Estimating General Variable Acceptance Sampling Plans by Bootstrap Methods

Herwig Friedl[*], Erwin Stampfer[†]

Technical University Graz

Steyrergasse 17

A–8010 Graz, Austria

Abstract

We consider variable acceptance sampling plans that control the lot or process fraction defective, where a specification limit defines acceptable quality. The problem is to find a sampling plan, fullfilling some conditions usually on the operation characteristic. Its calculation heavily depends on distributional properties, that in practice might be doubtful. If prior data is already available, we propose to estimate the sampling plan by means of bootstrap methods. Bias and standard error of the estimated plan can be assessed easily by Monte Carlo approximation to the respective bootstrap moments. This resampling approach does not require strong assumptions and, furthermore,

[*]Dr. Friedl is an Associated Professor in the Department of Statistics. His email address is friedl@stat.tu-graz.ac.at

[†]Dr. Stampfer is an Assistant Professor in the Department of Statistics. His email address is stampfer@stat.tu-graz.ac.at

it is a flexible method that can be extended to any statistic that might be informative for the fraction defective in a lot.

Introduction

Variable acceptance sampling is the assessment of the quality of a lot (some finite population of objects) based on a sample randomly drawn from it. Assume X to be a random variable measuring some quantity of interest on a product. Let further u be a given upper specification limit for X. Quality can be described by $p = P(X > u)$, giving the chance of a product being defective, often called fraction defective or non-conforming. Now from a lot of sufficiently large size we draw a random sample $X_1, \ldots, X_n$ of independent and identically distributed (iid) elements X_i, $i = 1, \ldots, n$, steming from the same distribution as X. By use of this sample we wish to decide wether the lot should be accepted or not, whereby we want to control for p. A widely used criterion for this decision is $\bar{X} + kS \leq u$, where $\bar{X}$ denotes the sample mean, S^2 is the empirical variance and k is some constant. If $\bar{X} + kS \leq u$, we accept the lot, otherwise it is rejected. We denote a sampling plan like this by $s = (n, k)$.

Let us assume for a moment that X stems from a normal distribution with parameter (μ, σ^2), where σ^2 is fixed but usually unkown. We write Φ_{μ,σ^2} for its distribution function and simply use the notation Φ for its standardized version $(\mu, \sigma^2) = (0, 1)$. Note that under this assumption

$$p = P(X > u|\Phi_{\mu,\sigma^2}) = 1 - \Phi\left(\frac{u - \mu}{\sigma}\right)$$

can be regarded as a one-to-one function of μ with inverse $\mu(p) = u - \sigma\Phi^{-1}(1 - p)$.

This motivates us to introduce the OC-function giving the probability of acceptance as a function of p. For $X_i \overset{iid}{\sim} F_p$ with $F_p(u) = 1 - p$, i.e. X_i has fraction defective p, we define

$$L(F_p, n, k) = P(\bar{X} + kS \leq u|F_p) = P\left(\frac{u - \bar{X}}{S} \geq k|F_p\right).$$

For the normal distribution as mentioned above, F_p is identical to $\Phi_{\mu(p),\sigma^2}$. Be aware that in general p does not identify the parameters of the distribution of X. In this case some idea of unification should come in, e.g. assuming that σ is fixed. But other strategies might be as reasonable, depending on the considered type of distribution.

The problem now is to choose a particular value of s, say $s_0 = (n_0, k_0)$, that fullfills some restrictions on the OC–function. A reasonable approach is to require $L(F_{p_0}, n_0, k_0) = 1 - \alpha$ and $L(F_{p_1}, n_0, k_0) = \beta$ for prespecified probabilities p_0 and p_1. Often p_0 is identical with the acceptable quality level, and p_1 with the rejectable quality level. In the normal case the solution may be found in many textbooks like in Montgomery (1996; p. 652).

The main drawback of variable acceptance sampling is that the distribution of X must be known. If the presumed probability model deviates from the true model, one has to accept the risk of wrong decisions, especially for small values of p. On the other hand its strength lies in the reduction of the sample size compared to the result of an attributive sampling plan, where one uses only the information, that an observation is smaller or larger than u. This leads to criterions following a binomial or a hypergeometric law.

Therefore, many results tackle this problem under various distributional assumptions. Eearlier works on this topic are Owen (1969) and Takagi (1972). The latter used normal approximation of $\bar{X} + kS$ to derive sampling plans for the Gamma, Weibull and Burr distribution. Suresh and Ramanathan (1997) gave similiar results for a class of symmetric distributions. Aminzadeh (1996) thoroughly discussed acceptance sampling based on the inverse gaussian distribution.

However, in all these references distributional assumptions come into play. In order to avoid misspecified distributional models we propose the bootstrap as nonparametric method to estimate sampling plans in a data driven manner. One of the obvious advantages of the bootstrap approach is that it is not necessary to specify completely a distribution for X, and that it enables also the possibility to use any other suitable criterion statistic. Of course, to make the boostrap work, we need a sample, which is already available, e.g. from any secondary data source or from a pilot study. If data consists of heterogeneous chunks, problems of how to combine them into a usable data set may emerge. This article does not tackle this problem, but assumes the existence of a prior homogeneous data set.

The bootstrap as a tool to determine the sample size has been discussed by Efron and Tibshirani (1993; p. 374) in the context of proving bioequivalence, and by Collings and Hamilton (1988) or Hamilton and Collings (1991), who used the bootstrap to estimate the power of the two-sample Wilcoxon test.

This paper is organized as follows. First we describe a real data problem that motivates the use of the bootstrap in this context. Then we explain our methodology and propose various bootstrap techniques to construct estimates for the sampling plan. Those can be assessed by means of simple Monte Carlo approximations to its bootstrap bias and its bootstrap standard error. After comparing the performances of the different resampling strategies in a simulation study, we finally come back to the example and apply the different approaches to the data.

