ACTEX CAS EXAM 3 STUDY GUIDE
FOR MATHEMATICAL STATISTICS

TABLE OF CONTENTS

INTRODUCTORY NOTE

This study guide is designed to help in the preparation for the mathematical statistics part of Exam 3 of the Casualty Actuarial Society. The study guide consists of a section of review notes, examples and problem sets, and a section with 10 practice tests.

Many of the examples in the notes, problems in the problem sets and questions on the practice tests are from CAS or SOA exams on the relevant topics. ACTEX gratefully acknowledge the permission granted by the CAS and SOA allowing the use of exam questions in this study guide.

The practice tests each have 8 questions. Recent exams have had about 8 or 9 questions on the mathematical statistics topic. The approximate time for a practice test is 48 minutes, which follows the average of 6 minutes per question on an actual exam.

Because of the time constraint on the exam, a crucial aspect of exam taking is the ability to work quickly. I believe that working through many problems and examples is a good way to build up the speed at which you work. It can also be worthwhile to work through problems that have been done before, as this helps to reinforce familiarity, understanding and confidence. Working many problems will also help in being able to more quickly identify topic and question types. I have attempted, wherever possible, to emphasize shortcuts and efficient and systematic ways of setting up solutions.

In order for the review notes in this study guide to be most effective, you should have some background at the junior or senior college level in probability and statistics. It will be assumed that you are reasonably familiar with differential and integral calculus.

Of the various calculators that are allowed for use on the exam, I think that the BA II PLUS is probably the best choice. Along with a multi-line display it has several memories. I think that the TI-30X IIS would be the second best choice.

If you have any questions, comments, criticisms or compliments regarding this study guide, you may contact me at the address below. I apologize in advance for any errors, typographical or otherwise, that you might find, and it would be greatly appreciated if you would bring them to my attention. ACTEX will be maintaining a website for errata that can be accessed from www.actexmadriver.com . It is my sincere hope that you find this study guide helpful and useful in your preparation for the exam. I wish you the best of luck on the exam.

Samuel A. Broverman
Department of Statistics
University of Toronto
100 St. George Street
Toronto, Ontario CANADA M5S 3G3
E-mail: sam@utstat.toronto.edu or 2brove@rogers.com

October, 2009

NOTES

AND

PROBLEM SETS

SECTION 1 - POINT ESTIMATION

The suggested time frame for covering this section is 3 hours.

Random Samples And The Sampling Distribution

A **random sample of size N** from the distribution of the random variable X is a collection of independent observations from the distribution of X, say $X_1, X_2, ..., X_N$. We can interpret the sample in two ways:

(i) it is a collection of numerical values, and

(ii) it is a collection of independent random variables.

Each interpretation is useful depending upon the way in which the random sample is being used.

The **sample mean** of the random sample is an **estimator** of the distribution mean μ_X,

and is $\overline{X} = \frac{X_1+X_2+\cdots+X_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$.

Under the two interpretations of random sample mentioned above, the sample mean $\overline{X}$ is either

(i) a number (representing an estimate of the mean of the distribution of X), or

(ii) a random variable (if each of the X_i's are interpreted as random variables)

Interpreted as being a random variable, the sample mean $\overline{X}$ has a **sampling distribution**. In particular, $\overline{X}$ has a mean and variance given by the following expressions:

$E[\overline{X}] = \mu_X = E[X]$ ($\overline{X}$ is an **unbiased estimator** of μ_X, and $\overline{X}$ has the same mean as X)

$Var[\overline{X}] = \frac{\sigma_X^2}{N}$ (this is the **variance of the sample mean**)

The **sample variance** of the random sample (not to be confused with the variance of the sample mean) is an **unbiased** estimator of σ_X^2 and has **$N-1$ degrees of freedom**:

$s_X^2 = V\widehat{ar}[X] = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2$ (also denoted s^2 or $\widehat{\sigma}_X^2$) .

As with the sample mean, the sample variance can be interpreted as a numerical value under interpretation (i) of a random sample, and can be interpreted as a random variable under interpretation (ii). To say that the estimator s_X^2 is unbiased is to say that when regarding s_X^2 as a random variable, the mean of s_X^2 is $Var[X] = \sigma_X^2$; in other words, $E[s_X^2] = \sigma_X^2$.

