Isaac N. Gibra

Stevens Institute of Technology

PROBABILITY AND STATISTICAL INFERENCE FOR SCIENTISTS AND ENGINEERS

Prentice-Hall, Inc.

Englewood Cliffs, New Jersey

Englewood Cliffs, N.J.

10 9 8 7 6 5 4 3 2 1

ISBN: 0-13-711622-5
Library of Congress Catalog Card No. 78-107395

Printed in the United States of America

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

cups is the manufacture of small arms cartridge cases. One critical dimension is of paramount significance, namely, the variation of wall thickness around the periphery denoted by X. Suppose management has set the following decision rule: Accept the lot if the expected variation in wall thickness around the periphery [denoted by $E(X)$] is $\leq .03$ mm; reject it otherwise. Obviously the decision management wishes to make is either to *accept* or *reject* the lot. The rule set by management for acceptance of a lot is

$$\text{Accept if} \quad E(X) \leq .03 \text{ mm}$$
$$\text{Reject if} \quad E(X) > .03 \text{ mm}$$

A naive approach would be to measure the variation in wall thickness of the whole lot, record these measurements and calculate their arithmetic mean $\bar{X}$, and then apply the decision rule set by management. Is this a practical approach? Obviously not: If we inspect each cup our inspection cost will be tremendous. Consequently, the cost of a cartridge case will be excessive. In the light of this analysis it is evident that the decision to accept or reject the lot will be based on the *result of an experiment* in which, say, a sample of size n is selected at random from the lot and then each cup is inspected and $\bar{X}$ is calculated. We expect that $\bar{X}$ will be too close to $E(X)$, and hence we are inclined to conclude that $E(X) \leq .03$ if and only if $\bar{X} \leq$ a prescribed constant, which should be $> .03$. How to determine the value of this constant is discussed later.

Actually the decision to accept or reject the lot will be based on the result of a specified experiment. In other words, our decision will be based on *statistical inference. Hence statistical inference can be defined as making inference about the population on the basis of samples.*

Now is this specified experiment the best? Or, in other words, is the choice of a sample of size n at random and observing the sample mean $\bar{X}$ as a criterion for decision the procedure that leads to the optimal decision? If the answer is yes, what is the value of n? If the answer is no, what other alternative procedures might be used in order to reach an optimal decision? An alternative procedure might be to inspect a sample of size n at random and to observe the largest measurement, X_{max}. If $X_{max} \leq$ a prescribed constant, conclude that $E(X)$ is $\leq .03$ and accordingly accept the lot or reject otherwise. Suppose this alternative procedure is better than the previous one. What is the value of n? In general, what is the basis for the selection of an *optimal procedure*? We shall answer these questions as we proceed.

8.3 Statistical Decision Theory

Considering again our introductory example, we have two *statistical hypotheses*. The first hypothesis is $E(X) \leq .03$, and the second hypothesis is $E(X) > .03$. The procedure by which a choice is made between these two statistical hypotheses is called *statistical hypothesis testing.*

The hypothesis that is tested $[E(X) \leq .03]$ is called the *null hypothesis* and is denoted by H_0; the other $[E(X) > .03]$ is called the *alternative hypothesis* and is denoted by H_1. Testing of statistical hypotheses involves *rejection* or *acceptance* of the null hypothesis. In other words, we wish to determine whether the null hypothesis is *true* or *false*. In symbols we write

$$H_0: \quad E(X) \leq .03$$
$$H_1: \quad E(X) > .03$$

The decision to accept or to reject the null hypothesis will be based on the outcome of our experiment. Suppose a sample of size n is drawn at random and the sample mean $\bar{X}$ is calculated. Furthermore, suppose that the following decision rule is specified: Accept the null hypothesis if and only if $\bar{X} \leq .035$ and reject otherwise. Accordingly, we shall reject the null hypothesis if and only if the observed outcome is greater than .035.

According to this decision rule we shall reject the null hypothesis if the value of the sample mean $\bar{X}$ falls in the *critical* region. *The critical region (sometimes known as the rejection region) is specified by the set of values of* $\bar{X}$ *that is greater than* .035. To simplify the analysis let us suppose that this critical dimension is normally distributed with unknown mean $E(X)$ and known standard deviation $\sigma = .006$. Accordingly, $\bar{X}$ is a random variable that is normally distributed with mean $E(X)$ and standard deviation $\sigma/\sqrt{n}$. Let us analyze further the outcomes based on this decision rule. If the mean of the lot under consideration is actually equal to .03, as shown in Fig. 8.1, then

1. The null hypothesis is *accepted* whenever the value of $\bar{X}$ does not fall in the critical region.
2. The null hypothesis is *rejected* whenever the value of $\bar{X}$ falls in the critical region.

If the mean of the lot under consideration is actually equal to .04, as shown in Fig. 8.2, then

3. The null hypothesis is *accepted* whenever the value of $\bar{X}$ does not fall in the critical region.
4. The null hypothesis is *rejected* whenever the value of $\bar{X}$ falls in the critical region.

Table 8.1 summarizes the outcomes based on this decision rule. Now in 1 and 4 we have made correct decisions whereas in 2 and 3 the decisions made are incorrect.

Obviously in case 2 we rejected the lot despite the fact that $E(X) = .03$. Thus we have committed an *error*. This error is known as an *error of the first*

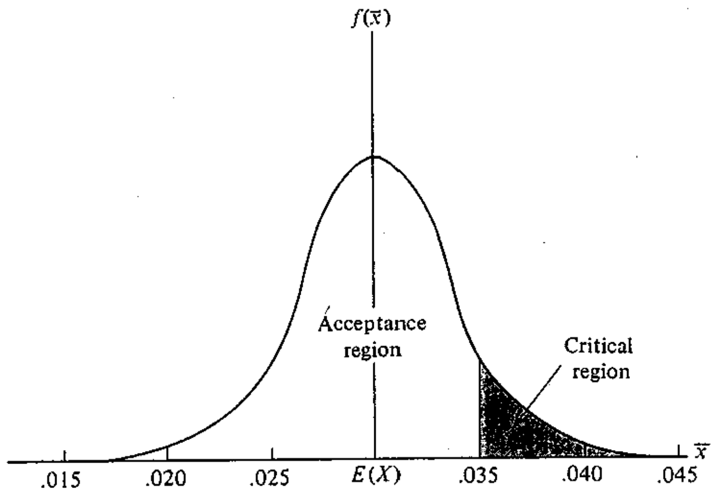

Fig. 8.1 Graphical description of critical and acceptance regions for the given decision rule [$E(X) = .03$].

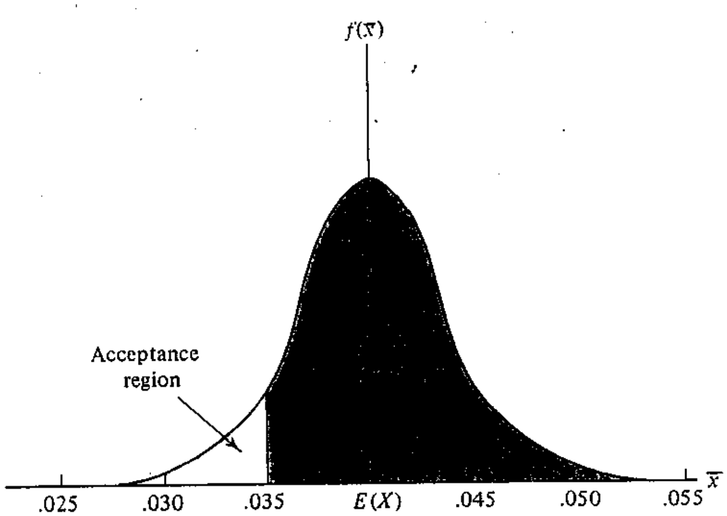

Fig. 8.2 Graphical description of critical and acceptance regions for the given decision rule [$E(X) = .04$].

