Acceptance Sampling Using Judgmental and Randomly Selected Samples

LH Sego   JE Wilson
SA Shulman   BA Pulsipher
KK Anderson   WK Sieber

September 2010

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*

DISCLAIMER

This document was printed on recycled paper.
(9/2003)

PNNL-19315

Acceptance Sampling Using Judgmental and Randomly Selected Samples

LH Sego[1]   JE Wilson[1]
SA Shulman[2]   BA Pulsipher[1]
KK Anderson[1]   WK Sieber[2]

1. Pacific Northwest National Laboratory, Richland, WA
2. National Institute of Occupational Safety and Health, Cincinnati, OH

September 2010

Pacific Northwest National Laboratory
Richland, Washington 9952

Acceptance Sampling Using Judgmental and Randomly Selected Samples

Landon H. Sego[1], Stanley A. Shulman[2], Kevin K. Anderson[1]

John E. Wilson[1], Brent A. Pulsipher[1], W. Karl Sieber[2]

September 2010

1: Statistics and Sensor Analytics Group, Pacific Northwest National Laboratory, Richland, WA
2: National Institute for Occupational Safety and Health, Cincinnati, OH

Abstract

We present a Bayesian model for acceptance sampling where the population consists of two groups, each with different levels of risk of containing unacceptable items. Expert opinion, or judgment, may be required to distinguish between the high and low-risk groups. Hence, high-risk items are likely to be identified (and sampled) using expert judgment, while the remaining low-risk items are sampled randomly. We focus on the situation where all observed samples must be acceptable, where the objective of the statistical inference is to quantify the probability that a large percentage of the unsampled items in the population are also acceptable. We demonstrate that traditional (frequentist) acceptance sampling and simpler Bayesian formulations of the problem are essentially special cases of the proposed model. We explore the properties of the model in detail, and discuss the conditions necessary to ensure that the required sample size is a non-decreasing function of the population size. The methodology is applicable to a variety of acceptance sampling problems, including environmental sampling where the objective is to demonstrate the safety of reoccupying a remediated facility that has been contaminated with a lethal agent.

KEY WORDS: compliance sampling, upper tolerance limit, judgmental sampling, environmental sampling

Contents

1 Introduction

Consider a population of independent items, each of which, upon sampling, can be identified as either possessing or not possessing a particular attribute of concern. Based on this attribute, each item will be classified as being 'acceptable' or 'unacceptable.' Suppose also that the population consists of two groups, or sub-populations, of units, wherein the units from the first group have a higher propensity to reflect the attribute of concern than units from the second group. While the two groups have different levels of risk, we desire that information obtained from sampling one of the groups will influence our belief about the other, and vice versa. In other words, sample results from either group influence our belief about the entire population. We presume that the size of the high-risk group is considerably smaller than the low-risk group, and that all of the items in the high-risk group are correctly identified and sampled. A random sample is then drawn from the low-risk group to augment the results of the census of the high-risk group. In practice, it is likely that the professional judgment of a subject matter expert would be used to distinguish between the groups. For this reason, we refer to samples from the high-risk group as 'judgmental,' or 'targeted,' samples, and samples from the low-risk group as 'random' samples. After taking both judgmental and random samples, if none of the inspected items are unacceptable, we wish to calculate the probability that a high percentage of the remaining, unsampled items of the population are also acceptable. For example, we may desire a probability of 0.95 that 99% of the total population is acceptable, given that all the samples were acceptable. We present a Bayesian formulation of this problem, wherein required sample sizes may be calculated to satisfy a desired level of posterior confidence in the acceptability of a high percentage of the population items.

Wright (1992), Grieve (1994), and Axelrod (2005) present a Bayesian model which addresses a very similar problem to the one considered here, except the population is homogeneous, i.e. the risk is the same for all members of the population. Traditional frequentist approaches to acceptance sampling (Schilling and Neubauer, 2009; Bowen and Bennett, 1988) also do not consider sub-populations with different levels of risk. Graves et al. (2007) addresses a somewhat related problem in estimating a proportion using biased convenience and unbiased random samples. However, our objective differs in that we esteem the 'bias' incurred from judgmentally sampling the high-risk group (and finding those observations to be acceptable)

to be advantageous in demonstrating that a high fraction of the population is acceptable. After all, if the goal is to demonstrate that most items are acceptable, it only makes sense, as a starting point, to inspect the items that are most likely to be unacceptable.

There are a number of contexts in which the proposed methodology could be employed: acceptance sampling (Schilling and Neubauer, 2009), discovery or audit sampling (Guy, Carmichael, and Whittington, 1998), or even the sampling of network packets which may pose cyber-security threats (Valgenti, Shaikot, and Kim, 2010). The approach may also be applied in the context of indoor environmental sampling (Emanuel, Roos, and Niyogi, 2008). While this methodology ignores spatial correlation among the population members, the simplifying assumption of independence turns out to be conservative, a point we discuss in more detail in Section 4.2.

One of the principal motivations for this work stemmed from evaluation by the U.S. Government Accountability Office (GAO) of the sampling efforts that were conducted in response to the anthrax incidents in the United States during 2001 (GAO, 2005). The GAO reported that sampling efforts based on judgmental samples alone were insufficient to provide a quantified level of confidence when no contamination was detected. They addressed the need to use "probabilistic" sampling, that is, sampling designs based on statistical models that quantify the confidence associated with sample results. The work presented herein was developed to address that concern, while creating a framework that would permit investigators to sample high-risk locations judgmentally and still quantify the confidence in negative results with a statistical model. Furthermore, the model we present permits investigators to leverage their prior belief and their judgment regarding the locations where contamination is most likely to be present in order to reduce the total number of samples required to achieve the desired level of confidence. Naturally, quantifying the level of prior belief and the exercise of judgment can be subjective and prone to error. However, in many instances, carefully (and conservatively) accounting for these quantities can improve the feasibility of sampling plans by requiring fewer samples than frequentist approaches.

2   Methods

Suppose the population of interest has N items, consisting of n_h high-risk items and n_ℓ low-risk items ($n_h + n_\ell = N$), where we typically expect n_h to be considerably smaller than n_ℓ. We presume that members of the two sub-populations can be correctly distinguished prior to sampling. In the context of the indoor sampling problem, this infers that investigators can correctly identify the high-risk locations. (The consequences of less than perfect distinction between the high and low-risk sub-populations is a topic of current research). Given that all the samples are acceptable, we desire to calculate the probability C that a desired fraction of the population, λ, is acceptable (where $0 < \lambda \leq 1$). Under the presumption of perfect distinction between the two groups, the number of required samples will be minimized if all the n_h high-risk items are sampled, along with a random sample of size n from the n_ℓ low-risk items.

Let each of the N items be modeled as independent Bernoulli variates with probability θ_h or θ_ℓ of being unacceptable, depending on whether the item belongs to the high-risk or low-risk sub-population, respectively. Naturally, we expect θ_h, on average, to be greater than θ_ℓ. The prior distribution(s) for θ_h and θ_ℓ will involve the Beta distribution, which we discuss in detail below. Let the binomial variates $X_h|\theta_h$ and $X_\ell|\theta_\ell$ denote the numbers of unacceptable high-risk and low-risk items observed in the sample, respectively. Specifically, we write the distributions of these random variables as $X_h|\theta_h \sim \mathrm{Bin}(n_h, \theta_h)$ and $X_\ell|\theta_\ell \sim \mathrm{Bin}(n, \theta_\ell)$. The Bayesian confidence in the acceptability of a large fraction of the population, C, is calculated using the posterior predictive distribution of the number of unsampled, unacceptable low-risk items, denoted by Y_ℓ. The confidence function

$$C = P(Y_\ell \leq (1-\lambda)N \mid X_h = 0, X_\ell = 0) \tag{1}$$

may be solved for the number of random samples, n, required to achieve a desired level of confidence, C'.

The Beta-binomial prior, hypergeometric likelihood, and Beta-binomial posterior are the natural choice for a Bayesian model that describes sampling from a finite, dichotomous population without replacement. This approach was used by Wright (1992), Grieve (1994), and Axelrod (2005). We chose instead a model which uses Beta priors, binomial likelihoods, Beta posteriors, and the posterior predictive distribution, because this approach facilitates accounting for the two levels of risk in the same population. It is interesting to

note that for a homogeneous population with only one level of risk, our approach yields the same posterior-predictive distribution of the number of unobserved, unacceptable items, as that obtained by Wright (1992) and Grieve (1994) (See Appendix A for details). Because this posterior-predictive distribution is the foundation for inference and sample size determination, our approach and the Wright/Grieve approach are virtually identical when the population is homogeneous.

Let $\theta \sim \text{Beta}(\alpha, \beta)$ indicate that the random variable θ follows a Beta distribution with density $p(\theta) = [B(\alpha, \beta)]^{-1} \theta^{\alpha-1}(1 - \theta)^{\beta-1}$, where $B(a, b) = \int_0^1 t^{a-1}(1 - t)^{b-1}dt$ denotes the complete beta function. The density has support on $(0, 1)$ and $E\theta = \alpha/(\alpha + \beta)$. While the two shape parameters α and β of the Beta family of distributions provide a variety of density shapes, we are typically interested in convex, exponential-type shapes that place most of the weight near zero, reflecting a prior belief that a large fraction of the population will be acceptable. These shapes can be achieved with $\alpha < 1$ and $\beta = 1$ or by setting $\alpha = 1$ and $\beta > 1$. We prefer the latter of the two, because $\alpha < 1$ places so much weight near zero that the prior can easily overwhelm the likelihood. In addition, the fact that $\lim_{\theta \to 0^+} p(\theta) = \infty$ when $\alpha < 1$ can contribute to numerical instability in calculating C and n. For the $\text{Beta}(\alpha = 1, \beta \geq 1)$ family, specifying the a priori expected value of θ is a natural way to initialize the prior and choose a starting value for β.

A non-informative prior may be obtained by setting $\alpha = 1$ and $\beta = 1$, resulting in a uniform distribution for θ. While some investigators may prefer a non-informative prior, Axelrod (2005) makes a compelling argument for employing at least some prior belief in this type of acceptance sampling. Specifically, the uniform prior implies the assumption that $P(\theta_h < 0.01) = P(\theta_h > 0.99)$. If we really believed that a large value of θ_h was just as likely as a small value, why would we be trying to demonstrate that a large fraction of the population is acceptable, especially when that inference will be conditional on observing no unacceptable items in the sample? Given the context and objectives of the sampling, a uniform prior is likely to be unduly conservative.

Specifying either the joint or marginal prior distribution(s) of θ_h and θ_ℓ must be done carefully to ensure that information obtained from sampling the high-risk group will influence our belief about the low-risk group, and vice versa. For example, if all 20 high-risk items were found to be acceptable, we desire the updated marginal distributions of both θ_h and θ_ℓ to have more weight closer to zero than they did prior to

sampling. The same would be true if, say, 30 samples were selected from the low-risk sub-population and found to be acceptable. Hence, some type of linkage between θ_h and θ_ℓ is required. A joint prior for θ_h and θ_ℓ with the linkage induced by the positive correlation between θ_h and θ_ℓ would be desirable. Specifically, we would like this bivariate beta distribution to have a high correlation with marginal distributions of Beta(1, β_h) and Beta(1, β_ℓ).

However, the various bivariate beta distributions we have considered result in analytic or computational challenges that make it difficult to 1) calculate the required sample size, n, without Bayesian simulation techniques, and/or 2) specify the correlation between θ_h and θ_ℓ. For example, Gupta and Wong (1985) presented a bivariate beta distribution based on Morgenstern's (1956) system, but these variables can, at most, have a positive correlation coefficient of 1/3 (Schucany, Parr, and Boyer, 1978), which is too low for our purposes. Lee (1996) describes another type of bivariate beta distribution based on the Sarmanov family (Sarmanov, 1966). However, we observed that the correlation (given by Danaher and Hardie (2005)), of the joint density given by Lee (1996) that has Beta(1, β_h) and Beta(1, β_ℓ) marginals is typically small (< 0.50). Shubina and Lee (2004) discussed the maximum attainable correlation for Sarmanov families, and this approach, subject to further investigation, might lead to the properties we desire for a bivariate beta density.

Olkin and Liu (2003) proposed a bivariate beta distribution constructed from a transformation of three gamma variates. While this approach can give rise to high correlation between θ_h and θ_ℓ, it does not provide a mechanism to specify the correlation, and it leads to a posterior distribution that does not appear to be readily integrable, thus requiring Bayesian simulation techniques, i.e. Markov Chain Monte Carlo (Gelman, Carlin, Stern, and Rubin, 2004) in order to estimate posterior probabilities. This, in turn, would make calculation of the sample size complex and unwieldy for practitioners. Last of all, Magnussen (2002) presented an algorithm to draw from a beta bivariate distribution with known marginal distributions and a specified correlation–but once again, simulation would complicate the objective of determining sample sizes for practitioners. Despite these challenges, the use of a bivariate beta prior is an appealing and natural choice for a Bayesian model such as this, and would provide an interesting subject for future research.