HCB Study

The organization of Styrian pumpkin seed growers in Austria conducted a toxicity study, where they monitored the Hexachlorobenzene (HCB) concentration (in μg/kg) in pumpkin seed. Because of the well known carcinogenic activity of this pesticide, it is strictly forbidden to use HCB as a plant protective agent. The aim of this study was to get a picture of the extent of the hazardous HCB load in the soil/plant system. A sample of size $m = 647$ was taken and inspected; only 10 out of them exceed the limit. For subsequent years this organization wishes to implement an reliable quality control system that guaranties its products to meet the quality standard defined by legal regulations, i.e. that the HCB concentration may not at all exceed the very strict upper limit of $u = 200\mu$g/kg.

We were asked to design a suitable variable acceptance sampling plan allowing to reject a lot of pumpkin seed on the basis of as small samples as possible. To calculate the necessary sample size one has to evaluate the OC–function of a testing procedure for varying values of n, where the distribution function F_p is considered at two prespecified values $p = p_0$ and $p = p_1$. The task is to find the minimal sample size fullfilling the conditions imposed on the OC–function for both of these values.

One of the most interesting characteristics of the prior sample is that the empirical distribution function shows up an extremely skewed shape and a very long right tail. Standard parametric models, e.g. an exponential, Weibull or lognormal model, etc. do not fit this data well. The empirical quantile plot in Figure 1 shows this clearly in the case when we assume an underlying exponential distribution. Hence, this a nonparametric approach seems to be advisable, at least to check parametric results.

***** Figure 1 about here *****

Bootstrap Methods for Estimating and Assessing Sampling Plans

A Family of Alternatives

Here we introduce a suitable family of distribution functions, whose members are determined by the fraction defective p. This family should reflect the stochastic mechanism generating the data for alternative quality levels.

Denote the quantile function of F by $q(p, F)$. Let H be an operator that maps any distribution function F and a probability p to a distribution function $F_p = H(p, F)$ with the property that the p-quantile equals $q(p, F) = u$. Because $H(p, F)$ generates distribution functions with fraction defective equal to p, we call it generating operator. In the sequel $H(p, F)$ for fixed F and $p \in (0, 1)$ will serve as a parametric model for X. Notice that beyond this the generating operator H is left completely unspecified. The choice of H relies on the problem at hand. The adequacy of this approach depends on whether $F_p = H(p, F)$ suitably describes the alternative distribution of X with fraction defective p.

We are going to discuss two types of generating operators, producing location and scale alternatives. Hence, we generally assume that X stems from a location-scale family with parameter (ν, λ), i.e. $X \sim F_{(\nu,\lambda)}$ where $F_{(\nu,\lambda)}(x) = F\left((x - \nu)/\lambda\right)$. Here F denotes an arbitrary distribution function reflecting mean μ and variance σ^2. Let

$$p(\nu, \lambda) = P(X > u | F_{(\nu,\lambda)}). \tag{1}$$

Now we have to put restrictions on either ν or λ to ensure that (1) can be solved for the remaining free parameter. Two possible choices are considered in the following:

(a) $\lambda = \lambda_0 > 0$ and $\nu \in \mathbb{R}$,

implying that $E(X) = \lambda_0 \mu + \nu$ and that $Var(X) = \lambda_0^2 \sigma^2$ is a constant.

(b) $\nu = \nu_0 = 0$ and $\lambda \in \mathbb{R}^+$,

implying that $E(X) = \lambda \mu$ and that $Var(X) = \lambda^2 \sigma^2$, or in other words the standard deviation of X equals the constant σ/μ times the mean $E(X)$.

In both cases we can solve for the free parameter, giving

$$p = p(\nu, \lambda) = \begin{cases} 1 - F\left((u - \nu)/\lambda_0\right), & \nu = u - \lambda_0 F^{-1}(1-p) \quad \text{for (a)} \\ 1 - F\left(u/\lambda\right), & \lambda = u/F^{-1}(1-p) \quad \text{for (b).} \end{cases} \tag{2}$$

Note that p only depends on ν in the first case whereas in the second case it is a function of λ. Result (2) motivates the use of the following generating operator

$$H(p, F) = \begin{cases} F((x - \nu)/\lambda_0), & \nu = u - \lambda_0 q(1-p, F) \quad \text{for (a)} \\ F(x/\lambda), & \lambda = u/q(1-p, F) \quad \text{for (b).} \end{cases} \tag{3}$$

This operator reflects a reasonable family of distribution functions parameterized by the fraction p. Given p, it is now easy to calculate the corresponding values of ν or λ, thereby determining the member of $\{H(p, F) | p \in (0, 1)\}$ with fraction defective p. Both, $H(p_0, F)$ and $H(p_1, F)$, will later be used to estimate the sample size.

Bootstrap Estimates

For $p \in (0, 1)$ we define the OC-function $L(F_p, n, k) = P(T_n \geq k | F_p)$, with the criterion statistic $T_n = T(X_1, \ldots, X_n; u)$. This can be any suitable statistic, e.g. $(u - \tilde{X})/IQR$, $\tilde{X}$ being the sample median and IQR its interquartile range, instead of $(u - \bar{X})/S$. Allowing for specified type I and type II errors α and $(1 - \beta)$, respectively, we have to determine $s_0 = (n_0, k_0)$ such that

$$\begin{aligned} L(F_{p_0}, n_0, k_0) &= P\left(T_{n_0} \geq k_0 | F_{p_0}\right) \geq 1 - \alpha \\ L(F_{p_1}, n_0, k_0) &= P\left(T_{n_0} \geq k_0 | F_{p_1}\right) \leq 1 - \beta, \end{aligned} \tag{4}$$

and n_0 is minimal. We write $s_0 = s_0(F_{p_0}, F_{p_1}, \alpha, \beta)$ in order to point at the dependence of (n_0, k_0) on F_{p_0} and F_{p_1} additionally to α and β.

The crucial point in formula (4) is that the probabilty calculations have to be done with respect to the distribution functions F_{p_0} and F_{p_1}, both usually being unknown. The idea of the bootstrap is to replace F_{p_j}, $j \in \{0, 1\}$, by proper estimates $\hat{F}_{p_j}$ and to do the probability calculations with respect to these. The remaining question is concerned with how to choose $\hat{F}_{p_j}$ given the prior sample $Y_1, \ldots, Y_m$ from F. This distribution function will be estimated by, say, $\hat{F}$. Various proposals for the definition of $\hat{F}$ will be resumed later (see Subsection 3.3).

