The Lecture Contains:

- Application of renewal theory and renewal theory concepts
- Sequential sampling methodologies
- Two-stage sampling procedure
- Purely sequential sampling procedure
- Three-stage sampling procedure
- Accelerated sequential sampling procedure
- Batch sequential sampling procedure
- Jump crawl (JC) sequential sampling technique

Previous Next

Application of renewal theory and renewal theory concepts

Example 5.1

A very simple example and illustration for renewal theory which has wide application is in sampling, and the area is known as sequential sampling. So first let us illustrate the concept of sequential sampling methodology or also called multi-stage sampling techniques.

Before that, let us illustrate the concept of bounded risk and its implications with a simple example. This concept of bounded risk will facilitate our understanding of sequential sampling methodology in a much better way. Furthermore we will illustrate through detailed examples four different distributions using two different loss functions, which we are sure will help the reader appreciate the application of renewal theory through its use in sequential sampling procedure.

Consider a normal distribution with the probability distribution function (p.d.f.) given by $f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma < \infty$, with both mean (μ) and variance (σ^2) unknown. Suppose one is interested to find the point estimate of m (location parameter) subject to a squared error loss (SEL) function, given by L(Tn,q) = $(Tn-q)^2$. After recording X_1, X_2, ….., X_n observations, we find the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, which is the estimator of m. The corresponding associated risk is given by $R(\bar{X}_n,\mu)=E[L(\bar{X}_n,\mu)]=\frac{\sigma^2}{n}$. Next suppose we require this risk to be such that it does not exceed a pre-assigned known value, w (> 0). We immediately see that if σ^2 (square of scale parameter) is known then the optimal sample size, n, is $D=\left\lceil \frac{\sigma^2}{w} \right\rceil$. But with σ^2 unknown the problem cannot be solved with any fixed sampling techniques and hence we can take recourse to the use of sequential sampling procedures which are briefly discussed below.

Previous Next

Sequential sampling methodologies

One is aware that as σ^2 is unknown, $D = \left\lceil \frac{\sigma^2}{w} \right\rceil$ is also unknown. Hence using fixed sampling rule will not help us to solve our problem of finding the minimum sample size. This implies, one has to take the recourse of some multistage or adaptive sampling techniques to solve this problem. We discuss few of the multistage or adaptive sampling methodologies used in literature to circumvent such bounded risk estimation problem.

Two-stage sampling procedure: Consider a two-stage sampling procedure, where at the first stage a sample of size $m(\geq 2)$ is drawn to estimate the unknown quantity D which is calculated using N. Here $N = \max\left\{m, \left\lceil \frac{S_m^2}{w} \right\rceil\right\}$, is the estimate of the number of observations, needed to satisfy the bound placed by w. The methodology works as follows. Start with $X_1, X_2, \ldots, X_m$ observations in a single batch and determine N. If N = m, then we stop and do not take any more observations in the second stage. However if N > m, then one samples an additional (N - m) observations in the second stage. Based on the total observations $X_1, X_2, \ldots, X_N$, the estimator, $\overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ is calculated. We must remember since S_m^2 is the result of a random procedure, N, the sample size is also a random number.

Purely sequential sampling procedure: Next consider a purely sequential methodology, which starts with a sample of size $m(\geq 2)$ and one continues to take one observation at a time until, $N = \inf\left\{n \geq m : n \geq \frac{\sigma^2}{w}\right\}$, where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ is the best estimate of σ^2 which is recalculated each time the sample size, n, changes. In other words, the estimator is updated at each stage with the arrival of each new observation, until the stopping rule is met for the very first time. Once sampling stops the value of m is evaluated using its estimator $\overline{X}_N$. Thus it establishes the superiority of the purely sequential sampling procedure over the two-stage procedure from a statistical asymptotic viewpoint and not from the practical perspective.

