THE ROYAL STATISTICAL SOCIETY

2006 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

STATISTICAL THEORY AND METHODS

PAPER II

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

Note. In accordance with the convention used in the Society's examination papers, the notation log denotes logarithm to base *e*. Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2006. Question 1

$$P(X=x)=\frac{e^{-\theta}\theta^{x}}{x!}, \qquad x=0,1,2,\ldots .$$

(i) The likelihood is $L = \prod_{i=1}^{n} P(X=x_i) = \dfrac{e^{-n\theta}\theta^{\Sigma xi}}{\prod_{i=1}^{n}(x_i!)} = \underbrace{e^{-n\theta}\theta^{t}}_{g(t,\theta)} \underbrace{\prod_{i=1}^{n}\frac{1}{x_i}!}_{h(\mathbf{x})}$,

i.e. a product of a function of θ and T with a function of $\boldsymbol{x}$. Hence by the factorisation theorem, T is sufficient for θ.

(ii) $\log L = -n\theta + t\log\theta + \log\prod_i \frac{1}{x_i!}$.

$\therefore \dfrac{d}{d\theta}(\log L) = -n + \dfrac{t}{\theta}$.

Setting this equal to zero gives $n\hat{\theta} = t$, i.e. $\hat{\theta} = \dfrac{t}{n}$.

It may easily be verified (e.g. by considering the second derivative) that this is indeed a maximum, and so the maximum likelihood estimator of θ is T/n.

Maximum likelihood estimators are function-invariant, so $\widehat{e^{-\theta}} = e^{-\hat{\theta}} = e^{-T/n}$.

(iii) T has the Poisson distribution with parameter $n\theta$.

$$\therefore E\left[\left(\frac{n-1}{n}\right)^{T}\right] = \sum_{t=0}^{\infty}\left(\frac{n-1}{n}\right)^{t}\frac{e^{-n\theta}(n\theta)^{t}}{t!}$$

$$= e^{-\theta}\sum_{t=0}^{\infty} e^{-(n-1)\theta}\{(n-1)\theta\}^{t}/t!$$

The sum here is the sum of probabilities for the Poisson($(n-1)\theta$) distribution and is therefore 1

$= e^{-\theta}$.

So $\left(\dfrac{n-1}{n}\right)^{T}$ is an unbiased estimator of $e^{-\theta}$.

(iv) Given that $\dfrac{n-1}{n} < e^{-1/n}$, we have $\left(\dfrac{n-1}{n}\right)^{T} \le e^{-T/n}$, and the inequality is strict so long as $T > 0$; note that, since T has a Poisson distribution, $P(T > 0) > 0$.

From (iii) we have $E\left[\left(\dfrac{n-1}{n}\right)^{T}\right] = e^{-\theta}$, so $E\left(e^{-T/n}\right) > e^{-\theta}$.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2005. Question 2

(i) The Cramér-Rao result relies on regularity conditions of the density function. For the uniform distribution, the range of values for which the density is non-zero depends directly on the parameter being estimated. Hence the regularity conditions (e.g. the ability to change the order of integration and differentiation) will not hold.

(ii)

$\hat{\theta}_1 = 2\bar{X}$.

$$E\left[\hat{\theta}_1\right] = \frac{2}{n}E\left[\sum X_i\right] = \frac{2}{n}\sum E[X_i] = \frac{2}{n}.n.\frac{\theta}{2} = \theta. \quad \text{So } \hat{\theta}_1 \text{ is unbiased.}$$

$$\text{Var}\left(\hat{\theta}_1\right) = \frac{4}{n^2}\sum \text{Var}\left(X_i\right) \quad \text{[note independence of } \{X_i\}\text{]} \quad = \frac{4}{n^2}.n\frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

$\hat{\theta}_2 = Y = \max X_i$. [Note that the pdf of Y, $g(y)$, is given in the question.]

$$E\left[\hat{\theta}_2\right] = \int_0^\theta yg\left(y\right)dy = \frac{n}{\theta^n}\int_0^\theta y^n dy = \left[\frac{n}{\theta^n}.\frac{y^{n+1}}{n+1}\right]_0^\theta = \frac{n\theta}{n+1}.$$