Note the following identity that may be useful in writing an expression for the sample variance

$$\sum_{i=1}^{N}(X_i - \overline{X})^2 = \sum_{i=1}^{N} X_i^2 - N\overline{X}^2$$

The notion of degrees of freedom in an estimator arises in a number of situations (usually in an estimator of a variance). In general, the degrees of freedom of an estimate is equal to the number of independent data points used in the estimate minus the number of parameters estimated at intermediate steps in the estimation of the parameter itself. For instance, the estimator s_X^2 given above uses N independent data points, but the mean of X is estimated to be $\overline{X}$ at an intermediate step. Therefore the degrees of freedom in the estimator s_X^2 is $N-1$ (the number of independent data points N, minus 1, for the one parameter (the mean) estimated at the intermediate step).

Bias And Consistency In Estimators

If a parameter β from a probability distribution is estimated using the estimator $\widehat{\beta}$ ($\widehat{\beta}$ itself will be a random variable and will have a sampling distribution), then the **estimator is called unbiased** if $E[\widehat{\beta}] = \beta$. The sample mean and unbiased form of the sample variance found from a random sample are examples of unbiased estimators. If the estimator $\widehat{\beta}_n$ is based on n sample points, and if $\lim_{n\to\infty} E[\widehat{\beta}_n] = \beta$, the estimator is called **asymptotically unbiased.**

The **bias associated with an estimated parameter** is $\mathbf{Bias}[\widehat{\boldsymbol{\beta}}] = \boldsymbol{E}[\widehat{\boldsymbol{\beta}}] - \boldsymbol{\beta}$.

The **mean square error** of the estimator $\widehat{\beta}$ is $\boldsymbol{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2] = (\mathbf{Bias}[\widehat{\boldsymbol{\beta}}])^2 + \boldsymbol{Var}[\widehat{\boldsymbol{\beta}}]$. If an estimator is unbiased, then the mean square error of the estimator is equal to the variance of the estimator, since the bias is 0. This will be the case for the sample mean taken as an estimate of the distribution mean and the (unbiased) sample variance taken as an estimate of the distribution variance.

$\widehat{\beta}$ is an **efficient unbiased estimator** if the variance of $\widehat{\beta}$ is smaller than the variance of any other unbiased estimators of β based on the same sample size. It might also be referred to as a uniformly minimum variance unbiased estimator (UMVUE).
If $\widehat{\beta}$ satisfies the relation $\lim_{N\to\infty} \text{Prob}(|\beta - \widehat{\beta}| < \delta) = 1$ for any number $\delta > 0$, then $\widehat{\beta}$ is called a **consistent estimator** of β. When the limit is satisfied, we say that **plim $\widehat{\boldsymbol{\beta}}$ is $\boldsymbol{\beta}$**. If $\widehat{\beta}_n$ is asymptotically unbiased and if $\lim_{n\to\infty} Var[\widehat{\beta}] = 0$, then $\widehat{\beta}_n$ is a consistent estimator.

A **robust estimator** is one that retains its usefulness even when one or more of its assumptions is violated.

Example ST-1: Suppose that the random variable X has variance σ^2. The "population variance" estimator of σ^2 for a random sample $X_1, ..., X_N$ is $\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2$. We find the bias in this estimator.
Solution: The unbiased estimator of σ^2 is $\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2$, and therefore
$E[\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2] = \frac{N-1}{N} \cdot E[\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2] = \frac{N-1}{N} \cdot \sigma^2$.
The bias is $\frac{N-1}{N} \cdot \sigma^2 - \sigma^2 = -\frac{1}{N}\sigma^2$. □

The "Loss Models" book gives a number of good examples in Section 9.2 to illustrate the concepts described above. The following is a summary of some of the examples presented there.

- for any distribution, the sample mean is an unbiased estimator of the distribution mean

- for the uniform distribution on the interval $(0, \theta)$, and sample $X_1, X_2, ..., X_n$, the estimator $\widehat{\theta} = max(X_1, X_2, ..., X_n)$ has mean $E[\widehat{\theta}\,] = \frac{n\theta}{n+1}$ and is a biased estimator of θ but it is asymptotically unbiased and consistent (the bias is $E[\widehat{\theta}\,] - \theta = \frac{n\theta}{n+1} - \theta = \frac{-\theta}{n+1}$)

Moment Estimation Applied to Parametric Distributions

For any random variable, the mean and variance (skewness, etc.) are sometimes referred to as "parameters" of the distribution (some parameters can be infinite) that can be estimated when a sample of data is available.