Three-stage sampling procedure: Even though from the theoretical standpoint, purely sequential sampling procedure satisfies the asymptotic second-order efficiency property, yet one immediately realizes that taking one observation at a time, as is done in the purely sequential sampling scheme, is practically inconvenient. Hence in order to save sampling operations and at the same time maintain the second-order property, different authors have considered the three-stage sampling procedure. This methodology is as follows. Let $m = O\left(D^{\frac{1}{r}}\right)$, for some r > 1. That is, the starting sample size m is allowed to grow, but in a manner that $\left(\frac{m}{D}\right) \to 0$ as $w \to 0$, which implies that m is allowed to increase at a slower rate than D as w becomes smaller. After having fixed $0 < \rho < 1$ and with the starting sample size of $m(\geq 2)$, let

$$T = \max\left\{m, \left[\rho\left(\frac{S_m^2}{w}\right)\right]\right\}$$

$$N = \max\left\{T, \left[\frac{S_T^2}{w}\right]\right\}$$

Here T estimates $\rho \times D$, which is a fraction of D. If T = m, then we do not sample any more in the second stage, but if T > m, one samples the difference (T - m) in one single batch. Based on the observations {X_1, X_2, ….., X_T} one now proceeds to find N which is the estimator of D. If now N = T, then we do not take any more sample in the third stage, but if N > T, the difference (N - T) of observations are taken in the third stage to find the value of N. After the sampling procedure terminates,the estimator $\overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ determines the estimated value of μ (the location parameter for the normal distribution). One must remember that even if there is a huge amount of variability in the last (N - T) observations, still we are certain to terminate our sampling procedure following the same stopping criteria. If the variability in the last (N - T) observations is appreciable, then the number of observations one needs to take in the third, i.e., the last stage would be quite high. It is observed that such a three-stage procedure apart from obeying the asymptotic consistency and asymptotic first order efficiency properties, also obeys the asymptotic second order efficiency property.

Previous Next

Module :Understanding of applications of renewal theory, Stationary Process with discrete and continuous parameters

Lecture 19:Application of Renewal Theory

Accelerated sequential sampling procedure: Another variation of the purely sequential sampling methodology cuts down on the cost by accelerating the original sequential procedure. Here also one starts with a sample size of $m(\geq 2)$ and after having fixed $0 < \rho < 1$ and with $S_n^{*2} = \left(\frac{n}{n-1}\right) S_n^2$ define

$$R = \inf\left\{n \geq m : n \geq \rho \frac{S_n^{*2}}{w}\right\}$$

$$T = \left\lceil \frac{S_R^{*2}}{w} \right\rceil$$

$$N = \max\{R, T\}$$

Thus one first samples purely sequentially obtaining X_1, X_2,....., X_R such that R estimates $\rho \times D$ and then proceeds to estimate D by N. If T = R, then we do not take any more samples, but if T > R, then one samples (T - R) observations in one single batch thus curtailing sampling operations and comes up with the estimate of the location parameter μ. The asymptotic second-order properties of such accelerated sequential procedures have also been developed by different authors.

Batch sequential sampling procedure: In batch sequential sequential sampling procedure, we first consider $0 < \rho_1 < \rho_2 < \rho_3 < \cdots . < \rho_{k-1} < 1.$. We also specify $r_1 \geq r_2 \geq r_3 \geq ... \geq r_k \geq 1$ and t_i's, where r_i (i = 1, 2, ..., k) denotes the minimum number of observations one takes at each and every step in the i^{th} batch, while t_i, is the number of such steps one is required to take in that ith batch. The connotation of minimum number of observations means the number of observations or individuals one takes at one go. Thus if we have k number of batches, then for the ith (i = 1, 2, ..., k) batch, the number of observations one would take is $t_i \times r_i$, and for the whole batch sequential sampling procedure it is $\left(m + \sum_{i=1}^{k} t_i \times r_i\right)$, where $m(\geq 2)$ is the number of observations required to initiate the batch sequential sampling procedure. Remember, this $m(\geq 2)$ number of observations is taken at one go in the first step which is literally the zero batch. One should also remember that r_1, r_2, r_3,....., r_k and $t^i \in Z^+$. . The procedure works as follows. Start with a sample size of $m(\geq 2)$ and for each batch follow the sampling methodology according to the rule give below

