Chapter 1

Preliminaries

1.1 Introduction to Sequential Procedures

Sequential procedures differ from other statistical procedures in that the sample size is not fixed in advance. The experimenter has the option of looking at a sequence of observations one (or a fixed number) at a time and decide whether to: stop sampling and take a decision; or to continue sampling and make a decision some time later. The order of the sequence of observations which the experimenter will take is specified in advance. Decision problems in which the experimenter may sequentially vary the treatments is of a higher order of difficulty and is called the *sequential design problem.* For example, consider the following problem.

Problem 1.1.1 *If we wish to compare several drugs or treatments (as in sequential screening of cancer drugs), then it should be possible to drop some drugs out of the trials at an early stage if the results from these are very poor when compared with the others.*

Thus, an essential feature of a sequential procedure is that the number of observations required to terminate the experiment is a random variable since it depends on the outcome of the observations. Sequential procedures are of interest because they are economical in the sense that we may reach a decision earlier via a sequential procedure than via a fixed-sample size procedure. In sequential experiments we need to specify:

1. the initial sample size
2. a rule for termination of the experiment
3. the additional number of observations to take if the experiment is to be continued; and

4. a terminal decision rule.

Notice that (2) and (3) can be combined into a single rule. Experiments in which only the number of observations is sequentially dependent, require simpler theory and will be of general applicability, than the sequential design problem in which not only the number of trials but also the number of treatments will be sequentially dependent.

If the experiment has been continued until we observe $X_1, X_2, ..., X_m$, a sequential test is completely defined by specifying the disjoint subsets R_m^0, R_m^1, and R_m^c of m-dimensional Euclidean space $\mathbb{R}_m$ for $m = 1, 2, ...$ If $(X_1, X_2, ..., X_m)$ belongs to R_m^0 we accept the hypothesis H, we reject H when it belongs to R_m^1 and we continue sampling if it falls within region R_m^c. Since the above sets are mutually exclusive and have union $\mathbb{R}_m$, it suffices to specify any two of the three sets. The basic problem is a suitable choice of these sets. The criteria for the choice of these sets will be dictated by the *operating characteristic* (OC) and the *average sample number* (ASN) functions which will be elaborated in the following.

Suppose that the underlying distribution function is indexed by a real-valued parameter, and suppose that the statistician has to choose between two hypotheses, H_0 and H_1. The OC(θ) is defined as *the probability of accepting H_0 when θ is the value of the parameter.* It is desirable that the OC function should be high for values of θ that are consistent with H_0 and low for values of θ that are consistent with H_1. For instance, one may require $\mathrm{OC}(\theta) \geq 1 - \alpha$ for all θ in H_0 and $\mathrm{OC}(\theta) \leq \beta$ for θ in H_1, where α and β denote the error probabilities. A sequential test S is said to be *admissible* if its OC function meets the above criteria.

As noted earlier, the number of observations required by a sequential procedure is a random variable and of much interest is its expected value when θ is the true value of the parameter. This expected value is typically a function of θ, and is called the ASN function. It is desirable to have a small ASN function for given α and β. We also desire the expected sample size to be smaller than those required by the fixed-sample size procedure. Let $\nu(\theta|D)$ denote the expected sample size of procedure D when θ is the true value of the parameter. If D_0 is admissible and if $\nu(\theta|D_0) = \min_D \nu(\theta|D)$ then D_0 is considered to be a "*uniformly best*" test. However, in general, no uniformly best test exists. It is possible to find an optimal sequential procedure when H_0 and H_1 are simple hypotheses. Wald's sequential probability ratio test (SPRT) gives the minimum ASN at both H_0 and H_1. The efficiency of a procedure D at θ is defined as the ratio of the minimum expected sample number at θ to the expected sample number of D at θ. Wald's SPRT has efficiency equal to 1 at both H_0 and H_1.