Therefore the bias is $\frac{n\theta}{n+1} - \theta = -\frac{\theta}{n+1}$.

$$E\left[\hat{\theta}_2{}^2\right] = \int_0^\theta y^2 g\left(y\right)dy = \frac{n}{\theta^n}\left[\frac{y^{n+2}}{n+2}\right]_0^\theta = \frac{n\theta^2}{n+2}.$$

$$\therefore \text{Var}(Y) = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n\left(n+1\right)^2 - n^2\left(n+2\right)}{\left(n+1\right)^2\left(n+2\right)}\theta^2 = \frac{n\theta^2}{\left(n+1\right)^2\left(n+2\right)}.$$

$\hat{\theta}_3 = \frac{n+1}{n}Y$.

$$E\left[\hat{\theta}_3\right] = \frac{n+1}{n}E\left[\hat{\theta}_2\right] = \theta. \quad \text{So } \hat{\theta}_3 \text{ is unbiased.}$$

$$\text{Var}\left(\hat{\theta}_3\right) = \left(\frac{n+1}{n}\right)^2 \text{Var}\left(\hat{\theta}_2\right) = \frac{\theta^2}{n\left(n+2\right)}.$$

Solution continued on next page

(iii) In each case, variance and bias $\to 0$ as n $\to \infty$, so all three estimators are consistent.

(iv) The efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_3$ is $\dfrac{\text{Var}\left(\hat{\theta}_3\right)}{\text{Var}\left(\hat{\theta}_1\right)} = \dfrac{3}{n+2}$. For large *n*, this $\to 0$, i.e. the efficiency declines towards zero as *n* grows large. This is because $\text{Var}\left(\hat{\theta}_3\right)$ is of order $\dfrac{1}{n^2}$ while $\text{Var}\left(\hat{\theta}_1\right)$ is of order $\dfrac{1}{n}$.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2006. Question 3

$$p(x;\theta)=\frac{(x+1)\theta^2}{(1+\theta)^{x+2}}, \qquad x=0,1,2,..., \qquad \theta>0.$$

(i) $E[X]=\sum_{x=0}^{\infty}\frac{x(x+1)\theta^2}{(1+\theta)^{x+2}}$. To find this sum, consider (among other methods) $w=(1+\theta)^{-x}$. We have $\frac{dw}{d\theta}=\frac{-x}{(1+\theta)^{x+1}}$ and $\frac{d^2w}{d\theta^2}=\frac{x(x+1)}{(1+\theta)^{x+2}}$.

Thus $E[X]=\sum_{x=0}^{\infty}\theta^2\frac{d^2w}{d\theta^2}$, where $w=(1+\theta)^{-x}$, and we see that

$$E[X] = \theta^2\sum_{x=0}^{\infty}\frac{d^2}{d\theta^2}(1+\theta)^{-x} = \theta^2\frac{d^2}{d\theta^2}\sum_{x=0}^{\infty}(1+\theta)^{-x}$$

$$= \theta^2\frac{d^2}{d\theta^2}\left\{\frac{1}{1-\frac{1}{1+\theta}}\right\} = \theta^2\frac{d^2}{d\theta^2}\left\{1+\frac{1}{\theta}\right\} = \theta^2.2\theta^{-3} = \frac{2}{\theta}.$$

Thus $E[\bar{X}]=\frac{2}{\theta}$, and so the method of moments estimator of θ is $\tilde{\theta}=\frac{2}{\bar{X}}$.

