Chapter 10
Operating Characteristics Curves

10.1 Operating Characteristics and Power Curves

Typically, for every sampling plan there is an operating characteristics (OC) curve which shows how the plan will perform as lots of different quality levels are submitted to it. In the background of OC curves is statistical inference which helps in determining critical points corresponding to the risk, or risks, under study.

The seminal work in the sense of World War II which led to practical applications of OC curves has left a legacy of tables, the best among them being MIL-STD-105A [1]. Say, as an example that a sample of size n is taken and inspected. Depending on the value of the percent defective, p (Chap. 14), the lot is:

- Accepted if there are up to l defective items, and
- Rejected if there are more than l, which is the *acceptance number*.

An application with acceptance numbers has been brought to the reader's attention in Chap. 9. Evidently $l \leq n$. Let us choose for l the values of 0, 1. When the number in the lot is large compared with that in the sample, the probability of acceptance can be computed from the theoretic sampling distribution (more on this later).

The contribution of an OC curve is to indicate the likelihood of rejection of a lot with acceptable percent defective, for instance $l = 1$, while it should have been accepted. Also the likelihood of acceptance of a lot with, say, $l = 3$, while it should have been rejected. Because of the dynamics of sampling inspection:

- The lot under inspection might be rejected while overall it is of acceptable quality.

This is known as Type I error or producer's risk. It is shown in Fig. 10.1 as α. The statistic α is the measure of our confidence that something will happen; it is a threshold permitting the quantification of risk (see also Sect. 10.4).

- The lot under inspection might be accepted while overall its quality is not acceptable; hence, it should have been rejected.

D. N. Chorafas, *Quality Control Applications*, Springer Series in Reliability Engineering,
DOI: 10.1007/978-1-4471-2966-0_10,

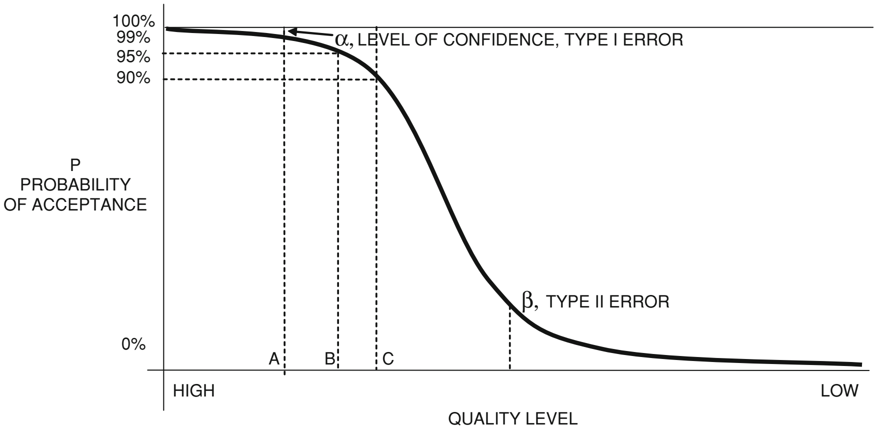

Fig. 10.1 The OC curve of a sampling plan's statistical distribution

This is known as Type II error or consumer's risk. It is shown in Fig. 10.1 as β.[2] Beta is another type of error committed when an existing effect remains undetected in spite of having defined the acceptance/rejection threshold. *Detection* is the keyword and the power of a statistical test is defined as the probability that it will correctly accept or reject the null hypothesis H_0 (see Chap. 8). The rejection of the null hypothesis is represented by $1-\beta$.

Type I error is embedded into the stochastic system. α is usually set at the level of 0.01, 0.05, or 0.10. With $\alpha = 0.01$ there is a 1 in 100 chance of incorrectly rejecting H_0. With $\alpha = 0.05$ this probability of rejection of H_0 increases to 5%. With $\alpha = 0.10$ the probability of rejecting of H_0, while it should have been accepted rises to 10%. In scientific research $\alpha = 0.10$ values border the ridiculous, still they are widely employed in economies and finance which talks volumes of the seriousness of studies providing a level of significance of only 90%.

In Fig. 10.1 the Type I error is shown at 99, 95 and 90% level with corresponding projections A, B, C on the abscissa which identifies the quality level. As a visual inspection can confirm the 90% level on the OC curve and its projection C on the abscissa is far out toward a lower quality level—which, as we will see later on, is the acceptable quality level (AQL) of a production process.

A weak point in power curve analysis is that β is often ignored by researchers. This is wrong because by doing so one disregards the important message conveyed by the OC curve. If β is properly considered, it can assure that a statistical test will have sufficient power to detect whether the phenomenon being examined is characterized by a large Type II error (most often because the sample size n is too small).

[2] Not to be confused with β which stands for volatility.

The reader should as well appreciate that α and β are not independent of one another. They are connected by the OC curve. With α set, the value of β will be constrained which affects the power of a test. A good criterion linking the Type I to Type II error is: β/α.

- If $\alpha = 0.05$ and $\beta = 0.30$, then $30/5 = 6$.
- Hence, the rejection of H_0 is six times more likely than erroneously accepting it.

In this example, the *power* of the test is $1-\beta = 1-0.30 = 0.70$. The power of the statistical test becomes particularly important when the null hypothesis H_0 is not rejected. In principle, the lower the power of the test the less likely H_0 is accepted correctly. The test results are ambiguous because the effect which is being examined has not been fully demonstrated.

If we test for the hypotheses of no difference between two populations with respectively mean parameters μ_1 and μ_2, but a common parameter σ for standard deviation, then the *sample size effect* can be computed as:

$$\gamma = \frac{\mu_2 - \mu_1}{\sigma}$$

where γ is an index.

Power analysis of experimental data permits an estimation of the effect of a size index, which can be used for calculating the power unit of the dependent variable by dividing it by the standard deviation of the measures in their respective populations. The null hypothesis H_0 assures $\mu_1 = \mu_2$ while the alternative hypothesis H_1 assures $\mu_2 > \mu_1$ for one-tailed distributions. Correspondingly for two-tailed distributions (nondirectional test) with two independent samples having the same standard deviation the algorithm is:

$$\gamma = \frac{|\mu_1 - \mu_2|}{\sigma}$$

where μ_1 and μ_2 are the means of the populations; and σ is the standard deviation of either population (assuming they are equal). Furthermore, the computation of a power level requires the values of α and of sample size *n*. When these values are available, the power can be easily calculated through the formula:

$$\delta = \gamma \,.\, \mathrm{f(n)}$$

δ combines the size effect and sample size into a single index that can be used with α to obtain the power level from statistical tables. The symbols γ and δ are used in this text in connection to statistical tests. They should not be confused with γ and δ respectively for second and first derivatives of underlying functions in the study of derivatives.