1.2 Sampling Inspection Plans

The earliest sequential procedure is the double sampling plan of Dodge and Romig (1929) for sampling inspection. A lot consisting of n items and rejecting (accepting) the lot if the number of defectives in the sample is $\geq$ ($<$) c. The drawback of this scheme is that we might have had more than c defectives earlier than sample size n. An alternative scheme is: sample one item at a time, reject the lot as soon as the number of defectives in the sample is $\geq c$, and accept the lot as soon as the number of effective items in the sample is $\geq n - c + 1$. The required sample size is at least c and is at most n. This scheme is called *curtailed inspection.*

1.2.1 Sample Size Distribution

Let N denote the random sample size required to terminate the experiment. Then

$$P_\theta\left(N = c \text{ and reject } H_0\right) = \theta^c, \tag{1.2.1}$$

$$P_\theta\left(N = c + r \text{ and reject } H_0\right) = \binom{c+r-1}{c-1}\theta^c\left(1-\theta\right)^r, \; r = 1, 2, .., n - c, \tag{1.2.2}$$

$$\begin{aligned} P_\theta\left(N = n - c + 1 + s \text{ and accept } H_0\right) &= \binom{n-c+s}{s}\theta^s\left(1-\theta\right)^{n-c+1}, \\ s &= 0, 1, .., c-1. \end{aligned} \tag{1.2.3}$$

Now

$$E_\theta\left(N\right) = \sum_{m=1}^{n} m p_m,$$

where p_m denotes the probability that a decision is reached at the m^{th} trial. (Note that $P\left(N = m | \text{reject } H_0\right) = 0$ for $m < c$, and $P\left(N = m | \text{accept } H_0\right) = 0$ for $m < n - c + 1$). Further

$$\begin{aligned} p_m &= P_\theta\left(\text{reject at stage } m\right) + P\left(\text{accept at stage } m, m \geq c\right) \qquad (1.2.4) \\ &= \binom{m-1}{c-1}\theta^c\left(1-\theta\right)^{m-c} + \binom{m-1}{n-c}\left(1-\theta\right)^{n-c+1}\theta^{m-(n-c+1)}. \end{aligned}$$

Hence

$$\begin{aligned} E_\theta(N) &= c\theta^c \sum_{m=1}^{n} \binom{m}{c} (1-\theta)^{m-c} + (n-c+1)(1-\theta)^{n-c+1} \\ &\quad \times \sum_{m=n-c+1}^{n} \binom{m}{n-c+1} \theta^{m-(n-c+1)} \qquad (1.2.5) \\ &= c\theta^c \sum_{r=0}^{n-c} \binom{r+c}{c} (1-\theta)^r + (n-c+1)(1-\theta)^{n-c+1} \\ &\quad \times \sum_{r=0}^{c-1} \binom{n-c+1+r}{r} \theta^r. \qquad (1.2.6) \end{aligned}$$

One should prefer the curtailed sampling plan to an equivalent single sampling plan because $E(N|\theta)$ for the former lies below the sample size for the single sampling plan. Consider the case $c = 1$. Then

$$\begin{aligned} E(N|\theta) &= \theta \sum_{r=0}^{n-1} (r+1)(1-\theta)^r + n(1-\theta)^n \\ &= (1-y) \sum_{r=0}^{n-1} (r+1) y^r + ny^n, \quad y = 1-\theta \\ &= \sum_{r=0}^{n-1} (r+1) y^r - \sum_{j=1}^{n} jy^j + ny^n = \sum_{r=0}^{n-1} (r+1) y^r - \sum_{j=0}^{n-1} jy^j \\ &= \sum_{r=0}^{n-1} y^r = \frac{(1-y^n)}{1-y}, \qquad (1.2.7) \end{aligned}$$

which is increasing in y. Hence $E(N|\theta)$ is decreasing in θ when $c = 1$. However, this is not true for $c > 1$. (See Table 1.2.1 and the case $c = 4$).

Table 1.2.1 Giving $E(N|\theta)$ for various values of n, c, and θ

	$c=1$			$c=2$			$c=4$		
θ	$n=10$	20	25	10	20	25	10	20	25
.01	9.56	18.20	22.22	9.07	19.06	24.01	7.07	17.17	22.22
.10	6.51	8.78	9.28	8.76	14.73	16.49	7.74	18.10	22.58
.20	4.46	4.94	4.98	7.45	9.58	9.84	8.34	16.15	18.02
.30	3.24	3.33	3.33	6.03	6.64	6.66	8.50	12.77	13.17
.40	2.48	2.50	2.50	4.86	5.00	5.00	8.13	9.94	9.99
.50	2.00	2.00	2.00	3.97	4.00	4.00	7.39	8.00	8.00

