that for $k \geq 2.5$. For instance, the probability values are inversely proportional to γ for $k = 1$ and directly proportional to γ for $k = 2.5$ and 4. In addition, P_{est}(1 var) underestimates the probabilities when k is small and overestimates the probabilities of passing the joint acceptance criterion when k is large.

It is clear from Tables 1 and 2 that all standard errors for P_{est}(1 var) are less than 0.005. For the estimation of P_{est}(2 var) using at least 30 lots of data, the standard errors are also often below 0.005 (16/18 in Table 1 and 32/36 in Table 2). This indicates that most of the time the estimated probabilities would be within $\pm 2(0.005) = 0.01$. In general, the amount of historical data for most products will be extensive enough to obtain at least 30 lots worth of data to estimate the variance components. However, if the historical data is limited, the 10 lot simulation results show that 23 out of 27 estimates of P_{est}(2 var) in Tables 1 and 2 have standard errors that are about 0.01 or less, and the maximum is 0.018. That is, one should be cautious to go below 10 lots for the estimation of the variance components and resulting probabilities.

In summary, the simulation study shows that when there are truly multiple variance components we need to account for them. In the case when there is truly just a single variance component and we model two, we do not lose much accuracy or precision of the estimated probability in regions that most companies would be operating in ($k \geq 2.5$).

Example 4. Constantine, Field, and Robinson (2000)—hereafter referred to as CFR—evaluated the probability of sample acceptance under the joint acceptance criteria (1) and (2) using an implicit assumption that the sample standard deviation (S) and the number of defective items (Y_c) are independent. Here, we use their wine bottle filling process example to assess the impact of this assumption on the accuracy of the calculated acceptance probability. Since the methodology of CFR only applies when there is one source of variability, namely σ_ϵ^2, we

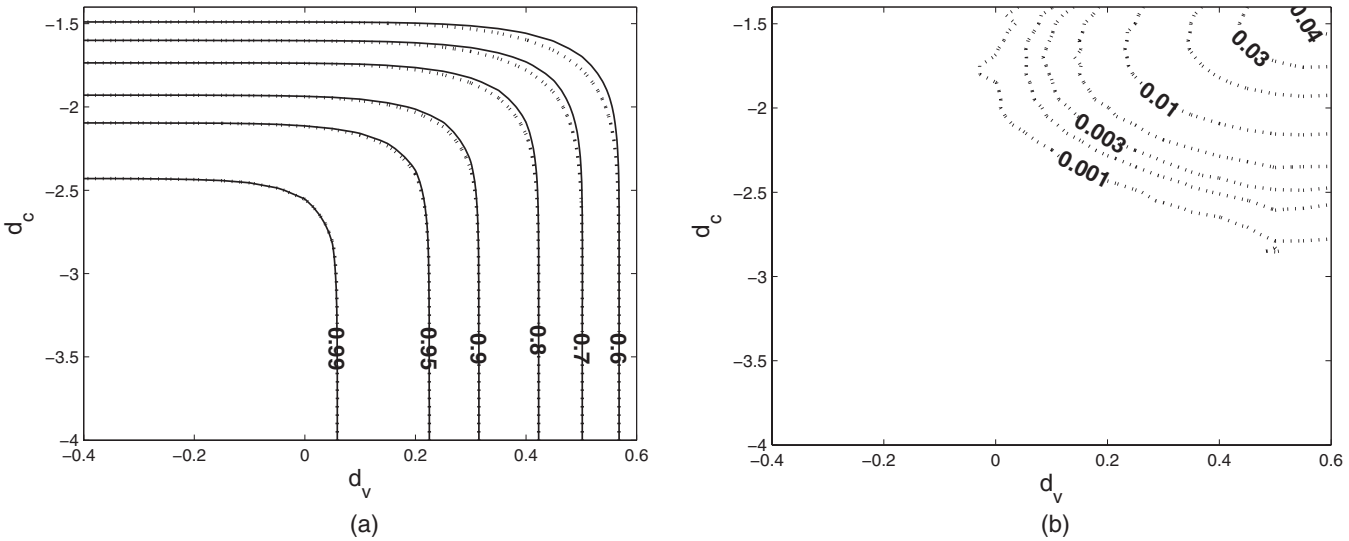

Figure 5. Probability of acceptance contours for the wine bottle example. Comparison of CFR methodology and Monte Carlo simulation results. (a) Solid: CFR; dashed: simulation, (b) dashed: discrepancy with simulation.