Define $\hat{q}_j = q(1 - p_j, \hat{F})$, the $(1 - p_j)$-quantile of $\hat{F}$. Depending on the problem at hand the alternative may be given by location (a) or scale (b) variants of $\hat{F}$. Replacing F by $\hat{F}$ in (3) results in estimates

$$\hat{F}_{p_j}(x) = \begin{cases} \hat{F}((x - u)/\lambda_0 + \hat{q}_j) & \text{for (a)} \\ \hat{F}(x\hat{q}_j/u) & \text{for (b).} \end{cases} \tag{5}$$

In the sequel we assume that $\hat{F}$ correctly reflects the variability of X. Hence, we choose $\lambda_0 = 1$ for the location alternative. However, other choices would be also possible, e.g. whenever the variability of the prior sample is biased.

The main principle of the nonparametric bootstrap technique is to estimate distributional properties of a functional, that depends on the sample and its distribution function, by the respective properties of its boostrap version, depending on a resample from any nonparametric estimate of this distribution function (see Efron (1979) for details). We now apply this idea to the sampling plan. Because s_0 is defined in terms of F_{p_0} and F_{p_1}, its bootstrap version has to refer to $\hat{F}_{p_0}$ and $\hat{F}_{p_1}$ given in (5). Therefore, $\hat{s}_0 = s_0(\hat{F}_{p_0}, \hat{F}_{p_1}, \alpha, \beta)$ and solves a bootstrap analogue to the restrictions (4). The criterion statistic T_n has to be replaced there by $T_n^* = T(X_1^*, \ldots, X_n^*; u)$, where $X_1^*, \ldots, X_n^*$ is an iid resample from $\hat{F}_{p_0}$ or from $\hat{F}_{p_1}$, respectively, i.e. the bootstrap translation of (4) reads

$$
\begin{aligned}
L(\hat{F}_{p_0}, \hat{n}_0, \hat{k}_0) &= P\left(T^*_{\hat{n}_0} \geq \hat{k}_0 | \hat{F}_{p_0}\right) \geq 1-\alpha \\
L(\hat{F}_{p_1}, \hat{n}_0, \hat{k}_0) &= P\left(T^*_{\hat{n}_0} \geq \hat{k}_0 | \hat{F}_{p_1}\right) \leq 1-\beta,
\end{aligned}
\tag{6}
$$

again with minimal $\hat{n}_0$.

We solve for the conditions in (6) by applying a Monte Carlo approximation. The following algorithm **MC–BSP** determines the bootstrap sampling plan:

1. For fixed n draw random samples $X^{*b} = X_1^{*b}, \ldots, X_n^{*b} \overset{iid}{\sim} \hat{F}_{p_0}$, where $b = 1, \ldots, B$, and calculate $T_n^{*b} = T(X^{*b}; u)$. Compute $k = q(1-\alpha, \hat{F}_{T^*})$, where $\hat{F}_{T^*}$ is the empirical distribution function of the bootstrap replicates T_n^{*b}.

2. Now draw other random samples $X^{*b} = X_1^{*b}, \ldots, X_n^{*b} \overset{iid}{\sim} \hat{F}_{p_1}$, $b = 1, \ldots, B$, and calculate $T_n^{*b} = T(X^{*b}; u)$ with these resamples. Approximate $L(\hat{F}_{p_1}, n, k)$ by its Monte Carlo value $L^* = \sharp\{T_n^{*b} \geq k\}/B$.

3. If $L^* > 1-\beta$, then increment n and go to the first step. Otherwise set $\hat{n}_0 = n$, $\hat{k}_0 = k$, and report these values as the estimated sampling plan $\hat{s}_0$.

Various simple ideas can be used to make this algorithm somewhat more efficient, e.g. search strategies for $\hat{n}_0$. We also like to mention that we have not distinguished between a notation for the exact bootstrap estimator $\hat{s}_0$ and for its Monte Carlo approximation, where the latter results from the above algorithm.

Additionally, we utilize the bootstrap technique again to construct estimates for $bias\left(s_0(\hat{F}_{p_0}, \hat{F}_{p_1}, \alpha, \beta) | F_{p_0}, F_{p_1}\right)$ and for $var\left(s_0(\hat{F}_{p_0}, \hat{F}_{p_1}, \alpha, \beta) | F_{p_0}, F_{p_1}\right)$. The respective bootstrap estimates are their plug-in versions

$$
bias\left(s_0(\hat{F}^*_{p_0}, \hat{F}^*_{p_1}, \alpha, \beta) | \hat{F}_{p_0}, \hat{F}_{p_1}\right), \qquad var\left(s_0(\hat{F}^*_{p_0}, \hat{F}^*_{p_1}, \alpha, \beta) | \hat{F}_{p_0}, \hat{F}_{p_1}\right), \tag{7}
$$

where $\hat{F}^*_{p_j}$, $j \in \{0,1\}$, denotes the estimated distribution function of an iid sample of size m drawn from $\hat{F}_{p_j}$. More accuratelly, denote the estimated distribution function of $Y_1^*, \ldots, Y_m^* \overset{iid}{\sim} \hat{F}$ by $\hat{F}^*$ and evaluate (5) for $\hat{F} = \hat{F}^*$, i.e.

$$\hat{F}^*_{p_j}(x) = \begin{cases} \hat{F}^*((x-u)/\lambda_0 + \hat{q}_j^*) & \text{for (a)} \\ \hat{F}^*(x\hat{q}_j^*/u) & \text{for (b),} \end{cases}$$

again with $\lambda_0 = 1$ and $\hat{q}_j^* = q(1-p_j, \hat{F}^*)$.

However, since the theoretical bootstrap estimates (7) cannot be evaluated analytically, their Monte Carlo approximations are again helpful alternatives. The respective algorithm relies on **MC-BSP** and can be decribed as follows:

1. Draw a random sample $Y^{*r} = Y_1^{*r}, \ldots, Y_m^{*r} \overset{iid}{\sim} \hat{F}$, $r = 1, \ldots, R$.

2. Apply **MC-BSP** to $\hat{F}^{*r}_{p_j}$ resulting in estimated sampling plans $\hat{s}_0^{*r} = (\hat{n}_0^{*r}, \hat{k}_0^{*r})$.