Let

$$\begin{aligned} P_1(\theta) &= P(\text{accept lot using the fixed sample procedure } |\theta) \\ &= \sum_{r=0}^{c-1} \binom{n}{r} \theta^r (1-\theta)^{n-r} \end{aligned} \tag{1.2.8}$$

and

$$\begin{aligned} P_2(\theta) &= P(\text{accept lot using the sequential rule } |\theta) \\ &= \sum_{m=n-c+1}^{n} P(\text{accept lot and } N = m|\theta) \\ &= \sum_{m=n-c+1}^{n} \binom{m-1}{n-c} \theta^{m-1-(n-c)} (1-\theta)^{n-c} (1-\theta) \\ &= (1-\theta)^{n-c+1} \sum_{r=n-c}^{n-1} \binom{r}{n-c} \theta^{r-(n-c)} \\ &= (1-\theta)^{n-c+1} \sum_{r=0}^{c-1} \binom{r+n-c}{r} \theta^r. \end{aligned} \tag{1.2.9}$$

Then we have following Lemma.

Lemma 1.2.1 *$P_1(\theta) = P_2(\theta)$ for all n and c.*

Proof. For $c = 1$, $P_1(\theta) = P_2(\theta) = (1-\theta)^n$.
For $c = 2$, $P_1(\theta) = P_2(\theta) = (1-\theta)^n + n\theta(1-\theta)^{n-1}$.
Now assume it is true for any c and consider the case $c+1$. That is, assume

$$\sum_{k=0}^{c-1} \binom{n}{k} \theta^k (1-\theta)^{n-k} = (1-\theta)^{n-c+1} \sum_{r=0}^{c-1} \binom{r+n-c}{r} \theta^r \tag{1.2.10}$$

and we wish to show that

$$\sum_{k=0}^{c} \binom{n}{k} \theta^k (1-\theta)^{n-k} = (1-\theta)^{n-c} \sum_{r=0}^{c} \binom{r+n-c-1}{r} \theta^r. \tag{1.2.11}$$

Subtract (1.2.10) from (1.2.11) and cancelling a common factor $(1-\theta)^{n-c}$, it suffices to show that

$$\binom{n}{c} \theta^c = \sum_{r=0}^{c} \binom{r+n-c-1}{r} \theta^r - (1-\theta) \sum_{r=0}^{c-1} \binom{r+n-c}{r} \theta^r,$$

or

$$\left[\binom{n}{c}-\binom{n-1}{c}\right]\theta^c = \sum_{r=0}^{c-1}\left\{\binom{r+n-c-1}{r}-\binom{r+n-c}{r}\right\}\theta^r + \sum_{r=0}^{c-1}\binom{r+n-c}{r}\theta^{r+1},$$

or

$$\binom{n-1}{c-1}\theta^c = -\sum_{r=0}^{c-1}\binom{r+n-c-1}{r-1}\theta^r + \sum_{s=1}^{c}\binom{s+n-c-1}{s-1}\theta^s,$$

or

$$0 = -\sum_{r=0}^{c-1}\binom{r+n-c-1}{r-1}\theta^r + \sum_{s=1}^{c-1}\binom{s+n-c-1}{s-1}\theta^s,$$

which is obviously true. ■

Remark 1.2.1 Lemma 1.2.1 can also be established by showing that all the sample paths leading to accepting the lot are exactly the same in both the sampling schemes.

1.3 Stein's Two-stage Procedure

In this section we present a certain hypothesis-testing problem for which fixed-sample and meaningful procedures do not exist. However, a two-stage procedure has been given for the same problem. Consider the following problem.

Let X be distributed as normal with mean θ and variance σ^2, where θ and σ^2 are both unknown. We wish to test $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta > \theta_0$; this is known as *Student's hypothesis.*

It is well-known that given a random sample $X_1, X_2, ..., X_n$, the uniformly most powerful unbiased test of H_0 against H_1 is to reject H_0 when

$$T = \frac{\left(\overline{X}-\theta_0\right)\sqrt{n}}{s} > t_{n-1,1-\alpha} \tag{1.3.1}$$

where $\overline{X}$ and s denote the mean and the standard deviation of the observed X_i's and $t_{n-1,1-\alpha}$ denotes the $100(1-\alpha)^{th}$ percentile of the t-distribution with $n-1$ degrees of freedom. If $1-\pi(\theta,\sigma)$ denotes the power of the test in (1.3.1), then $\pi(\theta_0,\sigma) = 1-\alpha$, irrespective of the value of σ. However, when one is planning an experiment, one is interested in knowing the probability with which the statistical test will detect a difference in the mean when it actually exists. However, the power function of "Student's" test depends on σ which is unknown. Hence, it is of interest to devise a test of H_0 versus H_1, the power of which does not depend on σ. However, Danzig (1940) has shown the nonexistence of meaningful fixed-sample test procedures for this problem. Stein (1945) proposed a two-sample (or two-stage) test having the above desirable property, where the size of the second sample depends on the outcome of the first.