Null hypothe
Null hypothe

kind or a *type*
an error of the

Let us suppose
Then the error

This means (gi
.0485. α is som
In case 3 w
have committe
as well the prob
by β). Then we

Obviously the e
the value taken

Usually it is
rule by definin
the test is the p
is true. Then

7206

TABLE 8.1

	Accept Null Hypothesis	*Reject Null Hypothesis*
Null hypothesis true	1	2
Null hypothesis false	3	4

kind or a *type I error.* Now let us evaluate the probability of occurrence of an error of the first kind (denoted by α). Then we write

$$\begin{aligned}\alpha &= P\{\bar{X} > .035 \mid E(X) \leq .03\}\\ &\leq P\left\{\frac{\bar{X} - E(X)}{\sigma/\sqrt{n}} > \frac{.035 - .03}{\sigma/\sqrt{n}}\right\}\\ &\leq P\left\{Z > \frac{.005}{.006/\sqrt{n}}\right\}\end{aligned}$$

Let us suppose that $\bar{X}$ is evaluated on the basis of a sample of size $n = 4$. Then the error of the first kind becomes

$$\begin{aligned}\alpha &\leq P\left\{Z > \frac{.010}{.006}\right\}\\ &\leq .0485\end{aligned}$$

This means (given the decision rule) the maximum error of the first kind is .0485. α is sometimes known as the *level of significance.*

In case 3 we accepted the lot despite the fact that $E(X) = .040$. Thus we have committed an *error of the second kind* or a *type II error*. Let us evaluate as well the probability of occurrence of an error of the second kind (denoted by β). Then we write

$$\beta\{E(X)\} = P\{\bar{X} \leq .35 \mid E(X) > .03\}$$

Obviously the error of the second type is not a constant then but depends on the value taken by $E(X)$. If $E(X) = .04$, we have

$$\begin{aligned}\beta\{E(X) = .04\} &= P\left\{\frac{\bar{X} - E(X)}{\sigma/\sqrt{n}} \leq \frac{.035 - .04}{\sigma/\sqrt{n}}\right\}\\ &= P\left\{Z \leq \frac{-.010}{.006}\right\}\\ &= .0485\end{aligned}$$

Usually it is more convenient to study the characteristics of the decision rule by defining a new function π called the *power of the test. The power of the test is the probability of rejecting the null hypothesis when the alternative is true.* Then we write

$$\begin{aligned}\pi\{E(X)\} &= 1 - \beta\{E(X)\}\\ &= P\{\bar{X} > .035 \mid E(X) > .03\}\end{aligned}$$

Table 8.2 shows the values of the power function for each possible value of the parameter $E(X)$ for the given decision rule. The graph of this power function is shown in Fig. 8.3. It can readily be seen from the graph that the chance of accepting a lot having an average variation in wall thickness greater than .03 mm decreases as this average value increases for the given decision rule.

TABLE 8.2

$E(X)$	.030	.033	.036	.039	.042	.045	.048
$\pi\{E(X)\}$	.0485	.2527	.6305	.9087	.9902	.9996	.9999

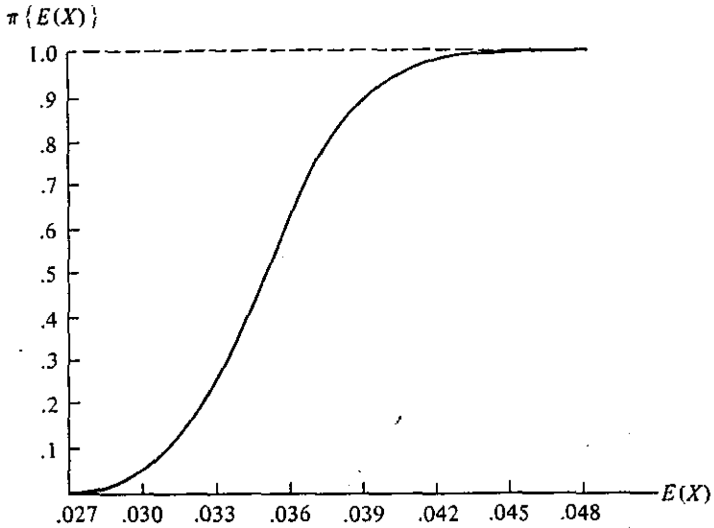

Fig. 8.3 Power curve for the given rule.

From our previous discussion the reader can see that we arbitrarily decided that $\bar{X}$ should be less than or equal to .035 and $n = 4$ in order to test the hypothesis given by management. Usually *this will not* be the appropriate procedure of testing statistical hypotheses. The inappropriateness of this procedure stems from the fact that the rule set by management was too *rigid*. Should management have decided that the *maximum probability of rejecting* a lot having $E(X) \leq .03$ is, say, .02, then we could find the appropriate test procedure in this case. In fact, then we would test the following hypothesis:

$$H_0: E(X) \leq .03 \quad \text{against} \quad H_1: E(X) > .03 \quad \text{given } \alpha_{max} = .02$$

Then we write

$$\alpha_{max} = P\{\bar{X} > k \mid E(X) = .03\}$$

where k is a co
lowing equatio

Thus

Consequently,
what is the sa
size will be tre
of the *sample*
assume that n
on this assum

Hence

Similarly, we
$n = 16$. Valu
in Table 8.3.
8.4.

$E(X)$
$\pi\{E(X)\}$

The risk
increases. Tc
a chance of
when $n = 4$,
100 when n
$= .039$ have

where k is a constant to be determined. The constant k must satisfy the following equation:

$$
\begin{aligned}
.02 &= P\left\{\frac{\bar{X} - E(X)}{\sigma/\sqrt{n}} > \frac{k - .03}{\sigma/\sqrt{n}}\right\} \\
&= P\left\{Z > \frac{k - .03}{\sigma/\sqrt{n}}\right\} \\
&= 1 - \Phi\left(\frac{k - .03}{\sigma/\sqrt{n}}\right)
\end{aligned}
$$

Thus

$$k = .03 + \frac{2.055\sigma}{\sqrt{n}} = .03 + \frac{.01233}{\sqrt{n}}$$

Consequently, reject the null hypothesis if $\bar{X} > .03 + .01233/\sqrt{n}$. Now what is the sample size? The optimal procedure for determining the sample size will be treated in the next chapter; however, we will analyze the effect of the *sample size* on controlling risk of an error of the second kind. Let us assume that n can be either 4, 9, or 16 and evaluate the power function based on this assumption. If $n = 4$, then

$$\beta\{E(X)\} = P\left\{\bar{X} \leq .03 + \frac{.01233}{2} \mid E(X) > .03\right\}$$

Hence

$$\pi\{E(X)\} = 1 - P\{\bar{X} \leq .03616 \mid E(X) > .03\}$$

Similarly, we can evaluate the value of the power function for $n = 9$ and $n = 16$. Values of the power function for these three sample sizes are given in Table 8.3. The graphs for the three power functions are shown in Fig. 8.4.