2.1   An initial model for combining judgmental and random samples

We consider instead an explicit link between the high and low-risk sub-populations by defining $\theta_\ell = \theta_h/\rho$, for $\rho \geq 1$. Thus, θ_h and θ_ℓ have densities

$$p_1(\theta_h) = \beta(1 - \theta_h)^{\beta-1} I_{\{0,\ 1\}}(\theta_h)$$

$$p_1(\theta_\ell) = \rho\beta(1 - \rho\theta_\ell)^{\beta-1} I_{\{0,\ 1/\rho\}}(\theta_\ell)$$

where the subscript '1' designates the initial model. The value of β can be obtained by using historical data or expert opinion to estimate the a priori expected value of θ_h and solve for β as follows:

$$\beta := (1 - E\theta_h)/E\theta_h \tag{2}$$

We can interpret $E\theta_h$ as the rate of unacceptable items we expect in the high-risk sub-population. We may interpret ρ as a factor indicating how much more likely a high-risk item is unacceptable than a low-risk item.

We note parenthetically that it may be desirable to model β and ρ as random variables with their own distributions. For example, β and ρ could each be modeled as independent Gamma variates with support on $[1, \infty)$. This would make it possible to account for the uncertainty in specifying a priori values for β and ρ. However, we elected not to do this primarily to avoid added complexity for the practitioner. In addition, adding distributions for β and ρ would induce two additional integrations in the confidence functions (5) and (14), as we would need to average over the measures of β and ρ.

Expressing the likelihoods only in terms of θ_h gives $X_h|\theta_h \sim \text{Bin}(n_h, \theta_h)$ and $X_\ell|\theta_h \sim \text{Bin}(n, \theta_h/\rho)$. We have by Bayes' rule and the independence of X_h and X_ℓ:

$$p_1(\theta_h \mid X_h = 0, X_\ell = 0) \propto P(X_h = 0 \mid \theta_h)P(X_\ell = 0 \mid \theta_h)p_1(\theta_h)$$

$$\propto (1 - \theta_h)^{n_h}(1 - \theta_h/\rho)^n(1 - \theta_h)^{\beta-1} \tag{3}$$

and normalizing the posterior density gives

$$p_1(\theta_h \mid X_h = 0, X_\ell = 0) = \frac{(1 - \theta_h)^{n_h+\beta-1}(1 - \theta_h/\rho)^n}{\int_0^1 (1 - \theta_h)^{n_h+\beta-1}(1 - \theta_h/\rho)^n d\theta_h} \tag{4}$$

Because $Y_\ell \sim \text{Bin}(N - n_h - n, \theta_h/\rho)$, we can then write the Bayesian confidence (1) as

$$C = \int_0^1 P(Y_\ell \leq (1-\lambda)N \mid \theta_h) \, p_1(\theta_h \mid X_h = 0, X_\ell = 0) \, d\theta_h$$

$$= \int_0^1 \sum_{y=0}^{\lfloor (1-\lambda)N \rfloor} \binom{N - n_h - n}{y} \left(\frac{\theta_h}{\rho}\right)^y \left(1 - \frac{\theta_h}{\rho}\right)^{N - n_h - n - y} p_1(\theta_h \mid X_h = 0, X_\ell = 0) \, d\theta_h \qquad (5)$$

Sample sizes can then be determined by numerically searching (5) for the smallest value of n that achieves a desired level of confidence, C'. This initial model was first discussed by Sego et al. (2007).

2.2 Drawbacks of the initial model

While the initial model provides a good starting point for the problem at hand, the approach has four immediate drawbacks. First, the integrals in (4) and (5) are difficult to accurately calculate with standard numerical integration routines because their integrands are mostly constant (near 0) over the unit interval. This is especially true for large values of β. Second, defining $\theta_h = \rho\theta_\ell$ results in truncating the support of the prior distribution of θ_ℓ to $(0, 1/\rho)$, which may not be desirable. Third, because the distribution function of Y_ℓ is discrete, the required sample size can oscillate strongly for increasing values of N, as illustrated in Figure 1.

Fourth, in situations where the prior evidence is strong (i.e. which is reflected in large values of β and/or ρ) relative to the desired fraction of acceptability, λ, the confidence function (5) is not necessarily a decreasing function of N. The practical implication is that more samples may be required for a smaller population than for a larger population to achieve the same level of confidence—an undesirable and non-intuitive result. This phenomenon results from an interesting characteristic of the binomial distribution function. To illustrate, let $g(N) = P(W \leq (1-\lambda)N)$ for $W \sim \text{Binomial}(N, \varphi)$. Then, for values of φ that are small relative to $1 - \lambda$, we observe $g(N)$ is not a decreasing function of N. However, larger values of φ result in a decreasing function of g. This phenomenon is illustrated in Figure 2. One of the factors that induces this phenomenon is the dependence of the distribution of W on N. Thus, one way to avoid it would be to make inference to a specific number of unacceptable items, rather than a fraction of the population size. For example, we could desire a high probability that no more than, say, 10 items in the population be unacceptable. Nonetheless, in many applications, it is often desirable to make inference to a quantile of the population (in this case,
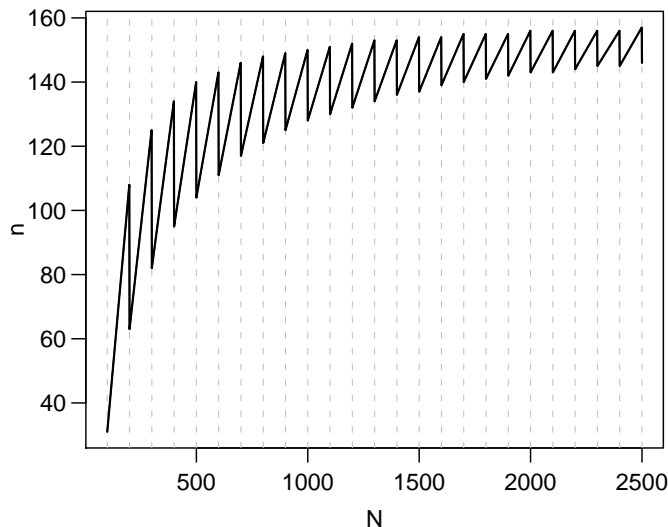
Figure 1: Number of random samples, n, required by the initial model, versus the population size, N, when $C' = 0.95$, $n_h = 20$, $E\theta_h = 0.02$ ($\beta = 49$), $\rho = 2$, and $\lambda = 0.99$. The random sample size was calculated using (5). Vertical, broken lines appear at values of N evenly divisible by 100. Note how the required sample size drops at these vertical lines, because $\lambda = 0.99$ results in an additional unacceptable item being tolerated at each increment of 100 items in the population.

λN), especially because it provides a standard level of confidence that would be applicable when sampling populations of different sizes. We note that this phenomenon affects the models we propose here, as well as those discussed by Wright (1992), Grieve (1994), and Axelrod (2005) when inference is made to λN (as opposed to a specified number of unacceptable items). In Section 3.2, we discuss this phenomenon in greater detail, along with an approach to avoid it.

2.3   A refined model

We therefore propose an alternative to the initial model discussed in Section 2.1 which addresses these four concerns. This refined model is also a reasonable approximation to the initial model and provides similar random sample sizes, which we discuss in Section 3.3. Hereafter, for simplicity, we refer to this refined model as the Combined Judgmental and Random (CJR) model. We begin by specifying independent prior distributions $\theta_h \sim \text{Beta}(1, \beta)$ and $\theta_\ell \sim \text{Beta}(1, \phi(\beta))$, with densities $p_2(\theta_h)$ and $p_2(\theta_\ell)$, where the subscript '2' designates the refined model and

$$\phi(\beta) = \rho(\beta + 1) - 1 \tag{6}$$

10
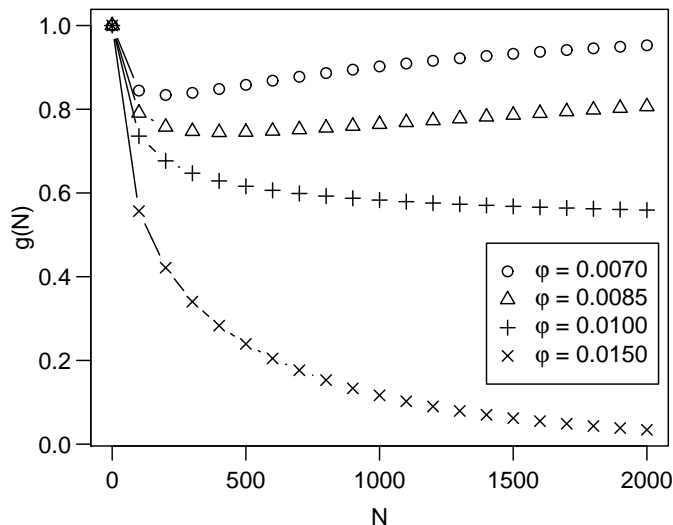
Figure 2: Plot of $g(N) = P(W \leq (1 - \lambda)N)$ versus N where $\lambda = 0.99$ and $W \sim \text{Binomial}(N, \varphi)$. Note how the smaller values of φ can cause $g(N)$ to increase. Points are plotted at values where $(1 - 0.99)N$ is integer-valued (increments of 100) to avoid oscillation that results from the discrete distribution function of W.

for $\rho \geq 1$. Note that (6) is the solution to

$$E\theta_h = \rho E\theta_\ell \tag{7}$$

a relationship we assume holds both before and after sample collection. As it is used in (7), we may interpret ρ as a factor indicating how much more likely, *on average*, a high-risk item is unacceptable than a low-risk item. We refer to ϕ as the linkage function which expresses the relationship between the high and low-risk items. Note that θ_h and θ_l share a common parameter, β. This approach addresses the first concern discussed in Section 2.2 by yielding conjugate Beta posterior distributions which result in a posterior predictive distribution function for Y_ℓ that is easily calculated using gamma functions. The second concern is also addressed because the prior on θ_ℓ is Beta with support on the entire unit interval. While it is not necessary to collect the data in stages—with high-risk judgmental samples first, followed by low-risk random samples—it is illustrative to consider two stages of sampling to motivate the approach.

Observing $X_h = 0$ in the first stage and applying Bayes' theorem, we write the resulting posterior density

of θ_h without the constant of integration, where f denotes the binomial likelihood:

$$p_2(\theta_h \mid X_h = 0) \propto p_2(\theta_h) f(X_h = 0 \mid \theta_h)$$

$$\propto (1 - \theta_h)^{\beta-1}(1 - \theta_h)^{n_h} \tag{8}$$

$$= (1 - \theta_h)^{n_h + \beta - 1}$$

which gives the well-known (Gelman et al., 2004) conjugate posterior of $\theta_h|(X_h = 0) \sim \text{Beta}(1, n_h + \beta)$. The information obtained by observing $X_h = 0$ is reflected in the updated value of the shape parameter: $n_h + \beta$. By inserting this updated value into the prior on θ_l, we allow the outcome of $X_h = 0$ to increase our prior belief in the acceptability of the low-risk samples. Therefore, substituting the updated shape parameter into the prior for θ_l via (6) gives

$$\theta_\ell \sim \text{Beta}(1, \phi(n_h + \beta)) \tag{9}$$

We abuse the conditioning notation and denote this updated prior density of θ_ℓ as $p_2(\theta_\ell \mid X_h = 0)$ to reflect the information obtained by the n_h acceptable judgmental items. Alternatively, we could view the first stage sampling of the high-risk items as an approach to elicit the prior distribution for θ_ℓ before sampling the low-risk group.

Upon observing $X_\ell = 0$ in the second stage, we arrive at the posterior distribution for θ_ℓ by combining the binomial likelihood of n acceptable low-risk items with the updated prior, $p_2(\theta_\ell \mid X_h = 0)$, in a manner similar to (8). This gives $\theta_\ell|(X_h = 0, X_\ell = 0) \sim \text{Beta}(1, \ \beta')$ where

$$\beta' := n + \phi(n_h + \beta) = n + \rho(n_h + \beta + 1) - 1 \tag{10}$$

Note that β' has a natural interpretation: the n_h acceptable, judgmentally sampled items are weighted ρ times more than the n acceptable, randomly sampled items.

