9. POINT AND INTERVAL ESTIMATION

Study time: 40 minutes

Aim - you will be able to

- explain the properties of the point estimation
- construct interval estimations for mean, standard deviation and variance

Explication

9.1. Introduction

The estimation problem is distinguished from hypothesis testing. In hypothesis testing we had a preference towards null hypothesis and only rejected it in face of strong contrary evidence. In the case of estimation, all parameter values or potential hypotheses are equal and we want to choose as our estimates those values which are supported by or consistent with the data. An estimate by itself is just a number. Anyone can make an estimate. To be usable, the accuracy of the estimate must also be known. Therefore in addition to deriving estimates, we must also make some assessment of the error of estimation.

9.2. Interval Estimation

The objective of interval estimation is to find an interval of values which have a high likelihood or probability of containing the true parameter values. The strategy used is to find those values which would have a large p-value if they had been chosen as the null hypothesis, i. e. those parameter values which are not inconsistent with the data. In order to give a probability interpretation to the data, we usually choose a fixed p-value, either one-sided or two-sided depending on whether we want a one or two sided interval, and then include in our interval all parameter values whose p-value for the observed data exceeds the chosen minimum p-value, α. The probability of the sample having a p-value which exceeds the selected p-value is $1-\alpha$, and therefore the probability that the interval so constructed will include the true parameter value is also $1-\alpha$. We call the value $1-\alpha$ the confidence level of the interval.

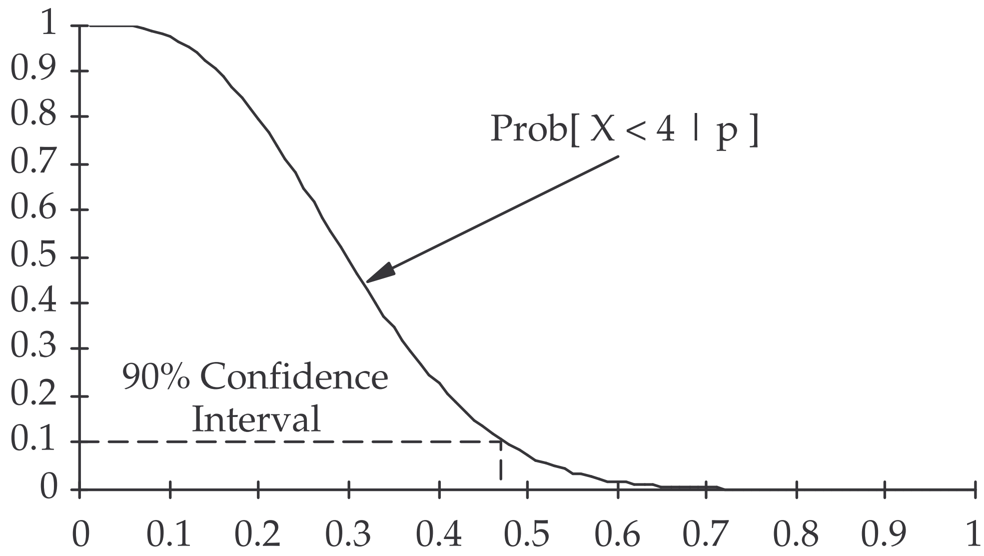
Consider the example of sampling semi-conductor devices to determine the proportion of defective devices produced. In this case, suppose it is a new process and we wish to estimate a maximum value for the proportion of defectives. A sample of 12 devices is selected and tested. Three are found to be defective. Since we are interested in an upper bound, we ask how large the true proportion of defectives can be before our observed sample has a very small probability. For some proportion, p, of defective devices, the probability of obtaining less than 4 defectives in a sample of 12 is

$$\Pr ob\left[\, X < 4 \mid p \,\right] = \sum_{x=0}^{3} \binom{12}{x} p^x (1-p)^{12-x}$$

To obtain a (1-α) upper for p, we find the value of p such that the p-value is exactly α .

$$\alpha = \sum_{x=0}^{3} \binom{12}{x} p^x (1-p)^{12-x}$$

The following diagram illustrates the p-value as a function of the proportion of defectives, finds the value of p whose p-value is α = 0.1, and identifies the 90% upper confidence limit for p.