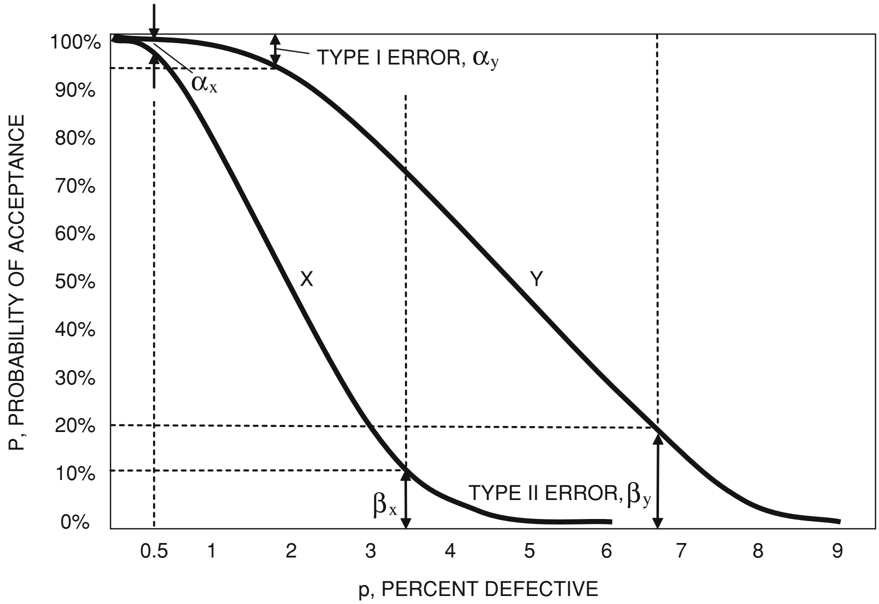

Fig. 10.2 OC curves for two different sampling plans X and Y

10.2 Improving the Shape of an OC Curve

A good way of improving the operating characteristic of the test is to decrease the standard deviation of the sample's statistics being tested. Different plans exist for this purpose; one of them is to take a lot size half as large as the original. Another, which ends up to the same result in terms of percentages is to double the size of a sample. Notice, however, that:

- The OC curve will steepen and β will shrink,
- But there will always be present α and β, albeit smaller ones.

Figure 10.2 compares the OC curves of two sampling plans for percent defective: P, X and Y. The OC curve of X is steeper. $\alpha_x < \alpha_y$ and $\beta_x < \beta_y$. In regard to both Type I and Type II error, sampling plan X is better than sampling plan Y in terms of producer's risk and consumer's risk.

Nevertheless, as Sect. 10.1 brought to the reader's attention, the Type I and Type II errors continue to exist because the percent defective in a sample may be more (or less) than the actual proportion of defective items in the lot. Given this variation, any lot-by-lot inspection plan based on sampling will include a certain amount of risk. What is important is that:

- With statistical quality control the errors are quantifiable and known.

- By contrast, with 100% the errors are present but unknown; therefore, it is like inspecting by the seat of the pants.

In the opinion of people who do not believe in the power of statistical tests, the existence of Type I and II errors in connection to sampling poses a question about the advantages of such a method as contrasted to a 100% inspection. Cost, fatigue of the inspectors (hence errors), and other reasons see to it that a 100% inspection would not provide 100% assurance.

Another significant advantage of OC curves is that they help in calibrating *sample size* in regard to *lot size*, by visualizing the effects of lot size and sample size on OC. People with experience in statistical quality control (SQC) know how to calibrate the sample n when they know (or decide) the size of the population N. The task is straightforward: it is possible to reduce the variance by:

- Decreasing N, while holding n constant
- Increasing n, while holding N constant
- Increasing n and N, while holding constant the ratio n/N.

In the general case, the effect of varying the sample size n is more important than the effect of varying N. In addition, the absolute size of the sample is more important than its size relative to the size of the population. These are two easy rules that should be always remembered.

Once the right SQC plan has been established, we are much better in control of outgoing quality. In addition, this plan becomes integral part of procedures put in place for quality assurance. The prerequisites to SQC are by no means complex, and the same is true of the aftermath. With sampling, the number of lots that would be accepted and the number that would be rejected would depend upon both:

- The nature of the inspection plan used, and
- The actual percentage of defective items in the submitted lots.

Therefore, in selecting a sample plan among alternatives it is good to have specific knowledge of how each of the available plans differentiates between good and bad lots (see also Chap. 9 on discovery sampling). Such information can best be presented as an operating characteristic curve where each point shows on the ordinate, the frequency of accepted lots giving the corresponding rating on the abscissa.

In the majority of cases, this study of alternatives should include the assumption that lots rejected by the sampling plan will be sorted out for control action. The fact of facilitating management control makes the SQC plan an important instrument in *quality analysis*, and it can provide a precious feedback to engineering design (Chap. 7). This feedback can be quantified by using two notions very important in quality control:

- AQL, (which the careful reader will recall from previous references), and
- Lot tolerance percent defective (LTPD, see also Sect. 10.3).

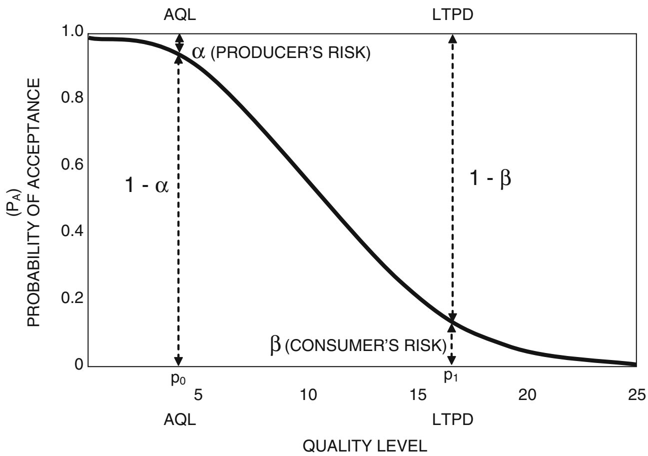

Fig. 10.3 Acceptable quality level and lot tolerance percent defective

As shown in Fig. 10.3, the AQL is equal to $1-\alpha$, and it identifies the percent of defective items in an inspection lot which are considered below the level of lot rejection. An $\alpha = 0.01$ at AQL gives a 1% chance of rejecting a submitted good lot containing a 0.5% defective. This corresponds to the 99% level of confidence.