TABLE 8.3

$E(X)$		.030	.033	.036	.039	.042	.045	.048
$\pi\{E(X)\}$	$n = 4$	.02	.15	.48	.82	.97	.99	~ 1
	$n = 9$	.02	.29	.82	.99	~ 1		
	$n = 16$	.02	.48	.97	~ 1			

The risk of making an error of the second kind decreases as the sample size increases. To illustrate, if the incoming lots have $E(X) = .039$ there will be a chance of accepting such lots 18 times out of 100 as having $E(X) \leq .030$ when $n = 4$, whereas there will be a chance of accepting such lots once out of 100 when $n = 9$. Referring to Table 8.2, we find that incoming lots with mean $= .039$ have a chance of being accepted 10 times out of 100 when $n = 4$ and

7209

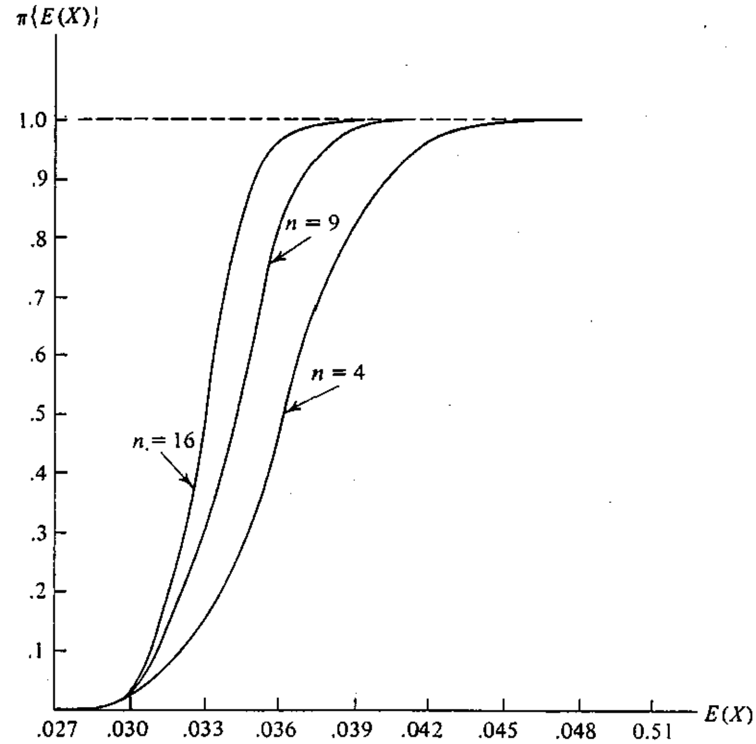

Fig. 8.4 Power functions for selected sample sizes (α = .02).

α_{max} = .0485. This means that *as the error of the first kind increases, the error of the second kind decreases and vice versa.* It follows that by *varying the sample size we can exercise control on the error of the second kind.*

Now if we plot $\beta\{E(X)\}$ against the true average variation in wall thickness for a fixed $\alpha = .02$ and sample sizes 4, 9, and 16, the resulting plot is called an *operating characteristic* (denoted by *OC*) *curve*. The result is represented by the curves in Fig. 8.5.

The level of significance and the sample size uniquely determine the OC curve for the given decision rule. It is evident that by increasing the sample size for a given level of significance the error of the second kind decreases. In practice a balance must be struck between the cost of additional observations and the advantage of decreasing the error of the second kind. In many situations it is not feasible to assess explicitly the cost parameters associated with alternative testing procedures. In the absence of knowledge of these cost parameters the criterion by which we can assess and compare tests of statistical hypothesis is found in the *OC* curves or power functions.

We have discussed so far a decision procedure that associated with it the outcome of random variable $\bar{X}$, sample size n, and acceptance region ($\bar{X} \leq$ a

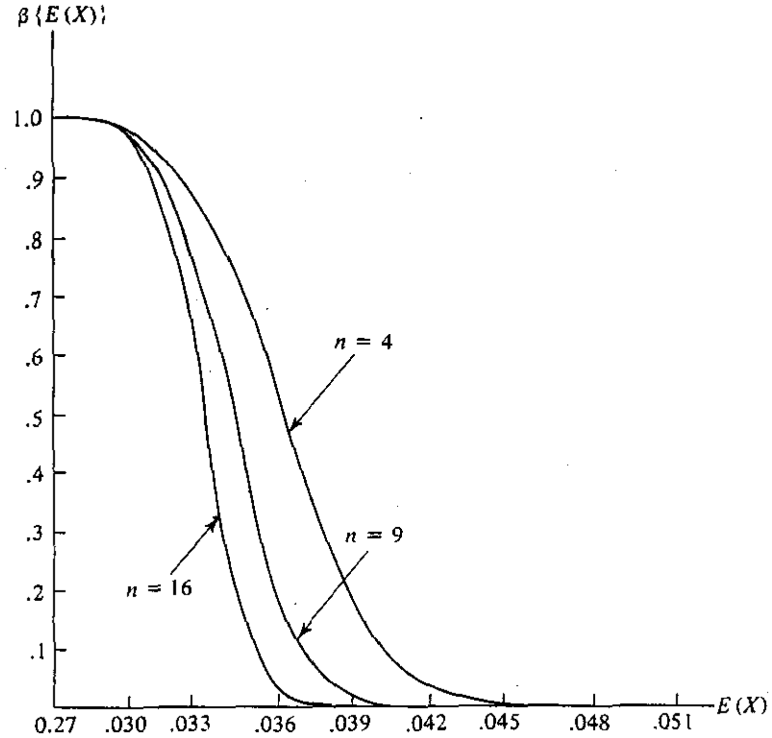

Fig. 8.5 *OC* curves for selected sample sizes ($\alpha = .02$).

prescribed constant). We shall refer to this decision procedure as the $\bar{X}$ procedure. Now we may raise the following question: Is this procedure superior to the X_{max} procedure? Note that the X_{max} procedure is specified by: draw at random a sample of size n; observe the largest measurement, X_{max}; if $X_{max} \leq$ a prescribed constant conclude that $E(X)$ is less than or equal to .03; otherwise conclude that $E(X)$ is greater than .03.

To compare these two procedures ($\bar{X}$ and X_{max}) we will fix the level of significance at .02 and the sample size at 4 for both procedures and use the *OC* curves to provide a criterion of comparison. In other words, we will be comparing both procedures in probabilistic terms.