There is also the notion of a **parametric distribution**, which means that the random variable X has a pdf (or cdf) which is formulated in terms of algebraic parameters. For example,

(i) the uniform distribution on the interval $[\,0\,,\,\theta\,]$ has pdf $f(x) = \begin{cases} 1/\theta & \text{if } 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}$; this is a parametric distribution with parameter $\theta > 0$ (the mean is $\frac{\theta}{2}$, and other distribution quantities such as variance, skewness, etc. are formulated in terms of the parameter θ).

(ii) the normal distribution has parameters μ (mean) and $\sigma > 0$ (standard deviation) and has pdf $f(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}$ for $-\infty < x < \infty$.

(iii) the Poisson distribution with parameter $\lambda > 0$ is a discrete, non-negative integer-valued random variable with pf $p(x) = P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0, 1, 2, ...$

There are a large number of parametric distributions in the Exam 3 exam tables.

Given a data set, and the assumption that the data comes from a particular parametric distribution, the objective is to estimate the value(s) of the distribution parameters. For instance, if the distribution is exponential with parameter θ then the objective is to estimate θ, or if the distribution is Pareto with parameters α and θ then the objective is to estimate α and θ.

For a distribution defined in terms of r parameters $(\theta_1, \theta_2, ..., \theta_r)$, the method of moments estimator of the parameter values is found by solving the r equations:

theoretical j-th moment $=$ empirical j-th moment , $j = 1, 2, ..., r$,

(or equivalently $E[X^j|\theta] = E[\widehat{X}^j]$, $j = 1, 2, ..., r$) .

If sample data is given in the form of a random sample $x_1, x_2, ..., x_n$ then the j-th empirical moment is $\frac{1}{n}\sum_{i=1}^{n} x_i^j$, for $j = 1, 2, ..., r$.

If the distribution has only one parameter θ, then we solve for θ from the equation

theoretical distribution first moment = empirical distribution first moment ,
(so that $E[X|\theta] = \frac{1}{n}\sum_{i=1}^{n} x_i$ if a random sample was given).

If the distribution has two parameters θ_1 and θ_2, then we solve two equations,

$\boldsymbol{E[X|\theta] = E[\widehat{X}]}$ (empirical first moment)

and $\boldsymbol{E[X^2|\theta] = E[\widehat{X}^2]}$ (empirical second moment) .

The second equation can be replaced by an equivalent moment equation that is sometimes more convenient to use:

theoretical distribution variance = empirical distribution variance .

(i.e., $Var[X|\theta] = Var[\widehat{X}]$). Note that if a random sample is given, the variance of the empirical distribution is $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, which is the biased form of the sample variance , so that the method of moments sets the theoretical variance equal to the biased sample variance.

Example ST-2: A random sample of $n = 8$ values from distribution of X is given:

$3, 4, 8, 10, 12, 18, 22, 35$.

(i) For the exponential distribution, estimate θ using the method of moments.

(ii) For the lognormal distribution, estimate μ and σ^2 using the method of moments.

(iii) For the Poisson distribution, estimate λ using the method moments.

(iv) For the negative binomial distribution, estimate r and β using the method of moments.

Solution:

(i) The exponential distribution has one parameter θ, so if it is assumed that the sample comes from an exponential distribution with parameter θ, then the method of moments estimate of θ is based on one equation: $\theta = E[X] = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(3 + 4 + \cdots + 35) = 14 \rightarrow \widehat{\theta} = 14$.

(ii) If it is assumed that the sample comes from a lognormal distribution with parameters μ, σ ($r = 2$ parameters), then there will be 2 moment equations when applying the method of moments as the estimation procedure (the expressions for the moments of a lognormal distribution are found in the Exam 3 Table of Distributions).

Example ST-2 continued

The two moment equations are

$exp(\mu+\frac{\sigma^2}{2})=E[X]=\frac{1}{n}\sum_{i=1}^{n}x_i=14$ and $exp(2\mu+\frac{4\sigma^2}{2})=E[X^2]=\frac{1}{n}\sum_{i=1}^{n}x_i^2=295.75$

These equations become $\mu+\frac{\sigma^2}{2}=ln\,14=2.639$ and $2\mu+\frac{4\sigma^2}{2}=5.690$.

Solving these equations results in $\widehat{\mu}=2.433$, $\widehat{\sigma}^2=.411$.