3. Obtain the Monte Carlo approximations to (7) by

$$\widehat{bias}\,(\hat{s}_0^*) = \frac{1}{R}\sum_{r=1}^{R} \hat{s}_0^{*r} - \hat{s}_0, \tag{8}$$

$$\widehat{Var}\,(\hat{s}_0^*) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{s}_0^{*r} - \frac{1}{R}\sum_{r=1}^{R}\hat{s}_0^{*r}\right)^T\left(\hat{s}_0^{*r} - \frac{1}{R}\sum_{r=1}^{R}\hat{s}_0^{*r}\right), \tag{9}$$

   respectively.

Remember that $\hat{s}_0$ in (8) denotes the sample based bootstrap estimate of the sampling plan, i.e. the result of **MC-BSP** applied to the original prior data $Y_1, \ldots, Y_m$.

Obviously, this algorithm can be very time consuming. One has to draw R prior data resamples of size m and to apply **MC-BSP** to each of them. The latter requires drawing B resamples of size n iteratively until the specifications on the OC-function are met.

Estimating F

In the following we describe different variants of estimators $\hat{F}$ for the distribution function F of the prior data. Given this estimate, we calculate $\hat{F}_{p_j}$, $j \in \{0,1\}$, by applying formula (5). Furthermore, we will characterize $\hat{F}$ in terms of random variables, which will help us to conveniently simulate from them.

Empirical Distribution Function (NP): The simplest approach is to estimate F by the empirical distribution function of the prior data, i.e.

$$\hat{F}(x) = \hat{F}_{NP}(x) = \frac{1}{m}\sum_{i=1}^{m} \chi_{(-\infty,x]}(y_i),$$

where $\chi_{(-\infty,x]}(y)$ denotes the indicator function of the intervall $(-\infty,x]$. This is the straightforward application of the nonparametric bootstrap technique. The bootstrap variable may be defined by y_{I^*}, where I^* is a discrete random variable uniformly distributed over the indices $\{1,\ldots,m\}$. Bootstrap resamples from $\hat{F}_{p_j}$ are simply realized through $y_{I*} + u - q(1-p_j, \hat{F}_{NP})$. A subsequent application of **MC-BSP** results in the estimate $\hat{s}_0$.

Linear Interpolation Method (LI1, LI2): We describe two slightly different variants here. The first one (LI1) has been used by Collings and Hamilton (1988). Let $y_{(1)},\ldots,y_{(m)}$ be the ordered sample and define $y_{(0)} = 2y_{(1)} - y_{(2)}$ and $y_{(m+1)} = 2y_{(m)} - y_{(m-1)}$. The estimate $\hat{F}_{LI1}(x)$ is given by assigning mass $1/(m+1)$ uniformly to each intervall (y_i, y_{i+1}), $i = 0,\ldots,m$. In terms of random variables this may be stated as follows. Define $d_i = y_{(i+1)} - y_{(i)}$ and let $y_{I^*} + U \cdot d_{I^*}$ be the bootstrap variable, where I^* is now uniform on $\{0,\ldots,m\}$ and U is uniform on $(0,1)$.

The second variant, denoted by $\hat{F}_{LI2}(x)$, distributes the resamples uniformly around the observations y_i. Let $c_i = (y_{(i+1)} - y_{(i)})/2$, $i = 1,\ldots,m-1$. Furthermore set $c_0 = c_1$ and $c_m = c_{m-1}$. Now assign mass $1/m$ uniformely to each intervall $(y_{(i)} - c_{i-1}, y_{(i)} + c_i)$, $i = 1,\ldots,m$. The random mechanism with this distribution can be described by $y_{I^*}+U$, where I^* and U are uniformly distributed on $\{1,\ldots,m\}$ and on $(-c_{I^*-1},c_{I^*})$, respectively.

Smoothing the Distribution Function (SB1, SB2): Both variants, which are known as (shrunk) smooth bootstrap, are described in Davison and Hinkley (1997; p. 79) and in Efron and Tibshirani (1993; p. 231). The first uses the estimate $\hat{F}_{SB1}(x)$, defined as the distribution function to the density estimate

$$\hat{f}_h(x) = \frac{1}{mh}\sum_{i=1}^{m}\phi\left(\frac{x-y_i}{h}\right),$$

where the gaussian kernel $\phi(\cdot)$ is defined as the density function of a standard normal distribution. Since $y_{I^*}+h\epsilon$, where I^* is uniform on $\{1,\ldots,m\}$ and $\epsilon\sim N(0,1)$, has distribution function $\hat{F}_{SB1}(x)$, drawing resamples is quite simple. Different choices of h are used in the example.

Note that the boostrap variable has mean $\bar{y}$ but variance $s_y^2+h^2$, i.e. larger than the sample variance s_y^2 of the prior data. A modification to the above estimate remedies this drawback. Let $\hat{F}_{SB2}(x)$ denote the distribution function to the density estimate

$$\hat{f}_h(x) = \frac{1}{mhb}\sum_{i=1}^{m}\phi\left(\frac{x-a-by_i}{hb}\right),$$

where $a=(1-b)\bar{y}$ and $b=\left(1+h^2/s_y^2\right)^{-1/2}$. Again, we consider $y'_{I^*}+h'\epsilon$, with transformed observations $y'_i=a+by_i$ and smoothing parameter $h'=hb$. Obviously this variable has mean $\bar{y}$ and variance $b^2(s_y^2+h^2)=s_y^2$, thus correctly reflecting the observed first two sample moments. This resampling strategy is known as the shrunk smooth bootstrap.