The alternative decision rule is to accept the null hypothesis if $X_{max} \leq k$, hence

$$P\{X_{max} \leq k\} = 1 - \alpha = .98$$

Since these measurements are independent, identically distributed, normal variates, we write

$$P\{X_1 \leq k\} P\{X_2 \leq k\} P\{X_3 \leq k\} P\{X_4 \leq k\} = .98$$

Hence

$$\int_{-\infty}^{\frac{k-.030}{.006}} \frac{e^{-(1/2)z^2}}{\sqrt{2\pi}}\,dz = (.98)^{1/4} = .995$$

whence $k = .0454$. This means the following: Reject the null hypothesis whenever X_{max} is greater than .0454. The probability of occurrence of the error of the second kind is given by

$$\beta\{E(X)\} = P\{X_{max} \leq .0454 \mid E(X) > .03\}$$

Suppose now that $E(X) = .036$; then

$$\begin{aligned}\beta\{.036\} &= P\{X_{max} \leq .0454 \mid E(X) = .036\} \\ &= \left[\int_{-\infty}^{\frac{.0454-.036}{.006}} \frac{1}{\sqrt{2\pi}} e^{-(1/2)z^2}\,dz\right]^4 \\ &= .78\end{aligned}$$

Similarly, we can calculate the probability of occurrence of the error of the second kind for possible values of the true average $\{E(X)\}$. The result is tabulated in Table 8.4. The *OC* curves for the $\bar{X}$ and X_{max} procedures are as indicated in Fig. 8.6.

TABLE 8.4

$E(X)$	.030	.033	.036	.039	.042	.045	.048
$\beta\{E(X)\}$	.98	.92	.78	.53	.26	.06	.01

Obviously, the procedure based upon $\bar{X}$ gives better protection than the X_{max} procedure for a whole range of alternatives. Thus the $\bar{X}$ procedure is preferred to the X_{max} procedure. Now we proved that the $\bar{X}$ procedure is superior to the X_{max} procedure. Does this imply that the $\bar{X}$ procedure is the *optimal* one? On what basis is a procedure said to be optimal? *The procedure is said to be optimal if the rejection region for a fixed sample size and level of significance minimizes the probability of occurrence of the error of the second kind for a whole range of alternatives.* The optimal procedure is sometimes called a *uniformly most powerful test over a range of alternatives.* There may be, in some situations, more than one optimal procedure. The procedure based upon $\bar{X}$ is in fact the optimal procedure. This fact is proved by the Neyman–Pearson lemma, where justification for the use of the *likelihood ratio* has been established.

We have developed this example in order to introduce some fundamental concepts of statistical inference in decision making. In Chapter 9 we will take up the Neyman–Pearson lemma for testing statistical hypotheses about a single parameter.

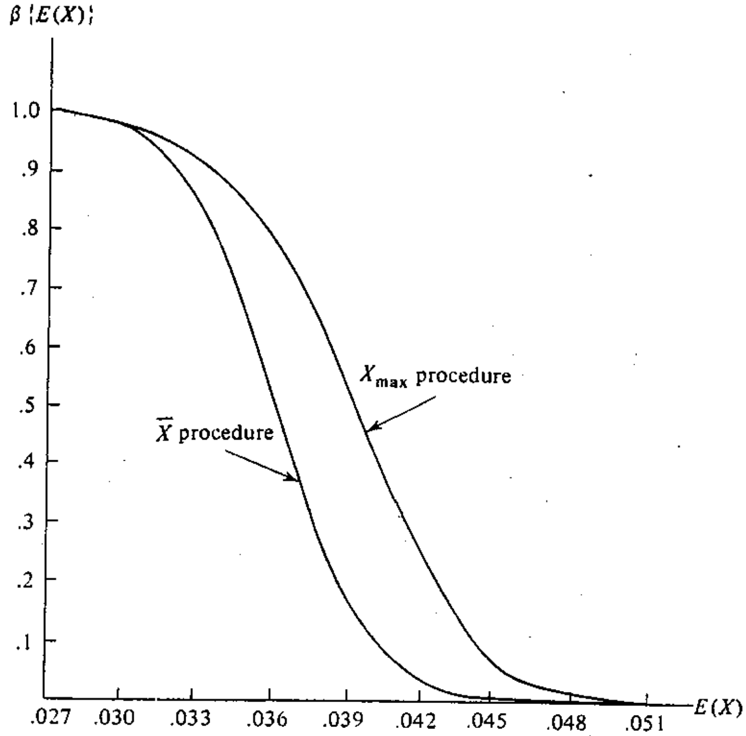

Fig. 8.6 Comparison of the *OC* curves of the $\bar{X}$ and X_{max} procedures for a fixed n and α.

8.4 Kinds of Tests

If a statistical hypothesis specifies the values of all the parameters of the distribution of the random variable under study it is called *simple hypothesis;* otherwise it is called a *composite hypothesis.* Suppose, for example, the random variable under study is normally distributed with mean μ and standard deviation σ. Hence the normal distribution is completely specified by two parameters, μ and σ. Testing if the mean is equal to, say, 100, given that σ is known and equal to, say, 5, the null hypothesis is $H_0 : \mu = 100$. It is called a *simple hypothesis.*

Another illustration is testing if the mean is greater than 100 given that σ is equal to 5. The null hypothesis in this case is given by $H_0 : \mu > 100$ is called a *composite hypothesis.*

Suppose the random variable under study has a Poisson distribution. The Poisson distribution is completely specified by a single parameter λ.

Now if we are interested in testing if $\lambda = 10$, then the null hypothesis H_0: $\lambda = 10$ is a simple hypothesis, whereas $H_0: \lambda < 10$ is a composite hypothesis. The study of testing hypotheses is usually classified in terms of the null hypothesis H_0 and alternate hypothesis H_1. Thus, if the variate studied has a Poisson distrubition, then the hypotheses

$$H_0: \lambda = 10 \qquad H_1: \lambda = 12$$

is a simple against simple, whereas

$$H_0: \lambda = 10 \qquad H_1: \lambda > 10$$

is a simple against composite. Finally

$$H_0: \lambda > 10 \qquad H_1: \lambda < 10$$

is a composite against composite. Simple hypotheses can be resolved, whereas some of the composite hypotheses defy analytical solution.

SUGGESTED REFERENCES

See the references given at the end of Chapter 10.

S

9.1 The L

The *like*
in testing
knowing wh
than 1 ohm
$N(\mu, \sigma_0^2)$, w
jecture, one
of the resist
normal dist
one having
tion having
hypothesis

If the mea
vidual mea
Since th

is discrete, the values of k' that correspond exactly to the usual specified values of α will not always exist. In that event we choose the percentile that corresponds to the level closest to α. The probability of accepting H_0 when H_1 is true is

$$\beta(\lambda_1) = P\{r \geq k' \mid \lambda = \lambda_1\} = \sum_{v=k'}^{\infty} \frac{e^{-\lambda_1 t}(\lambda_1 t)^v}{v!} \tag{9.43}$$

When $\lambda_0 t > 10$ Eqs. (9.42) and (9.43) can be approximated by the normal distribution.

Example 9.7 Suppose that the number of unexcused absences per week in a certain plant follows the Poisson distribution with parameter $\lambda = 7$. Management improved the working conditions for a period of 3 weeks. At the end of this period (3 weeks), the observed number of absences was found to be equal to ten.

(a) Would you infer from this result that management's action has reduced the number of absences? Assume $\alpha = .025$.

(b) What is the power of the test if $\lambda_1 = 5$?