$$R_1 = \inf\left\{n \ge (m + r_1 \times t_1) : n \ge \rho_1 \frac{S_n^2}{w}\right\} \qquad batch\,\#\,1 \qquad estimates\ \rho_1 D$$

$$R_2 = \inf\left\{n \ge (R_1 + r_2 \times t_2) : n \ge \rho_2 \frac{S_n^2}{w}\right\} \qquad batch\,\#\,2 \qquad estimates\ \rho_2 D$$

M

$$R_{k-1} = \inf\left\{n \ge (R_{k-2} + r_{k-1} \times t_{k-1}) : n \ge \rho_{k-1} \frac{S_n^2}{w}\right\} \qquad batch\,\#\,(k-1) \qquad estimates\ \rho_{k-1} D$$

$$N = \inf\left\{n \ge (R_{k-1} + r_k \times t_k) : n \ge \frac{S_n^2}{w}\right\} \qquad batch\,\#\,k \qquad estimates\ D$$

Previous Next

Module :Understanding of applications of renewal theory, Stationary Process with discrete and continuous parameters

Lecture 19:Application of Renewal Theory

Jump crawl (JC) sequential sampling technique: To illustrate the jump crawl sequential sampling methodology we first give the schematic diagram of the procedure in Figure 5.1.

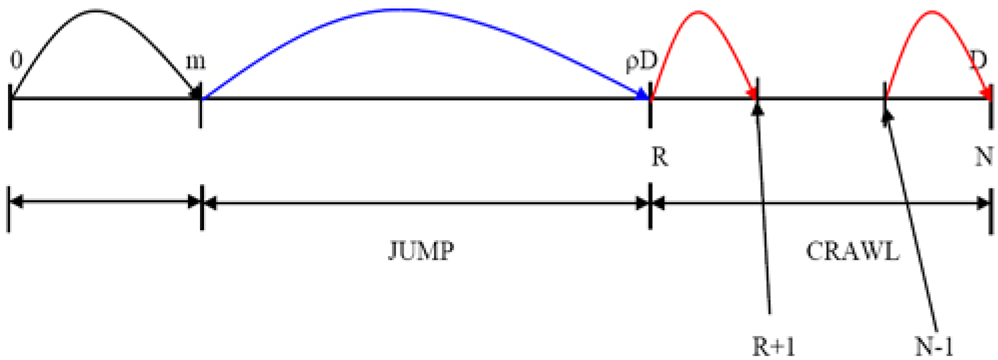

Figure 5.1: Jump crawl (JC) sequential sampling technique

The methodology works as follows. We start with an initial sample of size, $m(\geq 2)$ and also choose a value of $\rho\,(0 < \rho < 1)$. After that we jump by collecting a large sample of observations at one go to estimate $\rho \times D$ (using R), keeping in mind that, $R = \max\left\{m, \left[\rho\left(\frac{S_m^2}{w}\right)\right]\right\}$. Once $\rho \times D$ is estimated we check whether we have completed our sampling procedure. If not, we proceed purely sequentially, i.e., take one observation at a time following the rule $T = \inf\left\{n \geq R : n \geq \left(\frac{S_n^2}{w}\right)\right\}$. Finally the random sample, $N = \max\{R, T\}$ is found which estimates, D. One should be aware that the choice of ρ depends on the compromise one makes between efficiency of the result versus ease of sampling. A high value of ρ means we reduce our sampling effort but at a cost which results in a larger value of the estimated sample size, N. On the other hand for a low ρ, the reverse holds true where N as well as the estimate results are close to the optimal value but the effort one spends in conducting the experiment is high.

Previous Next