Graduate Lectures and Problems in Quality Control and Engineering Statistics: Theory and Methods

To Accompany

Statistical Quality Assurance Methods for Engineers

by

Vardeman and Jobe

Stephen B. Vardeman

V2.0: January 2001

© Stephen Vardeman 2001. Permission to copy for educational purposes granted by the author, subject to the requirement that this title page be affixed to each copy (full or partial) produced.

Chapter 5

Sampling Inspection

Chapter 8 of V&J treats the subject of sampling inspection, introducing the basic methods of acceptance sampling and continuous inspection. This chapter extends that discussion somewhat. We consider how (in the fraction nonconforming context) one can move from single sampling plans to quite general acceptance sampling plans, we provide a brief discussion of the effects of inspection/measurement error on the real (as opposed to nominal) statistical properties of acceptance sampling plans, and then the chapter closes with an elaboration of §8.5 of V&J, providing some more details on the matter of economic arguments in the choice of sampling inspection schemes.

5.1 More on Fraction Nonconforming Acceptance Sampling

Section 8.1 of V&J (and for that matter §8.2 as well) confines itself to the discussion of single sampling plans. For those plans, a sample size is fixed in advance at some value n, and lot disposal is decided on the basis of inspection of exactly n items. There are, however, often good reasons to consider acceptance sampling plans whose ultimate sample size depends upon "how the inspected items look" as they are examined. (One might, for example, want to consider a "double sampling" plan that inspects an initial small sample, terminating sampling if items look especially good or especially bad so that appropriate lot disposal seems clear, but takes an additional larger sample if the initial one looks "inconclusive" regarding the likely quality of the lot.) This section considers fraction nonconforming acceptance sampling from the most general perspective possible and develops the OC, ASN, AOQ and ATI for a general fraction nonconforming plan.

Consider the possibility of inspecting one item at a time from a lot of N, and after inspecting each successive item deciding to 1) stop sampling and accept

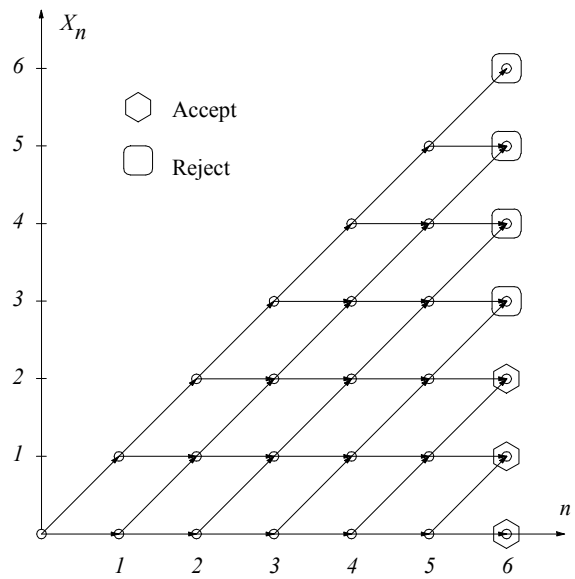

Figure 5.1: Diagram for the $n = 6$, $c = 2$ Single Sampling Plan

the lot, 2) stop sampling and reject the lot or 3) inspect another item. With

$$X_n = \text{the number of nonconforming items found among the first } n \text{ inspected}$$

a helpful way of thinking about various different plans in this context is in terms of possible paths through a grid of ordered pairs of integers (n, X_n) with $0 \leq X_n \leq n$. Different acceptance sampling plans then amount to different choices of "Accept Boundary" and "Reject Boundary." Figure 5.1 is a diagram representing a single sampling plan with $n = 6$ and $c = 2$, Figure 5.2 is a diagram representing a "doubly curtailed" version of this plan (one that recognizes that there is no need to continue inspection after lot disposal has been determined) and Figure 5.3 illustrates a double sampling plan in these terms.

Now on a diagram like those in the figures, one may very quickly count the number of permissible paths from $(0, 0)$ to a point in the grid by (working left to right) marking each point (n, X_n) in the grid (that it is possible to reach) with the *sum* of the numbers of paths reaching $(n-1, X_n - 1)$ and $(n-1, X_n)$ provided neither of those points is a "stop-sampling point." (No feasible paths leave a stop-sampling point. So path counts to them do not contribute to path counts for any points to their right.) Figure 5.4 is a version of Figure 5.2 with permissible movements through the (n, X_n) grid marked by arrows, and path counts indicated.