Solution

(a) Here we wish to test the hypothesis

$$H_0: \lambda = \lambda_0 \quad \text{against} \quad H_1: \lambda = \lambda_1 < \lambda_0$$

Hence, from Eq. (9.42) we have

$$\sum_{v=0}^{k'-1} \frac{e^{-21}(21)^v}{v!} = .025$$

Consulting Table I in the Appendix we find that $k' - 1 = 12$, i.e., $k' = 13$. Thus we reject H_0 when the number of observed absences is less than 13. Since the number of observed absences falls in the critical region (0, 12), we conclude that management's action has reduced the number of absences.

(b) The power of the test is

$$\pi(\lambda_1) = 1 - \sum_{v=13}^{\infty} \frac{e^{-15}(15)^v}{v!} = .268$$

The procedure of deriving the optimum rejection region and the power of the test for other alternatives $H_1: \lambda = \lambda_1 > \lambda_0$ and $H_1: \lambda = \lambda_1 \neq \lambda_0$ will be left as an exercise for the reader.

9.7.3 Tests Concerning the Parameter (p) of the Binomial Distribution

In this case the explicit hypothesis to be tested is

$$H_0: p = p_0 \quad \text{against} \quad H_1: p = p_1 < p_0$$

The probability function of the binomial distribution is

$$P_X(r) = \binom{n}{r} p^r (1-p)^{n-r} \qquad r = 0, 1, \ldots, n \qquad 0 \leq p \leq 1 \tag{9.44}$$

The nature of the optimum rejection region, then, is

$$\frac{L_1(r)}{L_0(r)} = \frac{p_1^r(1-p_1)^{n-r}}{p_0^r(1-p_0)^{n-r}} > k \tag{9.45}$$

We note that the likelihood ratio is a monotonically decreasing function for increasing r as long as $p_1 < p_0$. *Hence the optimum rejection region is equivalent to the set of values of r less than some other constant k'.* Accordingly, *we reject H_0 when*

$$r < k'$$

The probability of type I error, then, is

$$P\{r < k' \mid p = p_0\} = \sum_{v=0}^{k'-1} \binom{n}{v} p_0^v(1-p_0)^{n-v} = \alpha \tag{9.46}$$

The value of k' can be obtained from the tables of the cumulative binomial distribution (Table J in the Appendix). It should be noted that the exact value of k' for every α will not always exist because the binomial variate has a discrete distribution. In such cases we choose the percentile that corresponds to the level closest to α.

The probability of accepting H_0 when H_1 is true is

$$\beta(p_1) = P\{r \geq k' \mid p = p_1\} = \sum_{v=k'}^{\infty} \binom{n}{v} p_1^v(1-p_1)^{n-v} \tag{9.47}$$

When n is large, Eqs. (9.46) and (9.47) can be approximated by the normal distribution.

Example 9.8 A manufacturer claims that his product (submitted in large lots) is less than 25% defective. A random sample of size 20 is drawn from a large lot. The number of defective items observed in the sample was one.

(a) Would you substantiate or refute the manufacturer's claim? Use $\alpha = .025$.
(b) Find the probability of acceptance if the submitted lot is 10% defective.
(c) How large a sample is needed to make the answer in (b) equal .10? Use normal approximation.

Solution

(a) From Eq. (9.46) we have

$$\sum_{v=0}^{k'-1} \binom{20}{v}(.25)^v(.75)^{20-v} = .025$$

From Table J we find $k' = 2$. Actually, $k' = 2$ corresponds to the .0243 level of significance, which is close enough to the specified level. Hence, we reject H_0 when the number of defective items in the sample is less than two. Since the number of observed defectives falls into the critical region [0, 1], the null hypothesis can be rejected at the specified level of significance. This supports the manufacturer's claim.

(b)
$$\beta(p_1 = .10) = \sum_{v=2}^{\infty} \binom{20}{v}(.10)^v(.90)^{20-v} = .6083$$

That is, the probability that this test will accept H_0 when actually $p_1 = .10$ is .6083.

(c) From Eqs. (9.46) and (9.47) we have

$$\sum_{v=0}^{k'-1} \binom{n}{v}(.25)^v(.75)^{n-v} = .025$$

$$\sum_{v=k'}^{\infty} \binom{n}{v}(.1)^v(.9)^{n-v} = .10$$

By trying successive values of n we can find from the tables of the cumulative binomial distribution the values of smallest n and k' that satisfy the above equations. However, since the largest sample size given in Table J is 20, we shall use the normal approximation to the binomial. Hence, we have

$$\sum_{v=0}^{k'-1} \binom{n}{v}(.25)^v(.75)^{n-v} \cong \int_{-\infty}^{k'-1} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] dt$$

where $\mu = np_0 = .25n$ and $\sigma = \sqrt{np_0(1-p_0)} = .433\sqrt{n}$. Therefore

$$\Phi\left[\frac{(k' - 1 + \frac{1}{2}) - .25n}{.433\sqrt{n}}\right] \cong .025$$

whence

(9.48) $$k' \cong \tfrac{1}{2} + .25n - (1.96)(.433)\sqrt{n}$$

Similarly,

$$\Phi\left(\frac{.1n - k' + \frac{1}{2}}{.3\sqrt{n}}\right) \cong .10$$

whence

(9.49) $$k' \cong \tfrac{1}{2} + .1n + (.3)(1.28)\sqrt{n}$$

Solving Eqs. (9.48) and (9.49), we obtain

$$n \cong \left[\frac{(1.96)(.433) + (.3)(1.28)}{.15}\right]^2 = 67$$

In general, the required sample size is given by

(9.50) $$n \cong \left(\frac{\sqrt{p_0(1-p_0)}\,Z_{1-\alpha} + \sqrt{p_1(1-p_1)}\,Z_{1-\beta}}{p_1 - p_0}\right)^2$$

Thus, if we take a sample of 67 items we can detect an alternative $p_1 = .10$ with probability 90%.

In a similar way the optimum rejection region and the power of the test can be found for the alternatives $H_1: p = p_1 > p_0$ and $H_1: p = p_1 \neq p_0$.

PROBLEMS

9.1. Compute the error of the second kind if you wish to test the following hypothesis:

$$H_0: \mu = \mu_0 = 10 \quad \text{against} \quad H_1: \mu = \mu_1 = 11$$

at the 5% level of significance. Assume that measurements are normally distributed with $\sigma = 2.5$ and a sample of size 16 is taken.

17

ACCEPTANCE SAMPLING

17.1 Introduction

Most likely the manufacturer who buys his product (parts, subassemblies, material, etc.) in lots of considerable size, from one or more suppliers, desires to know whether the quality characteristic within each lot conforms to his specification. Obviously the manufacturer would like to accept submitted lots if their percent defective does not exceed the specified acceptable quality level. Therefore, each lot must be inspected to determine whether it is acceptable. More precisely, if a lot of size N items is submitted, every item in the lot will be inspected and classified as defective or satisfactory. The lot will be accepted if the number of defective items in the lot is less than or equal to an allowable number; otherwise it will be rejected. When the lot size N is exceptionally large, 100% inspection will be costly and time-consuming. Moreover, 100% inspection may not be feasible or advisable on the following grounds:

1. The loss incurred due to a defective item is very low. In some cases no inspection at all is the most economical course of action.
2. 100% inspection is impossible when inspection is destructive. For instance, a lot of small caliber ammunition is accepted as satisfactory if 99% of the shots fall within a specified distance from the center of a target at a given range. Hence the decision to accept or reject the lot will be reached after destroying the entire lot.