Having observed $X_h = 0$ and $X_\ell = 0$, the number of low-risk, unsampled items which are unacceptable, Y_ℓ, is Binomial with $N - n_h - n$ trials and probability $\theta_\ell|(X_h = 0, X_\ell = 0)$. Letting $F_{Y_\ell}(y \mid \theta_\ell)$ denote the distribution function of $Y_\ell|\theta_\ell$, i.e.

$$F_{Y_\ell}(y \mid \theta_\ell) = P(Y_\ell \leq y \mid \theta_\ell) = \sum_{i=0}^{y} \binom{N - n_h - n}{i} (\theta_\ell)^i (1 - \theta_\ell)^{N - n_h - n - i} \tag{11}$$

we write the Bayesian confidence as

$$C = P(Y_\ell \leq (1 - \lambda)N \mid X_h = 0, X_\ell = 0)$$

$$= \int_0^1 F_{Y_\ell}((1 - \lambda)N \mid \theta_\ell) \, p_2(\theta_\ell \mid X_h = 0, X_\ell = 0) \, d\theta_\ell \qquad (12)$$

We address the third drawback discussed in Section 2.2 by using the relation of the incomplete beta function to the binomial expansion (Abramowitz and Stegun, 1972, eq. 6.6.4) to provide a continuous approximation of F_{Y_ℓ}:

$$F_{Y_\ell}(y \mid \theta_\ell) \approx 1 - [B(y + 1, N - n_h - n - y)]^{-1} \int_0^{\theta_\ell} t^y (1 - t)^{N - n_h - n - y - 1} \, dt := F_{Y_\ell}^\star(y \mid \theta_\ell) \qquad (13)$$

Note that

$$F_{Y_\ell}(y \mid \theta_\ell) = F_{Y_\ell}^\star(y \mid \theta_\ell) \quad \text{for} \quad y \in \mathbb{N}$$

$$F_{Y_\ell}(y \mid \theta_\ell) \leq F_{Y_\ell}^\star(y \mid \theta_\ell) \quad \text{for} \quad y \in \mathbb{R}$$

Informally, this approximation of F_{Y_ℓ} results in not requiring the number of unacceptable items, $(1 - \lambda)N$, to be a whole number. Replacing F_{Y_ℓ} with $F_{Y_\ell}^\star$ and applying some calculus, (12) can be written as

$$C \approx \int_0^1 F_{Y_\ell}^\star((1 - \lambda)N \mid \theta_\ell) \, p_2(\theta_\ell \mid X_h = 0, X_\ell = 0) \, d\theta_\ell$$

$$= 1 - (\Gamma(A_1)\Gamma(A_2)) / (\Gamma(A_3)\Gamma(A_4)) \qquad (14)$$

where

$$A_1 = N - n_h - n + 1$$

$$A_2 = \lambda N - n_h + \rho(n_h + \beta + 1) - 1$$

$$A_3 = \lambda N - n_h - n \qquad (15)$$

$$A_4 = N - n_h + \rho(n_h + \beta + 1)$$

A detailed derivation of (14) is provided in Appendix B. In what follows, it will be useful to refer to the approximation of the confidence function in (14) as a function of a single variable with the remaining variables held constant. Specifically, let

$$h(\cdot) = 1 - \exp\left(\log \Gamma(A_1) + \log \Gamma(A_2) - \log \Gamma(A_3) - \log \Gamma(A_4)\right) \qquad (16)$$

generically represent this function, where, for example, $h(N)$ denotes the confidence as a function of N, $h(n)$ denotes the confidence as a function of n, etc. Note that (14) and (16) apply only for sensible values of the variables, which we refer to as the parameter space

$$\mathcal{P} = \{N, \lambda, \rho, \beta, n_h, n \; : \; N \geq 1, \; 0 < \lambda \leq 1, \; \rho \geq 1, \; \beta \geq 1, \; \text{and } 0 \leq n_h + n \leq \lambda N\} \tag{17}$$

As discussed by Axelrod (2005), the Bayesian confidence, C, for an infinite population can be calculated directly from the posterior of θ_ℓ itself, without using the posterior predictive distribution of Y_ℓ. Specifically, because

$$\theta_\ell | (X_h = 0, X_\ell = 0) \sim \text{Beta}(1, \beta') \tag{18}$$

the confidence for an infinite population is given by

$$P(\theta_\ell \leq 1 - \lambda \mid X_h = 0, X_\ell = 0) = \int_0^{1-\lambda} \beta'(1 - \theta_\ell)^{\beta'-1} \; d\theta_\ell = 1 - \lambda^{\beta'} \tag{19}$$

It is straightforward to show that an application of Stirling's approximation (Abramowitz and Stegun, 1972, eq. 6.1.39) to $h(N)$ and some algebra corroborate the result from (19):

$$h(N = \infty) := \lim_{N \to \infty} h(N) = 1 - \lambda^{\beta'} \tag{20}$$

2.4   Sample size determination

To determine the required number of random samples for the CJR model, let C' denote the desired confidence and let $h(n)$ denote the Bayesian confidence in (16) expressed as a function of n. The required random sample is given by the smallest integer value of n which satisfies $h(n) \geq C'$. When solving for n, note that if $h(0) \geq C'$, no random samples are required, as the prior belief and the outcome of the n_h acceptable judgmental samples already provide sufficient probability that $(1 - \lambda) \times 100\%$ of the population is acceptable. Otherwise, because $h(n)$ is continuous, it is straightforward to solve for n numerically and then round up the result to the nearest integer. More formally,

$$n = \begin{cases} 0 & \text{if } h(0) \geq C' \\ \lceil h^{-1}(C') \rceil & o.w. \end{cases} \tag{21}$$

where $h^{-1}(C')$ is the solution to $h(n) = C'$. When $\lambda = 1$, it is straightforward to derive a closed form solution for $h(n) = C'$:

$$n = \max\left(0, \; \lceil N - n_h - (1 - C')(N - n_h + \rho(n_h + \beta + 1) - 1)\rceil\right) \tag{22}$$

It is interesting to note that the value of n given by (22) agrees exactly with Grieve's solution (1994) if the population has homogeneous risk, i.e., $n_h = 0$ and $\rho = 1$. Details are provided in Appendix C. A closed form solution for the infinite population case is easily obtained by solving (19) for n:

$$n = \max\left(0, \; \left\lceil \frac{\log(1 - C')}{\log \lambda} - \rho(n_h + \beta + 1) + 1)\right\rceil\right), \quad 0 < \lambda < 1 \tag{23}$$

When $\lambda = 1$, we have found that (22) works nicely even for very large N.

3 Analysis

3.1 Properties of the refined model

Analysis of the confidence function given by (16) demonstrates the refined CJR model behaves as we would expect. For example, $h(n_h)$ should be increasing, which we confirm with the partial derivative:

$$\frac{\partial h}{\partial n_h} = (h - 1)\left(-\Psi(A_1) + (\rho - 1)\Psi(A_2) + \Psi(A_3) - (\rho - 1)\Psi(A_4)\right) \tag{24}$$

where $\Psi(x) = \frac{d}{dx}\log\Gamma(x)$ is the digamma function. Because $\log\Gamma(x)$ is convex, $\Psi(x)$ is increasing, which implies $\Psi(A_3) - \Psi(A_1) < 0$ and $(\rho - 1)\left(\Psi(A_2) - \Psi(A_4)\right) \leq 0$. This, combined with the fact that $0 \leq h \leq 1$, implies that $\frac{\partial h}{\partial n_h} \geq 0$ over $\mathcal{P}$. In like manner, it is straightforward to show that $\frac{\partial h}{\partial n} \geq 0$, $\frac{\partial h}{\partial \rho} \geq 0$, $\frac{\partial h}{\partial \beta} \geq 0$, and $\frac{\partial h}{\partial \lambda} \leq 0$, also over $\mathcal{P}$. These results are consistent with how we would expect the confidence function to behave. However, the behavior of $h(N)$ is not consistently increasing or decreasing, due to the characteristics of the binomial distribution function discussed in Section 2.2 and illustrated in Figures 2 and 3. Note that the partial derivative of h with respect to N is

$$\frac{\partial h}{\partial N} = (h - 1)\left(\Psi(A_1) + \lambda\Psi(A_2) - \lambda\Psi(A_3) - \Psi(A_4)\right) \tag{25}$$

15

For $\lambda = 1$, we can apply the recurrence relation, $\Psi(x+1) = \Psi(x) + 1/x$, (Abramowitz and Stegun, 1972, eq. 6.3.5) to show

$$\frac{\partial h}{\partial N} = (h-1)\left(\Psi(A_3) + \frac{1}{A_3} - \Psi(A_3) + \Psi(A_2) - \Psi(A_2) - \frac{1}{A_2}\right)$$

$$= (h-1)\left(\frac{1}{N - n_h - n} - \frac{1}{N + n_h(\rho - 1) + \rho(\beta + 1) - 1}\right) \leq 0 \qquad (26)$$

over the subset of $\mathcal{P}$ induced by $\lambda = 1$. This finding is intuitive: given the same prior belief (reflected by β), the same number of acceptable samples n_h and n, and the same ρ, we would expect, for a population of size 100, the confidence that 100% of the items being acceptable should be greater than for a population of one million items. However, for $\lambda < 1$, the behavior of $h(N)$ depends upon other variables in the parameter space, $\mathcal{P}$, and in particular, it is most sensitive to β.
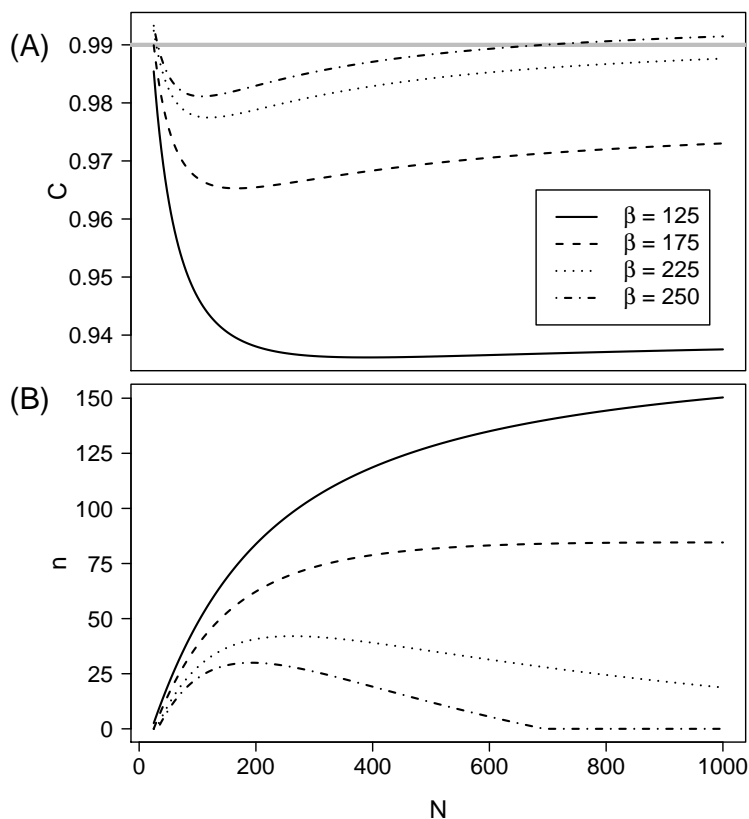
Figure 3: Plot (A): The confidence, $C = h(N)$, versus N for various levels of β, where $n = 0$, $\lambda = 0.99$, $n_h = 15$, and $\rho = 2$, calculated using (16). Plot (B): The required random sample size, n, versus N for the same values of β where $\lambda = 0.99$, $n_h = 15$, $\rho = 2$, and $C' = 0.99$. The random sample size was given by (21). As a reference for comparison, the target confidence level, $C' = 0.99$, is indicated by the horizontal gray line in plot (A).

16

Figure 3A demonstrates the influence of β on $h(N)$ when no random samples are taken ($n = 0$). Figure 3B shows how the required sample size, n, varies as a function of N, with a desired confidence of $C' = 0.99$, for the same values of λ, β, n_h, and ρ as in Figure 3A. Note how the curves in the two plots are somewhat reflective of each other. In particular, notice in Figure 3B how n can actually decrease as N increases for higher values of β, a non-intuitive and undesirable result. This undesirable behavior in $h(N)$ is likely to occur when the posterior expectation of the fraction of acceptable items is too large relative to λ. This typically occurs when the prior belief in the acceptability of the high-risk items is high (corresponding to large values of β), and/or when ρ is large. It is illustrative to consider that before any samples are taken, our prior belief in the fraction of the population that is acceptable, λ_p, is determined by β, n_h, ρ, and N. Specifically,

$$\lambda_p = E\left(1 - \frac{X_h + X_\ell + Y_\ell}{N}\right) = 1 - \frac{N + n_h(\rho - 1)}{N\rho(\beta + 1)} \tag{27}$$

Details of the derivation of (27) are provided in Appendix D. When λ_p is too large relative to λ, the prior belief, in a sense, already satisfies the intent to demonstrate (through sampling) that $\lambda \times 100\%$ of the population is acceptable. Hence, it is not surprising that large values of λ_p (relative to λ) are likely to result in undesirable behavior in $h(N)$.

3.2 Ensuring the Bayesian confidence is a non-increasing function of the population size

Because $h(n)$ is increasing, it follows that when $h(N, n = 0)$ is decreasing, the number of random samples given by (21) will be an increasing function of N. Consequently, choosing a 'viable' value of λ large enough to ensure that $h(N)$ is non-increasing will also ensure that, all things being equal, larger populations will require more samples than smaller populations. This viable value of the fraction of acceptable items, λ_v, could then be used to calculate the desired random sample size, n. The way in which λ influences the shape of the confidence function is illustrated in Figure 4. Note how increasing the value of λ improves the profile of $h(N)$ until, at $\lambda_v = 0.99$, it is a non-decreasing function of N. Furthermore, for $\lambda = 0.992$, (in fact, for all $\lambda \geq 0.99$), $h(N)$ is also non-increasing.
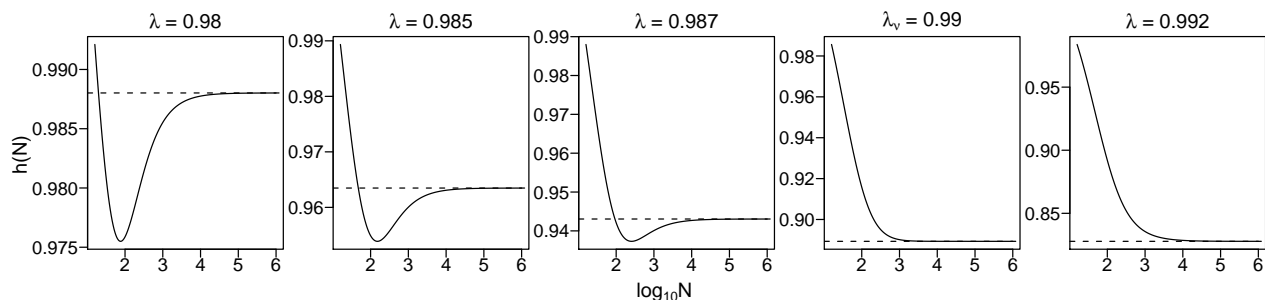
Figure 4: Plots of $h(N)$ for various values of λ where $n_h = 10$, $\rho = 2$, and $E\theta_h = 0.01$ ($\beta = 99$). The broken horizontal lines represent the asymptotic confidence, $h(N = \infty)$.