restrict our attention to the proposed comparable approximation given in (12) which uses the binomial approximation (10). For this example, the desired process features were $V = 750$ mL, $c = 735$ mL, $n = 20$, $r = 1$, and $b = t_{0.995,19}/\sqrt{20} = 0.640$, and the goal was to calculate the probability of sample acceptance for a variety of μ and σ_ϵ^2 values. We consider how accurately our proposed approximation given in (12) provides parameter (μ, σ_ϵ) settings that correspond to a prespecified probability of sample acceptance, and how accurately it gives the probability of sample acceptance for a given set of parameter (μ, σ_ϵ) values.

Following CFR, define $d_V = (V - \mu)/\sigma_\epsilon$ and $d_c = (c - \mu)/\sigma_\epsilon$. Without loss of generality, assume that $X_1, \ldots, X_n$ are $N(0, 1)$ distributed and subject to regulatory constraints $Y_{d_c} \leq r$ and $\bar{X}_n + bS \geq d_V$. Note that any combination (d_V, d_c) translates directly to the parameter setting $(\mu = V - \sigma_\epsilon d_V, \sigma_\epsilon = (c - V)/(d_c - d_V))$ in the original scale of the problem. The standardization facilitates calculating probabilities of sample acceptance corresponding to production processes on any scale of measurement. Figure 5(a) gives contours of the probability of sample acceptance calculated using the CFR method in solid lines, and similarly those obtained using Monte Carlo simulation in dashed lines for a variety of values of d_V and d_c. Alternatively, Figure 5(b) shows the contours of the discrepancy, P (acceptance estimated using CFR method) $-P$(acceptance estimated using Monte Carlo simulation), that speak to inaccuracies in parameter tuning for prespecified probabilities of sample acceptance.

It can be seen in Figure 5 that when d_V is large and negative (i.e., the process target μ is many standard deviations above the label V) so that $P(\bar{X}_n + bS \geq d_V) \approx 1$, or alternatively d_c is negative enough that $P(Y_{d_c} \leq r) \approx 1$ (this is the case in Example 1), then the independence assumption is moot and there is no approximation error. On the other hand, in the subset of the parameter space where neither of these marginal probabilities are essentially 1 (at the bend of the contours in Figure 5(a)), the independence assumption results in a systematic overestimation of the probability of passing inspection for a given parameter setting. Correspondingly, if the CFR method is used to find the parameter settings that yield a prespecified probability of passing, the necessary amount of overfilling is underestimated. While the error in the probability calculation can be as much as a few percent, the implications are relatively minor since they primarily occur in regions of the parameter space with low probabilities of sample acceptance, which producers will naturally avoid anyway. It is enough of a concern, however, to motivate the development of a new methodology for the multiple variance component setting.

In contrast, probability contour plots calculated using (12) are given in Figure 6. Figure 6(a) shows that the probability of acceptance contours estimated by (12) and by Monte Carlo simulation are visually indistinguishable. A closer look at the contours of the discrepancy between (12) and simulation, P(acceptance estimated using (12)) $-P$ (acceptance estimated using simulation), shows that the binomial approximation error for a given parameter setting is at most only a small fraction of a percentage point (see Figure 6(b)).