The reason that one cares about the path counts is that for any stop-sampling

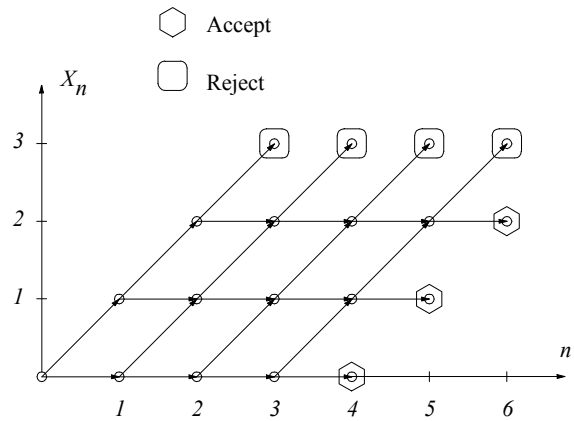

Figure 5.2: Diagram for Doubly Curtailed $n = 6$, $c = 2$ Single Sampling Plan

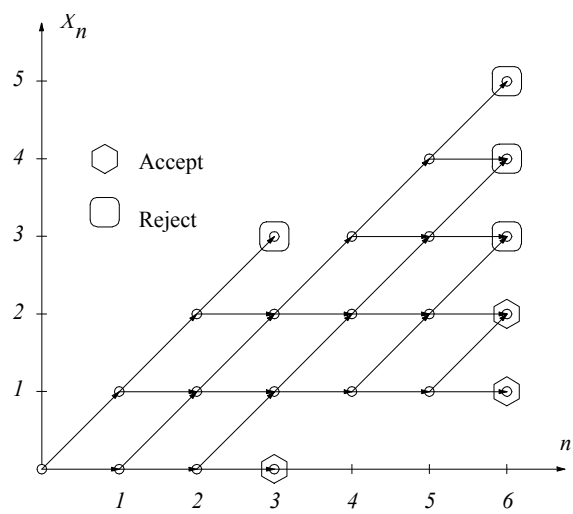

Figure 5.3: Diagram for a Small Double Sampling Plan

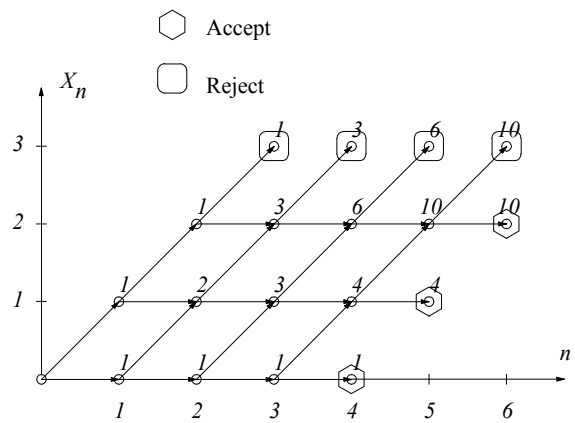

Figure 5.4: Diagram for the Doubly Curtailed Single Sampling Plan with Path Counts Indicated

point (n, X_n), from perspective A

$$P[\text{reaching } (n, X_n)] = (\text{path count from (0,0) to } (n, X_n)) \frac{\binom{N-n}{Np-X_n}}{\binom{N}{Np}} ,$$

while from perspective B

$$P[\text{reaching } (n, X_n)] = (\text{path count from (0,0) to } (n, X_n))\, p^{X_n}(1-p)^{n-X_n} .$$

And these probabilities of reaching the various stop sampling points are the fundamental building blocks of the standard statistical characterizations of an acceptance sampling plan.

For example, with A and R respectively the acceptance and rejection boundaries, the OC for an arbitrary fraction nonconforming plan is

$$Pa = \sum_{(n,X_n)\in \mathrm{A}} P[\text{reaching } (n, X_n)] . \tag{5.1}$$

And the mean number of items sampled (the Average Sample Number) is

$$ASN = \sum_{(n,X_n)\in \mathrm{A}\cup\mathrm{R}} nP[\text{reaching } (n, X_n)] . \tag{5.2}$$

Further, under the rectifying inspection scenario, from perspective B

$$AOQ = \sum_{(n,X_n)\in \mathrm{A}} (1-\frac{n}{N})pP[\text{reaching } (n, X_n)] , \tag{5.3}$$

from perspective A

$$AOQ = \sum_{(n,X_n)\in \mathrm{A}} (p-\frac{X_n}{N})P[\text{reaching } (n, X_n)] \tag{5.4}$$

and

$$ATI = N(1 - Pa) + \sum_{(n,X_n)\in \mathrm{A}} nP[\text{reaching } (n, X_n)] \ . \tag{5.5}$$

These formulas are conceptually very simple and quite universal. The fact that specializing them to any particular choice of acceptance boundary and rejection boundary might have been unpleasant when computations had to be done "by hand" is largely irrelevant in today's world of plentiful fast and cheap computing. These simple formulas and a personal computer make completely obsolete the many many pages of specialized formulas that at one time filled books on acceptance sampling.

Two other matters of interest remain to be raised regarding this general approach to fraction nonconforming acceptance sampling. The first concerns the difficult mathematical question "What are good shapes for the accept and reject boundaries?" We will talk a bit in the final section of this chapter about criteria upon which various plans might be compared and allude to how one might try to find a "best" plan ("best" shapes for the acceptance and rejection boundaries) according to such criteria. But at this point, we wish only to note that Abraham Wald working in the 1940s on the problem of sequential testing, developed some approximate theory that suggests that parallel straight line boundaries (the acceptance boundary below the rejection boundary) have some attractive properties. He was even able to provide some approximate two-point design criteria. That is, in order to produce a plan whose OC curve runs approximately through the points (p_1, Pa_1) and (p_2, Pa_2) (for $p_1 < p_2$ and $Pa_1 > Pa_2$) Wald suggested linear stop-sampling boundaries with

$$slope = \frac{\ln\left(\frac{1-p_1}{1-p_2}\right)}{\ln\left(\frac{p_2(1-p_1)}{p_1(1-p_2)}\right)} \ . \tag{5.6}$$