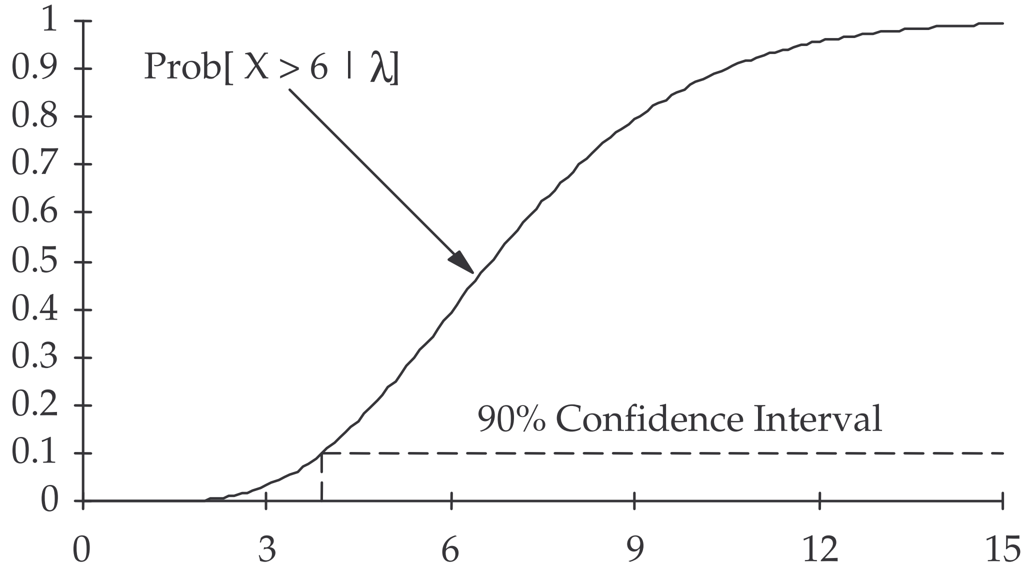


The probability that a population with 47% defectives will have less than 4 defectives in a sample of 12 is 10%. Thus, we are 90% confident that the true proportion of defectives is less than 47%. Another way of expressing this idea is to say that 90% of confidence intervals calculated by this methodology will include the true proportion of defectives.

Consider a second example. A firm which assembles PC's from basic components, loads the software, and tests the system before delivery is interested in estimated how long it takes a worker to complete preparation of a PC for delivery. They observe a worker for 4 hours, one half of his daily work period. In that time, the worker completes 7 PC's. If we assume that the time to complete a single PC is exponentially distributed then the number of PC's completed in 4 hours will have a Poisson distribution. The firm is interested in estimating an upper bound for the mean time to complete a PC or equivalently a lower bound for the rate at which PC's are completed. Therefore we ask how low the rate λ can be before the probability of our sample result, more than 6, has a very small probability. For a given value of λ , the p-value of our sample is

$$\Pr ob\left[X > 6 \mid \lambda \right] = \sum_{x=7}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!}.$$

The following diagram illustrates the p-value as a function of the completion rate λ, finds the value of p whose p-value is α=0.1, and identifies the 90% upper confidence limit for λ.



If the rate of preparing PC's for delivery is per 4 hour period, then the probability of completing 7 or more PC's in 4 hours is 10%. Therefore, we are 90% confident that the true rate of preparation is at least 3.9 computers per 4 hours, or slightly less than one computer per hour. Alternatively the mean time to complete each PC is no more than 61 minutes 32.3 seconds. This is obviously a conservative estimate since in our sample, computers were completed at the rate of 7 per 4 hours or 1.75 per hour with an equivalent mean preparation time of 34 minutes 17 seconds. To obtain a less conservative estimate at the same confidence level, a larger sample size is required.

This analysis depends on the assumption that the time to complete a PC is exponentially distributed. In practice this is unlikely to be a very good model because in theory according to the exponential distribution, the PC could be completed instantaneously. The Poisson process is a more appropriate model for events which occur randomly such as traffic accidents.

Now consider an example where we wish to estimate both an upper and lower bound for the parameter. In this case, we use the p-value for testing hypotheses against two-sided alternatives. The 1-α confidence interval is the set of all parameter values having a p-value greater than α.