A method for identifying λ_v is provided by the following theorem, whose proof is given in Appendix E.

Theorem 1. *Assume N, n_h, and n are integer valued. Consider the Bayesian confidence h as a function of N and n, with all other parameters (λ, n_h, ρ, and β) held fixed. Let $\mathcal{V} = \{\lambda \; : \; \frac{\partial h}{\partial N} \leq 0$ for all $N \in \mathcal{P}$ and for all $n \in \mathcal{P}\}$ denote the set of 'viable' λ values. Furthermore, let $k = \frac{1}{2}\rho(n_h + \beta + 1) - n_h$, and let $\rho(n_h + \beta + 1)$ be integer valued. Then (i) for $k \leq 1$, $\mathcal{V} = (0, \; 1]$, and (ii) for $k > 1$, $\mathcal{V} = [1 - \frac{1}{k}, \; 1]$.*

Consequently, for integer valued $\rho(n_h + \beta + 1)$ and $k > 1$,

$$\lambda_v = 1 - \frac{2}{\rho(n_h + \beta + 1) - 2n_h} \tag{28}$$

Note that any $\lambda \geq \lambda_v$ will result in $h(N)$ being non-increasing for all $N \geq 1$ and all $n \in \mathcal{P}$. For $k \leq 1$, any value of $\lambda \in (0, 1]$ is viable. When $\rho(n_h + \beta + 1)$ is not integer valued, one could proceed conservatively by rounding $\rho(n_h + \beta + 1)$ up to the nearest integer. However, it is best to avoid overestimating λ_v because large values of λ have the greatest impact on the random sample size, as we discuss in Section 4.1. Simply using a non-integer value of $\rho(n_h + \beta + 1)$ in (28) also gives an accurate estimate of λ_v, which we denote $\lambda_v^{(1)}$. The larger the value of k, the better the approximation, because the exact incremental values of λ_v (induced by integer valued $\rho(n_h + \beta + 1)$) grow closer together as k increases.

Another way to estimate λ_v is to set $\frac{\partial h}{\partial N}\big|_{N=10^6}$ equal to 0 and numerically solve for λ. We denote this estimate as $\lambda_v^{(2)}$. Note that $\lambda_v^{(2)}$ is determined by numerically solving a continuous function (25) and does not depend on the assumption that $\rho(n_h + \beta + 1)$ be integer valued. It does depend, however, on the value of N. Evaluating $\frac{\partial h}{\partial N}$ at $N = 10^6$ ensures that the outcome will be close to the asymptotic result that gives
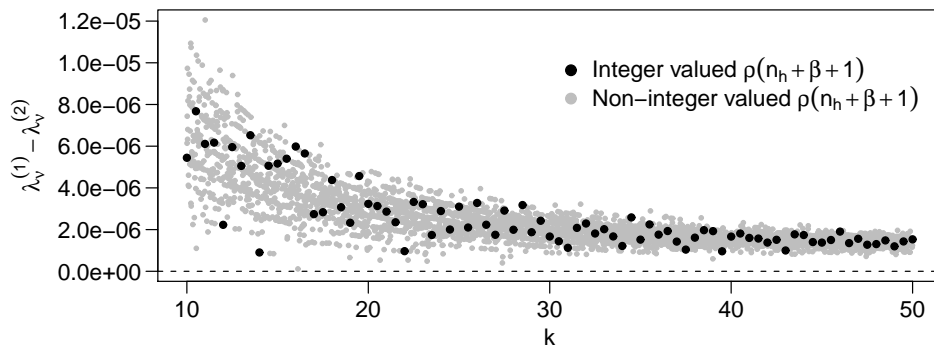
Figure 5: Plot of $\lambda_v^{(1)} - \lambda_v^{(2)}$ versus 4,001 values of k.

rise to the definition of k in Theorem 1 (see the proof in Appendix E for more details). Figure 5 illustrates how $\lambda_v^{(1)}$ and $\lambda_v^{(2)}$ differ for increasing values of k. The small amount of noise in the plot arises from solving (25) numerically. As k increases, an asymptotic, systematic difference between the two estimates of about 2×10^{-6} occurs, due to the finite value of $N = 10^6$ used to calculate $\lambda_v^{(2)}$. The principal conclusion we draw from Figure 5 is that the exact solutions of λ_v (shown as solid black circles), appear to be randomly scattered in the noise of the plot. This suggests that estimating λ_v with non-integer valued $\rho(n_h + \beta + 1)$ via (28) does not systematically differ from estimates obtained by solving (25), thereby corroborating the accuracy of $\lambda_v^{(1)}$.

To calculate the random samples size or calculate the Bayesian confidence, when $k = \rho(n_h + \beta + 1)/2 - n_h > 1$, and regardless of whether $\rho(n_h + \beta + 1)$ is integer valued, we recommend using a value of $\lambda \geq 1 - 2/\left(\rho(n_h + \beta + 1) - 2n_h\right)$ in order to avoid the fourth concern discussed in Section 2.2, i.e., the confidence not being a decreasing function of N, which can result in more samples being required for smaller populations than for larger ones.

3.3   Agreement of the initial and refined CJR models

To investigate how well the CJR model discussed in Section 2.3 approximates the initial model presented in Section 2.1, we calculated and compared the required sample sizes obtained from the two approaches for over 1.2 million cases. The cases were constructed using all feasible combinations of

- $\lambda = 0.95,\ 0.975,\ 0.99,\ 0.995,\ 0.999,\ 1,$

- $C' = 0.90,\ 0.95,\ 0.99$,

- eleven values of N ranging log-linearly from 1000 to 1.001 million,

- thirty different values of $E\theta_h$, increasing log-linearly on the interval $(0.0001, 0.5]$,

- twenty-nine values of ρ: $1,\ 1.5,\ \ldots,\ 9.5,\ 10,\ 11,\ \ldots,\ 19,\ 20$, and

- eight values of n_h: 0, 10, 25, 50, 100, 250, 500, and 750. When $\rho = 1$, there is no distinction between the high and low risk locations, and thus n_h was set to 0 for those cases.

The values of N and λ were chosen so that $(1 - \lambda)N$ would be integer valued, thus ensuring that $F_{Y_\ell} = F_{Y_\ell}^\star$. For these cases, we did not use λ_v to calculate the sample sizes because it would have made it considerably more difficult to ensure that $(1 - \lambda)N$ was integer valued. However, simply using the six values of λ listed above still allowed us to compare how the confidence functions given by (5) and (14) give rise to the required random sample sizes.

For each of these cases, the random sample size given by the initial model, n_{Init}, was obtained by identifying the value of n such that the Bayesian confidence given by (5) was at least as great as C'. For the approximation approach of Section 2.3, the random sample size, n_{CJR}, was obtained from (22) when $\lambda = 1$, from (23) when $N > 10^6$, and from (21) for the rest (and vast majority) of the cases that were considered. For roughly 70% of these cases, n_{Init} and n_{CJR} were both 0, indicating that prior belief represented by $E\theta_h$ and ρ combined with the outcomes from the n_h acceptable judgmental samples provided sufficient evidence to achieve the desired confidence level. The cumulative distribution of the absolute difference between the sample sizes, $\Delta_n = |n_{CJR} - n_{Init}|$, is shown in Figure 6. Note that 96.5% of the cases had random sample sizes that differed by no more than 5, and 98.6% of the cases had random sample sizes that differed by no more than 10.

Figure 7 illustrates how the various input parameters λ, C', $E\theta_h$, ρ, n_h, and N influence Δ_n. Increasing values of λ (except $\lambda = 1$) seem to be associated with increasing values of Δ_n. Likewise, increasing values of $E\theta_h$, increasing values of ρ, and increasing values of N also seem to be associated with increasing values of Δ_n. The most striking patterns occur for ρ and $E\theta_h$, the two parameters which appear to have the greatest
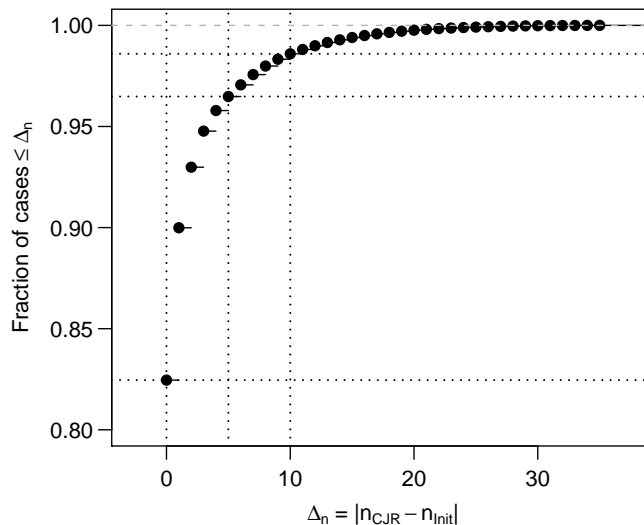
Figure 6: Cumulative distribution of the absolute difference between the random sample sizes required by the initial model discussed in Section 2.1 and the refined CJR model discussed in Section 2.3.

influence on the shape of the posterior distribution $\theta_\ell | (X_h = 0, X_\ell = 0)$.

For practical purposes, the refined model and the initial model behave very similarly in terms of the required random sample size, especially for smaller values of $E\theta_h$ and smaller values of ρ. However, the refined CJR model has desirable properties that eliminate the four drawbacks of the initial model which are discussed in Section 2.2.

3.4   Comparison to traditional acceptance sampling

Traditional acceptance sampling based on the hypergeometric or binomial distributions is discussed by (Schilling and Neubauer, 2009) and (Bowen and Bennett, 1988). This frequentist approach was designed for homogeneous populations and makes no prior assumptions regarding the fraction of unacceptable items. Using the hypergeometric distribution to model the number of unacceptable items in the population, we can identify the smallest sample size such that the probability of observing zero unacceptable items, given that at least $\lambda \times 100\%$ of the items are acceptable, is at least as large as a given confidence threshold, C_a. While certainly not equivalent, C_a is the conceptual analog of the Bayesian probability, C, given by (1). We present this methodology in a way that differs somewhat from the explanation given by Bowen and Bennett (1988), whose approach is slightly conservative (resulting in an achieved confidence that is higher than the desired
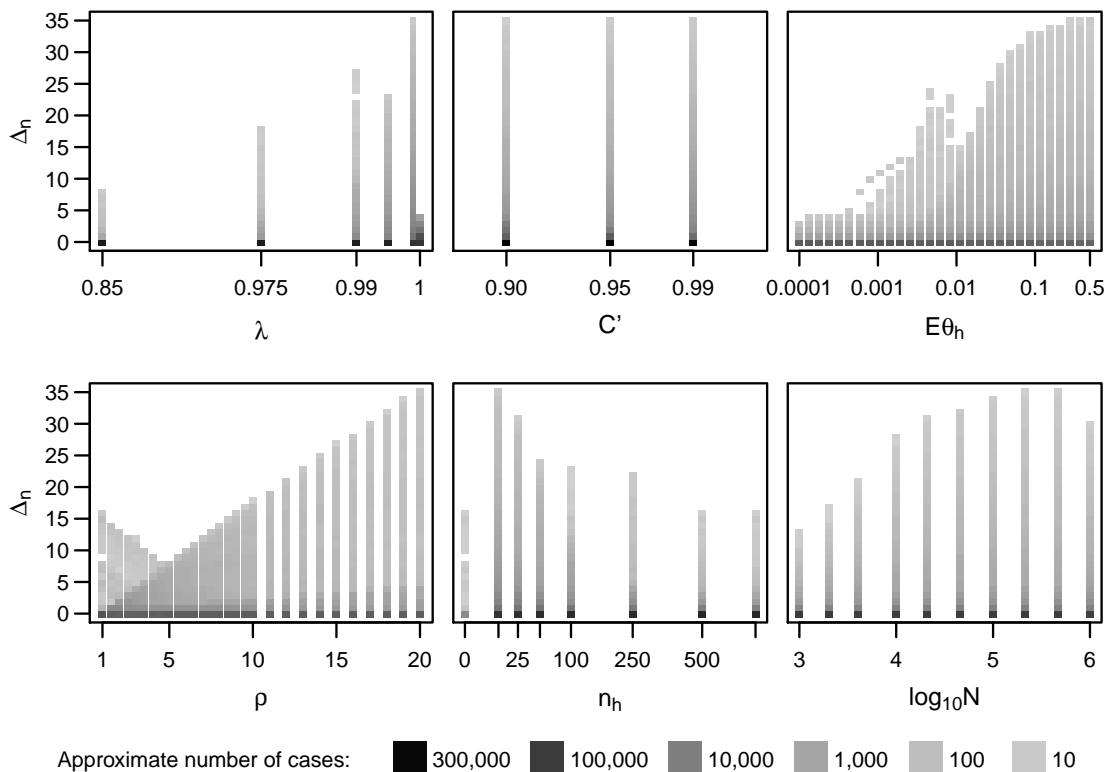
Figure 7: Influence of the input parameters on $\Delta_n = |n_{CJR} - n_{Init}|$ for the 1.2^+ million cases that were considered. Each plot represents the marginal counts of all the cases. That is, the counts of the cases shown in an individual plot are the summation across the five other input parameters. The number of cases represented by each point on the plot is indicated by the intensity of the shading, with darker points indicating more cases.