(iii) The Poisson parameter λ is the distribution mean, so that the moment equation is

$\lambda=E[X]=\frac{1}{n}\sum_{i=1}^{n}x_i=14\rightarrow\widehat{\lambda}=14.$

(iv) The negative binomial parameters are r and β , with mean $r\beta$ and variance $r\beta(1+\beta)$.

The moment equations are $r\beta=\frac{1}{n}\sum_{i=1}^{n}x_i=14$

and $r\beta(1+\beta)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\overline{x}^2=295.75-14^2=99.75$.

Then $1+\widehat{\beta}=\frac{99.75}{14}\rightarrow\widehat{\beta}=6.125$ and $\widehat{r}=\frac{14}{\widehat{\beta}}=\frac{14}{6.125}=2.29$. □

Applying the Method of Moments to Distributions in Exam 3 Table of Distributions

The following comments summarize the application of the method of moments and method of percentile matching to the distributions in the "Table of Distributions".

Two Parameter Distributions

Pareto α , θ: $E[X]=\frac{\theta}{\alpha-1}$, $E[X^2]=\frac{2\theta^2}{(\alpha-1)(\alpha-2)}$.

θ and α can be found from the first and second empirical moments from $\frac{(E[X])^2}{E[X^2]}=\frac{\alpha-2}{2(\alpha-1)}$.

Using the 8-point random sample of Example ST-2 $(3,4,8,10,12,18,22,35)$, this equation becomes $\frac{\alpha-2}{2(\alpha-1)}=\frac{14^2}{295.75}=.6627$, from which we get $\widehat{\alpha}=-2.07$ and the substituting into the equation for $E[X]$ we get $\frac{\theta}{\alpha-1}=14$ (the sample mean), from which it follows that $\widehat{\theta}=-43$. The negative values suggest this distribution is not a good model for the data.

  © ACTEX 2010

Gamma α , θ: $E[X] = \theta\alpha$, $E[X^2] = \theta^2(\alpha+1)\alpha$, $Var[X] = \theta^2\alpha$.
α can be found from the first and second empirical moments from
$\frac{E[X^2]}{(E[X])^2} = \frac{\alpha+1}{\alpha} = 1 + \frac{1}{\alpha}$. Then θ can be found from $\theta = \frac{E[X]}{\alpha}$.
Alternatively, $\frac{(E[X])^2}{Var[X]} = \alpha$ might be a more efficient way to get α if the sample mean and sample variance are given. Using the 8-point data set of Example ST-2 again, we get $\widehat{\alpha} = 1.96$, $\widehat{\theta} = 7.125$.

Lognormal μ , σ: The method of moments was illustrated in Example ST-2(ii).

One Parameter Distributions

Exponential θ: Example ST-2(i) illustrates the application of method of moments to the exponential distribution.

Single parameter Pareto with θ given and α unknown: $E[X] = \frac{\alpha\theta}{\alpha-1}$.
Using the data of Example ST-2, if θ is given to be 6, then we solve $\frac{6\alpha}{\alpha-1} = 14$ to get $\widehat{\alpha} = 1.75$ as the method of moments estimate.

Example ST-3: The random variable X has the density function

$$f(x) = .5\,\frac{1}{\lambda_1}\,e^{-x/\lambda_1} + .5\,\frac{1}{\lambda_2}\,e^{-x/\lambda_2}\ ,\ 0 < x < \infty\,,\ 0 < \lambda_1 < \lambda_2\ .$$

A random sample taken of the random variable X has mean 1 and variance k.
If k is 3/2, determine the method of moments estimate of λ_1.
Solution: The density is a weighted average (mixture) of exponential random variables with means λ_1 and λ_2 and mixing weights of .5 for each part of the mixture. .
Then, $E[X^n] = .5E[X_1^n] + .5E[X_2^n]$, so that
$E[X] = (.5)[\lambda_1 + \lambda_2]$ and $E[X^2] = (.5)[2\lambda_1^2 + 2\lambda_2^2] = \lambda_1^2 + \lambda_2^2$.
If the mean of the sample is 1 and the variance is k, then the second moment of the sample is $k + 1$(variance + mean^2).
By the method of moments we get the equations $(.5)[\lambda_1 + \lambda_2] = 1$ and $\lambda_1^2 + \lambda_2^2 = \frac{3}{2} + 1$.
Using $\lambda_2 = 2 - \lambda_1$, the second equation becomes $\lambda_1^2 + (2-\lambda_1)^2 = \frac{5}{2}$, with solutions $\lambda_1 = \frac{1}{2}$, $\frac{3}{2}$. Since $0 < \lambda_1 < \lambda_2$, the only feasible solution is $\lambda_1 = \frac{1}{2}$, $\lambda_2 = \frac{3}{2}$. $\square$