- *If* the lot acceptance plan accomplishes its full purpose, *then* the process average will become at least as small as p_0. The probability of this happening is α.
- AQL contrasts to LTPD which is equal to $1-\beta$ and identifies a lot of sufficient bad quality that we do not wish to accept more often than a small portion over time.

The way to interpret the $1-\beta$ in Fig. 10.3 is that if lots of 0.5% defective are submitted to this sampling plan, the consumer has a 12% risk that bad lots will pass. This 12% level, which is admittedly unacceptable, can be improved by steepening the OC curve—which, as we have seen, can be done by increasing the sample for the same population. Two more terms need to be defined in regard to practical applications of OC curves.

- Average outgoing quality (AOQ), and
- Average outgoing quality limit (AOQL).

AOQ is the expected fraction defective after substituting good items for bad ones in rejected lots (or correcting the identified errors), and in samples taken from accepted lots. AOQL represents the value of AOQ for lots that result in the largest average outgoing quality. Or, the best average quality that can result over a period of time under the chosen sampling plan. AQL and AOQL correlate.

The flexibility in terms of control action afforded with sampling plans and OC curves is another reason why SQC presents a better protection; one which is measurable and costs less than 100% inspection. The cases examined in this section help to confirm that the range between outgoing quality level and LTPD is an excellent quality assurance solution.

While these examples come from manufacturing where there exists today a very significant experience on the implementation of OC curves, the concepts underpinning them are just as applicable in finance. For instance in loans, like the example with RAROC in Chap. 9.

Whether in industrial applications or in banking, a hard-hitting successful program using OC curves must be carefully planned, simple, and clear-cut. It should as well be accompanied by the understanding and appreciation of opportunities and limits of statistical testing. Another "must" is to train in stochastic thinking the personnel and have its positive participation in the implementation of any statistical plan.

10.3 Using OC Curves: A Methodology

On many occasions, we are interested in comparing the performance of several acceptance sampling plans over a range of different (or likely) quality levels of submitted products. The use of OC curves is one of the best ways possible for attaining this goal. This opportunity comes from the fact that:

- OC curves serve in estimating the probability of accepting lots from a flow of products with fraction defective p, and
- For any given fraction defective p in a submitted lot, the OC curve shows the probability P_A that such a lot will be accepted by a given sampling plan.

The plan in Fig. 10.3 (Sect. 10.2) has been devised from statistics taken from a production line of a manufacturing firm. The quality of submitted inspection lots, p, in percent defective, is shown on the abscissa. The probability of accepting a lot if quality p is shown on the ordinate. This probability P_A presents the percent of lots accepted by the chosen sampling plan when many lots of quality p are submitted. The curve may be checked at a few points to determine the system's behavior.

A lot with no defectives should always be accepted. From the OC curve, we see that when $p = 0$, the probability of acceptance is equal to 1 ($P_A = 1$), indicating that all lots would be accepted. A lot "all defective" should never be accepted. Indeed, for $p = 1$, $P_A = 0$, the OC curve tells that none of the lots would be accepted. Between these two extreme points, there is a certain risk which should be taken into account.

The company whose statistics have been used in this case study did not employ SQC only for its own production lines. When giving a contract to a vendor, this contract specified in quantitative terms the quality level which the vendor must meet: For instance, all lots must be of quality p_0, or better. Lots of quality p_0, or

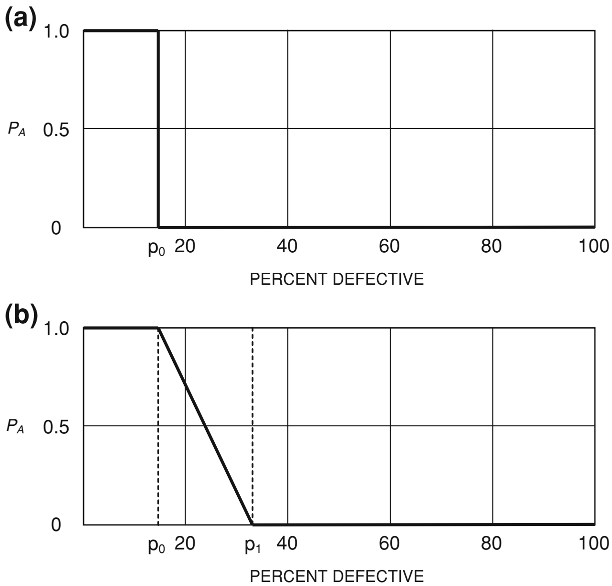

Fig. 10.4 Ideal OC Curves

better, were accepted by this plan all the time—a policy described by the OC curve in Fig. 10.4a.

A different firm, which also used sampling plans and OC curves for purchased goods, had the policy that it will:

- Buy all lots of quality p_0, or better;
- Buy none of the lots of quality p_1, or worse, and
- Accept a portion of the lots whose quality is between p_0 and p_1.

This policy is described by the sampling plan in Fig. 10.4b. Notice that the straight lines also represent an ideal condition not attainable whether with sampling or 100% inspection. One can easily observe that the OC curve in Fig. 10.3 is better than the "ideal" one of Fig. 10.4b. The power curve based on a SQC leads to the acceptance of:

- More lots when p is low, and
- Less lots as the percent defective increases.

By applying an OC curve, most of the lots with quality p_0, or better, are accepted. Most of the lots with quality p_1, or worse, are rejected. A portion of the lots with quality worse than p_0 but better than p_1 are rejected.