Example 5. The standardization of (V, c) to (d_V, d_c) can similarly be applied in the two-variance-component setting. Define $d_V = (V - \mu)/\sigma_{\text{total}}$ and $d_c = (c - \mu)/\sigma_{\text{total}}$. For a fixed proportion of noise variability, γ, plots similar to Figures 5 and 6 can be drawn. A slight inconvenience is that this plot must be redrawn for different proportion γ to be considered. In comparison with the standardized settings, it would typically be more practical to focus on visualization of such contours in the original unit of a particular problem (as in Example 2). However, the standardization is useful for gaining insight into the problem and confirming that the impact of the binomial approximation is minimal under a wide variety of parameter settings (we considered $r = 0$, 1 and $\gamma = 0.5, 0.7, 1$).

For each of these six parameter combination settings, we computed the probability of acceptance using (11) in the region $\{(d_V, d_c) : -0.5 \leq d_V \leq 1, -4 \leq d_c \leq -1.5\}$ and compared with the simulation results. The contours of the

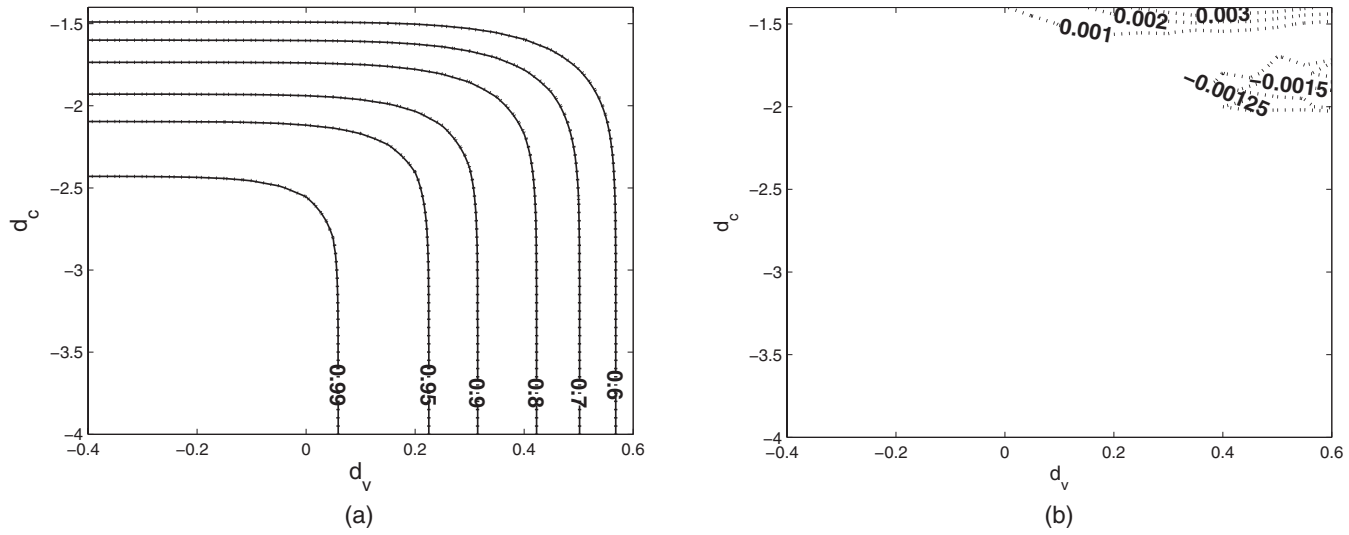

Figure 6. Probability of acceptance contours for the wine bottle example. Comparison of approximation given by (12) and Monte Carlo simulation results. (a) Solid: evaluation of (12); dashed: simulation, (b) dashed: discrepancy with simulation.

discrepancy between the probability of acceptance estimated using (11) and simulation look similar to that shown in Figure 6 and thus are omitted from the article. Nonetheless, there are a few points worth noting.