Discrete Distributions

Poisson Distribution, λ : $E[N] = \lambda$.
The probability function of the Poisson random variable N with parameter λ is
$p_k = P[N = k] = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k \geq 0$ $E[N] = Var[N] = \lambda$.
A data sample of observations from the Poisson distribution would consist of series of integer values. We let n_k be the number of observations which were the integer value k.
Then the total number of observations is $n = n_0 + n_1 + \cdots + n_j$, where j is the largest observed integer value.

Method of moments estimate of λ

The sample mean is $\frac{1}{n}[0 \cdot n_0 + 1 \cdot n_1 + \cdots + j \cdot n_j] = \frac{1}{n}\sum_{k=0}^{j} k \cdot n_k$, and the distribution mean is λ, so the method moments estimate is

$$\boldsymbol{\widehat{\lambda} = \frac{1}{n}\sum_{k=0}^{j} k \cdot n_k = \frac{\text{total number of claims from all } n \text{ observations}}{n}}$$

Negative Binomial Distribution, r, β : $E[N] = r\beta$, $Var[N] = r\beta(1+\beta)$.
For estimation we assume that the data available is similar to the Poisson case. The total number of observations is $n = n_0 + n_1 + \cdots + n_j$, where j is the largest observed integer value, and n_k is the number of observations for which $N = k$.

Method of moments estimates of r and β

The moment equations are

$$r\beta = \frac{1}{n}\sum_{k=0}^{j} k \cdot n_k \text{ and } r\beta(1+\beta) = \frac{1}{n}\sum_{k=0}^{j} k^2 \cdot n_k - \left(\frac{1}{n}\sum_{k=0}^{j} k \cdot n_k\right)^2 .$$

Dividing the second equation by the first results in the moment estimate of $1+\beta$
If $\frac{1}{n}\sum_{k=0}^{j} k^2 \cdot n_k - \left(\frac{1}{n}\sum_{k=0}^{j} k \cdot n_k\right)^2 < \frac{1}{n}\sum_{k=0}^{j} k \cdot n_k$ (sample variance smaller than sample mean), then $\widehat{\beta} < 0$, which is an indication that the negative binomial model is not a good representation for the data.

Binomial Distribution, m, q : $E[N] = mq$, $Var[N] = mq(1-q)$.
If an experiment is performed m times, independent of one another, and q is the probability of a "successful event" on a particular trial, then p_k is the probability of exactly k successes in the m trials. If the sample variance is larger than the sample mean, then the binomial distribution will not be a good model for the data.

A data sample would consist of $x_1, x_2, ..., x_j$, where each x_i is a number from 0 to m, indicating the number of "successes" in the i-th group of m trials. As an example, consider the number of heads occurring in 10 flips of a fair coin; this number has a binomial distribution with $m = 10$ and $q = .5$. Each x_i would be a number from 0 to 10, indicating the number of heads that turned up in the i-th group of 10 coin flips. If the 10-coin flip process was repeated 25 times, the data would be $x_1, x_2, ..., x_{25}$. We could then determine the number of groups in which 0 heads turned up, n_0, the number of groups in which 1 head turned up, n_1, ... , and the number of groups in which 10 heads turned up, n_{10} . Also, $n_0 + n_1 + \cdots + n_{10} = 25$, the total number of groups of 10 flips that were conducted. The total number of heads from all of the $(10)(25) = 250$ coin flips is total number of heads observed $= (0)n_0 + (1)n_1 + \cdots + (10)n_{10} = \sum_{k=0}^{10} kn_k$.
As before, n_k is the number of observations for which $N = k$.

<u>Estimation if m is known</u>: If m is known, then the data set is $n_0, n_1, \ldots, n_m$, and the **moment estimate** of q is equal to $\widehat{q} = \frac{1}{m} \frac{\sum_{k=0}^{m} kn_k}{\sum_{k=0}^{m} n_k} = \frac{\sum_{i=1}^{j} x_i}{mj}$.