(ii) The likelihood is $L=\prod_{i=1}^{n}\frac{(x_i+1)\theta^2}{(1+\theta)^{x_i+2}}=\theta^{2n}\prod_{i=1}^{n}(x_i+1)/\prod_{i=1}^{n}(1+\theta)^{x_i+2}$.

$$\therefore \log L = 2n\log\theta+\sum_i\log(x_i+1)-\left(2n+\sum_i x_i\right)\log(1+\theta).$$

$$\therefore \frac{d}{d\theta}(\log L) = \frac{2n}{\theta}-\frac{2n+\sum x_i}{1+\theta} = \frac{2n-\theta\sum x_i}{\theta(1+\theta)}.$$

Setting this equal to zero gives $\hat{\theta}\sum x_i=2n$, i.e. $\hat{\theta}=\frac{2}{\bar{x}}$. It may be verified (e.g. by considering the second derivative – see part (iii) below) that this is indeed a maximum, and so the maximum likelihood estimator of θ is $\hat{\theta}=\frac{2}{\bar{X}}$.

Solution continued on next page

(iii) $\dfrac{d^2}{d\theta^2}\left(\log L\right) = -\dfrac{2n}{\theta^2}+\dfrac{2n+\sum x_i}{\left(1+\theta\right)^2}$.

$$\therefore E\left[-\frac{d^2}{d\theta^2}\log L\right] = \frac{2n}{\theta^2}-\frac{\left(2n+\dfrac{2n}{\theta}\right)}{\left(1+\theta\right)^2} = \frac{2n\left(1+\theta\right)^2-\theta^2\left(2n+\dfrac{2n}{\theta}\right)}{\theta^2\left(1+\theta\right)^2}$$

$$= \frac{2n\left(1+2\theta+\theta^2\right)-2n\theta^2-2n\theta}{\theta^2\left(1+\theta\right)^2} = \frac{2n\left(1+\theta\right)}{\theta^2\left(1+\theta\right)^2} = \frac{2n}{\theta^2\left(1+\theta\right)}.$$

Therefore the Fisher information is $\dfrac{\theta^2\left(1+\theta\right)}{2n}$ (or $\dfrac{\theta^2\left(1+\theta\right)}{2}$ if found as the Fisher information per observation).

(iv) An approximate (large-sample) 90% confidence interval for θ is

$\hat{\theta}\pm 1.645\sqrt{\text{Est Var}\left(\hat{\theta}\right)}$ or $\hat{\theta}\pm 1.645\sqrt{\dfrac{\hat{\theta}^2\left(1+\hat{\theta}\right)}{2n}}$, which is

$$2.5\pm 1.645\sqrt{\frac{\left(2.5\right)^2\left(3.5\right)}{200}} = 2.5\pm 0.544, \qquad \text{i.e. (1.96, 3.04).}$$

Graduate Diploma, Statistical Theory & Methods, Paper II, 2006. Question 4

(i) When $\sigma = 1$, the two parts $x \le 0$ and $x > 0$ combine to give N(0, 1).

(ii) In general, the likelihood for a random sample of n observations $x_1, x_2, \ldots, x_n$ is

$$\prod_{x_i \le 0} A e^{-x_i^2/2} \prod_{x_i > 0} A e^{-x_i^2/2\sigma^2}, \quad \text{where } A = \frac{\sqrt{2}}{(1+\sigma)\sqrt{\pi}}.$$

Thus, if L_1 and L_0 are the likelihoods under H_1 and H_0 [H_0 is $\sigma = 1$, H_1 is $\sigma = \sigma^*$], we have

$$\frac{L_1}{L_0} = \frac{\left(\frac{\sqrt{2}}{(1+\sigma^*)\sqrt{\pi}}\right)^n \prod_{x_i>0} e^{-x_i^2/2\sigma^{*2}}}{\left(\frac{\sqrt{2}}{2\sqrt{\pi}}\right)^n \prod_{x_i>0} e^{-x_i^2/2}} = \frac{2^n}{(1+\sigma^*)^n} e^{-\frac{\sum_+ x_i^2}{2\sigma^{*2}} + \frac{\sum_+ x_i^2}{2}}$$

$$= \frac{2^n}{(1+\sigma^*)^n} e^{-\frac{1}{2}\sum_+ x_i^2 \left(\frac{1}{\sigma^{*2}} - 1\right)}.$$

The Neyman-Pearson approach is to reject H_0 when L_1/L_0 is large. Since $\sigma^* > 1$, the likelihood ratio here is an increasing function of $\Sigma_+ x_i^2$. So we reject H_0 for $\Sigma_+ x_i^2 > k$, where the value of k is chosen according to the required significance level for the test.