By specifying p_0 (AQL), p_1 (LTPD), α (alpha) and sample size which impacts on β (beta), the OC curve is specified. In choosing the sampling plan, the person in charge of quality control should remember that his critical decisions beyond p_0, p_1, and α are: n, the sample size, and c (specified by statistical tables) the limit of defective items, c is the *acceptance number*.

A quality control plan should never be adopted prior to being tested for the behavior of its OC curve. This is necessary to assure that it has the wanted characteristics.

For any practical purpose, an SQC plan is a *quality assurance* plan based on statistical inference. In its simplest form, a sample of given size, for instance n, is taken and inspected. As we have seen in a previous reference, depending on the value of the percent defective, p, the lot is accepted or rejected, the probability of acceptance being P_A.

$$P_A = (1-p)^n, \quad \text{for} \quad c = 0,$$
$$P_A = (1-p)^n + np(1-p)^{n-1}, \quad \text{for} \quad c = 1,$$
$$P_A = (1-p)^n + np(1-p)^{n-1} + \frac{n(n-1)}{2}p^2(1-p)^{n-2}, \quad \text{for} \quad c = 2,$$

On the basis of n and c, can be calculated the OC curves. Some sample curves are presented in Fig. 10.5, for $\alpha = 0.05$. In Fig. 10.5a, the sample size is kept constant and c takes values c_1, c_2, c_3, where $c_1 > c_2 > c_3$. Conversely in Fig. 10.3b the acceptance number c is kept constant and the sample size takes the values n_1, n_2, n_3, where $\mathrm{n}_1 > n_2 > c_3$.

One of the interesting possibilities provided by this methodology is that by means of statistical analysis based on test data, quality assurance information can move upwards the manufacturing hierarchy—and from there all the way to the design source (we have discussed this issue in Chaps. 6 and 7 in connection to reliability engineering). However, the reader should be aware of the fact that while statistical theory provides measures for errors which give guidance,

- It does not remove uncertainty.
- To the contrary, statistical inference is based on uncertainty and the reader should learn to live with it.

Take a simple situation as an example. A producer offers a lot which the consumer either accepts or rejects. This action is the result of inspection and it is often seen as being a simple choice between the alternatives of acceptance and rejection. This is, however, the wrong way of thinking.

As documented by the practical example in this section, the factors underpinning product assurance (Chap. 1) are much more complex than what is revealed by superficial approach to quality control in manufacturing. A dry number of rejected devices or lots provides no way for understanding whether the production process is *in control*.

Many advantages in quality assurance can be derived from the fact that the OC curve of a sampling plan offers a complete statistical description of the consequences of variation in outgoing quality.

The probability of accepting a lot of items can be read directly from the diagram we have seen in the preceding figures. If the population mean and sample

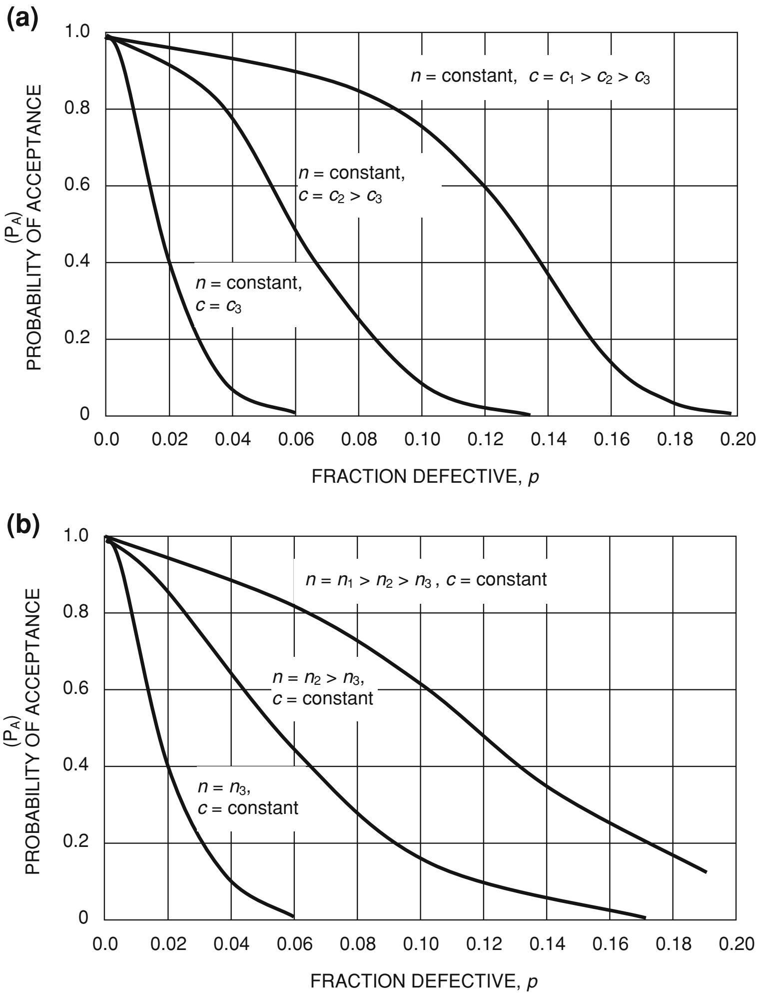

Fig. 10.5 **a** Alternative OC curves by keeping sample size constant, while varying the acceptance number. **b** Alternative OC curves by keeping the acceptance number constant and varying the sample size

size are known; the probability of rejection is one minus the probability of acceptance. Moreover, this complete statistical description can be invaluable in reliability engineering.[3] As it has been already discussed in Chaps. 6 and 7, reliability should not be confused with quality control.

[3] Nevertheless, in considering OC curves it would be of advantage to give an example from reliability engineering.

As with other quality control problems, in reliability engineering there exist two statistical risks. The first risk is that good equipment will be considered bad (producer's risk). The second risk is that bad equipment will be considered good (consumer's risk). Product acceptability as judged by a sampling plan is commonly established by statistically estimating the fraction of the total lot which is defective.

There exist as well limits of the accuracy of measurement often referred to as "confidence intervals" which is unfortunate since they may be confused with statistical confidence limits (Sect. 10.4). For instance, in response to the request by its customer on mean life of its equipment, a company stated that the mean time between failures (MTBF)[4] was estimated to be 600 h $\pm 15\%$ accuracy of measurement. Such a statistic was evidently unacceptable (see in Sect. 10.4 the discussion on accuracy).