1.3.1 The Procedure

A random sample of size n_0 observations $X_1, X_2, ..., X_{n_0}$, is taken and the variance σ^2 is then estimated by

$$s^2 = \frac{1}{n_0 - 1}\left[\sum_{i=1}^{n_0} X_i^2 - \frac{1}{n_0}\left(\sum_{i=1}^{n_0} X_i\right)^2\right]. \tag{1.3.2}$$

Then calculate n as

$$n = \max\left\{\left[\frac{s^2}{z}\right] + 1, n_0 + 1\right\}, \tag{1.3.3}$$

where z is a previously specified constant and $[y]$ denotes the largest integer less than y, and draw additional observations $X_{n_0+1}, X_{n_0+2}, ..., X_n$. Evaluate, according to any specified rule that depends only on s^2, real numbers a_i $(i = 1, 2, ..., n)$ such that

$$\sum_{i=1}^{n} a_i = 1, \; a_1 = a_2 = ... = a_{n_0}, \; s^2 \sum_{i=1}^{n} a_i^2 = z. \tag{1.3.4}$$

This is possible since

$$\min \sum_{i=1}^{n} a_i^2 = \frac{1}{n} \leq \frac{z}{s^2} \tag{1.3.5}$$

by (1.3.3), the minimum being taken subject to the conditions $a_1 + a_2 + ... + a_n = 1, a_1 = a_2 = ... = a_{n_0}$.

Define T' by

$$\begin{aligned} T' &= \frac{\sum_{i=1}^{n} a_i X_i - \theta_0}{\sqrt{z}} \\ &= \frac{\sum_{i=1}^{n} a_i \left(X_i - \theta\right) + \left(\theta - \theta_0\right)}{\sqrt{z}}. \end{aligned} \tag{1.3.6}$$

Then

$$U = \frac{\sum_{i=1}^{n} a_i \left(X_i - \theta\right)}{\sqrt{z}}$$

is such that

$$\begin{aligned} var\left(U|s\right) &= \left(\sum_{i=1}^{n} a_i^2\right)\left(\frac{\sigma^2}{z}\right) \\ &= \frac{\sigma^2}{s^2}. \end{aligned}$$

Also, it is well-known that $V = (n_0 - 1)\, s^2/\sigma^2$ is distributed as central chi-square with $n_0 - 1$ degrees of freedom.

$$U|s^2 \stackrel{d}{=} \text{normal}\left(0, \sigma^2/s^2\right).$$

Hence

$$U\frac{s}{\sigma}|s^2 \stackrel{d}{=} \text{normal}\,(0,1)\,.$$

Since the distribution of $U\frac{s}{\sigma}|s^2$ does not involve s, we infer that $U\frac{s}{\sigma}$ is unconditionally distributed as normal$(0,1)$ and is independent of s^2. Consequently

$$\frac{Us/\sigma}{\left[(n_0 - 1)\, s^2/\sigma^2\, (n_0 - 1)\right]^{1/2}} = U = t_{n_0-1}. \tag{1.3.7}$$

If $f(x,y)$ denotes the joint density of Us/σ and s^2, then

$$f\,(x,y) = f\,(x|y)\, f_{s^2}\,(y) = g\,(x)\, f_{s^2}\,(y)\,,$$

where $g\,(x)$ is the density of Us/σ because

$$\int f\,(x,y)\, dy = g\,(x)\,.$$

So, Us/σ and s^2 are stochastically independent. i.e., U has the t-distribution with $n_0 - 1$ degrees of freedom irrespective of the value of σ. Hence,the test based on T' is unbiased and has power free of σ. Then in order to test for the one-sided alternative $\theta > \theta_0$, the critical region of size α is defined by

$$\frac{\left(\sum_{i=1}^{n} a_i X_i - \theta_0\right)}{\sqrt{z}} > t_{n_0-1,1-\alpha}. \tag{1.3.8}$$

The power function is then

$$1 - \pi\,(\theta) = P\left\{t_{n_0-1} > t_{n_0-1,1-\alpha} + (\theta_0 - \theta)\,/\sqrt{z}\right\}. \tag{1.3.9}$$

Analogues critical region, with similar power function independent of σ, holds for the two-sided alternative: $\theta \neq \theta_0$.