3. 100% inspection is not 100% perfect since manual or mechanical inspection is subject to some margin of error.

In light of the previous discussion, can a receiver use better inspection to assure the quality of product or work submitted by a producer? The answer is to use *acceptance sampling plans*. That is, the decision to accept or reject a lot will be based on a series of samples drawn at random from the submitted lot. Sampling plans are not only economical but also are as effective as 100% inspection. In many instances a well-designed sampling plan may produce better results than 100% inspection.

It should be borne in mind that an acceptance sampling plan may accept occasional lots with a much higher fraction of defective items than the consumer is willing to tolerate. However, if submitted lots differ in quality, the sampling plan will accept the good lots more frequently than the bad lots, and as a result a long-range average quality level, consistent with the quality specified, can be maintained.

Sampling plans may be based on two different kinds of measurements. Inspection may be performed by grading the product as defective or nondefective or as good or bad, e.g., checking the size of cylindrical male parts by go and not-go ring gages. Inspection also may be performed by measuring the product to verify whether it conforms to specification, e.g., measuring the pitch diameter of a screw with a thread micrometer. When related to sampling inspection, the first is known as sampling by *attributes* whereas the second is known as sampling by *variables*. In general, inspection by attributes is less expensive than by variables. However, inspection by variables is more informative than attributes and requires smaller sample size to provide the same protection against accepting lots of poor quality.

SAMPLING BY ATTRIBUTES

17.2 Single Sampling Plans

A sampling plan based on one sample is called a *single sampling plan*. It is characterized by two numbers (n, c), where n is the sample size and c is the acceptance number.

A sample of size n is drawn from the lot and inspected by attributes. The lot is accepted if the number of defectives (d) in the sample does not exceed the acceptance number (c). That is, accept the lot if $d \leq c$ and reject the lot if $d > c$. Now on what basis can one determine the values of n and c? Obviously the optimal selection of n and c should be based on economic considerations. However, the formulation of an economic model which includes relevant cost parameters is complicated. Therefore the values of n and c are determined so that the sampling plan will discriminate between good and bad lots with specified odds for any level of fraction defective in the submitted lots.

17.2.1 The Operating Characteristic (OC) Curve—Lot Quality

Let p denote the fraction defective in a submitted lot. Suppose that the consumer will accept a submitted lot if its fraction defective is less than or equal to 1% and invariably will reject a lot of poorer quality. A plan that would discriminate perfectly between lots with $p \leq 1\%$ and lots with $p > 1\%$ would have the *operating characteristic* (*OC*) curve shown in Fig. 17.1.

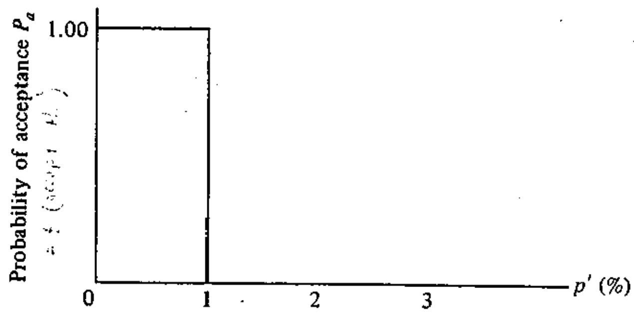

Fig. 17.1 Ideal *OC* curve for a sampling plan.

This ideal *OC* curve can be achieved only with 100% inspection, provided that 100% inspection is infallible. Unfortunately no sampling plan will have an ideal *OC* curve as such. A well-designed sampling plan, however, can approach such a curve. Now if the consumer will reject a submitted lot whenever its fraction defective exceeds 1% (using 100% inspection), the producer will have to screen the rejected lot to eliminate defectives. This means that both the consumer and producer will sustain excessive inspection cost. Consequently, it seems necessary to seek a more realistic approach to this problem, an approach by which it would be feasible to reduce the prohibitive cost of inspection. This dilemma has been solved by instituting acceptance sampling plans.

In practice, the producer and consumer reach an agreement on a sampling plan that is fair to both. Obviously the consumer wants to protect himself against accepting a poor quality lot having a sizable fraction of defectives. He must define the risk he is willing to take in having a poor quality lot accepted by the sampling plan. In other words, the consumer specifies the probability of the sampling plan accepting a lot that has a fraction defective p_2. This probability is usually denoted by β. Similarly, the producer specifies

7220

the probability of the sampling plan rejecting a lot that has a fraction defective p_1. This probability is usually denoted by α. Once the consumer and producer have come to agreement on the values of α, β, p_1, and p_2, a sampling

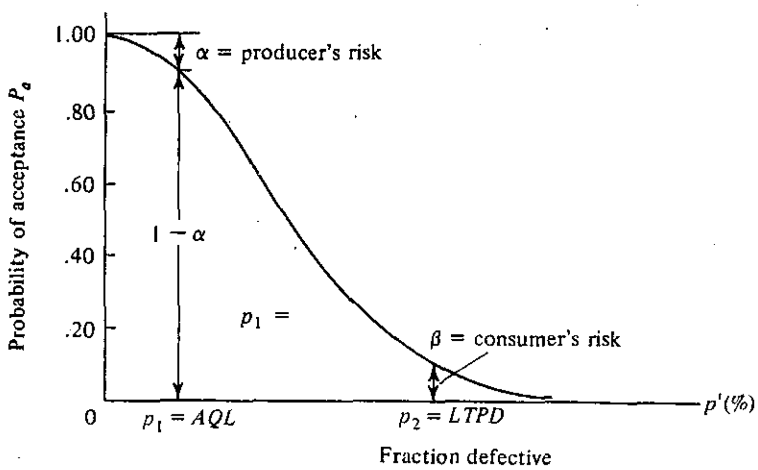

Fig. 17.2 *OC* curve for a single sampling plan.

plan is determined. The *OC* of this sampling plan *should* pass through the two points (p_1, α) and (p_2, β), as shown in Fig. 17.2. The following nomenclature will be adopted for these points:

α = producer's risk
β = consumer's risk
p_1 = acceptable quality level (denoted by AQL)
p_2 = lot tolerance percentage defective (denoted by LTPD) or sometimes called rejectable quality level (denoted by RQL)

The area between the AQL and LTPD is known as the *indifference zone*. From Fig. 17.2 it can be seen that if the quality of the submitted lot is better than p_1, the lot will be accepted with probability greater than $(1 - \alpha)$; if worse than p_2, the lot will be accepted with probability less than β. Thus the *OC* curve for any sampling plan will give the probability with which the plan will discriminate between good and bad or acceptable and unacceptable lots for any level of fraction defective.

Let us now derive the probability of accepting a lot submitted with fraction defective p'. If the incoming lot size is N and we are sampling without replacement, then the probability distribution of the number of defectives (k) in a sample of size n is hypergeometric. In symbols,

$$P_X(k) = \frac{\binom{Np'}{k}\binom{N - Np'}{n - k}}{\binom{N}{n}} \tag{17.1}$$

In practice, the lot size N runs into hundreds, thousands, or even larger. In Sec. 6.19.4 we proved that the hypergeometric distribution with parameters n, p', and N approaches the binomial distribution with parameters n and p. Consequently, Eq. (17.1) can be written

$$P_X(k) \cong \binom{n}{k} p'^k (1-p')^{n-k} \tag{17.2}$$

We proved in Sec. 6.2 that if $n \longrightarrow \infty$ and $p' \longrightarrow 0$ so that $np' = \lambda$, the limiting distribution of the binomial is a Poisson. Thus Eq. (17.2) becomes

$$P_X(k) \cong \frac{e^{-\lambda}\lambda^k}{k!} \tag{17.3}$$

where $\lambda = np'$.