The error induced by the binomial approximation (10), that is, the discrepancy P (acceptance estimated using (11)) $-P$ (acceptance estimated using simulation), never exceeds a small fraction of a percent anywhere in the specified (d_V, d_c) region for any given (r, γ) combination we considered. As r changes from 0 to 1, there is no difference in acceptance probabilities in the d_V direction, but higher passing probabilities are achieved with μ being fewer standard deviations above c (there are changes in the d_c direction). When one defective item is permitted in the sample, the process mean does not have to exceed the defective level as much as it does when no defective items are permitted. Moreover, as γ increases a producer has more flexibility in parameter tuning at high probabilities of acceptance. For example, when $\gamma = 0.5$, there are very few settings that yield a 99% probability of acceptance. There are more possibilities when $\gamma = 0.7$ or $\gamma = 1$. This parallels the earlier observation that less overfilling is required when the proportion of lot-to-lot variability decreases.

5. DISCUSSION

Random government inspections on net content hold manufacturers accountable to their customers, but it is mutually important for manufacturers to be able to run their processes efficiently and cost-effectively within acceptable bounds. Having a means to calculate the probability of passing inspections for given process parameters helps achieve this balance. In this article, we derive a new methodology for calculating the probability of sample acceptance with two primary innovations. First, by conditioning in the way *opposite* to Elder and Muse (1982), we are able to derive an approximation to the acceptance probability under the more complex current joint acceptance criteria. Second, by assuming a random effects model for the production process, we allow for the incorporation of multiple sources of variation. A byproduct of this random effects approach is that *how* items are sampled becomes a concern. Assuming all items come from one production lot is a conservative sampling scenario with computational benefits. Generalizations to more refined knowledge or prior beliefs about the sampling scheme can be incorporated as desired on a case-by-case basis.

One challenge associated with the use of a random effects model is that it is highly individualized. Though the theory is quite general, computations and visualizations are most useful when they are tailored to a specific problem. There is no convenient standardization for generating multiuse operating characteristic (OC) curves or tables of acceptance probabilities. This is not prohibitive, however, because although different applications may require their own study, the probability approximations we derived are relatively routine calculations under quite broad assumptions. The complexity that can be considered is really only limited by the ability to solve high-dimensional integrals (we used a Gaussian quadrature routine in Matlab with an error tolerance of 10^{-6}). Overall, we feel the theoretical results proposed here are a nice complement to simulation studies or other available tools for exploring process quality.

Finally, the methodology as it currently stands enables manufacturers to calculate the probability of passing for their given process parameters. It does not, however, provide an automatic means of recovering the inverse information, that is, finding all parameter combinations that yield a prespecified probability. As illustrated in Figures 5 and 6, plots can be constructed to help access this information. Having a more concise form for the inverse problem would be extremely helpful for economical process tuning and is future work.

APPENDIX A: DERIVATION OF (7) IN SECTION 3

The probability $P(\bar{X}_n + bS \geq V | a_L)$ can be derived by starting with the distributional assumption that $X_{Lj}|a_L \sim N(\mu + a_L, \sigma_\epsilon^2)$ for all j, and $\{X_{Lj}, j = 1, \ldots, n\}$ is a random sample from $N(\mu + a_L, \sigma_\epsilon^2)$. Thus, $\bar{X}_n|a_L \sim N(\mu + a_L, \sigma_\epsilon^2/n)$ is

conditionally independent of $\frac{(n-1)S^2}{\sigma_\epsilon^2}|a_L \sim \chi^2_{(n-1)}$. Then

$$P(\bar{X}_n + bS \geq V|\, a_L) = P\left(\frac{\bar{X}_n - V}{S} \geq -b|a_L\right) = P\left(\frac{\sqrt{n}(\bar{X}_n - V)/\sigma_\epsilon}{S/\sigma_\epsilon} \geq -b\sqrt{n}|a_L\right). \tag{A.1}$$

Letting $\delta = \sqrt{n}(\mu + a_L - V)/\sigma_\epsilon$, it follows that $\sqrt{n}(\bar{X}_n - V)/\sigma_\epsilon|a_L \sim N(\delta, 1)$. Since the desired probability in (A.1) is the ratio of a conditionally independent noncentral normal random variable and the square root of a chi-squared random variable divided by its degrees of freedom, results from Johnson and Kotz (1970) can be used to yield (7).