A different way of looking at this issue is to consider consumer's risk β in terms of how long shall the test continue when the equipment MTBF is so inaccurate that the incorrect decision might be made from a short test (For instance, one made under stress conditions). Precisely for this reason, in reliability practice certain limitations have been established with the objective of optimizing test procedures.

As an example in one of the projects I participated, it was decided to require by contract that the minimum acceptable MTBF should be by 50% greater than the actually desired minimum. This value is associated with a level of confidence $\alpha = 0.05$ (Sect. 10.4) and β which led to a recommended sampling plan.

Going back to the fundamentals of inspection, the effect of errors in manufacturing and in acceptance of purchased material is the likelihood of a region of poor discrimination among the lots which should be accepted and those which should be rejected. The greater the errors in inspection, the poorer the discrimination would be. This is of particular importance in quality testing because the number of samples and time available for failure rate testing is usually severely limited.

10.4 Level of Confidence

Level of confidence is the degree of protection observed in statistical inference against movements in the underlying measurements or observations, in regard to characteristics of a population under study. To appreciate the fine print of this definition we should return to what was stated about the normal distributions as well as asymmetries in Chap. 8.

The development of an OC curve is based on the hypothesis of the normal distribution which, as the careful reader will remember, is an approximation of real life situations. In other terms, the level of confidence we define (more exactly, we

[4] For a definition of MTBF see Chap. 7.

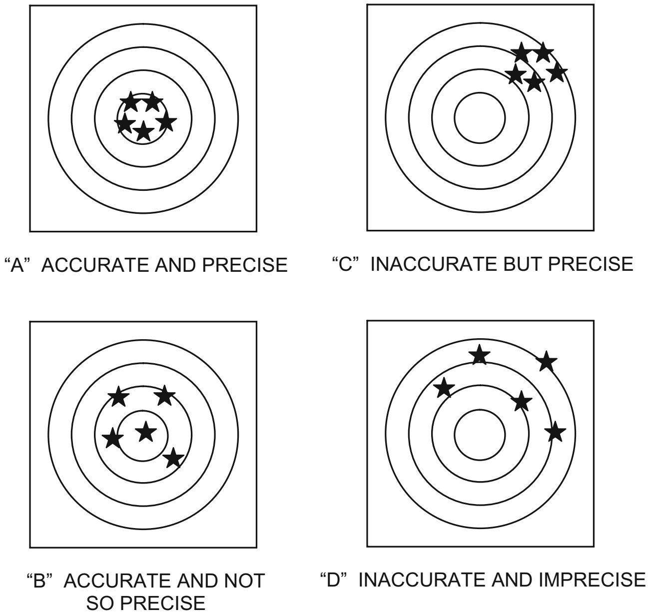

Fig. 10.6 Accuracy and precision are not at all the same thing. **a** Accurate and precise. **b** Accurate and so not precise. **c** Inaccurate but precise. **d** Inaccurate and imprecise

seek to define) will not be precise—but in a large number of cases it will be accurate enough for our job.

Many people, as well as some technical articles, tend to confuse *accuracy* and *precision*. This is wrong. Not only these two terms have different meanings but also they bite into one another. Maybe we like that our measurements are both accurate and precise, but usually we cannot have both accuracy and precision simultaneously:

- Something is *accurate* if it is correct and error-free. In statistics an accurate measurement is close to the expected value.
- *Preciseness* has different meanings which range from exactness to a state of being meticulous, critical, scrupulous, unambiguous, and unbending.

Figure 10.6 presents a graphical definition. Five people are shooting at the same target. In terms of outcome, "A" is highly skilled. His results are both accurate and precise; "B" is accurate though not so precise; "C" is inaccurate but precise; "D" is both inaccurate and imprecise.

In science, and most particularly in engineering, *if* we cannot have the results of "A" *then* we will go for those of "B". In other terms, accuracy is more important than precision. In terms of accuracy in measurements the three higher order non-zero digits will do. We usually, albeit not always, do not need a 7 or 10 digit precision. Interestingly enough, this is also true in business.

- The president of a $100 billion corporation should think in billions and hundreds of millions. If he counts down to cents his attention will be misdirected and his company will go to the rocks.

Accuracy relates to *materiality*. An amount of $1 million is not material for the $100 billion company, but it is highly material to the local firm which makes $10 million per year.

- By contrast, accounting must be precise all the way to dollars and cents. That is what the law demands in most lands, and the letter of the law has to be observed.

That holds all the way to statistical inference. When they are accurate, even if not quite precise, statistical confidence levels are an excellent way to reflect on the likelihood of events, observations, or measurements which will (or will not) occur with a specified degree of confidence.

Accuracy is necessary to be in charge of the variation of a given process, whether in engineering, finance, or other fields. When this is the case, we can utilize a confidence level in order to be certain, in terms of percentages, that a given event will not exceed a particular amount in the envelope of the level of confidence. Through confidence levels, an engineer or other scientist (as well as a financial analyst) is in a position of determining the differential between expected and unexpected events, observations, or measurement. An example is provided in Fig. 10.7 with 95, 99, and 99.9% confidence intervals. The statistics come from a financial study on the change of correlation co-efficients between product lines as a function of volatility. The variable in the abscissa is time. Or, more precisely, the change of volatility over time identified by time series.

As shown in Fig. 10.7, an upper confidence limit is a value larger than the statistic of the central tendency. The opposite is true for a lower confidence limit. The criterion is that in the long run a specified portion of observations, measurement, or test results—hence of the actual population values which interest us—will fall within the so-defined envelope.

- For a two-tailed test of confidence, as the one in Fig. 10.7, the interval between upper and lower limit is the *confidence interval*.
- But the test may also be one tailed. For instance we may be interested *only* in the upper or *only* in the lower confidence limit.

It needs no explaining that the expected population distribution is always important, and so is the sampling procedure. In addition, as it has been underlined on several occasions, we must make sure that the sample is statistically valid, the data are drawn from the same population (a fact which concerns us greatly in connection to confidence intervals), and this continues being so as the number of observation increases.