level). To formalize the hypergeometric acceptance sampling model, let ϑ denote the unknown number of unacceptable items in the population of size N, and let Z denote the hypergeometric variate which represents the number of unacceptable items observed in a sample of size n_a. We develop the confidence statement using a hypothesis test, where H_0 and H_1 denote the null and alternative hypotheses, respectively:

$$H_0 : \vartheta > (1 - \lambda)N$$
$$(29)$$
$$H_1 : \vartheta \leq (1 - \lambda)N$$

We wish to reject H_0 in favor of H_1 with $C_a \times 100\%$ confidence if we observe $Z = 0$ unacceptable items in the sample. We arrive at the required sample size via the expression for the desired Type I error rate for

this test:

$$1 - C_a \geq P(\text{Reject } H_0 \mid H_0 \text{ is true})$$

$$= P(Z = 0 \mid \vartheta > (1 - \lambda)N) \tag{30}$$

$$\geq P(Z = 0 \mid \vartheta = \vartheta_0)$$

where $\vartheta_0 = \lfloor (1 - \lambda)N \rfloor + 1$ denotes the smallest whole number of unacceptable items that may be in the population if H_0 is true. Rewriting the last line in (30) using the mass function of Z, we have

$$1 - C_a \geq \frac{\binom{\vartheta_0}{0}\binom{N - \vartheta_0}{n_a - 0}}{\binom{N}{n_a}} = \frac{(N - \vartheta_0)! \, (N - n_a)!}{(N - \vartheta_0 - n_a)! \, N!} \tag{31}$$

and thus the sample size is the smallest n_a which satisfies (31). Jaech (1973, pp 327) provided a convenient approximation for the mass function of Z evaluated at 0:

$$P(Z = 0 \mid \vartheta = \vartheta_0) = \frac{(N - \vartheta_0)! \, (N - n_a)!}{(N - \vartheta_0 - n_a)! \, N!} \approx \left(1 - \frac{2n_a}{2N - \vartheta_0 + 1}\right)^{\vartheta_0} \tag{32}$$

which, when solved for n_a, gives

$$n_a = \lceil 0.5(1 - (1 - C_a)^{1/\vartheta_0})(2N - \vartheta_0 + 1) \rceil \tag{33}$$

The CJR method (when the population is assumed to be homogeneous and a non-informative prior is used) requires virtually identical sample sizes as those required by the hypergeometric acceptance sampling (AS) method. This is illustrated in Table 1, which gives the number of random samples required to achieve a 95% probability (for the CJR method) or 95% confidence (for the AS method) that a high percentage (95% or 99%) of the items in the population are acceptable. Sample sizes for the CJR model were calculated using (21), with $\beta = 1$ (a uniform prior), $n_h = 0$, and $\rho = 1$. For each of the CJR cases shown in Table 1, the values of $\lambda = 0.95$ and 0.90 were viable, because $k = 1$ for these cases. Sample sizes for the AS approach were calculated using (33), which, for the cases we tried, proved to be identical to samples sizes given by solving (31).

Given the similarity in their sample sizes, it is not surprising that the confidence functions of the CJR and AS models are also closely related. If we assume λN is a whole number, $\vartheta_0 = N - \lambda N + 1$, and we can write (31) as

$$C_a = 1 - \frac{(\lambda N - 1)! \, (N - n_a)!}{(\lambda N - n_a - 1)! \, N!} = 1 - \frac{\Gamma(\lambda N) \, \Gamma(N - n_a + 1)}{\Gamma(\lambda N - n_a) \, \Gamma(N + 1)} \tag{34}$$

	$\lambda = 0.95$		$\lambda = 0.99$	
N	CJR	AS	CJR	AS
100	38	39	78	78
1,000	55	56	237	238
10,000	58	59	290	291
100,000	58	59	297	298

Table 1: Sample sizes required by the CJR method and the hypergeometric acceptance sampling (AS) method for $C' = C_a = 0.95$. For the CJR, $n_h = 0$, $E\theta_h = 0.5$ ($\beta = 1$), and $\rho = 1$.

which bears a close resemblance to (14) when $n_h = 0$, $\rho = 1$, and $\beta = 1$:

$$C = 1 - \frac{\Gamma(\lambda N + 1)\,\Gamma(N - n + 1)}{\Gamma(\lambda N - n)\,\Gamma(N + 2)} = 1 - \frac{\lambda N}{N + 1}\left[\frac{\Gamma(\lambda N)\,\Gamma(N - n + 1)}{\Gamma(\lambda N - n)\,\Gamma(N + 1)}\right] \tag{35}$$

and thus $1 - C$ and $1 - C_a$ differ only by a factor of $\lambda N/(N + 1)$.

4 Discussion

4.1 Why requiring "100% acceptable" is not acceptable

Certainly it is most desirable to state with high probability that the population has no unacceptable items, i.e. $\lambda = 1$. However, achieving that degree of certainty typically requires that almost the entire population be sampled—unless there is very strong prior evidence that the items are acceptable. Table 2 shows the number of random samples required for various levels of ρ, C', and λ. Note that larger samples sizes are required to achieve higher confidence (i.e. higher C' and/or higher λ) and that larger values of ρ result in fewer random samples. Most striking are the massive amounts of random samples required in order to achieve high probability that none of the items in the population are unacceptable.

	$\lambda = 0.994$		$\lambda = 1.00$	
ρ	$C' = 0.95$	$C' = 0.99$	$C' = 0.95$	$C' = 0.99$
1	345	578	4,721	4,925
2	231	471	4,714	4,923
3	118	363	4,708	4,922

Table 2: Values of n required by the refined model to achieve various levels of confidence for $N = 5000$, $n_h = 25$, and $E\theta_h = 0.01$. Note that $\lambda = 0.994$ is viable for the six cases shown on the left half of the table.

Even with very strong prior belief and a large difference in the risk between the two sub-populations, the number of random samples required to show 100% acceptability is still untenable. For example, when $\lambda = 1$, $E\theta_h = 0.0001$, $\rho = 15$, $C' = 0.99$, $n_h = 25$, and $N = 5,000$, the required sample size is 3,422.
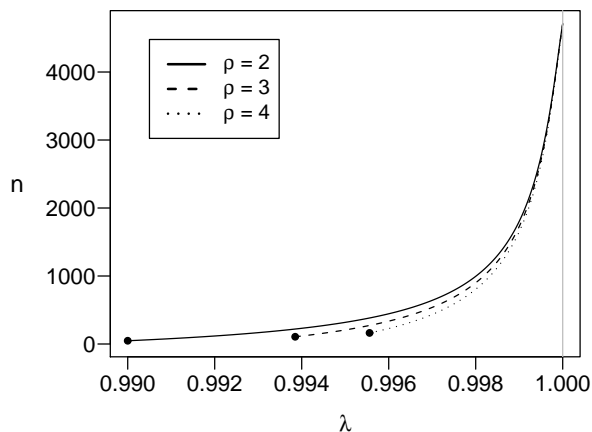
Figure 8: Random sample sizes, n, versus λ, for various values of ρ and for $C' = 0.95$, $E\theta_h = 0.01$, $n_h = 25$, and $N = 5,000$. The solid circles indicate λ_v (the lowest viable λ) for the three cases.

The required number of random samples as a function of λ for various values of ρ is shown in Figure 8. Note that as λ approaches 1, small increases in λ result in very large increases in n. In fact, among the six input parameters for the refined model, n is the most sensitive to λ, which underscores the importance of identifying the smallest viable λ possible, given by (28). While it would obviously be preferable to be able to state with high confidence that 100% of the population is acceptable, such large sample sizes are rarely (if ever) feasible in practice.

4.2 Application to environmental sampling

We now consider a population which consists of a defined spatial area, where units of the population can be characterized by their location within that area. Such is the case for many types of environmental sampling, where the objective is to demonstrate with high confidence that a large fraction of the decision area is acceptable, given all of the sample results are acceptable. This type of sampling plan is especially relevant in the context of the release of a lethal biological or chemical agent inside a building. The information obtained from sampling the building is used to assess the risk to public health, formulate and validate remediation plans, and to eventually certify the safety of the building for re-occupancy.

Discrete sampling units can be created by virtually superimposing a lattice of tessellating units on the surface(s) of the decision area. The shape of the units within the lattice is typically (but not necessarily) square. The size and shape of the unit would be chosen to accommodate the surface sampling methodology

(Emanuel et al., 2008, Ch. 5). In practice, sample locations are often chosen from the units of the lattice based on professional judgment from areas presumed most likely to be contaminated. For example, investigators at the National Institute of Occupational Safety and Health (NIOSH) successfully used professional judgment approaches to quickly identify the presence of *Bacillus anthracis* contamination during the 2001 anthrax incidents. However, decisions that are based solely upon the results of judgmental samples rely heavily on the accuracy of the professional judgment.

To achieve greater confidence that a decision area is acceptable, it is desirable to augment judgmental samples with samples that are taken from randomly selected locations. Hence, judgmental samples are presumed to come from the high-risk sub-population, whereas randomly selected samples are presumed to come from the low-risk sub-population. The initial model and refined CJR model presented in Sections 2.1 and 2.3 were originally developed for this type of environmental sampling. We envision that the CJR model would be especially applicable in situations where it is necessary to demonstrate the acceptability of a decision area, but there is little reason to expect the presence of contamination—either because the decision area is well-removed from the source of contamination, or perhaps because the area has been decontaminated.

We generally expect that data obtained from environmental sampling may exhibit spatial correlation, where samples from locations that are proximal to one another exhibit higher positive correlation than those that are separated by larger distances. If the structure of this correlation were known before sampling, it could be used to potentially reduce the required number of acceptable random samples. More precisely, if a sampled unit proves to be acceptable, it may be likely that adjacent units would also be acceptable, with the positive correlation diminishing for units that are further away. Consequently, in a well-defined spatial model, accounting for the correlation would reduce the risk for unsampled units that are proximal to the acceptable, sampled locations, thereby increasing the overall confidence and reducing the required number of random samples.

The CJR model does not account for spatial correlation, as all items (or units) in the population are modeled as independent Bernoulli observations. Not accounting for spatial variability is conservative and results in larger sample sizes than we would expect if some type of spatial correlation model were assumed for the decision area. Because the inference made via the CJR model is conditional on observing no unacceptable

items, modeling the structure of the spatial correlation (Cressie, 1993) from such data would be impossible, and a spatial model would have to be inferred from other information, if it were available. Furthermore, in cases when sampling is conducted to validate remediation efforts, the impact of the decontamination process on the spatial correlation is likely to be unknown (or unknowable)—in which case a sampling design that does not rely on the estimates of spatial correlation, such as the CJR model, would be appropriate.

4.3   Recommendations for practitioners

Sego et al. (2007) presents the initial model (from Section 2.1) in the context of environmental sampling, with additional background and explanations that may be helpful to those less familiar with statistical models. The CJR model is based on binary outcomes, such as 1) the presence or absence of a particular quality, 2) a quantitative sample result being acceptable or unacceptable as defined by an action level threshold or the limit of detection, 3) contamination being detected or not detected, etc. Consequently, it is important to clearly define the criterion whereby sample outcomes will be labeled 'acceptable' or 'unacceptable.'

In addition to choosing the desired confidence level (C' and λ), there are two parameters that must be specified by the investigator before sampling begins. The first is $E\theta_h$, the expected rate of unacceptable items in the high-risk sub-population. The second is ρ, the factor that indicates the ratio of the expected probability of an unacceptable judgmental sample to the expected probability of an unacceptable random sample.

In general, we recommend using values of $E\theta_h$ between 0.001 and 0.5. If the investigator prefers to make a neutral (non-informative) assertion regarding the prior belief on θ_h, setting $E\theta_h = 0.5$ results in the uniform prior. However, for acceptance sampling, the uniform prior is very conservative (and even pessimistic) because it essentially represents a prior belief where "the chance of a high-risk item being unacceptable is 1%" is just as likely as "the chance of a high-risk item being unacceptable is 99%."