An appropriate X_n-intercept for the acceptance boundary is approximately

$$h_{\mathrm{A}} = \frac{\ln\left(\frac{Pa_1}{Pa_2}\right)}{\ln\left(\frac{p_2(1-p_1)}{p_1(1-p_2)}\right)} \ , \tag{5.7}$$

while an appropriate X_n-intercept for the rejection boundary is approximately

$$h_{\mathrm{R}} = \frac{\ln\left(\frac{1-Pa_2}{1-Pa_1}\right)}{\ln\left(\frac{p_2(1-p_1)}{p_1(1-p_2)}\right)} \ . \tag{5.8}$$

Wald actually derived formulas (5.6) through (5.8) under "infinite lot size" assumptions (that also allowed him to produce some approximations for both the OC and ASN of his plans). Where one is thinking of applying Wald's boundaries in acceptance sampling of a real (finite N) lot, the question of exactly how to truncate the sampling (close in the right side of the "continue sampling region")

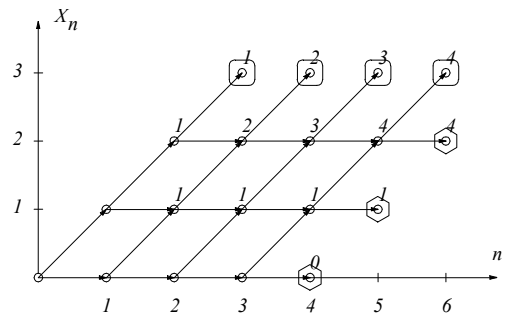

Figure 5.5: Path Counts from $(1,1)$ to Stop Sampling Points for the Plan of Figure 5.4

must be answered in some sensible fashion. And once that is done, the basic formulas (5.1) through (5.5) are of course relevant to describing the resulting plan. (See Problem 5.4 for an example of this kind of logic in action.)

Finally, it is an interesting side-light here (that can come into play if one wishes to estimate p based on data from something other than a single sampling plan) that provided the stop-sampling boundary has *exactly* one more point in it than the largest possible value of n, the uniformly minimum variance unbiased estimator of p for both type A and type B contexts is (for (n, X_n) a stop-sampling point)

$$\widehat{p}\,((n, X_n)) = \frac{\text{path count from (1,1) to } (n, X_n)}{\text{path count from (0,0) to } (n, X_n)} \ .$$

For example, Figure 5.5 shows the path counts from (1,1) needed (in conjunction with the path counts indicated in Figure 5.4) to find the uniformly minimum variance unbiased estimator of p when the doubly curtailed single sampling plan of Figure 5.4 is used.

Table 5.1 lists the values of $\widehat{p}$ for the 7 points in the stop-sampling boundary for the doubly curtailed single sampling plan with $n = 6$ and $c = 2$, along with the corresponding values of X_n/n (the maximum likelihood estimator of p).

5.2 Imperfect Inspection and Acceptance Sampling

The nominal statistical properties of sampling inspection procedures are "perfect inspection" properties. The OC formulas for the attributes plans in §8.1 and §8.4 of V&J and §5.1 above are really premised on the ability to tell with certainty whether an inspected item is conforming or nonconforming. And the OC formulas for the variables plans in §8.2 of V&J are premised on an assumption that the measurement x that determines whether an item is conforming or

Table 5.1: The UMVUE and MLE of p for the Doubly Curtailed Single Sampling Plan

Stop-sampling point (n, X_n)	UMVUE, $\widehat{p}$	MLE, X_n/n
$(3,3)$	$1/1$	$3/3$
$(4,0)$	$0/1$	$0/4$
$(4,3)$	$2/3$	$3/4$
$(5,1)$	$1/4$	$1/5$
$(5,3)$	$3/6$	$3/5$
$(6,2)$	$4/10$	$2/6$
$(6,3)$	$4/10$	$3/6$

Table 5.2: Perspective B Description of a Single Inspection Allowing fo Inspection Error

		Inspection Result G	Inspection Result D	
Actual	G	$(1-w_{\mathrm{G}})(1-p)$	$w_{\mathrm{G}}(1-p)$	$1-p$
Condition	D	pw_{D}	$p(1-w_{\mathrm{D}})$	p
		$1-p^*$	p^*	

nonconforming can be obtained for a given item *completely without measurement error*. But the truth is that real-world inspection is not perfect and the nominal statistical properties of these methods at best approximate their actual properties. The purpose of this section is to investigate (first in the attributes context and then in the variables context) just how far actual OC values for common acceptance sampling plans can be from nominal ones.

Consider first the percent defective context and suppose that when a conforming (good) item is inspected, there is a probability w_{G} of misclassifying it as nonconforming. Similarly, suppose that when a nonconforming (defective) item is inspected, there is a probability w_{D} of misclassifying it as conforming. Then from perspective B, a probabilistic description of any single inspected item is given in Table 5.2, where in that table we are using the abbreviation

$$p^* = w_{\mathrm{G}}(1-p) + p(1-w_{\mathrm{D}})$$

for the probability that an item (of unspecified actual condition) is classified as nonconforming by the inspection process.