Files transmitted via computer networks are often bundled into groups of files having similar network pathways. Into order to determine the optimal number of files to include in a single bundle, network engineers need some estimate of the distribution of file size. A sample of 15 files is taken with the following result. Sizes are in MB units

4.027	1.887	3.806	7.018	2.753
5.956	8.117	2.857	4.525	7.282
0.140	6.186	5.171	10.558	5.534

The summary statistics for these data are

Mean	5.055	Standard. Deviation.	2.646
Median	5.171	1.483*MAD	2.739
Shorth	3.806	to	7.018

For any hypothetical value of the true mean file size, we can compute the t-statistic for our observed sample.

$$t_{n-1}(\mu) = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

and its associated p-value. Since in this case, we want both upper and lower bounds for our estimate of μ, we use the two-sided definition of p-value.

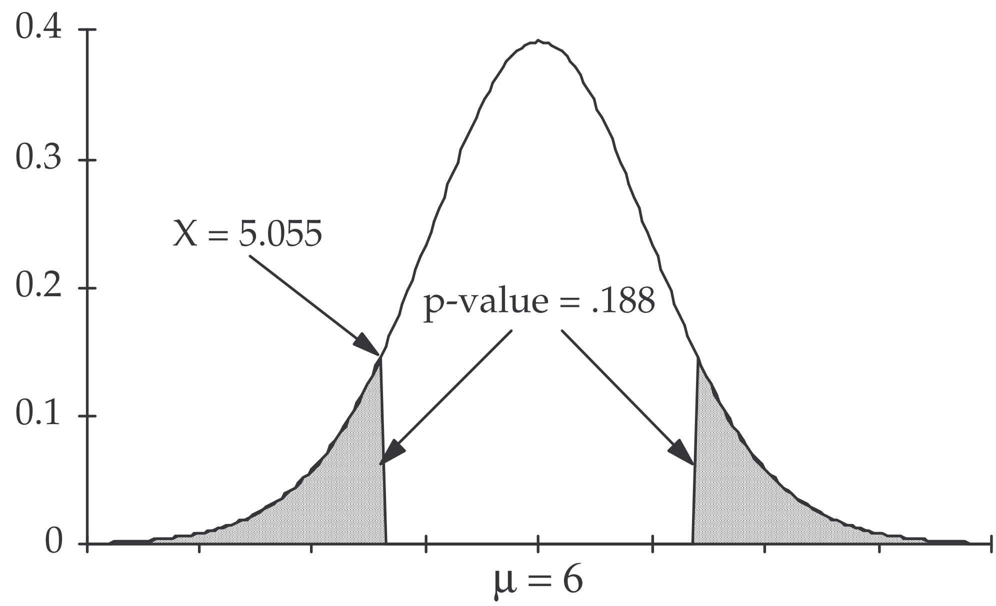
$$p(\bar{x}|\mu) = 2\min\{F_{n-1}[t(x|\mu)], 1-F_{n-1}[t(x|\mu)]\}$$

For example for a hypothetical value of μ = 6, the observed t-value is

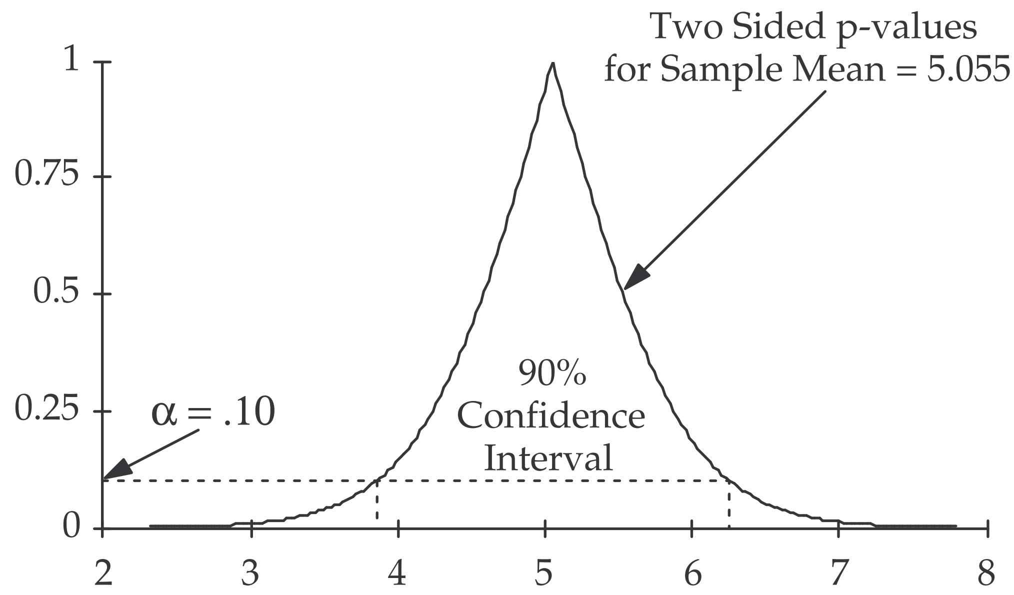
$$t_{14} = \frac{5.055-6}{2.646/\sqrt{15}} = -1.384$$