(iii) The form of this test does not depend on the actual value of σ^*, so the same test is appropriate for all σ^*. The test is therefore uniformly most powerful for testing against $H_2 : \sigma > 1$.

(iv) $H_0 : \sigma = 1$; $H_3 : \sigma \ne 1$. We have $n = 30$ and $\Sigma_+ x_i^2 = 80$. The likelihood ratio test of (ii) is used with $\hat{\sigma} = 2$.

The likelihood ratio is $\lambda = \frac{2^{30}}{3^{30}} e^{-\frac{1}{2} \cdot 80 \left(\frac{1}{4} - 1\right)} = \left(\frac{2}{3}\right)^{30} e^{30}$.

$\therefore 2\log\lambda = 35.67$. On H_0, $2\log\lambda \sim \chi^2_1$, so this result is very highly significant and there is very strong evidence against H_0.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2006. Question 5

(i) Let X_j be the number of faulty items in stage j (j = 1, 2, 3). X_j is approximately Poisson with parameter $\lambda = 30p$.

Let N be the number of items sampled in total.

$N = 30$ if $X_1 = 0$ (accept) or $X_1 \geq 2$ (reject).

$$\therefore P(N=30) = P(X_1 = 0 \text{ or } X_1 \geq 2) = 1 - P(X_1 = 1) = 1 - \lambda e^{-\lambda}.$$

$N = 60$ if $X_1 = 1$ and $X_2 = 0$ (accept) or $X_2 \geq 3$ (reject).

$$\therefore P(N=60) = \lambda e^{-\lambda}\left\{1 - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}\right\}.$$

$N = 90$ if $X_1 = 1$ and $X_2 = 1$ or 2 (i.e. if $N \neq 30$ or 60).

$$\therefore P(N=90) = 1 - \left\{1 - \lambda e^{-\lambda} + \lambda e^{-\lambda}\left(1 - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}\right)\right\}$$

$$= 1 - \left\{1 - \lambda e^{-\lambda} + \lambda e^{-\lambda} - \lambda^2 e^{-2\lambda} - \frac{\lambda^3 e^{-2\lambda}}{2}\right\} = \frac{1}{2}\lambda^2 e^{-2\lambda}(2+\lambda).$$

$$\therefore E(N) = 30\left(1 - \lambda e^{-\lambda}\right) + 60\left(\lambda e^{-\lambda} - \lambda^2 e^{-2\lambda} - \frac{\lambda^3 e^{-2\lambda}}{2}\right) + 90\frac{\lambda^2(2+\lambda)e^{-2\lambda}}{2}$$

$$= 30\left(1 - \lambda e^{-\lambda} + 2\lambda e^{-\lambda} - 2\lambda^2 e^{-2\lambda} - \lambda^3 e^{-2\lambda} + 3\lambda^2 e^{-2\lambda} + \frac{3}{2}\lambda^3 e^{-2\lambda}\right)$$

$$= 30\left(1 + \lambda e^{-\lambda} + \lambda^2 e^{-2\lambda} + \frac{1}{2}\lambda^3 e^{-2\lambda}\right).$$

Solution continued on next page

(ii) We seek to maximise $P(N = 90)$. We have

$$P(N=90) = \frac{1}{2}\lambda^2 e^{-2\lambda}(2+\lambda) = \left(\lambda^2 + \frac{1}{2}\lambda^3\right)e^{-2\lambda}.$$

$$\therefore \frac{dP}{d\lambda} = \left(\lambda^2 + \frac{1}{2}\lambda^3\right)\left(-2e^{-2\lambda}\right) + e^{-2\lambda}\left(2\lambda + \frac{3}{2}\lambda^2\right) = e^{-2\lambda}\left(2\lambda - \frac{1}{2}\lambda^2 - \lambda^3\right).$$

So $\frac{dP}{d\lambda} = 0$ if $2 - \frac{1}{2}\lambda - \lambda^2 = 0$ (since $\lambda e^{-2\lambda} \neq 0$),

i.e. for $2\lambda^2 + \lambda - 4 = 0$, $\lambda = -\frac{1 \pm \sqrt{33}}{4}$, i.e. for $\lambda = \frac{\sqrt{33}-1}{4}$ as $\lambda > 0$. [Should check that this is indeed a minimum.]