Now if the sampling plan is specified by (n, c) and the lot quality is p', then the probability of accepting the lot is

$$P_a \cong \sum_{k=0}^{k=c} \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=0}^{k=c} \frac{e^{-np'}(np')^k}{k!} \tag{17.4}$$

Equation (17.4) can be evaluated by using Table I in the Appendix. The product $np' = \lambda$ is used to enter Table I; in the column headed $k = c$, to find the P_a value. The following example will illustrate the use of Table I:

Example 17.1 A single sampling plan uses a sample size of 40 and an acceptance number of 1. The lot size is large in comparison with sample size. Use Table I to compute the probabilities of acceptance of lots .5, 1, 2, 3, 4, 5, 6, 7, 8, and 10% defective. Plot the *OC* curve for the sampling plan.

Solution. Here we have a sampling plan with $n = 40$, $c = 1$. That is, a sample of 40 items is drawn from the lot and inspected. The lot is accepted if the sample contains not more than one defective. If $p' = .5\%$, then $\lambda = .20$. The probability of accepting .5% defective lot is

$$P_a(p' = .5\%) \cong \sum_{k=0}^{k=1} \frac{e^{-.20}(.20)^k}{k!} = .982$$

(Note that .982 is read out of Table I with entries $\lambda = .2$ and $k = 1$.) Similarly, we can compute the values of P_a for the specified values of p', which are tabulated below:

p'	.5%	1%	2%	3%	4%	5%	6%	7%	8%	10%
P_a	.982	.938	.809	.663	.525	.406	.308	.231	.171	.092

The *OC* curve for the sampling plan is shown in Fig. 17.3.

Note that the probability of accepting a lot given a specified lot quality (p') depends solely on the sample size (n) and the acceptance number (c). Thus the two numbers n and c completely determine the *OC* curve. Let us now study the effect of n and c on the shape of the *OC* curve. Suppose in

ı larger.
'ameters
'ı and p.

limiting

ity is p',

dix. The
l $k = c$,
Table I:

ın accep-
Jse Table
7, 8, and

a sample
ıe sample
:obability

Similarly,
tabulated

	10%
	.092

ɔt quality
mber (c).
'e. Let us
ıppose in

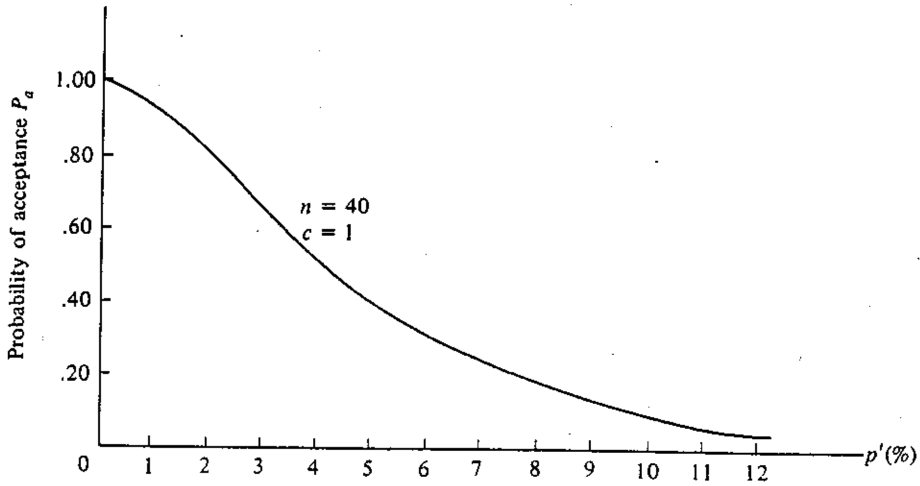

Fig. 17.3 *OC* curve for Example 17.1.

Example 17.1 that we keep the sample size ($n = 40$) constant and change the acceptance number c. The *OC* curves resulting from changing the acceptance number alone are shown in Fig. 17.4.

If we keep the acceptance number ($c = 1$) constant and change the sample size n we obtain the *OC* curves shown in Fig. 17.5. From Figs. 17.4 and 17.5 we draw the following conclusions:

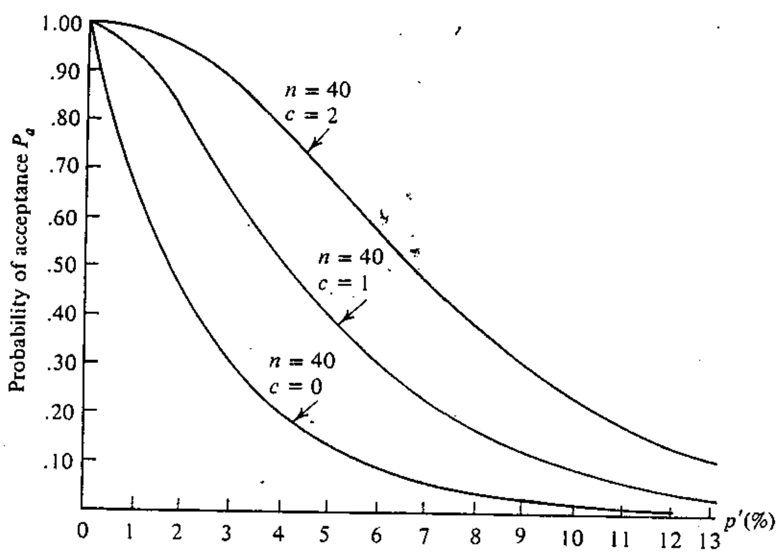

Fig. 17.4 Comparison of *OC* curves with different acceptance numbers.

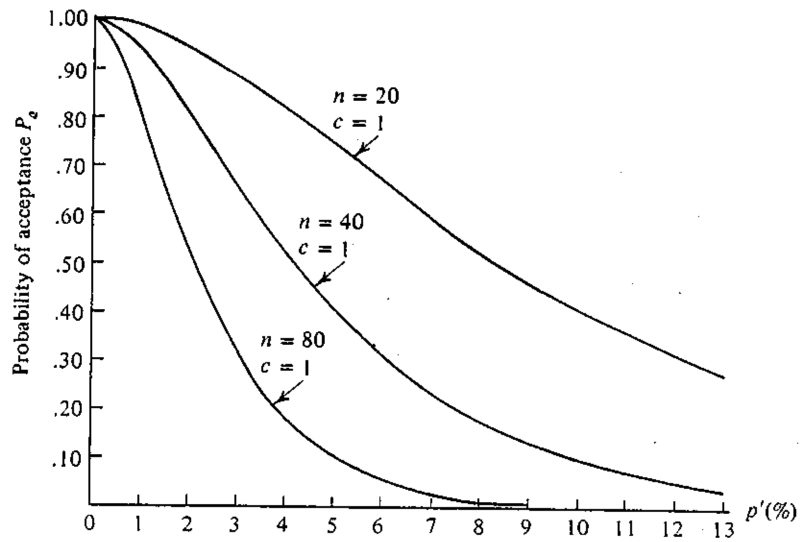

Fig. 17.5 Comparison of *OC* curves with different sample sizes.