As with every case of statistical inference, the right sampling procedures, associated to the analytical study we are doing, increases the dependability of the level of confidence and its implied intervals. In the longer run this level assures that a specified proportion of the distribution will fall between the expected value

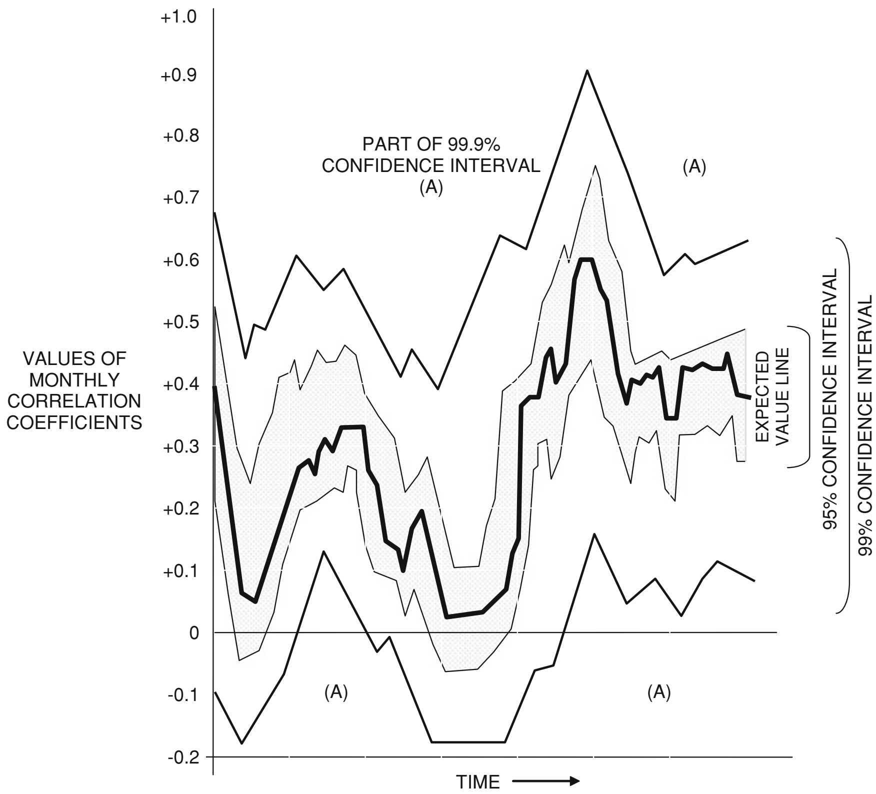

Fig. 10.7 Confidence intervals of correlation co-efficients of two principal variables of a financial risk model

(mean, central tendency) and the limit implied by the level of significance. Only the Type I error, α, will fall outside the so-created boundary condition.

- $\alpha = 0.1$ means that the confidence interval is 90%, and 10% of all cases may fall outside this envelope.
- $\alpha = 0.01$ means that the confidence interval is 99%, and the outliers are only 1%.

The level of confidence is a modeling tool, and as these examples demonstrate one of the major benefits we obtain through modeling is the proper identification and definition of *boundaries*. In all scientific studies, a significant part of the importance played by boundary conditions lies in the ambivalent role of dividing and connecting at the same time.

- Boundaries are places which mark the transition between different conditions, regimes, or functions.

- With this marking, they define different characteristics of the underlying, or its parts, and also reflect the tension which may exist in the limits.

A basic rule of *boundary conditions*, and processes, is fencing off, sealing off what is included in the boundary envelope. There is more homogeneity between points within the boundary, for instance being part of the 95% level of confidence, than across the boundary.

The background concept resembles to one of the famous paradoxes developed by the Greek philosopher Zeno of Elea in the fifth century B.C.: To go from point A to point B, a runner must first reach the midpoint between A and B, then the midpoint of the remaining distance, and so on ad infinitum. Because the process involves an infinite number of steps, Zeno argued, the runner will never reach the destination. The infinite sum 1/2 + 1/4 + 1/8 + 1/16 + … converges to the finite limit 1 but is not equal to 1.

For another examples, if we assume that domestic and foreign letters have on the average the same number of bits, then statistics on transmitted letters tell a story. The ratio of domestic to foreign mail tends to vary between 3 and 87, always significantly more than 1 [2].

In conclusion, it is important to appreciate the meaning of confidence intervals and of boundary conditions. "The 99.9% level is more prudent than the 95%," said a risk management officer in the course of a study, "because with 99% limits are considerably larger than with 95%." But another risk management officer was of the opposite opinion when he stated that "with 95% confidence level traders are more careful since they know the worst case will be exceeded with a frequency of 5%." The latter statement talks volumes about illiteracy in statistical inference, as well as about the wrong psychology associated to boundary conditions.

10.5 Tests of a System

Inspection sampling methods encounter cases where the homogeneity of samples drawn from a given population, or simply supposed to exist, comes into question. A frequently encountered challenge is that involving vendor and consumer. The former submits to the latter the result of a test based on a sample from the lot it delivers. But the consumer is not convinced that this is accurate; hence he, too, takes a sample from that lot and makes a test.

Essentially, what producer and consumer do is to test a system of exchange (goods versus money) from two different viewpoints identified by the now familiar Type I and Type II errors: α and β. The test of a system, of any system, involves the following seven steps:

1. Define the system
2. State the hypothesis (Chap. 8)
3. Select a typical portion: Random sample or representative part
4. Administer the appropriate experiment

5. Observe and record quantitative results
6. Subject these results to a statistical test procedure
7. Decide on the basis of outcome: Whether the system is or is not operating at an AQL.

As this and previous chapters have explained, there are several problems associated to the test of a system. They are ranging from sample size to the homogeneity of the background population under study and the exact methodology used in the two tests (producer and consumer).

- *If* the homogeneity is questionable,
- *Then* the quality level of the inspection is reduced and it may even be questionable.

This is the theme treated in the present section. Let us start with the assumption that two independent samples are drawn at random from a lot. The inspection is by attributes (Chap. 14). Each item in the samples is classified as go/no–go, which stands for conforming/nonconforming. The comparison of fractions defective p_1 and p_2 in the two samples is a *test of homogeneity* of the two samples.