It should thus be obvious that from perspective B in the fraction nonconforming context, an attributes single sampling plan with sample size n and acceptance number c has an actual acceptance probability that depends not only on p but on w_{G} and w_{D} as well through the formula

$$Pa(p, w_{\mathrm{G}}, w_{\mathrm{D}}) = \sum_{x=0}^{c} \binom{n}{x} (p^*)^x (1-p^*)^{n-x} \quad . \tag{5.9}$$

On the other hand, the perspective A version of the fraction nonconforming scenario yields the following. For an integer x from 0 to n, let U_x and V_x be

independent random variables,

$$U_x \sim \text{Binomial}\,(x, 1 - w_{\text{D}}) \;\text{ and }\; V_x \sim \text{Binomial}\,(n - x, w_{\text{G}})\ .$$

And let

$$r_x = P[U_x + V_x \leq c]$$

be the probability that a sample containing x nonconforming items actually passes the lot acceptance criterion. (Note that the nonstandard distribution of $U_x + V_x$ can be generated using the same "adding on diagonals of a table of joint probabilities" idea used in §1.7.1 to generate the distribution of $\overline{x}$.) Then it is evident that from perspective A an attributes single sampling plan with sample size n and acceptance number c has an actual acceptance probability

$$Pa(p, w_{\text{G}}, w_{\text{D}}) = \sum_{x=0}^{n} \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}}\, r_x\ . \tag{5.10}$$

It is clear that nonzero w_{G} or w_{D} change nominal OC's given in displays (8.6) and (8.5) of V&J into the possibly more realistic versions given respectively by equations (5.9) and (5.10) here. In some cases, it may be possible to determine w_{G} and w_{D} experimentally and therefore derive both nominal and "real" OC curves for a fraction nonconforming single sampling plan. Or, if one were *a priori* willing to guarantee that $0 \leq w_{\text{G}} \leq a$ and that $0 \leq w_{\text{D}} \leq b$, it is pretty clear that from perspective B one might then at least guarantee that

$$Pa(p, a, 0) \leq Pa(p, w_{\text{G}}, w_{\text{D}}) \leq Pa(p, 0, b) \tag{5.11}$$

and have an "OC band" in which the real OC (that depends upon the unknown inspection efficacy) is guaranteed to lie.

Similar analyses can be done for nonconformities per unit contexts as follows. Suppose that during inspection of product, real nonconformities are missed with probability m and that (independent of the occurrence and inspection of real nonconformities) "phantom" nonconformities are "observed" according to a Poisson process with rate λ_{P} per unit inspected. Then from perspective B in a nonconformities per unit context, the number of nonconformities observed on k units is Poisson with mean

$$k(\lambda(1 - m) + \lambda_{\text{P}})\ ,$$

so that an actual acceptance probability corresponding to the nominal one given in display (8.8) of V&J is

$$Pa(\lambda, \lambda_{\text{P}}, m) = \sum_{x=0}^{c} \frac{\exp\left(-k(\lambda(1-m) + \lambda_{\text{P}})\right)\left(k(\lambda(1-m) + \lambda_{\text{P}})\right)^x}{x!}\ . \tag{5.12}$$

And from perspective A, with a realized per unit defect rate λ on N units, let $U_{\lambda,m} \sim \text{Binomial}\,(k\lambda, \left(\frac{k}{N}\right)(1 - m))$ be independent of $V_{\lambda_{\text{P}}} \sim \text{Poisson}\,(k\lambda_{\text{P}})$.

Then an actual acceptance probability corresponding to the nominal one given in display (8.7) of V&J is

$$Pa(\lambda, \lambda_{\mathrm{P}}, m) = P[U_{\lambda,m} + V_{\lambda_{\mathrm{P}}} \leq c] \ . \tag{5.13}$$

And the same kinds of bounding ideas used above for the fraction nonconforming context might be used with the OC (5.12) in the mean nonconformities per unit context. Pretty clearly, if one could guarantee that $\lambda_{\mathrm{P}} \leq a$ and that $m \leq b$, one would have (from display (5.12))

$$Pa(\lambda, a, 0) \leq Pa(\lambda, \lambda_{\mathrm{P}}, m) \leq Pa(\lambda, 0, b) \tag{5.14}$$

in the perspective B situation.

The violence done to the OC notion by the possibility of imperfect inspection in an attributes sampling context is serious, but not completely unmanageable. That is, where one can determine the likelihood of inspection errors experimentally, expressions (5.9), (5.10), (5.12) and (5.13) are simple enough characterizations of real OC's. And where w_{G} and w_{D} (or λ_{P} and m) are small, bounds like (5.11) (or (5.14)) show that both the nominal (the $w_{\mathrm{G}} = 0$ and $w_{\mathrm{D}} = 0$, or $\lambda_{\mathrm{P}} = 0$ and $m = 0$ case) OC and real OC are trapped in a fairly narrow band and can not be too different. Unfortunately, the situation is far less happy in the variables sampling context.