We recommend that ρ be chosen conservatively, so that investigators err on the side of underestimating ρ. However, overly conservative estimates can result in large numbers of random samples required to obtain the desired confidence level. In general, we recommend choosing values of ρ between 1 and 5. Choosing a value of $\rho = 1$ makes the contribution of the judgmental samples equivalent to the random samples, in

which case we recommend setting $n_h = 0$. This ensures that the formulas presented in Section 2.3 collapse to remove the distinction between θ_h and θ_ℓ, thereby modeling a homogeneous population. Consequently, the prior would be elicited by specifying the fraction of the population that we expect to be unacceptable prior to sampling, and this value would be used in place of $E\theta_h$ in (2).
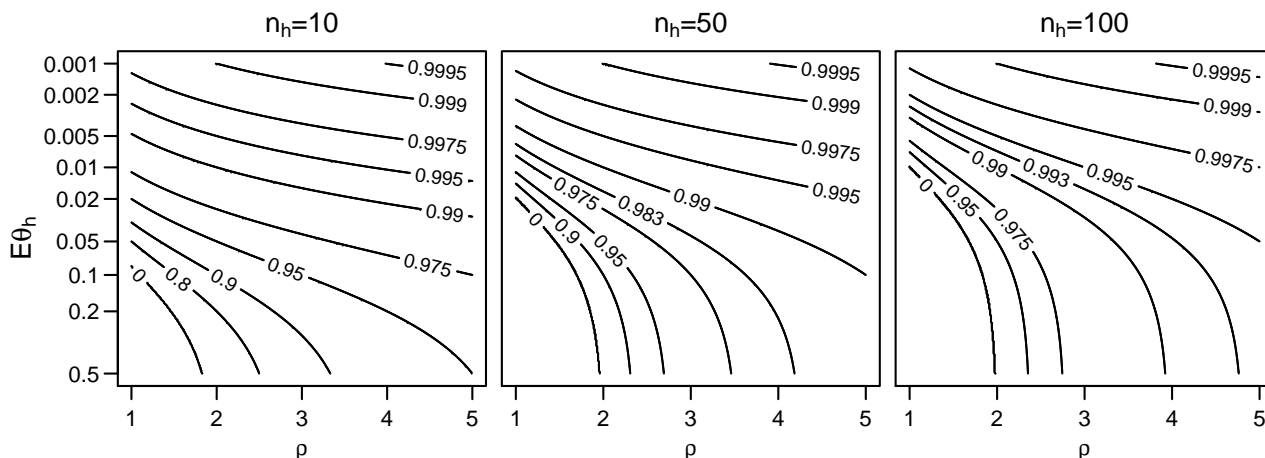

Figure 9: Contours of the minimum viable fraction of the population that is acceptable, λ_v, as a function of $E\theta_h$ and ρ for various values of n_h. The surface of λ_v falls very steeply and becomes flat (and equal to 0) in the lower left corners of the plots, indicating a region where any value of $\lambda \in (0, 1]$ may be used.

Of all the model parameters, the desired acceptable fraction of the population, λ, has the strongest impact on the random sample size (see Figure 8). Consequently, investigators may wish to consider reducing the choice of λ slightly. However, in order to ensure that larger populations require larger sample sizes, the value of λ must be at least as large as λ_v, the minimum viable λ. The way in which λ_v varies as a function $E\theta_h$ and ρ over the recommended range of these two model parameters is illustrated in Figure 9. If a small value of $E\theta_h$ and/or a large value of ρ are chosen, the resulting large value of λ_v will lead to much larger sample sizes (albeit with a stronger confidence statement) than would result from a smaller value of λ_v induced by larger values of $E\theta_h$ and/or smaller values of ρ. In other words, if λ_v results in an infeasible sample size, increasing $E\theta_h$ and/or decreasing ρ can lower the value of λ_v, which, in turn, will lead to smaller sample sizes (at the expense of a weaker confidence statement due to the smaller value of λ).

We also recommend, in general, that λ be greater than C', because it results in much stronger confidence statements than when $\lambda < C'$. For example, being '95% confident that 99% of the population is acceptable'

is a much stronger statement (and consequently, will require more samples) than being '99% confident that 95% of the population is acceptable.'
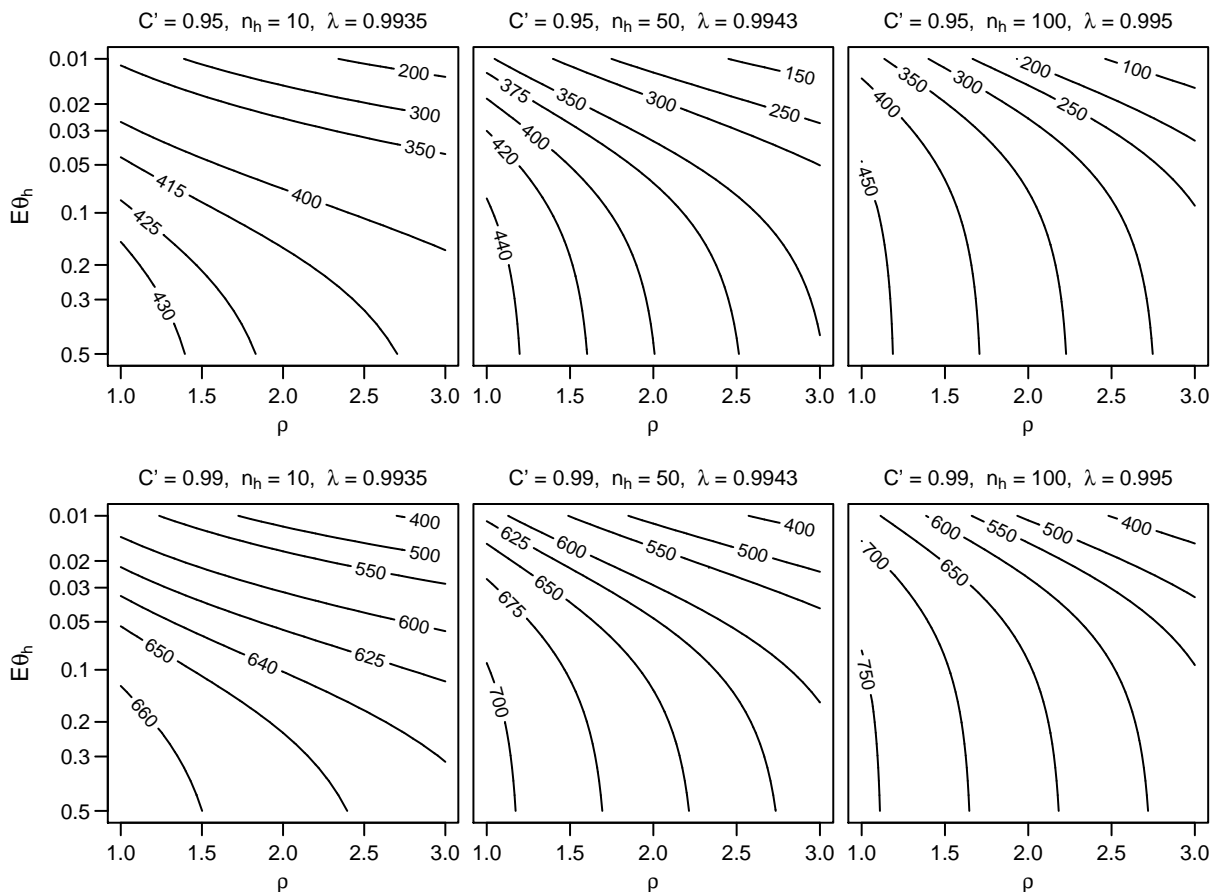
Figure 10: Contours of n as a function of $E\theta_h$ and ρ for $N = 10,000$ and various values of C' and n_h. For each of the six plots, the value of λ was chosen to be viable for all the values of $E\theta_h$ and ρ shown in the plot.

The relationship between $E\theta_h$ and ρ and the required sample size, n, is illustrated in Figure 10. Note that the impact of ρ relative to $E\theta_h$ becomes more pronounced the larger the value of n_h. Naturally, stronger prior evidence in the acceptability of the high-risk samples leads to smaller sample sizes, as do larger values of ρ, which effectively increases the contribution of the n_h acceptable judgmental samples to the overall confidence.

When the values of n_h, $E\theta_h$, and ρ result in no random samples being required, we recommend that practitioners increase C' and/or λ so as to augment the design with some random samples and thereby avoid making inference solely based upon judgmental sampling and the model assumptions.

4.4 Software implementation of the CJR method

The CJR methodology has been implemented in Visual Sample Plan (VSP), a freely-available software tool for the design and analysis of environmental sampling plans (VSP Development Team, 2010). VSP calculates the Bayesian confidence using (14) and sample sizes using (21), (22), or (23). It also calculates λ_v and requires that the user's choice of λ be at least as large as λ_v. The CJR method is presented in VSP in two different modules, one in the context of environmental sampling and the other in the context of discrete item sampling.

The algorithms for the implementation of the CJR method in VSP were extensively validated by comparing VSP calculations to those produced by routines that were separately coded in the statistical language R (R Development Core Team, 2010). In a manner similar to the study discussed in Section 3.3, over 163,000 cases were generated so as to span (as much as possible) the subset of $\mathcal{P}$ that might be encountered in practical use. The value of λ_v and the corresponding random sample size, n, were calculated for each case. The R and VSP calculations of λ_v we carried out to 5 decimal places of accuracy and the resulting values were found to be equivalent in all cases. The random sample sizes (rounded up to the nearest integer) computed by R and VSP were equivalent in all but a single case, where machine error resulted in a difference of 1 sample.

4.5 Ideas for further research

There are several opportunities for additional research. One of the fundamental premises of the CJR method is that the high and low-risk groups can be distinguished perfectly before sampling. As this distinction is likely to be made using professional judgment, it is incumbent to consider what happens when the judgment is partially, or even completely, wrong. There are a number of ways in which judgmental samples could be incorrectly chosen—this is briefly discussed and illustrated in Sego (2007, Figure 8). The sensitivity of the CJR method to misspecification of the high and low-risk groups is the subject of ongoing research.

It would also be advantageous to generalize the CJR method to accommodate more than two risk-based sub-populations, especially in the context of indoor environmental sampling. Allowing for more than two groups would make it possible to account for different levels of risk induced by the various factors that

influence the likelihood of detecting a contaminant at a particular location. These factors may include the spatial distribution of a contaminant, the persistence and interaction of the contaminant on a particular surface material, the impact of decontamination, and the sensitivity of the sampling methodology itself.

Investigation of the optimal allocation of samples among the high and low-risk groups (or among multiple groups) is another area of inquiry. MacQueen (2007) determined that the allocation of samples depends on the certainty with which items in the population can be classified into the correct risk group. If the risk groups can be determined with perfect certainty, allocating samples to the highest risk group (until it is completely sampled), then to the next highest group, and so on, is optimal, in the sense that it will require the fewest samples to achieve the desired confidence level. However, if the risk groups cannot be determined with certainty, other types of sample allocation among the groups would be preferable.

The CJR method could also be adapted to account for the probability of false negative measurements, perhaps using the approach discussed by Axelrod (2005). It would also be desirable to accommodate environmental samples obtained via different sampling methodologies (Emanuel et al., 2008, Ch. 5) in the same decision area. For example, wipe, swab, and vacuum samples each have sampling footprints of different sizes and different limits of detection.

5   Conclusion

We present a Bayesian model for acceptance sampling where the population consists of two groups, each with different levels of risk of containing unacceptable items. Expert opinion, or judgment, may be required to distinguish between the high and low-risk groups. The sampling scheme presumes that all the high-risk items are sampled (judgmentally) and that a random sample is taken from the low-risk items. For this reason, we call the approach the combined judgmental and random (CJR) method. Inference is made conditional on all the observed samples being acceptable, in which case we may conclude with a specified probability that a high percentage of the items in the population are acceptable.

Special consideration must be taken to ensure the CJR method has desirable analytical properties. In particular, the required sample size should be a non-decreasing function of the population size, which may not be the case if the prior belief in the acceptability of the population is so strong that it already satisfies

the intent to demonstrate (through sampling) that a high percentage of the population is acceptable. In these situations, the desired fraction of acceptable items must be increased until the Bayesian confidence function is non-increasing, which then ensures that larger populations will require larger sample sizes.

We demonstrated that the Bayesian acceptance sampling approach presented by Wright (1992) and Grieve (1994), as well as hypergeometric acceptance sampling, are, in a sense, special cases of the CJR method. We also illustrate how, even with strong prior belief in the acceptability of the population, attempting to demonstrate that 100% of the population is acceptable results in unrealistically large sample sizes.

While applicable to a broad range of applications, the CJR method was developed to provide a statistical model for environmental sampling, particularly inside buildings that may have been contaminated with a biological or chemical agent. Because the CJR methodology is based on the assumption that all units of the population are independent, any spatial correlation that may exist among the sampling units is ignored. However, developing an accurate spatial model when all the sampled items are expected to be acceptable may be difficult or even impossible. The assumption of independence is conservative, because accounting for spatial correlation could be used to reduce the number of required samples.

Careful consideration must be made in determining the input parameters for the CJR model. We recommend that the expected rate of unacceptable high-risk items $(E\theta_h)$ be chosen between 0.001 and 0.5. Likewise, we recommend that the ratio of the expected probability of an unacceptable judgmental sample to the expected probability of an unacceptable random sample, ρ, be chosen between 1 and 5. It is important to choose a fraction of the population that we wish to demonstrate is acceptable, λ, to be at least as great as a minimum value of λ that will ensure that the required random sample size increases as the population size increases. This minimum value is easily calculated as discussed in Section 3.2. The CJR methodology is implemented in Visual Sample Plan (VSP Development Team, 2010) software and is freely available.