Provided that sampling inspection and testing are uniformly accomplished, the concern will be whether the percentages of defective being observed would be occurring by chance selection on the reason is nonhomogeneity. The question to be answered; therefore, is whether the differences in sampling inspection results between vendor and consumer can be attributed to:

- The luck of the draw in selecting sample units at random from the lot, or
- Real differences distinguishing the two samples, or
- The difference finds its reason in varying inspection practice between vendor and consumer.

For instance, in the latter case among background factors may be improper use of inspection aids, misinterpretation of inspection standards, or failure to select random samples. (It is a sound practice to regard the inspection performed by the consumer as the standard against which the performance of the supplier will be judged).

The test of the system and of homogeneity of its contents is based on critical values for indicating discrepancies which should not be confused with the rejection numbers of sampling plans in determining acceptance of supplies. The decision regarding the dependability of inspection results is distinct from the decision to accept or reject a lot for quality reasons, even if the latter decision may be contingent upon the former.

In the case the test concerns difference in quality inspection practises between supplier and consumer, the consumer must ascertain an *action number* associated with the number of defectives observed and recorded by the vendor. He can then compare the number of defectives he found with that action number. If the number of defects observed equals or exceeds the action number, the consumer's inspector should adopt a course based on the premise that the discrepancy actually exists in the vendor's inspection system.

Notice that independently of issues regarding methodology there exists the case that the size of the sample used by vendor and consumer differs. The ratio of sample sizes used by the two parties may be equal to 1, 2, 3, or higher. However, for simplicity in the OC curve to be examined we will take this ratio equal to 1 (the two sample sizes are equal).

The frequency rate of the probability of *not* accepting the hypothesis of homogeneity is set with the aid of statistical tables [3]. These show the correspondence between likelihood of acceptance of verification tests. OC of the test for homogeneity are shown in Fig. 10.8 for two equal size samples by vendor and consumer. These are analogous to the OC curves of acceptance sampling plans which we have already studied.

- The ordinate is the probability of acceptance at consumer's side.
- The abscissa gives the ratio of fraction defective, and
- The key variable is the expected number of defects in the vendor's samples.

These curves demonstrate the relationship between a range of apparent quality differences brought about by differences in the vendor–consumer inspection systems and the probability of accepting the hypothesis of homogeneity. *If* the consumer can specify the tolerable ratio of the quality which should be detected as frequently as possible when it exists, *then* the appropriate sample size ratio can be selected—provided the expected number of defectives is in the vendor's samples can be estimated from his:

- Sample size, and
- Process average.

Operating characteristics curves can give good approximation of true probability of acceptance associated with the test for homogeneity. *If* the samples of the supplier and the consumer risk depleting the lot, *then* a special arrangement is required to permit valid comparison of the respective inspection results. For instance, the vendor retains his sample and does not return it to the lot purified of the defectives it contains until the consumer has drawn an independent sample.

In addition, since the incidence of defectives is a small lot is very low, the results from consecutive lots must be pooled until the expected number of acceptance within the desired range for required quality. Alternatively, the consumer can rely in part upon an engineering check of the quality control and inspection system of the vendor.

A double or multiple sampling procedure (Chap. 9) is also possible instead of single sampling. When the supplier elects to use it, some minor modifications are necessary in the verification methodology described in the preceding paragraphs. Check ratings are obtained only for the first sample from each lot, but the critical values of single sampling are applicable to each sample individually or collectively as predetermined.

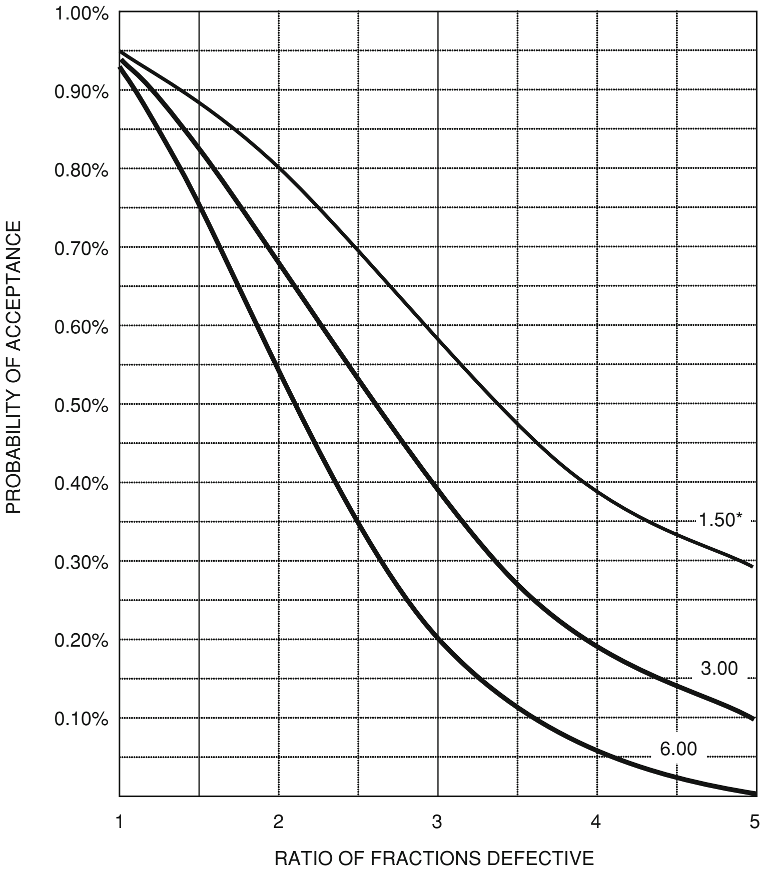

Fig. 10.8 OC curves of two-sample test for homogeneity with variable expected number of defects. * Expected number of defects in vendor's sample

Resubmitted lots may require a larger verification sample because the relative incidence of defects may be smaller than usual, or because the tolerable quality discrepancy ratio and associated risk may be modified to protect the consumer. Whether to pool results of resubmitted lots or not will depend upon the number to be submitted at every case and the expected number of defectives in the vendor's samples.

The test for homogeneity can be instrumental in avoiding arguments between suppliers and consumers, their inspectors and other affected personnel. It is therefore to the advantage of everyone involved, especially in large organizations, to proceed with well-planned statistical tests after standardizing form, content, defect definition, lot and other issues as well as AQL.