and the associated p-value is

$$p(\bar{x}|\mu) = \Pr ob[|t_{14}| \geq 1.384] = 0.188$$



Thus, we can determine the p-value associated with any value of μ. The (1-α) confidence interval will consist of all values of whose p-value is greater than α. The following chart shows the two-sided p-value at different values of μ. The p-value reaches its maximum value of 1 when μ = 5.055, the sample mean. If we include in our interval estimate all values of μ having a p-value of at least 10%, then we cane be 90% confident that the interval estimate contains the true mean value in the sense that 90% of interval estimates derived by this procedure contain the true value of the mean. We call such an interval a 90% confidence interval. In this example, the 90% confidence interval for mean file size



is the range (3.852, 6.258). Notice that μ = 6 with a p-value of 0.188 is included in the 90% confidence interval.

9.3. Construction of Confidence Intervals

There is a simple procedure for constructing confidence intervals for parameters whose test statistic has a symmetric distribution, such as the Student's t or the normal. This procedure eliminates the need to compute the p-value for every value of μ. To construct a (1-α) confidence interval, we need only determine the upper and lower bounds of the interval. The lower bound will be that value of μ less than the sample mean whose p-value is exactly α. Therefore the t-value of the sample mean with respect to the lower bound must be equal to the (1-α /2) percentage point of the Student's t distribution.

$$t_{n-1,\alpha/2} = \frac{\bar{x} - \mu_{Lower}}{s/\sqrt{n}}$$

Solving for μ_{Lower}, we have

$$\mu_{Lower} = \bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

Similarly μ_{Upper} must satisfy the equation

$$t_{n-1,\alpha/2} = \frac{\mu_{Upper} - \bar{x}}{s/\sqrt{n}}$$

Hence the (1 - α) confidence interval is given by

$$(\mu_{Lower}, \mu_{Upper}) \Leftrightarrow \bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

In our example of files sizes, the 90% confidence interval is

$$\bar{x} \pm t_{14,.05}\frac{s}{\sqrt{n}} \Leftrightarrow 5.055 \pm 1.761\frac{2.646}{\sqrt{15}} \Leftrightarrow 5.055 \pm 1.203$$

As previously determined graphically, this interval is (3.852, 6.258).

9.4. Sample Size Determination

As well as giving a range of reasonably good parameter values, an interval estimate also provides information about the quality of the estimated values. The quality of an estimate has two aspects,

1) Accuracy

2) Reliability

The accuracy of a interval estimate is equivalent to the length of the interval. The smaller the interval, the greater the accuracy. Reliability is given by the confidence level of the interval. However, as for Type I and Type II errors of hypothesis tests, accuracy and reliability of a confidence interval are in conflict. For a fixed sample size, the confidence level can only be increased by increasing the length of the interval thereby reducing its accuracy. Increasing both the accuracy and reliability can only be achieved by increasing the sample size.

Determining the sample size required to construct an interval estimate having some fixed reliability and accuracy is a problem which arises commonly in practice. Suppose it is necessary to estimate a mean to an accuracy of Δ with a reliability of (1-α). That is, we require a (1-α) confidence interval of length not exceeding 2Δ. Then the sample size must be chosen large enough to satisfy the following inequality.