Simulation Study

We evaluate the bootstrap approaches in a situation, where exact answers can also be given. This enables us to show that the resampling results at least for this case are trustworthy. We assume that $X \sim N(0,1)$, and that $\bar{X} + kS \leq u$, with $u = 1.64$, serves as the criterion for acceptance and consider three constellations C_1, C_2, C_3 of conditions on the OC-function. Table 1 contains those constellations along with the exact values of (n_0, k_0) and summarizes the simulation results. We generate $V = 200$ gaussian prior samples $y_1^v, \ldots, y_m^v$, $v = 1, \ldots, V$, of size m. Bootstrap estimates $\hat{s}_0^v = (\hat{n}_0^v, \hat{k}_0^v)$ are calculated for each setting. We report the Monte Carlo mean $\widehat{E}(\hat{s}_0) = V^{-1} \sum_v \hat{s}_0^v$ and standard errors $\widehat{se}(\hat{s}_0) = \mathrm{diag}(\widehat{var}(\hat{s}_0))^{1/2}$, where $\widehat{var}(\hat{s}_0) = (V-1)^{-1} \sum_v (\hat{s}_0^v - \widehat{E}(\hat{s}_0))(\hat{s}_0^v - \widehat{E}(\hat{s}_0))^T$. Actually, we did this for various sample sizes but report only the results for $m = 40$ and $m = 120$. The bandwidth h for both smooth bootstraps was determined by the adaptive estimate $\hat{h} = 1.06 \min(S, IQR/1.34) m^{-1/5}$ proposed by Silverman (1986; p. 47) for normal populations.

***** Table 1 about here *****

Before going into detail we like to summarize our global impressions. Since constellation C_2 equals C_3 up to a slightly different distance $p_1 - p_0$, all considered methods behave in a similar way but show up smaller standard errors the more p_1 is away from p_0. Setting C_1 describes the smallest difference between the fraction defective but allows for larger type I and type II errors. As expected, bias and standard errors get smaller the larger we chose the sample size m. For small m, $\hat{n}_0$ overestimates n_0 under C_1, whereas it is often too small under C_2 or C_3. In contrary, $\hat{k}_0$ seems to be almost always a bit too large in that case. Figures 2 and 3 show the boxplots of $\hat{s}_0^v$. Each panel contains the results under C_1, C_2, and C_3 for both sample sizes, e.g. the first two boxplots represent the estimates $\hat{n}_0^v$ (or $\hat{k}_0^v$) under C_1 for $m = 40$ and $m = 120$, etc. Clearly, they all are very skew for $\hat{n}_0$ (see Figure 2), sometimes showing extremely large values. This is mildest for NP and similar for $LI2$ and $SB2$, being rather ostentatious for $LI1$ and $SB1$. Figure 3 displays the respective boxplots for $\hat{k}_0$. In contrary to those for n_0, they show up a quite symmetric behaviour for all methods.

***** Figure 2 and 3 about here *****

Surprisingly, the NP method seems to be the best in terms of variability. In all combinations of settings and sample sizes this straightforward bootstrap estimate is among the best ones. In terms of bias there is evidence that method $SB2$ yields the most reliable estimates, whereas NP as well as $LI2$, both seems to underestimate the sample sizes for small m, especially under C_2 or C_3. These three methods give useful results, whereas $LI1$ and $SB1$ have a tendency to be more variable.

For constellation C_1 the methods NP, $LI2$, and $SB2$ lead to similiar estimates, whereas $LI1$ as well as $SB1$ tend to overestimate n_0 and have large standard errors.

In contrary, under C_2 or C_3 the methods NP and $LI2$ tend to underestimate n_0 for $m = 40$, which seems to be remedied for larger samples. Their standard errors of $\hat{n}_0$ are amongst the smallest, but k_0 is slightly overestimated by both. $LI1$ shows a slight upward bias for n_0 if m increases; but it always hits k_0. $SB1$ tends to slightly overestimate n_0 and has high variability, whereas $SB2$ is seemingly downward biased for n_0, but overestimates k_0.

The methods $LI1$ and $SB1$ tend to be too variable. For $LI1$ this results from distributing the resampled historical data beyond the minimum or the maximum uniformly over a span determined by the distance from $y_{(1)}$ to $y_{(2)}$ and $y_{(m)}$ to $y_{(m-1)}$, respectively. These distances are quite variable and the range of the resamples is artificially widened if they are extremely large. For $LI2$ this effect is somewhat milder, because it is based on only half the above distances. Also for $SB1$ the resamples are spread out to much. This drawback is remedied by $SB2$, which correctly reflects the sample variability by shrinking the smoothing parameter.

HCB Study Continued

As indicated above we like to control the probability p that the HCB load exceeds the tolerable concentration 200μg/kg, i.e. the fraction of non-conforming quality in pumpkin seed. Two questions arise here, (1) which family of distributions one should use to describe different quality levels and (2) what acceptance criterion to deploy. The first one can be answered either on subject matter grounds by an expert or by inspecting different historical samples coming from sources reflecting different quality levels. For the HCB study (and for concentration data in general), where data is nonnegative and frequently skewed, it seems to be reasonable to use a scale family as indicated by the second part of the generating operator (3). Under different quality levels the distribution function of the concentrations have the same support, namely all nonnegative values, but are more or less spread out. Application of a shift in location will result in an unsuitable data range, excluding either part of the positive or including part of the negative real axes. Domain specialists affirmed this approach being in accordance with their expertise.

The second and more technical problem is the choice of a powerful criterion. First we examined $\bar{X} + kS \leq u$. But this criterion led to estimated sampling plans, where inacceptable large sample sizes $\hat{n}_0$ were required to fullfill the restrictions imposed on the OC-function. This was due to the high right tail of the (empirical) distribution of the HCB load in the historical data set. Since this property is characteristic for the HCB concentration in pumpkin seed, this could not be handled by elimination of the extreme values. Because this estimated plan would also be too costly, we decided to design a more powerful sampling plan. It turns out that the application of the robust criterion $\tilde{X} + kIQR \leq u$ leads to estimates that are acceptable in terms of organizational effort and costs.

Table 2 summarizes the results for an upper limit of $u = 200\mu$g/kg and for two different restrictions D_1 and D_2 on the OC-function. The first constellation is less stringent whereas the second is supposed to meet acceptable quality standards. The Monte Carlo results are based on drawing $R = 300$ random samples from the historical data. For each of those we applied the algorithm **MC-BSP** with $B = 2500$.