Generally speaking the estimate of a lots acceptability is subject to errors. Therefore, there are benefits to be derived from a scientific method. The methodology described in this section was designed primarily for use in receiving inspection but it is useful in all manufacturing operations. Among other advantages it assists in reducing inspection costs while assuring dependable results.

10.6 Controlling the Hazard of Guesswork Through Experiments

As the reader will remember, some of the case studies in Chaps. 1 and 2 came from the Omega lamp manufacturing company and they mainly concerned wire quality. The provision of comprehensive information on quality is a recurrent problem, usually addressed by tons of paper. As Omega's vice president of Engineering said: "I have one kilo heavy reports. They don't provide me with any clue on quality problems. It is important for me to see in one page the exception. Working in the traditional way it is difficult to decide what is an exception, and that's why I keep on getting these one kilo reports".

Omega's CEO looked at this same problem under a somewhat different light. He wanted that his company establishes a unique quality control system which can be valid through its global operations, and which at the same time observes both international norms and specifications demanded by major local customers. In his words: "A subject like quality control can never be spoken of too much. Our aim should be to have a uniform quality, consistent with sales objectives and with the standards of our manufacturing equipment. This consistence will be the real mark of our products' high quality".

The VP Engineering interpreted the CEO's wish as fulfilling the company's marketing argument: "We are more expensive in our product but this is the best lamp one can find in the market". To make this argument stick, he outlined a complete list of factors which influence wire drawing quality and constancy by addressing the physical and chemical properties of the end product:

1. Wire drawing: Regularity of spooling as a condition for the regularity of wire drawing back tension and uniformity.
2. Elements of wire guidance: Error-free run, size of the angle.
3. Deposition of lubricant: Binding element between wire and lubricant, thickness of oxide layer; porosity of the oxide; sticking of oxide of the wire (more on this later).

Both the lubricant and its baking were signaled out as being important and calling for more attention than it received that far. This greater amount of attention included the lubricant's chemical composition, dispersion of graphite form and size of graphite particles, temperature of the lubricant, method of deposing, and mechanical reliability of the system. In connection to baking of the lubricant, critical factors have been length of furnace, average temperature of furnace, temperature profile in furnace, drawing speed.

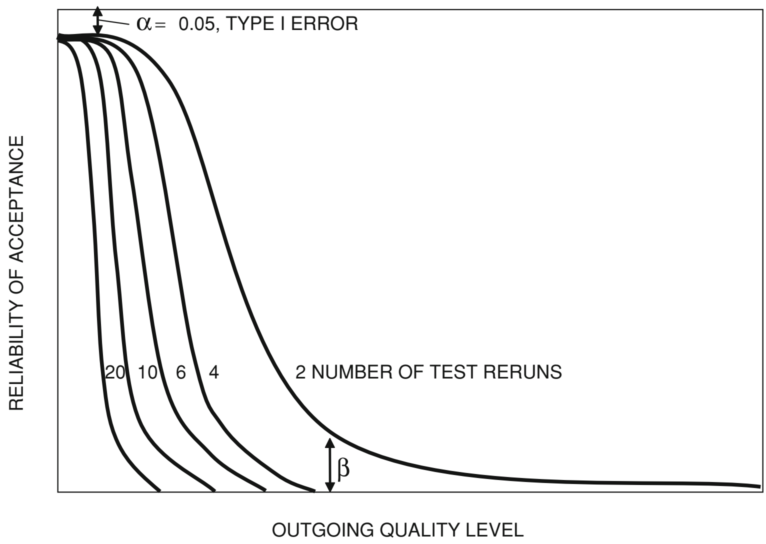

Fig. 10.9 OC curves helped in choosing the number of test reruns (sample size)

For problem No. 3 (the lubricant) was decided a level of significance $\alpha = 0.01$ and consumer's risk β (which cannot be communicated). Four variables were selected as being the most important:

- Binding element
- Thickness of oxide layer
- Porosity of oxide
- Sticking of the oxide of the wire.

The levels of each were fixed by the nature of the test. The big question has been: What level of incremental change, δ, justifies rejection of the null hypothesis? The question was answered by engineering which conducted experiments to document the level of critical difference.

4. Drawing process: Wire approach to the die, including temperature of wire, degree of dryness of lubricant, thickness of oxide layer, thickness of graphite layer.

Conditions specific to the die were: Geometric shape of the die, polishing, length of deforming part relative to diameter, heat transfer between diamond and casing, parallel between the geometric axis of the die and the direction of lateral transfer, lateral transfer of the draw itself, temperature of the interface between wire and die, roundness of the hole. Crucial in regard to drawing conditions were: Amount of reduction and drawing speed. Quality of spooling has been studied in regard to spooling tension, accuracy of spools, and mechanical stability of spools.

All the foregoing factors were proven to have important bearing on end quality, and so did the methodology selected for establishing an orderly approach to the study and analysis of the aforementioned critical factors. The chosen methodology has been experimental design (Chap. 11) and OC curves.

Knowing the number of variables and the range at which each can vary, facilitated the design of experiments. The total number of retests in an experiment was established by selecting the best OC curve among those shown in Fig. 10.9. Though 20 experiments were estimated to produce the steepest OC curve, it was judged that the accuracy of ten is acceptable.

This is not a choice to be made likely. In all experiments, the most important element is to discover main effects of those variables acting independently. Some experiments also called for an evaluation of interaction effects by variables which, in combination, affect the output. Both issues were present in this research which by means of experimental design (Chap. 11) thrust upon itself the goals of:

- Determining repeatability's accuracy
- Analyzing main effects, and
- Evaluating interaction effects.

Thickness of oxide layer and porosity of oxide were examined for interaction effects. Two-way tables gave the better answer through a comparison of average readings for a combination of two levels of two variables. Similar tables were made for each of the other combinations of variables. Variables which were found to be working together to change the output were treated through Latin squares (Chap. 11). The results were satisfactory as they went well beyond the change that would occur with either variable considered alone.

References

1. Military Standard Sampling Procedures and Tables for Inspection by Attributes (MIL-STD-105A) (1950) US Government Printing Office, Washington, DC
2. Miller JG (1978) Living systems. McGraw-Hill, New York
3. Sampling Procedures and Tables for Life and Reliability Testing (1960) Office of the Assistant Secretary of Defense (Supply and Logistics), Washington DC