APPENDIX B: DERIVATION OF (9) IN SECTION 3

Derivation of $P(X_{Lj} \leq c, \bar{X}_n + bS \geq V|\, a_L)$ starts with the same distributional assumptions as given in Appendix A. Let $f_{\bar{X}_n|a_L}(\cdot)$ and $f_{S|a_L}(\cdot)$ denote the sampling distributions of $\bar{X}_n|a_L$ and $S|a_L$, respectively. Then

$$\begin{aligned} &P(X_{Lj} \leq c, \bar{X}_n + bS \geq V|\, a_L) \\ &= \int_0^\infty P\left(X_{Lj} \leq c, \bar{X}_n + bs \geq V|s, a_L\right) f_{S|a_L}(s)ds \\ &= \int_0^\infty \int_{V-bs}^\infty P\left(\frac{X_{Lj} - \bar{X}_n}{S} \leq \frac{c - \bar{x}_n}{s}|\bar{x}_n, s, a_L\right) \\ &\quad \times f_{\bar{X}_n|a_L}(\bar{x}_n) f_{S|a_L}(s)d\bar{x}_n ds \\ &= \int_0^\infty \int_{V-bs}^\infty P\left(\frac{X_{Lj} - \bar{X}_n}{S} \leq \frac{c - \bar{x}_n}{s}|a_L\right) \\ &\quad \times f_{\bar{X}_n|a_L}(\bar{x}_n) f_{S|a_L}(s)d\bar{x}_n ds \end{aligned} \tag{B.1}$$

The last step follows from an application of Basu's Theorem (see, e.g., Casella and Berger 2001). Define $c_0 = \frac{1}{2}(1 + \frac{\sqrt{n}(c-\bar{x}_n)}{(n-1)s})$. Then the probability in the integrand of (B.1) can be written as

$$P\left(\frac{X_{Lj} - \bar{X}_n}{S} \leq \frac{c - \bar{x}_n}{s}|a_L\right) = P\left(\frac{1}{2}\left\{1 + \frac{\sqrt{n}}{n-1}\left(\frac{X_{Lj} - \bar{X}_n}{S}\right)\right\} \leq c_0|a_L\right).$$

Let

$$U = \frac{\sqrt{n}(X_{Lj} - \bar{X}_n)}{\sqrt{n-1}\sigma_\epsilon} \sim N(0,1) \quad \text{and}$$
$$V = \frac{(n-2)S^2_{(-j)}}{\sigma_\epsilon^2} \sim \chi^2_{n-2},$$

where

$$S^2_{(-j)} = \frac{1}{n-2}\sum_{k\neq j}(X_{Lk} - \bar{X}_{(-j)})^2 \quad \text{and}$$
$$\bar{X}_{(-j)} = \frac{1}{n-1}\sum_{k\neq j} X_{Lk}.$$

By noting that

$$(n-1)S^2 = (n-2)S^2_{(-j)} + \frac{n}{n-1}(X_{Lj} - \bar{X}_n)^2,$$

it can be seen that $\frac{(n-1)S^2}{\sigma_\epsilon^2} = V + U^2$. Define

$$T = \frac{\sqrt{n}}{n-1}\left(\frac{X_{Lj} - \bar{X}_n}{S}\right) = \frac{U}{\sqrt{V + U^2}}.$$

A change of variables from (U, V) to $(T, R = V + U^2)$ and integration over R gives the desired result that

$$\frac{1}{2}\left\{1 + \frac{\sqrt{n}}{n-1}\left(\frac{X_{Lj} - \bar{X}_n}{S}\right)\right\}|\, a_L = \frac{1}{2}(1+T)|\, a_L \sim \text{Beta}\left(\frac{n-2}{2}, \frac{n-2}{2}\right),$$

as used in (9).