<u>Estimation if m is unknown</u>
Moment estimation: The moment equations are

$$mq = \frac{1}{n}\sum_{k=0}^{j} k \cdot n_k \text{ and } mq(1-q) = \frac{1}{n}\sum_{k=0}^{j} k^2 \cdot n_k - \left(\frac{1}{n}\sum_{k=0}^{j} k \cdot n_k\right)^2 .$$

Dividing the second equation by the first results in the moment estimate of $1-q$, and once q is estimated, m can be calculated. This value of m will have to be adjusted to be an integer and then the estimate of q will be adjusted using $mq = \frac{1}{n}\sum_{k=0}^{j} k \cdot n_k$.

Example ST-4: The following 100 observations are made from a discrete non-negative integer-valued random variable:

k	0	1	2	3	4	5	6
n_k	6	17	26	21	18	7	5

(i) The Poisson distribution parameter estimate by the method of moments is

$\widehat{\lambda} = \frac{1}{n}\sum_{k=0}^{j} k \cdot n_k = \frac{1}{100}[(0)(6) + (1)(17) + (2)(26) + (3)(21) + (4)(18) + (5)(7) + (6)(5)]$

$= 2.69$, and $V\widehat{a}r[\widehat{\lambda}] = \frac{\widehat{\lambda}}{n} = .0269$.

(ii) The negative binomial distribution parameter estimate by the method of moments is found from $r\beta = \frac{1}{n}\sum_{k=0}^{j} k \cdot n_k = 2.69$ and

$r\beta(1+\beta) = \frac{1}{n}\sum_{k=0}^{j} k^2 \cdot n_k - \left(\frac{1}{n}\sum_{k=0}^{j} k \cdot n_k\right)^2 = 9.53 - (2.69)^2 = 2.294 \rightarrow \beta = -.147$.

This suggests that the negative binomial is not a good fit to the data.

(iii) The binomial distribution with $m = 6$ has method of moments estimate for q of

$\widehat{q} = \frac{1}{m} \frac{\sum_{k=0}^{m} k n_k}{\sum_{k=0}^{m} n_k} = \frac{1}{6} \cdot \frac{269}{100} = .4483$.

If m is unknown, then we can apply the method of moments to get $mq = 2.69$, and

$mq(1-q) = \frac{1}{n}\sum_{k=0}^{j} k^2 \cdot n_k - \left(\frac{1}{n}\sum_{k=0}^{j} k \cdot n_k\right)^2 = 2.294$, so that

$1 - \widehat{q} = \frac{2.294}{2.69} = .853$, $\widehat{q} = .147$, and $\widehat{m} = \frac{2.69}{\widehat{q}} = 18.3$.

We would then adjust m to be 18. □

Maximum Likelihood Estimation

Maximum likelihood estimation is a method that is applied to estimate the parameters in a parametric distribution. The general objective in maximum likelihood estimation is based on the following idea. Given a data set, we try to find the distribution parameter (or parameters) that result in the maximum density or probability of that data set occurring. The first step in applying the method of maximum likelihood estimation is the formulation of the **likelihood function** $L(\theta)$, where θ is the parameter (or parameters) to be estimated in a distribution with pdf/pf $f(x;\theta)$.

Complete Individual data likelihood function (based on a random sample):

$$L(\theta) = \prod_{j=1}^{n} f(x_j;\theta) \ , \ \text{random sample } x_1, x_2, \ldots, x_n$$

It is almost always more convenient to maximize the natural log of the likelihood function (loglikelihood), $l(\theta) = ln\ L(\theta)$, and this results in the same maximum likelihood estimate of θ.

As a practical consideration, it is useful to look at the form of $ln\, f(x;\theta)$ (the natural log of the density) before setting up the log-likelihood function $\ell(\theta)$. Any <u>additive term</u> in $ln\, f(x;\theta)$ that does involve the parameter θ can be ignored when setting up the log-likelihood function (it will not affect the value of θ that maximizes $\ell(\theta)$). This helps "clean up" the log-likelihood expression that will be maximized.

The usual way to estimate θ (in a one-parameter case) is to set $\frac{d}{d\theta}\, ln\ L(\theta) = 0$ and solve for θ. In general, the likelihood function $L(\theta)$ is product of factors (densities), and the loglikelihood $l(\theta)$ is a sum of factors (log-densities), so that the derivative of $l(\theta)$ is a sum of the derivatives of those factors. Any additive factor which does not involve θ (the parameter being estimated) in $l(\theta)$, will have derivative 0 with respect to θ, and therefore will disappear from the mle equation.