The origin of the difficulty with admitting there is measurement error when it comes to variables acceptance sampling is the fundamental fact that standard variables plans attempt to treat all (μ, σ) pairs with the same value of p equally. And in short, once one admits to the possibility of measurement error clouding the evaluation of the quantity x that must say whether a given item is conforming or nonconforming, *that goal is unattainable.* For any level of measurement error, there are (μ, σ) pairs (with very small σ) for which product variation can so to speak "hide in the measurement noise." So some fairly bizarre real OC properties result for standard plans.

To illustrate, consider the case of "unknown σ" variables acceptance sampling with a lower specification, L and adopt the basic measurement model (2.1) of V&J for what is actually observed when an item with characteristic x is measured. Now the development in §8.2 of V&J deals with a normal (μ, σ) distribution for observations. An important issue is "What observations?" Is it the x's or the y's of the model (2.1)? It must be the x's, for the simple reason that p is defined in terms of μ and σ. These parameters describe what the lot is really like, NOT what it looks like when measured with error. That is, the σ of §8.2 of V&J must be the σ_x of page 19 of V&J. But then the analysis of §8.2 is done essentially supposing that one has at his or her disposal $\bar{x}$ and s_x to use for decision making purposes, *while all that is really available are $\bar{y}$ and s_y*!!! And that turns out to make a huge difference in the real OC properties of the standard method put forth in §8.2.

That is, applying criterion (8.35) of V&J to what can really be observed (namely the noise-corrupted y's) one accepts a lot iff

$$\bar{y} - L \geq k s_y \ . \tag{5.15}$$

And under model (2.1) of V&J, a given set of parameters (μ_x, σ_x) for the x distribution has corresponding fraction nonconforming

$$p(\mu_x, \sigma_x) = \Phi\left(\frac{L-\mu_x}{\sigma_x}\right)$$

and acceptance probability

$$\begin{aligned} Pa(\mu_x, \sigma_x, \beta, \sigma_{\text{measurement}}) &= P\left[\frac{\bar{y}-L}{s_y} \geq k\right] \\ &= P\left(\frac{\frac{\bar{y}-\mu_y}{\sigma_y/\sqrt{n}} - \frac{L-\mu_y}{\sigma_y/\sqrt{n}}}{\frac{s_y}{\sigma_y}} \geq k\sqrt{n}\right) \end{aligned}$$

where σ_y is given in display (2.3) of V&J. But then let

$$\Delta = -\frac{L-\mu_y}{\sigma_y/\sqrt{n}} = -\frac{(L-\mu_x)/\sigma_x - \beta/\sigma_x}{\sqrt{1+\frac{\sigma^2_{\text{measurement}}}{\sigma_x^2}}/\sqrt{n}} \ , \tag{5.16}$$

and note that

$$\frac{\bar{y}-\mu_y}{\sigma_y/\sqrt{n}} \sim \text{Normal}\,(0,1)$$

independent of $\frac{s_y}{\sigma_y}$, which has the distribution of $\sqrt{U/(n-1)}$ for U a χ^2_{n-1} random variable. That is, with W a noncentral t random variable with noncentrality parameter Δ given in display (5.16), we have

$$Pa(\mu_x, \sigma_x, \beta, \sigma_{\text{measurement}}) = P[W \geq k\sqrt{n}] \ .$$

And the crux of the matter is that (even if measurement bias, β, is 0) Δ in display (5.16) is not a function of $(L-\mu_x)/\sigma_x$ alone *unless one assumes that* $\sigma_{\text{measurement}}$ *is EXACTLY* 0.

Even with no measurement bias, if $\sigma_{\text{measurement}} \neq 0$ there are (μ_x, σ_x) pairs with

$$\frac{L-\mu_x}{\sigma_x} = z$$

(and therefore $p = \Phi(z)$) and Δ ranging all the way from $-z\sqrt{n}$ to 0. Thus considering $z \leq 0$ and $p \leq .5$ there are corresponding Pa's ranging from

$$P[\text{a } t_{n-1} \text{ random variable } \geq k\sqrt{n}]$$

to

$$P[\text{a non-central } t_{n-1}(-z\sqrt{n}) \text{ random variable } \geq k\sqrt{n}] \ ,$$

(the nominal OC), while considering $z \geq 0$ and $p \geq .5$ there are corresponding Pa's ranging from (the nominal OC)

$$P[\text{a non-central } t_{n-1}(-z\sqrt{n}) \text{ random variable } \geq k\sqrt{n}] \ ,$$

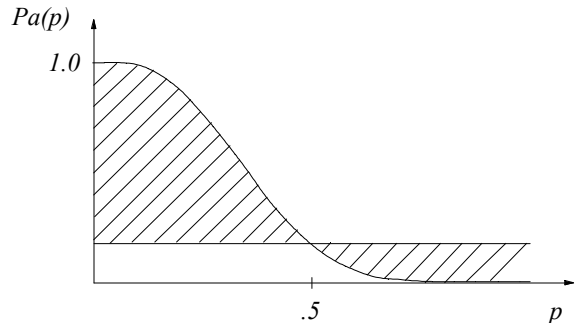

Figure 5.6: Typical Real OC for a One-Sided Variables Acceptance Sampling Plan in the Presence of Nonzero Measurement Error

to

$$P[\text{a } t_{n-1} \text{ random variable } \geq k\sqrt{n}] \ .$$

That is, one is confronted with the extremely unpleasant and (initially counter-intuitive) picture of real OC indicated in Figure 5.6.