Acknowledgments

for helpful conversations regarding some of the technical details. We also express our appreciation to Randall Smith (NIOSH) and Yan Jin (NIOSH) for their thorough and thoughtful reviews of the manuscript.

References

Abramowitz, M., and Stegun, I. (Eds.). (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Washington, D.C.: National Bureau of Standards.

Axelrod, M. (2005). *Using ancillary information to reduce sample size in discovery sampling and the effects of measurement error* (Tech. Rep. No. UCRL-TR-216206). Livermore, CA: Lawrence Livermore National Laboratory. Retrieved 24 August 2010, from `https://e-reports-ext.llnl.gov/pdf/324013.pdf`

Bowen, W. M., and Bennett, C. A. (1988). *Statistical methods for nuclear material management* (No. NUREG/CR-4604). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Cressie, N. (1993). *Statistics for spatial data* (Revised ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Danaher, P. J., and Hardie, B. G. S. (2005). Bacon with your eggs? Applications of a new bivariate beta-binomial distribution. *The American Statistician*, *59*, 282-286.

Emanuel, P., Roos, J. W., and Niyogi, K. (Eds.). (2008). *Sampling for biological agents in the environment.* Washington, D.C.: ASM Press.

GAO. (2005). *Anthrax detection: Agencies need to validate sampling activities in order to increase confidence in negative results* (Tech. Rep. No. GAO-05-251). Washington, D.C.: United States Government Accountability Office. Retrieved 24 August 2010, from `www.gao.gov/new.items/d05251.pdf`

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Graves, T., Hamada, M., Booker, J., Decroix, M., Chilcoat, K., and Bowyer, C. (2007). Estimating a proportion using stratified data from both convenience and random samples. *Technometrics*, *49*, 164-171.

Grieve, A. P. (1994). A further note on sampling to locate rare defectives with strong prior evidence. *Biometrika*, *81*, 787-789.

Gupta, A. K., and Wong, C. R. (1985). On three and five parameter bivariate beta distributions. *Metrika*, *32*, 85-91.

Guy, D. M., Carmichael, D. R., and Whittington, O. R. (1998). *Practitioner's guide to audit sampling.*

Hoboken, NJ: John Wiley & Sons, Inc.

Hoehn, L., and Niven, I. (1985). Averages on the move. *Mathematics Magazine*, *58*, 151-156.

Jaech, J. L. (1973). *Statistical methods in nuclear material control* (Tech. Rep. No. TID-26298). Springfield, VA: NTIS.

Lee, M.-L. T. (1996). Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics–Theory and Methods*, *25*, 1207-1222.

MacQueen, D. H. (2007). *Material-based stratification* (Tech. Rep. No. UCRL-TR-231455). Livermore, CA: Lawrence Livermore National Laboratory. Retrieved 24 August 2010, from `https://e-reports-ext.llnl.gov/pdf/348316.pdf`

Magnussen, S. (2002). An algorithm for generating positively correlated beta-distributed random variables with known marginal distributions and specified correlation. *Computational Statistics & Data Analysis*, *46*, 397-406.

Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitt. Math. Statist.*, *8*, 234-235.

Niven, I. (1981). *Maxima and minima without calculus*. USA: The Mathematical Association of America. Dolciani Mathematical Expositions No. 6.

Olkin, I., and Liu, R. (2003). A bivariate beta distribution. *Statistics & Probability Letters*, *62*, 407-412.

R Development Core Team. (2010). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. Available from `http://www.R-project.org` (ISBN 3-900051-07-0)

Sarmanov, O. V. (1966). Generalized normal correlation and two-dimensional Frechet classes. *Doklady (Soviet Mathematics)*, *168*, 596-599.

Schilling, E. G., and Neubauer, D. V. (2009). *Acceptance sampling in quality control* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Schucany, W. R., Parr, W. C., and Boyer, J. E. (1978). Correlation structure in Farlie-Gumbel-Morgenstern distributions. *Biometrika*, *65*, 650-653.

Sego, L., Anderson, K., Shulman, S., Matzke, B., Pulsipher, B., Sieber, K., et al. (2007). *An environmental sampling model for combining judgmental and randomly placed samples* (Tech. Rep. No. PNNL-16636). Richland, WA: Pacific Northwest National Laboratory. Retrieved 24 August 2010, from `http://www.pnl.gov/main/publications/external/technical_reports/PNNL-16636.pdf`

Shubina, M., and Lee, M.-L. T. (2004). On maximum attainable correlation and other measures of dependence for the Sarmanov family of bivariate distributions. *Communications in Statistics–Theory and Methods*, *33*, 1031-1052.

Valgenti, V. C., Shaikot, S. H., and Kim, M. S. (2010). Playing the odds: Sampling with Snort. *Submitted*.

VSP Development Team. (2010). *Visual Sample Plan: A tool for design and analysis of environmental sampling.* Pacific Northwest National Laboratory. Richland, WA. Available from `http://vsp.pnl.gov`

Wright, T. (1992). A note on sampling to locate rare defectives with strong prior evidence. *Biometrika*, *79*, 685-691.

A   Equivalence of the distribution of unobserved unacceptable items

Wright (1992) and Grieve (1994) use a Beta-binomial prior on the number of unacceptable items in the population and a hypergeometric likelihood for the number of observed unacceptable items. This yields a Beta-binomial posterior for the number of unacceptable items (Grieve, 1994, Eq. $1 \cdot 3$) and a Beta-binomial distribution of the number of *unobserved* unacceptable items (Grieve, 1994, Eq. $2 \cdot 1$). In our model we employ a Beta prior distribution and a Binomial likelihood, which gives rise to a Beta posterior. We demonstrate below that the resulting posterior predictive Beta-binomial distribution of the number of unobserved unacceptable items from our model is the same as that obtained by Wright and Grieve.

Let θ denote the probability of an item being unacceptable with prior distribution $\theta \sim \text{Beta}(\alpha, \beta)$. For a population of size N, let X denote the number of unacceptable items observed in a sample of size n, where $X|\theta \sim \text{Bin}(n, \theta)$. After observing $X = x$ unacceptable items in the sample, we have by the standard Beta/Binomial conjugacy (Gelman et al., 2004) that $\theta|(X = x) \sim \text{Beta}(x + \alpha, n - x + \beta)$, whose density we write as $p(\theta|X = x)$. Let Y denote the number of *unobserved* unacceptable items among the unsampled items of the population, where $Y|\theta \sim \text{Bin}(N - n, \theta)$. Equation $(2 \cdot 3)$ in Wright (1992) and Equation $(2 \cdot 1)$ in Grieve (1994) give the posterior probability of exactly $Y = y$ unobserved unacceptable items remaining in the population after observing $X = x$ unacceptable items. Note that our Y, y, X, and x and correspond to Wright and Grieve's $\widetilde{d}$, $\widetilde{d}_0$, d, and d_0, respectively. Using our model, we arrive at the same result obtained by Wright and Grieve:

$$
\begin{aligned}
P(Y = y \mid X = x) &= \int_0^1 P(Y = y \mid \theta) \; p(\theta|X = x) \; d\theta \\
&= \int_0^1 \binom{N-n}{y} \theta^y (1-\theta)^{N-n-y} \; [B(x+\alpha, \; n-x+\beta)]^{-1} \, \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1} \; d\theta \\
&= \binom{N-n}{y} [B(x+\alpha, \; n-x+\beta)]^{-1} \int_0^1 \theta^{y+x+\alpha-1} (1-\theta)^{N-y-x+\beta-1} \; d\theta
\end{aligned}
$$

$$= \binom{N-n}{y} \frac{B(y+x+\alpha,\ N-y-x+\beta)}{B(x+\alpha,\ n-x+\beta)}$$

B   Derivation of the approximate Bayesian confidence function

We provide here the details of the derivation of (14). To simplify these expressions, let

$$a = (1-\lambda)N + 1$$

$$b = \lambda N - n_h - n$$

$$\beta' = n + \rho(n_h + \beta + 1) - 1$$

and let A_1, A_2, A_3, and A_4 be defined as in (15). Then:

$$
\begin{aligned}
C &\approx \int_0^1 F_{Y_\ell}^\star((1-\lambda)N \mid \theta_\ell)\, p_2(\theta_\ell \mid X_h = 0, X_\ell = 0)\, d\theta_\ell \\
&= \int_0^1 \left(1 - [B(a,b)]^{-1} \int_0^{\theta_\ell} t^{a-1}(1-t)^{b-1}\, dt\right) p_2(\theta_\ell \mid X_h = 0, X_\ell = 0)\, d\theta_\ell \\
&= 1 - [B(a,b)]^{-1} \int_0^1 \left(\int_0^{\theta_\ell} t^{a-1}(1-t)^{b-1}\, dt\right) \beta'(1-\theta_\ell)^{\beta'-1}\, d\theta_\ell \\
&= 1 - [B(a,b)]^{-1} \int_0^1 \int_t^1 t^{a-1}(1-t)^{b-1}\beta'(1-\theta_\ell)^{\beta'-1}\, d\theta_\ell\, dt \\
&= 1 - [B(a,b)]^{-1} \int_0^1 t^{a-1}(1-t)^{b-1}(1-t)^{\beta'}\, dt \\
&= 1 - [B(a,b)]^{-1} \int_0^1 t^{a-1}(1-t)^{\beta'+b-1}\, dt \\
&= 1 - B(a, \beta' + b)/B(a,b) \\
&= 1 - (\Gamma(a+b)\Gamma(\beta'+b)) / (\Gamma(b)\Gamma(a+b+\beta')) \\
&= 1 - (\Gamma(A_1)\Gamma(A_2)) / (\Gamma(A_3)\Gamma(A_4))
\end{aligned}
$$

C   Sample size equivalence

When $\lambda = 1$, it is interesting to note that the value of n given by (22) agrees exactly with Grieve's solution (1994) if the population has homogeneous risk, i.e., $n_h = 0$ and $\rho = 1$. Rewriting (14) by substituting $n_h = 0$, $\rho = 1$, and $\lambda = 1$ yields:

$$
\begin{aligned}
C &\approx 1 - (\Gamma(N-n+1)\Gamma(N+\beta)) / (\Gamma(N-n)\Gamma(N+\beta+1)) \\
&= 1 - (N-n)/(N+\beta)
\end{aligned}
$$

For a desired confidence level C', we want $C' \leq 1 - (N-n)/(N+\beta)$ which can be expressed as

$$N - n \leq (N+\beta)(1-C') \tag{36}$$

Grieve's (1994) expression for the sample size (equation $2 \cdot 5$) is given as

$$N - n \leq (N+\beta)\{1 - (1-\epsilon)^{(1-\delta)/(\beta\delta)}\} \tag{37}$$

Grieve defines $\delta = \alpha/(\alpha+\beta)$. In our context, $\alpha = 1$, hence $\delta = 1/(1+\beta)$. Our desired confidence level, C', is equivalent to Grieve's $1 - \epsilon$. Thus, translating (37) to our notation demonstrates (36) is equivalent to (37).

D   Expected value of the fraction of acceptable items prior to sampling

To motivate the derivation of the expected value of the fraction of acceptable items prior to sampling, λ_p, we state below the framework of the refined CJR model presented in Section 2.3, which we assume holds prior to the collection of judgmental and random samples:

$$\theta_h \sim \text{Beta}(1, \beta)$$
$$\theta_\ell \sim \text{Beta}(1, \rho(\beta+1) - 1)$$
$$X_h | \theta_h \sim \text{Bin}(n_h, \theta_h)$$
$$X_\ell | \theta_\ell \sim \text{Bin}(n, \theta_\ell)$$
$$Y_\ell | \theta_\ell \sim \text{Bin}(N - n_h - n, \theta_\ell)$$

Thus,

$$
\begin{aligned}
\lambda_p &= E\left(1 - \frac{X_h + X_\ell + Y_\ell}{N}\right) \\
&= 1 - \frac{1}{N}\left[E[E(X_h|\theta_h)] + E[E(X_\ell|\theta_\ell)] + E[E(Y_\ell|\theta_\ell)]\right] \\
&= 1 - \frac{1}{N}\left[E(n_h\theta_h) + E(n\theta_\ell) + E((N - n_h - n)\theta_\ell)\right] \\
&= 1 - \frac{1}{N}\left[\frac{n_h}{1+\beta} + \frac{n}{\rho(\beta+1)} + \frac{N - n_h - n}{\rho(\beta+1)}\right] \\
&= 1 - \frac{N + n_h(\rho - 1)}{N\rho(\beta+1)}
\end{aligned}
$$

E   Determining the viable fraction of acceptable items

The viable fraction of acceptable items, λ_v, can be determined using Theorem 1, which is restated and proven below.