1. For the same sample size, increasing the c value serves to move the curve farther from the origin. Thus for a fixed sample size the sampling plan would give better discrimination among lots of different quality if the acceptance number is reduced. To illustrate, suppose a 4% defective lot is submitted for inspection; the lot would be accepted 20% of the time if $c = 0$, whereas the same lot would be accepted 78% of the time if $c = 2$.
2. For the same acceptance number, increasing the sample size causes the slope of the *OC* to become steeper. The steeper the curve, the better the protection against accepting lots of poorer quality.

17.2.2 Determination of Sampling Plan

Assume that the producer and the consumer have agreed to use a single sampling plan for attributes that will protect specified values of AQL and LTPD with specified values of α and β, respectively. What is now needed is an *OC* curve that will pass through the points (AQL, $1 - \alpha$) and (LTPD, β). This *OC* curve is uniquely determined by the numbers n and c. It should be noted that since n and c can take on integral values only, it is usually not possible to find an *OC* curve that will pass through these points exactly; however, it is possible to find a curve that will closely approach these points.

Example 17.2 Devise a single sampling plan that will provide the following protection: $\alpha = .05$, AQL $= p_1 = .02$ and $\beta = .05$, LTPD $= p_2 = .08$.

Solution. To find the sample size n and the acceptance number c, we first assume that $c = 0$ and then find the ratio p_2/p_1. If this ratio is equal to $.08/.02 = 4$, then

the acceptance number of tl
size can be determined. If p_2
way until we find the value o
To illustrate, if $c = 0$, then
(17.5)
whence $np_1 = .05$. Similarly,
in this way we obtain the fol

c	$np_1(1 -$
0	.C
1	.3
2	.8
3	1.3
4	1.9
5	2.6
6	3.2

Therefore, $c = 5$ and $n = 2$.
requirements. Actually, we ob

AQ

which is close indeed to the r
The same result can be
The following table was deve
single sampling schemes ($\alpha =$

R_0
58.
12.
7.5
5.7
4.6
4.0
3.6
3.3
3.1
2.7
2.37
2.03
1.81
1.61
1.51
1.335
1.251

7224

the acceptance number of the sampling plan is 0 and the corresponding sample size can be determined. If $p_2/p_1 > 4$ for $c = 0$, we try $c = 1$. We continue in this way until we find the value of c that yields a value for p_2/p_1 equal or closest to 4. To illustrate, if $c = 0$, then

$$P_a(p_1) = e^{-np_1} = .95 \qquad (17.5)$$

whence $np_1 = .05$. Similarly, $P_a(p_2) \equiv e^{-np_2} = .05$, whence $np_2 = 3.0$. Proceeding in this way we obtain the following results:

c	$np_1(1 - \alpha = .95)$	$np_2(\beta = .05)$	$\frac{p_2}{p_1}$
0	.05	3.0	60.0
1	.35	4.8	13.7
2	.80	6.3	7.9
3	1.36	7.8	5.7
4	1.95	9.17	4.7
5	2.61	10.5	4.02
6	3.20	11.9	3.71

Therefore, $c = 5$ and $n = 2.61/.02 \approx 131$ appear to correspond closely with the requirements. Actually, we obtain the following protection with $n = 131$ and $c = 5$

$$\text{AQL} = .02 \qquad \text{LTPD} = .08$$
$$\alpha = .0494 \qquad \beta = .05$$

which is close indeed to the requirements.

The same result can be obtained by using the Peach–Littauer [14] method. The following table was developed by P. Peach and S. B. Littauer for designing single sampling schemes ($\alpha = \beta = .05$):

R_0	c	np_1
58.	0	.05
12.	1	.36
7.5	2	.82
5.7	3	1.37
4.6	4	1.97
4.0	5	2.61
3.6	6	3.29
3.3	7	3.98
3.1	8	4.70
2.7	10	6.17
2.37	14	9.25
2.03	21	14.89
1.81	30	22.44
1.61	47	37.20
1.51	63	51.43
1.335	129	111.83
1.251	215	192.41

Directions for use of table:

1. Calculate $R_0 = \frac{P_2}{P_1}$.
2. Find R_0 in table. If it does not appear, use the next larger value shown.
3. Read directly the acceptance number c.
4. Davide np_1 by p_1 to get n, the sample size.

The authors [14] proved that if the number of defectives in a sample of n follows a Poisson law, then

$$\frac{\chi^2_{\alpha;\,2(c+1)}}{\chi^2_{1-\beta;\,2(c+1)}} = \frac{2np_1}{2np_2} = \frac{p_1}{p_2} \tag{17.6}$$

Applying Eq. (17.6) to Example 17.2 we obtain

$$\frac{\chi^2_{.05,\,2(c+1)}}{\chi^2_{.95,\,2(c+1)}} = \frac{1}{4} \tag{17.7}$$

From Table G in the Appendix we find that Eq. (17.7) is satisfied when $2(c + 1) = 12$; i.e., $c = 5$ and the corresponding sample size can be obtained from the equations

$$\chi^2_{\alpha;\,2(c+1)} = 2np_1 \quad \text{or} \quad \chi^2_{1-\beta;\,2(c+1)} = 2np_2 \tag{17.8}$$

On substitution, Eq. (17.8) becomes

$$\chi^2_{.05;\,12} = 5.226 = 2n(.02) \quad \text{or} \quad \chi^2_{.95;\,12} = 21.026 = 2n(.08)$$

whence $n \cong 131$, which agrees with our previous finding.

17.3 Average Outgoing Quality

Sampling plans also may be specified according to the *quality level* of lots that leave the inspection point. Suppose that lots of size N are being subjected to a single sampling plan specified by n and c. Furthermore, suppose that lots of but one quality level p' are submitted for inspection. If inspection is nondestructive and the lot size is very large compared to the sample size, then the sampling plan will reject p'% defective lot with probability

$$1 - \sum_{k=0}^{c} \frac{e^{-np'}(np')^k}{k!}$$

Now if rejected lots are 100% inspected and the defectives are removed and replaced by nondefectives, none of these lots will be rejected by the sampling plan. These lots are called *rectified lots* and the inspection scheme is known as *rectifying inspection.* Thus lots accepted by the sampling plan will contain either (1) approximately the percent defective submitted (p') although they will be slightly improved by the replacement of any defectives found in the

samples by nondefectives
are rectified. This mean
inspection point is a co
outgoing quality (denotec

(17.9) AOQ = (N –

If for large N and small
becomes

(17.10)

Suppose that the qua
not constant for all lots.
plan improves, the prob
sorting and screening will
hand, as the quality of su
and, as a result, more lot
quality will improve sin
situation graphically by c
going quality for various

p'(%)
.2
.5
1
2
3
4
5
6
7
8
9
10
12
14

Figure 17.6 illustrates the
The maximum value of A
leave the inspection poi
(AOQL). From Fig. 17.
quality level of all lots su
not exceed 2.1%. It shoul
in a particular instance;
exceed the AOQL. Refer

7226