It is important to understand the picture painted in Figure 5.6. The situation is worse than in the attributes data case. There, if one knows the efficacy of the inspection methodology it is at least possible to pick a single appropriate OC curve. (The OC "bands" indicated by displays (5.11) and (5.14) are created only by ignorance of inspection efficacy.) The bizarre "OC bands" created in the variables context (and sketched in Figure 5.6) do not reduce to curves if one knows the inspection bias and precision, but rather are intrinsic to the fact that unless $\sigma_{\text{measurement}}$ is exactly 0, different (μ, σ) pairs with the same p must have different Pa's under acceptance criterion (5.15). And the only way that one can replace the situation pictured in Figure 5.6 with one having a thinner and more palatable OC band (something approximating a "curve") is by guaranteeing that

$$\frac{\sigma_x^2}{\sigma^2_{\text{measurement}}}$$

is of some appreciable size. That is, given a particular measurement precision, one must agree to concern oneself only with cases where product variation cannot hide in measurement noise. Such is the only way that one can even come close to the variables sampling goal of treating (μ, σ) pairs with the same p equally.

5.3 Some Details Concerning the Economic Analysis of Sampling Inspection

Section 8.5 of V&J alludes briefly to the possibility of using economic/decision-theoretic arguments in the choice of sampling inspection schemes and cites the 1994 *Technometrics* paper of Vander Wiel and Vardeman. Our first objective

in this section is to provide some additional details of the Vander Wiel and Vardeman analysis. To that end, consider a stable process fraction nonconforming situation and continue the w_{G} and w_{D} notation used above (and also introduced on page 493 of V&J). Note that Table 5.2 remains an appropriate description of the results of a single inspection. We will suppose that inspection costs are accrued on a per item basis and adopt the notation of Table 8.16 of V&J for the costs.

As a vehicle to a very quick demonstration of the famous "all or none" principle, consider facing N potential inspections and employing a "random inspection policy" that inspects each item independently with probability π. Then the mean cost suffered over N items is simply N times that suffered for 1 item. And this is

$$\begin{aligned} \mathrm{E}Cost &= \pi\left(k_{\mathrm{I}} + (1-p)w_{\mathrm{G}}k_{\mathrm{GF}} + p(1-w_{\mathrm{D}})k_{\mathrm{DF}} + pw_{\mathrm{D}}k_{\mathrm{DP}}\right) + (1-\pi)pk_{\mathrm{DU}} \\ &= \pi(k_{\mathrm{I}} + w_{\mathrm{G}}k_{\mathrm{GF}} - pK) + pk_{\mathrm{DU}} \end{aligned} \tag{5.17}$$

for

$$K = (1-w_{\mathrm{D}})(k_{\mathrm{DU}} - k_{\mathrm{DF}}) + w_{\mathrm{D}}(k_{\mathrm{DU}} - k_{\mathrm{DP}}) + w_{\mathrm{G}}k_{\mathrm{GF}}$$

(as in display (8.50) of V&J). Now it is clear from display (5.17) that if $K < 0$, E$Cost$ is minimized over choices of π by the choice $\pi = 0$. On the other hand, if $K > 0$, E$Cost$ is minimized over choices of π

$$\text{by the choice } \pi = 0 \text{ if } p \leq \frac{k_{\mathrm{I}} + w_{\mathrm{G}}k_{\mathrm{GF}}}{K}$$

and

$$\text{by the choice } \pi = 1 \text{ if } p \geq \frac{k_{\mathrm{I}} + w_{\mathrm{G}}k_{\mathrm{GF}}}{K} \ .$$

That is, if one defines

$$p_{\mathrm{c}} = \begin{cases} \infty & \text{if} \quad K \leq 0 \\ \frac{k_{\mathrm{I}} + w_{\mathrm{G}}k_{\mathrm{GF}}}{K} & \text{if} \quad K > 0 \end{cases}$$

then an optimal random inspection policy is clearly

$$\pi = 0 \text{ (do no inspection) if } p < p_{\mathrm{c}}$$

and

$$\pi = 1 \text{ (inspect everything) if } p > p_{\mathrm{c}} \ .$$

This development is simple and completely typical of what one gets from economic analyses of stable process (perspective B) inspection scenarios. Where quality is poor, all items should be inspected, and where it is good none should be inspected. Vander Wiel and Vardeman argue that the specific criterion developed here (and phrased in terms of p_{c}) holds not only as one looks for an optimal *random inspection policy*, but completely generally as one looks *among all possible inspection policies* for one that minimizes expected total cost. But it is essential to remember that the context is a stable process/perspective B

context, where costs are accrued on a per item basis, and in order to implement the optimal policy *one must know* p! In other contexts, the best (minimum expected cost) implementable/realizable policy will often turn out to not be of the "all or none" variety. The remainder of this section will elaborate on this assertion.