As mentioned earlier, the above test is not used in practice. However, a simpler, and slightly more powerful, version of the test is available, as we now show. (Intuitively Stein's test wastes information in order to make the power of the test strictly independent of the variance.) Instead of (1.3.3), take a total of

$$n = \max\left\{\left[\frac{s^2}{z}\right] + 1, n_0\right\}, \tag{1.3.10}$$

observations and define

$$
\begin{aligned}
T'' &= \left[\frac{1}{n}\sum_{i=1}^{n} X_i - \theta_0\right]\frac{\sqrt{n}}{s} \\
&= n^{-1/2}\sum_{i=1}^{n}\frac{(X_i-\theta)}{s} + \frac{(\theta-\theta_0)\sqrt{n}}{s} \\
&= U' + \frac{(\theta-\theta_0)\sqrt{n}}{s}. \qquad (1.3.11)
\end{aligned}
$$

One can easily establish that U' has a t-distribution with $n_0 - 1$ degrees of freedom. Since $n > s^2/z$, we have $|(\theta-\theta_0)\sqrt{n}/s| > |(\theta-\theta_0)/\sqrt{z}|$. So, if we employ critical region $T'' > t_{n_0-1,1-\alpha}$ instead of (1.3.8) the power of the test will always be increased. Also, the number of observations will be reduced by 1 or left the same.

Suppose we want the power to be $1-\beta$ when $\theta = \theta_0 + \delta$ where δ is specified. Then power at $\theta_0 + \delta$ is

$$
\begin{aligned}
& P\left\{\frac{\sqrt{n}\left(\overline{X}-\theta\right)}{s} > t_{n_0-1,1-\alpha} - \frac{\delta\sqrt{n}}{s}\,\middle|\, \theta = \theta_0 + \delta\right\} \\
&= P\left\{t_{n_0-1} > t_{n_0-1,1-\alpha} - \frac{\delta\sqrt{n}}{s}\right\} \\
&= 1-\beta,
\end{aligned}
$$

provided $t_{n_0-1,1-\alpha} - \delta\sqrt{n}/s = -t_{n_0-1,1-\beta}$ where $\overline{X}$ denotes the sample mean. Now solving for n we obtain

$$
n = \frac{s^2\left[t_{n_0-1,1-\alpha} + t_{n_0-1,1-\beta}\right]^2}{\delta^2}. \qquad (1.3.12)
$$

Similarly in the two-sample case let $X \overset{d}{=} \text{normal}(\mu_1, \sigma^2)$, $Y \overset{d}{=} \text{normal}(\mu_2, \sigma^2)$ and X, Y be independent. Suppose we wish to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_2 > \mu_1$. Suppose we wish to have error probability α when H_0 is true and power $1-\beta$ when $\mu_2 - \mu_1 = \delta$. In the first stage we observe $(X_1, X_2, ..., X_{n_0})$ and $(Y_1, Y_2, ..., Y_{n_0})$ and compute

$$
s^2 = \frac{1}{2(n_0-1)}\left[\sum_{i=1}^{n_0}\left(X_i - \overline{X}_{n_0}\right)^2 + \sum_{i=1}^{n_0}\left(Y_i - \overline{Y}_{n_0}\right)^2\right]. \qquad (1.3.13)
$$

Then the total sample size to be drawn from each population is

$$
n = \max\left(n', n_0\right),
$$

where

$$n' = \frac{s^2\left[t_{2(n_0-1),1-\alpha} + t_{2(n_0-1),1-\beta}\right]^2}{\delta^2}. \tag{1.3.14}$$

Moshman (1958) has investigated the proper choice of the initial sample size n_0 in Stein's two-stage procedure and believes that an upper percentage point of the distribution of the total sample size, n when used in conjunction with the expectation of the sample size, is a rapidly computable guide to an efficient choice of the size of the first sample. However, the optimum initial sample that maximizes a given function involves an arbitrary parameter which has to be specified by the experimenter from non-statistical considerations.

If the initial sample size is chosen poorly in relation to the unknown σ^2, the expected sample size of Stein's procedure can be large in comparison to the sample size which would be used if σ^2 were known (which it is not). For example, this can occur if σ^2 is very small; then (if σ^2 were known) a small total sample size would suffice, but one may use n_0 much larger (hence being inefficient). However, this problem is not of practical significance.