APPENDIX C: DERIVATION OF p_{Wi} IN SECTION 3.2

Here we derive the probability p_{W1} and the derivation of p_{W2} follows similarly. When there is a mixture of two production lots represented in the sample, the sample can be written as $\mathbf{X} = (X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2})'$ where $\mathbf{X}|\boldsymbol{\theta}, \boldsymbol{\eta} \sim \text{MVN}(\mu\mathbf{1}_n + \mathbf{H}\boldsymbol{\theta}, \sigma_\epsilon^2\mathbf{I}_n)$ with $\boldsymbol{\theta} = (a_1, a_2)'$ and $\boldsymbol{\eta} = (n_1, n_2)'$. It follows that $\bar{X}_n|\boldsymbol{\theta}, \boldsymbol{\eta} \sim N(\mu + \bar{a}, \sigma_\epsilon^2/n)$ and is conditionally independent of $\frac{(n-1)S^2}{\sigma_\epsilon^2}|\boldsymbol{\theta}, \boldsymbol{\eta} \sim \chi^2_{n-1}(\lambda)$, where

$$\bar{a} = \frac{n_1 a_1 + n_2 a_2}{n} \quad \text{and} \quad \lambda = \frac{n_1(a_1 - \bar{a})^2 + n_2(a_2 - \bar{a})^2}{\sigma_\epsilon^2}.$$

The desired probability is

$$p_{W_1} = P(X_{1j} \leq c|\bar{X}_n + bS \geq V, \boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{P(X_{1j} \leq c, \bar{X}_n + bS \geq V|\boldsymbol{\theta}, \boldsymbol{\eta})}{P(\bar{X}_n + bS \geq V|\boldsymbol{\theta}, \boldsymbol{\eta})}, \tag{C.1}$$

The denominator in (C.1) is given by the cdf of a doubly noncentral t distribution (see the discussion in the text and Bulgren and Amos 1968). The numerator can be found using an extension of the results in Appendix B. As above,

$$\begin{aligned} &P(X_{1j} \leq c, \bar{X}_n + bS \geq V|\boldsymbol{\theta}, \boldsymbol{\eta}) \\ &= \int_0^\infty \int_{V-bs}^\infty P\left(\frac{X_{1j} - \bar{X}_n}{S} \leq \frac{c - \bar{x}_n}{s}|\boldsymbol{\theta}, \boldsymbol{\eta}\right) \\ &\quad \times f_{\bar{X}_n|\boldsymbol{\theta},\boldsymbol{\eta}}(\bar{x}_n) f_{S|\boldsymbol{\theta},\boldsymbol{\eta}}(s)d\bar{x}_n ds. \end{aligned}$$

To solve this integral, it is necessary to find

$$\begin{aligned} &P\left(\frac{X_{1j} - \bar{X}_n}{S} \leq \frac{c - \bar{x}_n}{s}|\boldsymbol{\theta}, \boldsymbol{\eta}\right) \\ &= P\left(\frac{\sqrt{n}(X_{1j} - \bar{X}_n)}{(n-1)S} \leq \frac{\sqrt{n}}{n-1}\left(\frac{c - \bar{x}_n}{s}\right)|\boldsymbol{\theta}, \boldsymbol{\eta}\right) \\ &= P(T_1 \leq c_1|\boldsymbol{\theta}, \boldsymbol{\eta}), \end{aligned} \tag{C.2}$$

where $c_1 = \frac{\sqrt{n}(c-\bar{x}_n)}{(n-1)s}$ and $T_1 = \frac{U_1}{\sqrt{U_1^2+V_1}}$. Random variables U_1 and V_1 are defined as follows. Let

$$U_1 = \frac{\sqrt{n}(X_{1j} - \bar{X}_n)}{\sqrt{n-1}\sigma_\epsilon} \quad \text{and} \quad \xi_1 = \frac{\sqrt{n}(a_1 - \bar{a})}{\sqrt{n-1}\sigma_\epsilon}$$

so that $U_1|\boldsymbol{\theta}, \boldsymbol{\eta} \sim N(\xi_1, 1)$. Also define

$$\begin{aligned}(n-2)S^2_{1(-j)} &= \sum_{k=1(\neq j)}^{n_1} (X_{1k} - \bar{X}_{1(-j)})^2 \\ &\quad + \sum_{k=1}^{n_2} (X_{2k} - \bar{X}_{1(-j)})^2 \text{ and } \bar{X}_{1(-j)} \\ &= \frac{1}{n-1}\left(\sum_{k=1(\neq j)}^{n_1} X_{1k} + \sum_{k=1}^{n_2} X_{2k}\right).\end{aligned}$$