For the balance of the section we will consider (Barlow's formulation) of what we'll call the "Deming Inspection Problem" (as Deming's consideration of this problem rekindled interest in these matters and engendered considerable controversy and confusion in the 1980s and early 1990s). That is, we'll consider a lot of N items, assume a cost structure where

$$k_1 = \text{the cost of inspecting one item (at the proposed inspection site)}$$

and

$$k_2 = \text{the cost of later grief caused by a defective item that is not detected}$$

and suppose that inspection is without error. (This is the Vander Wiel and Vardeman cost structure with $k_{\mathrm{I}} = k_1, k_{\mathrm{DF}} = 0$ and $k_{\mathrm{DU}} = k_2$, where both w_{G} and w_{D} are assumed to be 0.) The objective will be optimal (minimum expected cost) choice of a "fixed n inspection plan" (in the language of §8.1 of V&J, a single sampling with rectification plan). That is, we'll consider the optimal choice of n and c supposing that with

$$X = \text{the number nonconforming in a sample of } n \ ,$$

if $X \leq c$ the lot will be "accepted" (all nonconforming items in the sample will be replaced with good ones and no more inspection will be done), while if $X > c$ the lot will be "rejected" (all items in the lot will be inspected and all nonconforming items replaced with good ones). (The implicit assumption here is that replacements for nonconforming items are somehow known to be conforming and are produced "for free.") And we will continue use of the stable process or perspective B model for the generation of the items in the lot.

In this problem, the expected total cost associated with the lot is a function of n, c and p,

$$\begin{aligned} \mathrm{ETC}(n,c,p) &= k_1 n + (1 - Pa(n,c,p))k_1(N-n) + pPa(n,c,p)k_2(N-n) \\ &= k_1 N \left(1 + Pa(n,c,p)\left(1 - \frac{n}{N}\right)\left(p\frac{k_2}{k_1} - 1\right)\right) \ . \end{aligned} \tag{5.18}$$

Optimal choice of n and c requires that one be in the business of comparing the functions of p defined in display (5.18). How one approaches that comparison depends upon what one is willing to input into the decision process in terms of information about p.

First, if p is fixed/known and available for use in choosing n and c, the optimization of criterion (5.18) is completely straightforward. It amounts only to the comparison of numbers (one for each (n,c) pair), not functions. And the

solution is quite simple. In the case that $p > k_1/k_2$, $\left(p\frac{k_2}{k_1} - 1\right) > 0$ and from examination of display (5.18) minimum expected total cost will be achieved if $Pa(n, c, p) = 0$ or if $\left(1 - \frac{n}{N}\right) = 0$. That is, "all" is optimal. In the case that $p < k_1/k_2$, $\left(p\frac{k_2}{k_1} - 1\right) < 0$ and from examination of formula (5.18) minimum expected total cost will be achieved if $Pa(n, c, p) = 1$ and $\left(1 - \frac{n}{N}\right) = 1$. That is, "none" is optimal. This is a manifestation of the general Vander Wiel and Vardeman result. For known p in this kind of problem, sampling/partial inspection makes no sense. One is not going to learn anything about p from the sampling. Simple economics (comparison of p to the critical cost ratio k_1/k_2) determines whether it is best to inspect and rectify, or to "take one's lumps" in later costs.

When one may not assume that p is fixed/known (and it is thus unavailable for use in choosing an optimal (n, c) pair) some other approach has to be taken. One possibility is to describe p with a probability distribution G, average $\mathrm{ETC}(n, c, p)$ over p according to that distribution to get $\mathrm{E}_G\mathrm{ETC}(n, c)$, and then to compare numbers (one for each (n, c) pair) to identify an optimal inspection plan. This makes sense

1. from a Bayesian point of view, where the distribution G reflects one's "prior beliefs" about p, or

2. from a non-Bayesian point of view, where the distribution G is a "process distribution" describing how p is thought to vary lot to lot.

The program SAMPLE (written by Tom Lorenzen and modified slightly by Steve Crowder) available off the Stat 531 Web page will do this averaging and optimization for the case where G is a Beta distribution.

Consider what insights into this "average out according to G" idea can be written down in more or less explicit form. In particular, consider first the problem of choosing a best c for a particular n, say $(c_G^{\mathrm{opt}}(n))$. Note that if a sample of n results in x nonconforming items, the (conditional) expected cost incurred is

$$nk_1 + (N - n)k_2\mathrm{E}_G[p\,|X = x] \text{ with no more inspection}$$

and

$$Nk_1 \text{ if the remainder of the lot is inspected .}$$

(Note that the form of the conditional mean of p given $X = x$ depends upon the distribution G.) So, one should do no more inspection if

$$nk_1 + (N - n)k_2\mathrm{E}_G[p\,|X = x] < Nk_1 \ ,$$

i.e. if

$$\mathrm{E}_G[p\,|X = x] < \frac{k_1}{k_2} \ ,$$

and the remaining items should be inspected if

$$\mathrm{E}_G[p\,|X = x] > \frac{k_1}{k_2}\ .$$

So, an optimal choice of c is

$$c_G^{\mathrm{opt}}(n) = \max\left\{x \mid \mathrm{E}_G[p \mid X = x] \leq \frac{k_1}{k_2}\right\}\ . \tag{5.19}$$

(And it is perhaps comforting to know that the monotone likelihood ratio property of the binomial distribution guarantees that $\mathrm{E}_G[p\,|X = x]$ is monotone in x.)