$$\Delta \geq t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

and therefore the sample size must exceed

$$n \geq \left[\frac{t_{n-1,\alpha/2}\, s}{\Delta} \right]^2$$

In practice it is common to substitute a conservative estimate for s and substitute the z quantile for the t quantile on the assumption that the required sample size will be large enough that the applicable t distribution will be approximately normal. If we wished to estimate file size to an accuracy of 256 KB, or .25 MB with 90% confidence, using a conservative estimate of 3 for s, we would require a sample size of

$$n \geq \left[\frac{z_{\alpha/2}\, s}{\Delta} \right]^2 = \left[\frac{1.645 * 3}{.25} \right]^2 = 389.67$$

9.5. Point Estimation

Randomness is difficult and unpopular. We are used to have specific answers to questions. An interval of estimates can be unsatisfying. We may ask "What is the single best point or value in the interval?" Such a single value would be a point interval. The single best value is clearly the value which has the highest p-value for the observed data. This point estimate is called the maximum likelihood estimate. But notice that as the p-value increases and the size of the confidence interval decreases, the confidence level decreases accordingly. In most cases, the maximum p-value will be 1, so the confidence level will be zero. That is, the point estimate will never be exactly correct. Therefore in this case it is extremely important to estimate the error of estimation.

While choosing the single point estimate to be the parameter value assigning the maximum p-value to the observed data is a logical consequence of the method of constructing confidence intervals, it may have indeterminate or non-unique solutions in certain cases, particularly for discrete random variables. Therefore, point estimates are determined by a method similar in spirit to maximizing p-values, the method of maximum likelihood. Rather than maximizing the p-value, maximum likelihood point estimates seek the parameter value which maximizes the probability mass or probability density of the observed sample data.

9.6. Maximum Likelihood Estimation

Because statistics measure general distributional properties such as location and scale, means, medians, and standard deviations can be applied to any distribution. But estimators are associated with parameters for a specific distribution. Developing satisfactory estimators for every individual distributional form would become a daunting task without some general procedure or approach. Fortunately, the idea of likelihood offers such an approach. Intuitively, if the conditional probability or likelihood of the observed data is greater for one parameter value than another, then the first parameter value is a preferred estimate of the population parameter. By extension, the best choice of estimate for the population parameter should be the parameter value whose likelihood is maximum.

$$\hat{\theta} = \max_{\theta} \Pr ob\left[x \mid \theta\right] = \max_{\theta} f(x, \theta)$$

An estimator derived by this criterion is called a maximum likelihood estimator or MLE and is always denoted with a small cap over the parameter symbol as indicated.

Consider the case of sampling for an attribute. If X is the number of items in a sample of size n having the desired attribute, then X will have a binomial distribution with parameters n which is known from the sampling procedure and p which is unknown. The likelihood of p for the observed X is the conditional probability of X given p.

$$f(p \mid x) = \Pr ob[x \mid p] = \binom{n}{x} p^x (1-p)^{n-x}$$

We can find the maximum likelihood estimate of p by finding the point of the likelihood function having zero slope. That is, by solving the following equation.

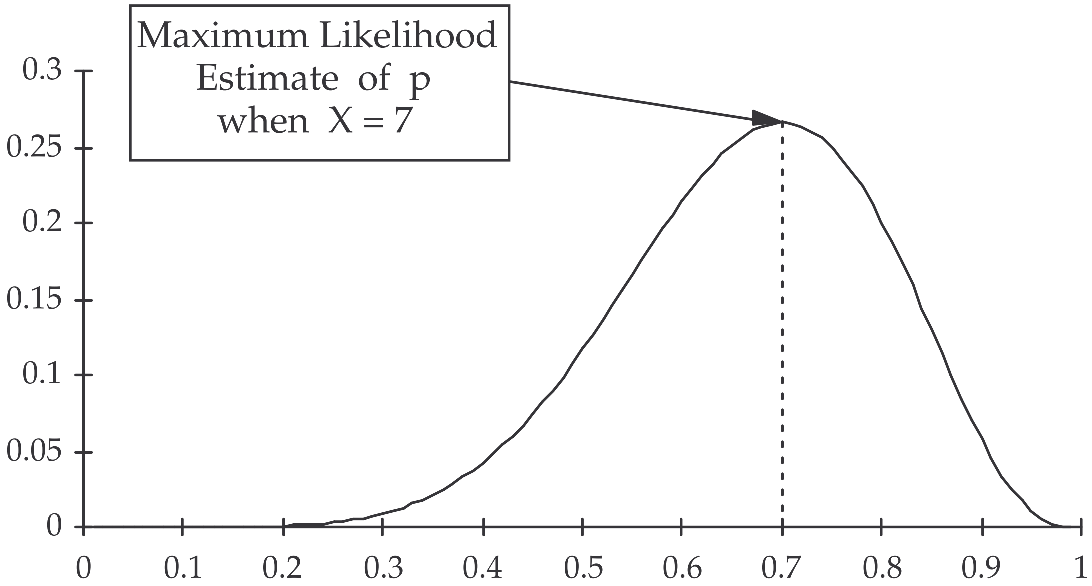
$$\frac{\partial f(p \mid x)}{\partial p} = \frac{\partial \binom{n}{x} p^x (1-p)^{n-x}}{\partial p}$$