***** Table 2 about here *****

Comparable $\hat{n}_0$ result from the methods NP, $LI1$, and $LI2$. Their bias corrected estimates advice to use a sample of size 62, 64 and 69 for D_1 or 75, 75 and 79 for D_2, respectively. The bandwidth h for the smoothing versions $SB1$ and $SB2$ was determined by unbiased cross-validation as described in Scott (1992; p. 160). Both of these approaches tend to give smaller estimates for n_0 (55, 52 for D_1, or 61, 60 for D_2). The estimated standard errors of $\hat{n}_0$ are larger for the nonparametric bootstrap results in this case than for those estimate based on smoothing or interpolation methods. Since $m = 647$ is rather large we may trust in the answer of the NP method, which is also theoretically well supported (see Beran (1986)).

Summary

We demonstrated that the bootstrap can be successfully adopted in the context of variable acceptance sampling if some prior data is already available. In our simulation study we compared the performance of the nonparametric bootstrap to some more sophisticated smoothed versions. Motivated by these results we like to recommend the nonparametric bootstrap for general use. Moreover, it is theoretically much better understood than the others (see e.g. Bickel and Friedman (1981) for a general discussion). The application of smoothing techniques, either by interpolation or smooth density estimation, is more complicated and it seems that there is no substantial gain in terms of bias or variance of the resulting sampling plan estimator.

The methodology applied to a real data example has proven to give reliable and useful results. In contrast to Hamilton and Collings (1991) who proposed bootstrap estimates for n_0 in the special case of the two-sample Wilcoxon test, we also had a focus on assessing the bias and the variability of these estimates. As they observed by simulation, sample size determination is far from being an exact science. However, since we applied Monte Carlo methods to assess the resampling estimates, we were able to correct them for bias and to check their standard errors.

The obvious power of the bootstrap does not only lie in the weak assumptions it makes, but also in the ease of extending this approach to a general criterion statistics by utilizing Monte Carlo techniques. This opens a wide field of applications in acceptance sampling methods, e.g. the selection of a powerful criterion out of a set of competing alternatives by bootstrapping their OC-functions. In our example the robust quantile criterion is more powerful than the respective moment based one and leads to much smaller estimates of n_0.

Generally, the bootstrap approach provides a measure of uncertainty, which should be taken into account. We like to stress the fact that the weakness of this approach is indeed its strength, instead of pretending illegitimate accuracy it quantifies the statistical error that is likely to occur here.

References

[1] Aminzadeh, M.S. (1996). "Inverse–Gaussian Acceptance Sampling Plans by Variables". *Communications in Statististics—Theory and Methods* 25, pp. 923–935.

[2] Beran, R. (1986). "Simulated Power Functions". *Annals of Statistics* 14, pp. 151–173.

[3] Bickel, P.J. and Freedman, D.A. (1981). "Some Asymptotic Theory for the Bootstrap". *Annals of Statistics* 9, pp. 1196–1217.

[4] Collings, B.J. and Hamilton, M.A. (1988). "Estimating the Power of the Two-Sample Wilcoxon Test for Location Shift". *Biometrics* 44, pp. 847–860.

[5] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Applications.* Cambridge University Press, Cambridge.

[6] Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife". *Annals of Statistics* 7, pp. 1–26.

[7] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, London.

[8] Hamilton, M.A. and Collings, B.J. (1991). "Determining the Appropriate Sample Size for Nonparametric Tests for Location Shift". *Technometrics* 33, pp. 327–337.

[9] MONTGOMERY, D.C. (1996). *Introduction to Statistical Quality Control*, 3rd edition, John Wiley & Sons, New York, NY.

[10] OWEN, D.B. (1969). "Summary of Recent Work on Variables Acceptance Sampling with Emphasis on Non–Normality". *Technometrics* 11, pp. 631–637.

[11] SCOTT, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, New York, NY.

[12] SILVERMAN, B.W. (1986). *Density Estimation.* Chapman & Hall, New York.

[13] SURESH, R.P. AND RAMANATHAN, T.V. (1997). "Acceptance Sampling Plans by Variables for a Class of Symmetric Distributions". *Communications in Statistics—Simulation and Computation* 26, pp. 1379–1391.

[14] TAKAGI, K. (1972). "On Designing Unknown–Sigma Sampling Plans Based on a Wide Class of Non–Normal Distributions". *Technometrics* 14, pp. 669–678.

Key Words: *Location–Scale Family, OC–Function, Monte Carlo Methods, Nonparametric Designs, Resampling.*

Figure 1: Empirical quantile plot of the observed HCB concentrations in pumpkin seed versus exponential distribution with rate 1.

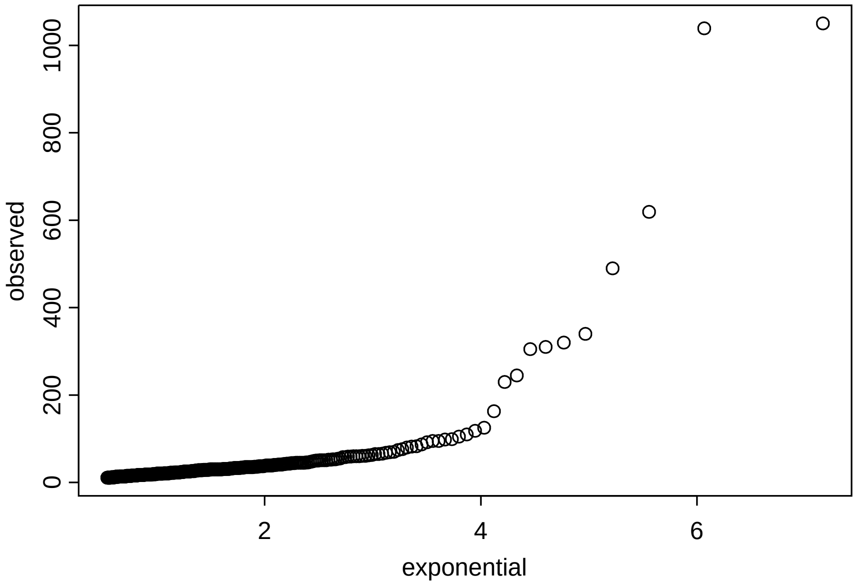

Table 1: Simulation results based on a normal population.