What is this saying? The assumptions 1) that $p \sim G$ and 2) that conditional on p the variable $X \sim$ Binomial (n, p) together give a joint distribution for p and X. This in turn can be used to produce for each x a conditional distribution of $p|X = x$ and therefore a conditional mean value of p given that $X = x$. The prescription (5.19) says that one should find the largest x for which that conditional mean value of p is still less than the critical cost ratio and use that value for $c_G^{\mathrm{opt}}(n)$. To complete the optimization of $\mathrm{E}_G\mathrm{ETC}(n, c, p)$, one then would then need to compute and compare (for various n) the quantities

$$\mathrm{E}_G\mathrm{ETC}(n, c_G^{\mathrm{opt}}(n), p)\ . \tag{5.20}$$

The fact is that depending upon the nature of G, the minimizer of quantity (5.20) can turn out to be anything from 0 to N. For example, if G puts all its probability on one side or the other of k_1/k_2, then the conditional distributions of p given $X = x$ must concentrate all their probability (and therefore have their means) on that same side of the critical cost ratio. So it follows that if G puts all its probability to the left of k_1/k_2, "none" is optimal (even though one doesn't know p exactly), while if G puts all its probability to the right of k_1/k_2, "all" is optimal in terms of optimizing $\mathrm{E}_G\mathrm{ETC}(n, c, p)$.

On the other hand, consider an unrealistic but instructive situation where $k_1 = 1$, $k_2 = 1000$ and G places probability $\frac{1}{2}$ on the possibility that $p = 0$ and probability $\frac{1}{2}$ on the possibility that $p = 1$. Under this model the lot is either perfectly good or perfectly bad, and *a priori* one thinks these possibilities are equally likely. Here the distribution G places probability on both sides of the breakeven quantity $k_1/k_2 = .001$. Even without actually carrying through the whole mathematical analysis, it should be clear that in this scenario the optimal n is 1! Once one has inspected a single item, he or she knows for sure whether p is 0 or is 1 (and the lot can be rectified in the latter case).

The most common mathematically nontrivial version of this whole analysis of the Deming Inspection Problem is the case where G is a Beta distribution. If G is the Beta(α, β) distribution,

$$\mathrm{E}_G[p\,|X = x] = \frac{\alpha + x}{\alpha + \beta + n}$$

so that $c_G^{\rm opt}(n)$ is the largest value of x such that

$$\frac{\alpha + x}{\alpha + \beta + n} \leq \frac{k_1}{k_2} \ .$$

That is, in this situation, for $\lfloor y \rfloor$ the greatest integer in y,

$$c_G^{\rm opt}(n) = \lfloor \frac{k_1}{k_2}(\alpha + \beta + n) - \alpha \rfloor = \lfloor \frac{k_1}{k_2} n - \alpha + \frac{k_1}{k_2}(\alpha + \beta) \rfloor \ ,$$

which for large n is essentially $\frac{k_1}{k_2} n$. The optimal value of n can then be found by optimizing (over choice of n) the quantity

$$\mathrm{E}_G \left(\mathrm{ETC}(n, c_G^{\rm opt}(n), p) \right) = \int_0^1 \mathrm{ETC}(n, c_G^{\rm opt}(n), p) \frac{1}{B(\alpha, \beta)} p^{\alpha - 1} (1 - p)^{\beta - 1} dp \ .$$

The reader can check that this exercise boils down to the minimization over n of

$$\left(1 - \frac{n}{N}\right) \sum_{x=0}^{c_G^{\rm opt}(n)} \binom{n}{x} \int_0^1 p^x (1 - p)^{n - x} \left(p \frac{k_2}{k_1} - 1 \right) p^{\alpha - 1} (1 - p)^{\beta - 1} dp \ .$$

(The SAMPLE program of Lorenzen alluded to earlier actually uses a different approach than the one discussed here to find optimal plans. That approach is computationally more efficient, but not as illuminating in terms of laying bare the basic structure of the problem as the route taken in this exposition.)

As two final pieces of perspective on this topic of economic analysis of sampling inspection we offer the following. In the first place, while the Deming Inspection Problem is not a terribly general formulation of the topic, the results here are typical of how things turn out. Second, it needs to be remembered that what has been described here is the finding of a cost-optimal *fixed* n inspection plan. The problem of finding a plan optimal among all possible plans (of the type discussed in §5.1) is a more challenging one. For G placing probability on both sides of the critical cost ratio, not only need it not be that case that "all" or "none" is optimal, but in general an optimal plan need not be of the fixed n variety. While in principle the methodology for finding an overall best inspection plan is well-established (involving as it does so called "dynamic programming" or "backwards induction") the details are unpleasant enough that it will not make sense to pursue this matter further.