Theorem 1. *Assume N, n_h, and n are integer valued. Consider the Bayesian confidence h as a function of N and n, with all other parameters (λ, n_h, ρ, and β) held fixed. Let $\mathcal{V} = \{\lambda \;:\; \frac{\partial h}{\partial N} \leq 0 \text{ for all } N \in \mathcal{P} \text{ and for all } n \in \mathcal{P}\}$ denote the set of 'viable' λ values. Furthermore, let $k = \frac{1}{2}\rho(n_h + \beta + 1) - n_h$, and let $\rho(n_h + \beta + 1)$ be integer valued. Then (i) for $k \leq 1$, $\mathcal{V} = (0, \; 1]$, and (ii) for $k > 1$, $\mathcal{V} = [1 - \frac{1}{k}, \; 1]$.*

Proof. The partial derivative of the Bayesian confidence function h with respect to N is $\frac{\partial h}{\partial N} = (h-1)M$ where

$$M = \Psi(A_1) + \lambda\Psi(A_2) - \lambda\Psi(A_3) - \Psi(A_4)$$

and A_1, A_2, A_3, and A_4 are defined by (15). Because $0 \leq h \leq 1$, if $M \geq 0$, then $\frac{\partial h}{\partial N} \leq 0$.

Let $\beta' = n + \rho(n_h + \beta + 1) - 1$. To further characterize the components of M, we note the relationship between $\Psi(A_4)$ and $\Psi(A_1)$:

$$
\begin{aligned}
\Psi(A_4) &= \Psi(\beta' + A_1) \\
&= \Psi\left((\beta' + 1) + (A_1 - 1)\right) \\
&= \frac{1}{\beta' + A_1 - 1} + \frac{1}{\beta' - 1 + A_1 - 1} + \cdots + \frac{1}{2 + A_1 - 1} + \frac{1}{1 + A_1 - 1} + \Psi(A_1) \quad (38) \\
&= \Psi(A_1) + \sum_{i=1}^{\beta'} \frac{1}{A_1 - 1 + i}
\end{aligned}
$$

where the third equality arises from the recurrence relation of the digamma function (Abramowitz and Stegun, 1972, eq. 6.3.6), so long as β' is integer valued. Thus,

$$\Psi(A_4) - \Psi(A_1) = \frac{\beta'}{H(\mathcal{S}_1)} \tag{39}$$

where $\mathcal{S}_1 = \{A_1, A_1 + 1, \ldots, A_1 + \beta' - 1\}$ and $H(\mathcal{S})$ denotes the harmonic mean of the elements of $\mathcal{S}$. Applying the same argument for $\Psi(A_2)$ and $\Psi(A_3)$ gives

$$\lambda(\Psi(A_2) - \Psi(A_3)) = \frac{\lambda\beta'}{H(\mathcal{S}_2)} \tag{40}$$

where $\mathcal{S}_2 = \{A_3, A_3 + 1, \ldots, A_3 + \beta' - 1\}$. Combining (39) and (40) gives

$$M = \beta'\left(\frac{\lambda}{H(\mathcal{S}_2)} - \frac{1}{H(\mathcal{S}_1)}\right) \tag{41}$$

Using (41) it is straightforward to show that $M \geq 0$ holds if and only if

$$\lambda H(\mathcal{S}_1) - H(\mathcal{S}_2) \geq 0 \tag{42}$$

Note that adding $(1 - \lambda)N + 1$ to each element of $\mathcal{S}_2$ gives $\mathcal{S}_1$. Consequently, we have from Hoehn (1985, eq. 8)

$$H(\mathcal{S}_1) - H(\mathcal{S}_2) > (1 - \lambda)N + 1 \tag{43}$$

Thus

$$
\begin{aligned}
\lambda H(\mathcal{S}_1) - H(\mathcal{S}_2) &= H(\mathcal{S}_1) - H(\mathcal{S}_2) + (\lambda - 1)H(\mathcal{S}_1) \\
&> (1 - \lambda)N + 1 + (\lambda - 1)H(\mathcal{S}_1) \\
&= (1 - \lambda)(N - H(\mathcal{S}_1)) + 1
\end{aligned}
\tag{44}
$$

where the inequality is implied by (43).

Let $A(\mathcal{S})$ denote the arithmetic mean of the elements in $\mathcal{S}$. Thus

$$
A(\mathcal{S}_1) = \frac{1}{\beta'} \sum_{i=1}^{\beta'} (N - n_h - n + i) = N + k_n
\tag{45}
$$

where $k_n = (\rho(n_h + \beta + 1) - 2n_h - n)/2$. From Niven (1981, Sec. 2.7) we have

$$
A(\mathcal{S}_1) > H(\mathcal{S}_1)
\tag{46}
$$

Continuing now with the last line of (44) we have from (45) and (46) that

$$
(1 - \lambda)(N - H(\mathcal{S}_1)) + 1 > (1 - \lambda)(N - A(\mathcal{S}_1)) + 1 = 1 - k_n(1 - \lambda)
\tag{47}
$$

Note that k_n is decreasing in n, and thus

$$
1 - k_n(1 - \lambda) \geq 1 - k(1 - \lambda)
\tag{48}
$$

where $k \equiv k_0$. By combining (44), (47), and (48), we have for all $N \in \mathcal{P}$ and $n \in \mathcal{P}$:

$$
\lambda H(\mathcal{S}_1) - H(\mathcal{S}_2) > 1 - k(1 - \lambda)
\tag{49}
$$

and thus any λ which satisfies

$$
1 - k(1 - \lambda) \geq 0
\tag{50}
$$

will be an element of $\mathcal{V}$.

When $k \leq 1$, (50) is satisfied for any $\lambda \in (0, 1]$, which proves case (i). When $k > 1$, (50) implies that any $\lambda \geq 1 - \frac{1}{k}$ will be in $\mathcal{V}$. To show that $(1 - \frac{1}{k}) = \inf \mathcal{V}$, we need the following lemma:

Lemma. *Using the following symbols as defined previously,*

$$
\lim_{N \to \infty} (\lambda H(\mathcal{S}_1) - H(\mathcal{S}_2)) = 1 - k_n(1 - \lambda) .
$$

We provide the proof of the Lemma below. Because (49) demonstrates that $(1 - k(1 - \lambda))$ is a lower bound for $(\lambda H(\mathcal{S}_1) - H(\mathcal{S}_2))$ which holds for all $N, n \in \mathcal{P}$, and (48) and the Lemma together demonstrate $(\lambda H(\mathcal{S}_1) - H(\mathcal{S}_2))$ achieves the lower bound in the limit, it follows that $(1 - \frac{1}{k})$ is the infimum of $\mathcal{V}$, thus proving case (ii). $\square$

Proof of Lemma. We begin with a result from Hoehn and Niven (1985, eq. 16)

$$\lim_{x \to \infty} [H(\mathcal{S} + x) - x] = A(\mathcal{S}) \tag{51}$$

where $\mathcal{S} + x$ indicates the set resulting from the addition of x to each element of $\mathcal{S}$. Let

$$\mathcal{S}_3 = \{-n_h - n, \ -n_h - n + 1, \ \ldots, \ -n_h - n + \beta' - 1\}$$

Note that

$$\begin{aligned} \mathcal{S}_1 &= \mathcal{S}_3 + (N + 1) \\ \mathcal{S}_2 &= \mathcal{S}_3 + \lambda N \end{aligned} \tag{52}$$

It is straightforward to show

$$A(\mathcal{S}_3) = k_n - 1 \tag{53}$$

We thus have by (51), (52), and (53):

$$\begin{aligned} \lim_{N \to \infty} H(\mathcal{S}_1) - (N + 1) &= k_n - 1 \\ \lim_{N \to \infty} H(\mathcal{S}_2) - \lambda N &= k_n - 1 \end{aligned} \tag{54}$$

By adding and subtracting certain terms, it follows that

$$\lambda H(\mathcal{S}_1) - H(\mathcal{S}_2) = \lambda \left(1 + H(\mathcal{S}_1) - (N + 1)\right) - (H(\mathcal{S}_2) - \lambda N) \tag{55}$$

By applying (54) and the well-known properties of limits, the limit of the right-hand side of (55) as $N \to \infty$ is given by

$$\lambda(1 + k_n - 1) - (k_n - 1) = 1 - k_n(1 - \lambda)$$

$$\square$$

F   Glossary of symbols

$A(\mathcal{S})$ The arithmetic mean of the elements of $\mathcal{S}$

α The first shape parameter of the Beta distribution; In the models considered here, α is typically 1

β The second shape parameter of the Beta distribution; Sometimes it has various subscripts to indicate high risk, h, or low risk, ℓ

β' The value of the second shape parameter for the posterior beta distribution of θ_ℓ; It is equal to $n + \rho(n_h + \beta + 1) - 1$

40

$B(a, b)$ The complete beta function, given by $\int_0^1 t^{a-1}(1-t)^{b-1}dt$

Beta   Indicates the Beta distribution

Bin   Indicates the Binomial distribution

C The confidence level (and for Bayesian models, the probability) that a high percentage of the population will be acceptable

C' The desired confidence level that is used to determine the random sample size

C_a The desired confidence level that is used to determine the random sample size for hypergeometric acceptance sampling (Section 3.4)

Δ_n The absolute difference of the random sample sizes prescribed by the initial model (Section 2.1) and the CJR model (Section 2.3)

$E\theta_h$ The expected value of θ_h which is equal to $(1 + \beta)^{-1}$; It is also the expected rate of unacceptable items in the high-risk group; Used to choose the value of β

f The Binomial likelihood (or mass) function, where $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for x successes in n independent Bernoulli trials, each with probability of success p

F_{Y_ℓ} The distribution function of Y_ℓ

$F_{Y_\ell}^\star$ A continuous approximation of the distribution function of Y_ℓ

$\Gamma(x)$ The Gamma function, given by $\int_0^\infty t^{x-1}e^{-t}dt$; Note that $\Gamma(x+1) = x\Gamma(x)$ for $x \in \mathbb{R}$ and $\Gamma(n+1) = n!$ for $n \in \mathbb{N}$

$h(\cdot)$ A generic representation of the Bayesian confidence function of the CJR model, defined in (16)

$H(\mathcal{S})$ The harmonic mean of the elements of $\mathcal{S}$

k A constant used in Theorem 1 to determine λ_v, equal to $\frac{1}{2}\rho(n_h + \beta + 1) - n_h$

λ The fraction of the population that we desire to be acceptable with $C' \times 100\%$ confidence

λ_p For the CJR model (Section 2.3), the expected value of the fraction of acceptable items in the population prior to sampling, defined by (27)

λ_v The minimum viable λ; The smallest value of λ that will ensure that $h(N)$ will be non-increasing for all possible n and N; When determining the sample size, choosing $\lambda \geq \lambda_v$ will ensure that the required sample size, n, is a non-decreasing function of N; It follows as a result of Theorem 1 and is given by (28) for $k > 1$.

log   The natural (base e) logarithm

$\log_{10}$ The base 10 logarithm

N The number of items (or units) in a population

n_h The number of high-risk items in the population; Also the number of judgmental samples

n_ℓ The number of low-risk items in the population

n The number of random samples from the low-risk sub-population; If clarification is needed, subscripts (described below) are added

n_a The number of random samples required by the hypergeometric acceptance sampling method described in Section 3.4

n_{CJR} The number of random samples required by the CJR method described in Section 2.3

n_{Init} The number of random samples required by the initial model as described in Section 2.1

$\mathcal{P}$ The parameter space of the CJR model which defines the acceptable ranges of the model parameters, defined in (17)

$p_1(\theta_h),\ p_1(\theta_\ell)$ The prior distributions used for θ_h and θ_ℓ in the initial model described in Section 2.1

$p_2(\theta_h),\ p_2(\theta_\ell)$ The prior distributions used for θ_h and θ_ℓ in the CJR model described in Section 2.3

$\phi(\beta)$ The linkage function that results from $E\theta_h = \rho E\theta_\ell$ in the CJR model (Section 2.3)

$\Psi(x)$ The digamma function, equal to $\frac{d}{dx}\log\Gamma(x)$

ρ For the initial model (Section 2.1), the ratio of θ_h to θ_ℓ; For the CJR model (Section 2.3), the ratio of $E\theta_h$ to $E\theta_\ell$

θ The probability that an item is unacceptable

θ_h The probability that a high-risk item is unacceptable, follows a Beta distribution

θ_ℓ The probability that a low-risk item is unacceptable, follows a Beta distribution

$\mathcal{V}$ For given values of β, n_h, and ρ, the set of λ that ensure that $h(N)$ is decreasing over all $N \geq 1$ and all $n \in \mathcal{P}$.

φ The probability of being unacceptable for the binomial variate W

ϑ For the hypergeometric acceptance sampling model (Section 3.4), the unknown number of unacceptable items in the population

ϑ_0 For the hypergeometric acceptance sampling model (Section 3.4), equal to $\lfloor (1 - \lambda)N \rfloor + 1$; It is the smallest whole number of unacceptable items that can exist in the population if the null hypothesis, $H_0 : \vartheta > (1 - \lambda)N$, is true

W A binomial variate used to illustrate the phenomenon that the Bayesian confidence may not be a decreasing function of N

X_h A binomial variate of the number of unacceptable items in the sample of size n_h from the high-risk group (the judgmental samples)

X_ℓ A binomial variate of the number of unacceptable items in the random sample of size n from the low-risk group

Y_ℓ A binomial variate of the number of unacceptable items in the unsampled, low-risk group of size $N - n_h - n$

Z A hypergeometric variate used in frequentist acceptance sampling (Section 3.4) indicating the number of unacceptable items in the sample