The mle has the following properties:
- it is asymptotically normal (as sample size n increases, $\widehat{\theta}_n$ approaches a normal distribution),
- it is asymptotically unbiased, meaning that as $n \to \infty$ the expected value of the estimator approaches the true parameter value,
- it has the smallest asymptotic variance of all estimators of θ that are asymptotically normal

Example ST-5: A random sample of $n = 8$ values from distribution of X is given:
$3\,,\,4\,,\,8\,,\,10\,,\,12\,,\,18\,,\,22\,,\,35$
Apply maximum likelihood estimation to estimate the parameter θ in the following two cases.
(i) Exponential distribution with parameter θ.
(ii) Uniform distribution on the interval $(\,0\,,\,\theta\,)$.
(iii) Single parameter Pareto distribution with $\theta = 1$, and estimation is applied to find α.

Solution:
(i) If we assume that the distribution of X is exponential with parameter θ then the pdf is $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$, and the likelihood function is $L(\theta) = \frac{1}{\theta}e^{-3/\theta}\cdots\frac{1}{\theta}e^{-35/\theta} = \frac{1}{\theta^8}e^{-112/\theta}$.
Then $ln\,L(\theta) = \,-\,8\,ln\,\theta\,-\,\frac{112}{\theta}$. To maximize $\ell(\theta) = ln\;L(\theta)$, (and equivalently, $L(\theta)$) we differentiate $\ell(\theta)$ with respect to θ and set equal to 0 and solve for θ:
$\frac{d}{d\theta}[\,-\,8\,ln\,\theta\,-\,\frac{112}{\theta}] = \,-\,\frac{8}{\theta} + \frac{112}{\theta^2} = 0 \rightarrow \widehat{\theta} = 14$.
Note that this maximum likelihood estimate of θ is the sample mean of the random sample, which is the same as the method of moments estimate of θ. For the exponential distribution, it is always the case that for individual data, the mle and the moment estimate of the parameter θ are the same (equal to the sample mean)..

(ii) If we assume that the distribution of X is uniform on the interval $(0,\theta]$, then the pdf is $f(x;\theta) = \frac{1}{\theta}$ for $x \in (0,\theta)$. The likelihood function is $L(\theta) = \frac{1}{\theta}\cdots\frac{1}{\theta} = \frac{1}{\theta^8}$.
We can see directly that $L(\theta)$ is maximized at the minimum feasible value of θ, which is $\widehat{\theta} = 35$. Any value of θ less than 35 would make the data point $x = 35$ impossible (i.e., $f(35;\theta) = 0$ for any $\theta < 35$, since any loss from the uniform distribution on the interval $(0,\theta)$ must be less than θ). Notice that in this case we maximized L directly, rather $ln\,L$.

(iii) The pdf of the single parameter Pareto distribution with $\theta = 1$ is $f(x;\alpha) = \frac{\alpha}{x^{\alpha+1}}$.
The likelihood function is $L(\alpha) = \frac{\alpha}{x_1^{\alpha+1}}\cdot\frac{\alpha}{x_2^{\alpha+1}}\cdot\cdot\cdot\frac{\alpha}{x_8^{\alpha+1}} = \frac{\alpha^8}{(\prod_{i=1}^{8}x_i)^{\alpha+1}}$.

The loglikelihood is

$$ln\,L(\alpha) = \ell(\alpha) = 8\,ln\,\alpha\,-(\alpha+1)\sum_{i=1}^{8}ln(x_i) = 8\,ln\,\alpha\,-(\alpha+1)[ln\,3\,+\,ln\,4\,+\cdots+\,ln\,35]\,.$$

Then $\frac{d}{d\alpha}\,\ell(\alpha) = \frac{8}{\alpha} - [ln\,3\,+\,ln\,4\,+\cdots+\,ln\,35] = 0$ results in $\widehat{\alpha} = 2.36$. □

Summary of maximum likelihood estimates for some common distributions

Suppose that we are given a random sample $x_1, ..., x_n$. The maximum likelihood estimates of parameters for some commonly used distributions are as follows.

Exponential distribution with mean θ: mle of θ is the sample mean $\overline{x}$ (same as the moment estimate).

Uniform distribution on the interval $(0, \theta)$: mle of θ is $max\{x_1, ..., x_n\}$.

Normal distribution with mean μ and variance σ^2: mle of μ is the sample mean $\overline{x}$, and the mle of σ^2 is $\frac{1}{n}\Sigma(x_i - \overline{x})^2$ (the biased from of the sample variance; these are the same as the moment estimates).