$$= \binom{n}{x} \left[\mathrm{x} p^{x-1} (1-p)^{n-x} - (n-x) p^x (1-p)^{n-x-1}\right] = 0$$

The solution is the sample proportion. We know from the sampling distribution of this estimator that it is unbiased. It is also consistent, sufficient, and efficient among unbiased estimators.

$$\hat{p} = \frac{x}{n}$$

The maximum likelihood estimate for n = 10 and X = 7 is illustrated below.



The maximum likelihood estimator of λ for the Poisson distribution is derived in the same fashion.

$$f(\lambda \mid x) = \mathrm{Prob}[x \mid \lambda] = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

Solving for the value of λ where the slope of the likelihood function is zero

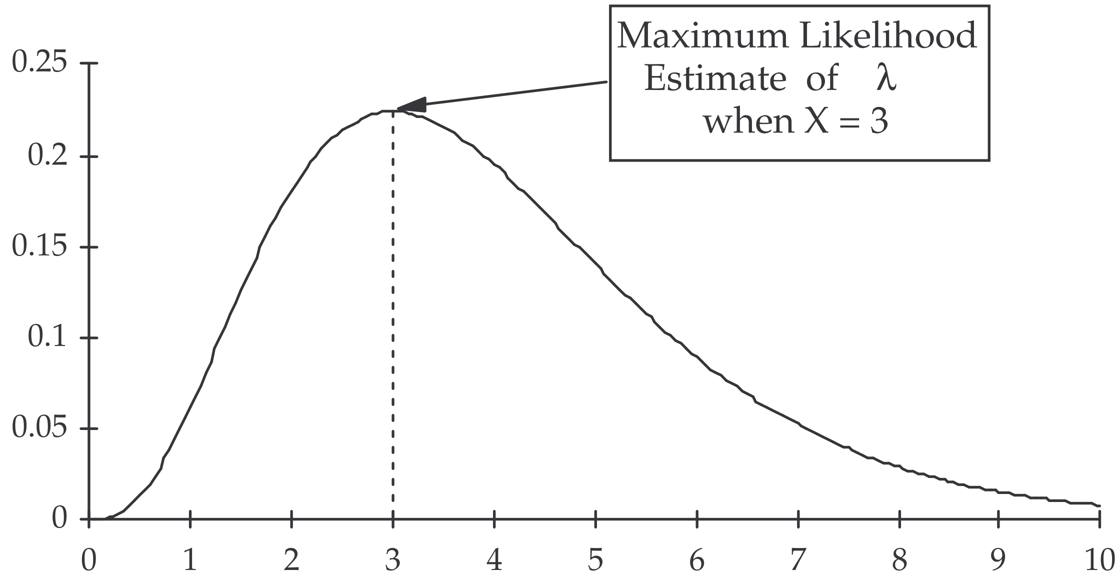
$$\frac{\partial f(\lambda \mid x)}{\partial \lambda} = \frac{\partial \frac{(\lambda t)^x e^{-\lambda t}}{x!}}{\partial \lambda}$$

$$= \frac{x(\lambda t)^{x-1} t e^{-\lambda t} - (\lambda t)^x t e^{-\lambda t}}{x!} = 0$$

we find

$$\hat{\lambda} = \frac{x}{t}.$$

The maximum likelihood estimate of λ when X = 3 and t = 1 is illustrated below.



The maximum likelihood estimator of λ is also unbiased, efficient, consistent, and sufficient for the Poisson distribution.

For the normal distribution, the maximum likelihood estimate of μ is obtained by minimizing the value in the exponent of the density.

$$\min_{\mu} \sum_{i=1}^{n} (x_i - \mu)^2$$

The solution is the sample mean.