Then

$$(n-1)S^2 = (n-2)S^2_{1(-j)} + \frac{n}{n-1}(X_{1j} - \bar{X}_n)^2$$

and

$$V_1 = \frac{(n-2)S^2_{1(-j)}}{\sigma^2_\epsilon}|\boldsymbol{\theta}, \boldsymbol{\eta} \sim \chi^2_{n-2}(\lambda_{(-j)})$$

where

$$\lambda_{(-j)} = \frac{(n_1-1)(a_1 - \bar{a}_{(-j)})^2 + n_2(a_2 - \bar{a}_{(-j)})^2}{\sigma^2_\epsilon} \quad \text{and}$$

$$\bar{a}_{(-j)} = \frac{(n_1-1)a_1 + n_2 a_2}{n-1}.$$

ACKNOWLEDGMENTS

The authors thank the AE and two anonymous referees for many useful comments and suggestions. We also thank the Procter and Gamble Company for providing a motivating example and data. This work was partially supported by Discovery grants from the Natural Sciences and Engineering Research Council of Canada and NSF DMS-0806106.

[Received March 2009. Accepted March 2012.]

REFERENCES

Altham, P. M. E. (1978), "Two Generalizations of the Binomial Distribution," *Applied Statistics*, 22, 162–167. [246]

Bahadur, R. R. (1960), "Some Approximations to the Binomial Distribution Function," *The Annals of Mathematical Statistics*, 31, 43–54. [246]

Bulgren, W. G., and Amos, D. E. (1968), "A Note on Representations of the Doubly Non-central t Distribution," *Journal of the American Statistical Association*, 63, 1013–1019. [254]

Casella, G., and Berger, R. L. (2001), *Statistical Inference*, Belmont, CA: Duxbury Press. [254]

Constantine, A. G., Field, J. B. F., and Robinson, N. I. (2000), "Probabilities of Failure in Mixed Acceptance Sampling Schemes," *Australian and New Zealand Journal of Statistics*, 42, 225–233. [244,251]

Elder, R. S., and Muse, D. H. (1982), "An Approximate Method of Evaluating Mixed Sampling Plans," *Technometrics*, 24, 207–211. [243,244,246,247,253]

Grubbs, F. E. (1950), "Sample Criteria for Testing Outlying Observations," *Annals of Mathematical Statistics*, 21, 27–58. [243]

Johnson, N. L., and Kotz, S. (1970), *Continuous Univariate Distributions-1*, New York, NY: Houghton Mifflin. [246,254]

Nair, K. R. (1948), "The Distribution of the Extreme Deviate From the Sample Mean and its Studentized Form," *Biometrika*, 35, 118–144. [243]

NIST Handbook 133. (2005), *Checking the Net Contents of Packaged Goods*, Department of Commerce, United States of America. [244]

Robert, C. P., and Casella, G. (2005), *Monte Carlo Statistical Methods* (2nd ed.), New York, NY: Springer. [245]

Schilling, E. G., and Dodge, H. F. (1969), "Procedures and Tables for Evaluating Dependent Mixed Acceptance Sampling Plans," *Technometrics*, 11, 341–372. [243]

Soon, S. Y .T. (1996), "Binomial Approximation for Dependent Indicators," *Statistica Sinica*, 6, 703–714. [246,248]

Vangel, M. G. (2002), "Lot Acceptance and Compliance Testing Using the Sample Mean and an Extremum," *Technometrics*, 44, 242–249. [243,244]