Lognormal distribution with parameters μ and σ^2: we let $y_i = ln\, x_i$ for $i = 1, ..., n$. and the mle of μ is $\overline{y}$, and the mle of $\sigma^2$is $\frac{1}{n}\Sigma(y_i - \overline{y})^2$ (not the same as the moment estimates).

Cramer-Rao Inequality

Suppose that X is a random variable with pdf or pf $f(x; \theta)$. Suppose that $\widehat{\theta}$ is any unbiased estimator of θ based on a random sample of size n. Then the variance of the estimator $\widehat{\theta}$ cannot be less than $\frac{-1}{n \cdot E[\frac{\partial^2}{\partial\theta^2}\, ln\, f(x;\theta)]}$. If an unbiased estimator has this variance, then it is an *unbiased minimum variance estimator*. In general, maximum likelihood estimators have variance that is approximately equal to the minimum possible variance.

For the exponential and normal distributions, the mle's are the minimum variance estimators. This is fairly easy to see for the exponential distribution. If $x_1, ..., x_n$ is a random sample from the exponential distribution with mean θ, then the mle of θ is $\widehat{\theta} = \frac{1}{n}\Sigma x_i$, and the variance of $\widehat{\theta}$ is $Var[\widehat{\theta}] = Var[\frac{1}{n}\Sigma x_i] = Var[\overline{x}] = \frac{Var[X]}{n} = \frac{\theta^2}{n}$, since the variance of the exponential random variable is the square of the mean. According to the Cramer-Rao lower bound for the variance of an unbiased estimator, we first find

$\frac{\partial^2}{\partial\theta^2}\, ln\, f(x;\theta) = \frac{\partial^2}{\partial\theta^2}\, ln[\frac{1}{\theta}e^{-x/\theta}] = \frac{\partial^2}{\partial\theta^2}\, [-\frac{x}{\theta} - ln\,\theta] = [-\frac{2x}{\theta^3} + \frac{1}{\theta^2}]$.

Then, $E[\frac{\partial^2}{\partial\theta^2}\, ln\, f(x;\theta)] = E[-\frac{2x}{\theta^3} + \frac{1}{\theta^2}] = -\frac{2\theta}{\theta^3} + \frac{1}{\theta^2}] = -\frac{1}{\theta^2}$.

The minimum variance is $\frac{-1}{n\cdot E[\frac{\partial^2}{\partial\theta^2}\, ln\, f(x;\theta)]} = \ = \frac{-1}{n\cdot(-\frac{1}{\theta^2})} = \frac{\theta^2}{n}$.

We see that the mle has the minimum possible variance of any unbiased estimator, and it is the "best" in that sense.

PROBLEM SET 1

Properties of Estimators, Moment and Maximum Likelihood Estimation

Questions 1 and 2 are based on the following random sample of 12 data points from a population distribution X: 7 , 15 , 15 , 19 , 26 , 27 , 29 , 29 , 30 , 33 , 38 , 53

1. Find both the biased and the unbiased sample variances.

2. Suppose that the distribution variance is 100.
(a) Find the mean square error of the sample mean as an estimator of the distribution mean.
(b) Find the bias in the biased form of the sample variance as an estimator of the distribution variance.

3. The random variable X has pdf $f(x) = 2x$ for $0 < x < 1$.
$X_1, ..., X_n$ is a random sample from the distribution of the continuous random variable X.
$\widehat{\theta} = \frac{1}{n-1}\sum_{i=1}^{n} X_i$ is taken as an estimate of the distribution mean θ. Find $MSE_\theta(\widehat{\theta})$.
A) $\frac{2}{3n}$ B) $\frac{n+2}{3(n-1)^2}$ C) $\frac{4}{9n}$ D) $\frac{n+8}{18(n-1)^2}$ E) $\frac{n+4}{9(n-1)^2}$

Questions 4 to 8 are based on the following random sample of 12 data points from a population distribution X 7 , 12 , 15 , 19 , 26 , 27 , 29 , 29 , 30 , 33 , 38 , 53

4. Determine the method of moments estimate of $P[X > 30]$ for the exponential distribution.
A) .30   B) .32   C) .34   D) .36   E) .38

5. Determine the method of moments estimate of $P[X > 30]$ for the normal distribution.
A) .30   B) .32   C) .34   D) .36   E) .38

 © ACTEX 2010