		$m = 40$				$m = 120$			
		$\widehat{E}(\hat{n}_0)$	$\widehat{se}(\hat{n}_0)$	$\widehat{E}(\hat{k}_0)$	$\widehat{se}(\hat{k}_0)$	$\widehat{E}(\hat{n}_0)$	$\widehat{se}(\hat{n}_0)$	$\widehat{E}(\hat{k}_0)$	$\widehat{se}(\hat{k}_0)$
C_1	NP	58.36	37.89	1.43	0.13	57.46	27.99	1.38	0.08
	LI1	69.34	52.75	1.34	0.14	63.60	33.10	1.33	0.08
	LI2	59.88	40.25	1.42	0.14	58.06	28.60	1.37	0.08
	SB1	70.37	49.06	1.29	0.12	63.55	31.70	1.28	0.07
	SB2	61.27	41.02	1.43	0.13	58.26	29.01	1.37	0.08
C_2	NP	41.90	27.61	1.71	0.18	44.98	15.93	1.59	0.10
	LI1	45.80	27.78	1.58	0.14	49.39	18.88	1.54	0.10
	LI2	42.19	28.41	1.69	0.17	45.55	16.26	1.59	0.10
	SB1	50.37	40.50	1.54	0.17	49.40	18.44	1.48	0.10
	SB2	44.79	34.28	1.70	0.18	45.50	16.87	1.59	0.10
C_3	NP	32.38	18.05	1.66	0.17	34.96	10.59	1.55	0.10
	LI1	35.62	18.13	1.54	0.13	38.27	12.56	1.49	0.10
	LI2	32.71	17.92	1.65	0.16	35.38	10.57	1.54	0.10
	SB1	38.20	25.60	1.49	0.16	38.42	12.26	1.44	0.09
	SB2	34.57	22.41	1.65	0.17	35.54	11.04	1.55	0.10

Restrictions $(p_0, p_1, \alpha, 1-\beta)$ on the OC-function, and resulting exact values (n_0, k_0):

$C_1 : (0.05, 0.15, 0.05, 0.05)$, $(56, 1.34)$

$C_2 : (0.025, 0.15, 0.025, 0.01)$, $(47, 1.54)$

$C_3 : (0.025, 0.175, 0.025, 0.01)$, $(37, 1.50)$

Figure 2: Boxplots of the sample size estimates. From the upper left panel, by row: NP, $LI1$, $LI2$, $SB1$, $SB2$. Within panel: C_l, $l = 1, 2, 3$, with $m = 40$ and $m = 120$.

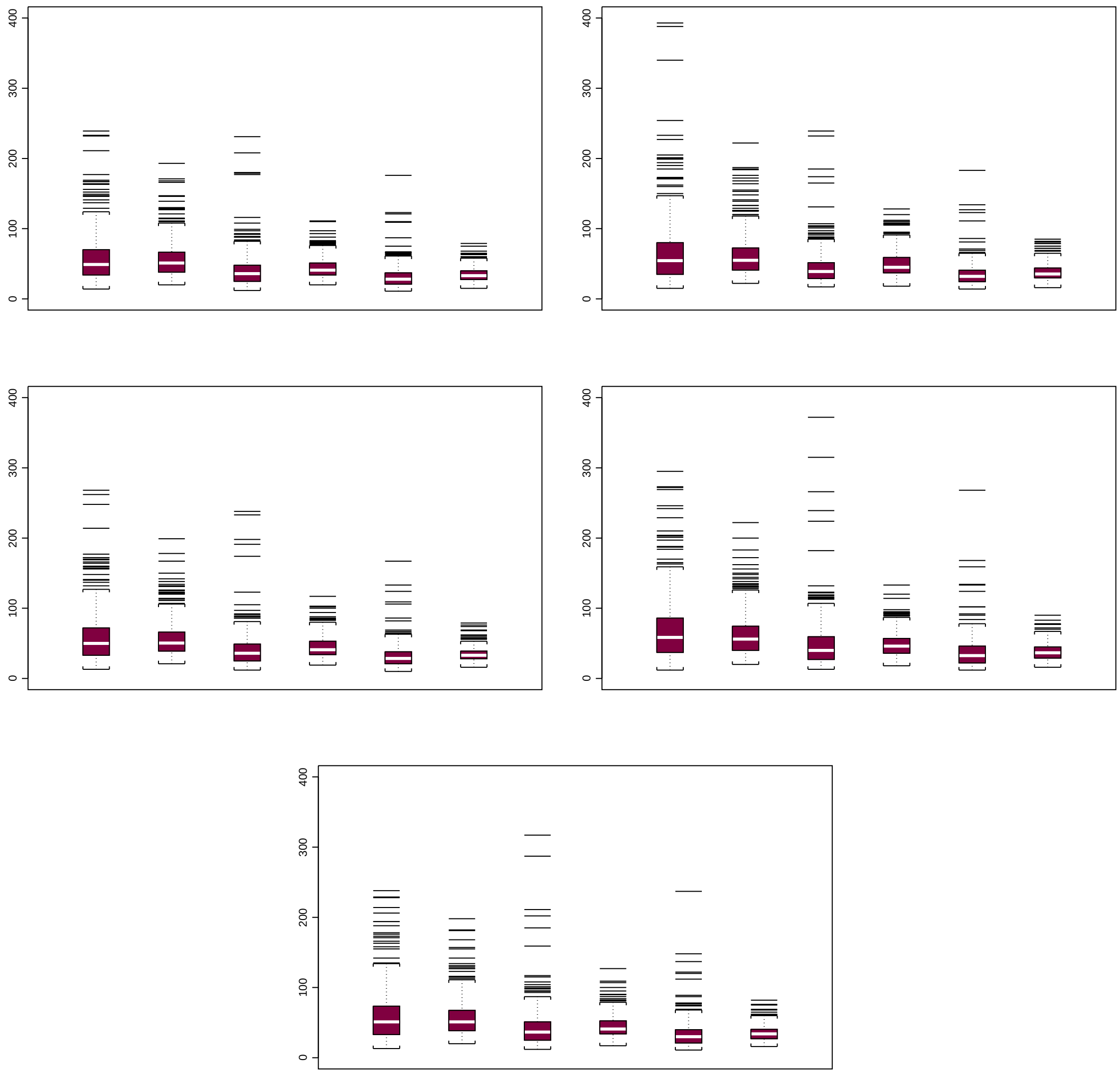

Figure 3: Boxplots of the estimates for k_0. From the upper left panel, by row: NP, $LI1$, $LI2$, $SB1$, $SB2$. Within panel: C_l, $l = 1, 2, 3$, with $m = 40$ and $m = 120$.