9.7. Estimation, Estimators, Properties of Estimator

What is an Estimate or an Estimator?

Formally, an estimator is a statistic defined on the domain of the sample data which is used as an estimate a population parameter. An estimate is the value of that statistic for a particular sample result. Since every statistic has a theoretical probability distribution for every hypothetical probability distribution of the population, it is possible to examine the properties of an estimator through its probability distributions. Since many estimators are based on likelihood functions, many of the properties of the likelihood function will also be exhibited in the probability distribution.

Because of the special requirements of an estimator, criteria particular to the problem of estimation have been proposed as means of evaluating the suitability of a statistic as an estimator, of comparing competing estimators, and of developing new and improved estimators. You will not that the following criteria can only be applied to statistics which are required to be close to some parameter, hence to estimators.

a) Consistency

Consistency is generally agreed to be an essential characteristic of an estimator. As the sample size increases, an estimator which is consistent will have smaller and smaller probability of deviating a specified distance from the parameter being estimated. In the limit of an infinitely large sample, the value of the estimator will be equal to the estimated parameter with probability one. An estimator which does not have this property would give more reliable results for smaller samples and therefore could not be using the information in the sample consistently.

b) Sufficiency

Every statistics is a reduction or summarization of the original sample data and as such discards some information contained in the original sample data. An estimator which is sufficient does not discard any information relevant to the parameter being estimated. This may seem at first to be a rather vague requirement but in fact sufficiency has a very specific probabilistic definition. Any statistic creates a partition of the sample space. All elements within any partition lead to the same value of the statistic. If the relative or conditional probabilities of the individual sample space elements are independent of the parameter being estimated, then the partition and the estimator which created it are sufficient. For example, the binomial random variable partitions the sample space of n Bernoulli trials into subsets all having the same number of successes. The probability of two elements having the same number of successes is always the same no matter what the value of p, the probability of success. Therefore no further information about p can be obtained by knowing which element in the partition actually occurred. Therefore, the sample proportion is a sufficient statistic or estimator of p. We always try to work with sufficient statistics.

c) Bias

Bias or unbiasedness concerns the expected or mean value of the statistic. The statistic should be close to the parameter being estimated and therefore its mean value should be near the parameter value. Bias is the difference between the mean of the estimator and the value of the parameter. Bias should be small. If bias is zero, we say the estimator is unbiased. Unbiased estimators are not always preferable to biased ones if they are not sufficient or have larger variance. An estimator need not be unbiased in order to be consistent.

d) Efficiency

As has be said several times and reinforced by the orientation of the preceding criteria, an estimator should be close to the parameter being estimated. Simply being unbiased will not insure that the estimator is close to the parameter. The variance of the sampling distribution of the estimator must be small as well. For two estimators, the estimator whose sum of variance and squared bias is smaller is said to be more efficient. If the two estimators are unbiased, then the estimator with smaller variance is more efficient. If an unbiased estimator has minimum possible variance for all unbiased estimators, then it is said to be efficient.

Summary of notions

From a methodological point of view we use two kinds of parameters estimations. It is a **point estimation** where distribution parameter is approximated by a number and so called **interval estimation** where this parameter is approximated by an interval where the parameter belongs with a high probability. **Unbiassed, consistent and efficient** estimations of parameters are the most important for a quality of point estimation.

In the case of interval estimation of a parameter we can search for **two-sided** or **one-sided** estimations.

Questions

1. What is the consistent estimation of parameter?
2. How we can describe 100.(1-α) % two-sided confidence interval for a θ-parameter?

Problems

Example 1: In random sample of chipsets there is a 10% not suitable for new quality demands. Find 95% confidence interval for a p-chipset proportion not suitable for a new norm if a sample size is:

a) n = 10
b) n= 25
c) n = 50
d) n = 200

{ *Answer:* a) $-0.06<p<0.26$, b) $0.00<p<0.20$, c) $0.03<p<0.17$, d) $0.07<p<0.13$}

Example 2: Four students were random selected from the first parallel group. Their exam results were 64, 66, 89 and 77 points. Three students were random selected from the second parallel group. Their exam results were 56, 71 and 53 points. Find 95% confidence interval for difference between means values of exam results.

{ *Answer:* (-4;32)}