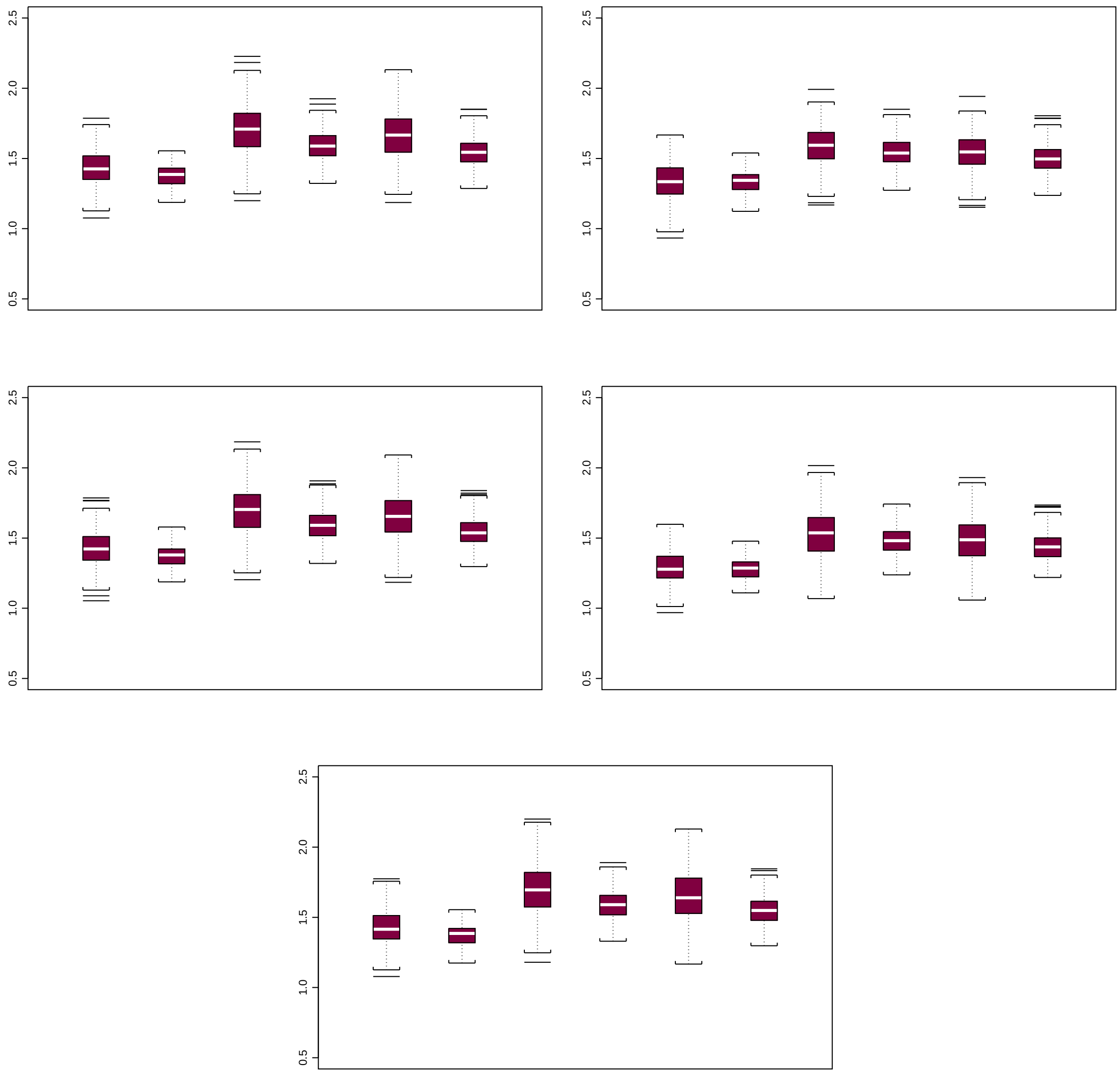

Table 2: Estimated acceptance sampling plans, their Monte Carlo bias and standard errors, for the HCB study.

		$\hat{n}_0$	$\widehat{bias}(\hat{n}_0^*)$	$\widehat{se}(\hat{n}_0^*)$	$\hat{k}_0$	$\widehat{bias}(\hat{k}_0^*)$	$\widehat{se}(\hat{k}_0^*)$
D_1	NP	68	6.1	26.30	1.67	-0.02	0.14
	LI1	70	1.2	20.50	1.66	0.04	0.15
	LI2	68	4.0	24.10	1.65	0.03	0.14
	SB1	55	-0.9	17.30	1.66	0.06	0.13
	SB2	54	2.3	15.10	1.85	-0.07	0.14
D_2	NP	79	4.3	40.70	2.64	0.07	0.40
	LI1	80	5.2	32.60	2.66	0.02	0.32
	LI2	82	3.6	33.00	2.66	0.03	0.35
	SB1	66	5.2	32.10	2.62	0.01	0.30
	SB2	63	3.8	24.00	2.93	-0.15	0.33

Restrictions $(p_0, p_1, \alpha, 1-\beta)$ on the OC-function:

$D_1 : (0.05, 0.15, 0.05, 0.05)$

$D_2 : (0.025, 0.10, 0.01, 0.01)$