Thus we have $\lambda = 1.186$, so $p = \lambda/30 = 0.040$.

(iii) P(accept batch)

$$= P(X_1 = 0) + P(X_1 = 1)\left\{P(X_2 = 0) + P(X_2 = 1)P(X_3 = 0 \text{ or } 1) + P(X_2 = 2)P(X_3 = 0)\right\}$$

$$= e^{-\lambda} + \lambda e^{-\lambda}\left\{e^{-\lambda} + \lambda e^{-\lambda}\left(e^{-\lambda} + \lambda e^{-\lambda}\right) + \frac{\lambda^2 e^{-\lambda}}{2}.e^{-\lambda}\right\}$$

$$= e^{-\lambda}\left\{1 + \lambda e^{-\lambda}\left(1 + \lambda e^{-\lambda} + \frac{3}{2}\lambda^2 e^{-\lambda}\right)\right\}.$$

Graduate Diploma, Statistical Theory & Methods, Paper II, 2006. Question 6

Part (a)

In Bayesian analysis, a parameter θ that is being estimated is assumed to be described by a probability distribution, based on existing knowledge – this is the prior distribution. A sample of data is then obtained, from a population indexed by θ, and combined with the prior distribution using Bayes' theorem to give a posterior distribution (see (b) for the method). If the posterior distribution is in the same family as the prior, we have a conjugate family of distributions.

For example, Normal prior + Normal data $\rightarrow$ Normal posterior. In (b) the gamma distribution is found conjugate for Poisson data.

Part (b)

θ has prior distribution $g(\theta) = \dfrac{\beta^{\alpha}\theta^{\alpha-1}e^{-\beta\theta}}{\Gamma(\alpha)}$.

(i) The likelihood function for the sample data is $L(\boldsymbol{x}|\theta) = \dfrac{e^{-n\theta}\theta^{\Sigma xi}}{\prod_i x_i!}$.

So the posterior distribution $\propto L(\boldsymbol{x}|\theta)\,g(\theta)$

i.e. $\propto \theta^{\alpha-1+\Sigma x_i}e^{-\theta(\beta+n)}$,

i.e. it is gamma with parameters $\alpha+\Sigma x_i$ and $\beta+n$. With the given α and β this is gamma $\left(2+\Sigma x_i,\ \frac{1}{2}+n\right)$.

(ii) Using the squared error loss function, the Bayes estimator is the posterior mean.

The moment generating function (quoted in the question) may conveniently be used to find the mean of a gamma distribution:-

$$M(t) = \left(\frac{\beta}{\beta-t}\right)^{\alpha}. \quad \therefore M'(t) = \frac{\alpha\beta^{\alpha}}{(\beta-t)^{\alpha+1}}.$$

$\therefore$ mean $= M'(0) = \dfrac{\alpha}{\beta}$. Hence the Bayes estimator of θ is $(2 + \Sigma x_i)/(\frac{1}{2} + n)$.

(iii) Given that $2\beta\theta \sim \chi^2_{2\alpha}$, the given case is $2\times\frac{7}{2}\theta \sim \chi^2_{2\times10}$ or $\theta \sim \frac{1}{7}\chi^2_{20}$. Upper and lower 2½% points for χ^2_{20} are 34.170 and 9.591, hence the required interval is $1.37 < \theta < 4.88$.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2006. Question 7

(a) Suppose that X follows a distribution whose pdf contains a parameter θ, and $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a sample of data from this distribution. A pivotal quantity is a function of the data and the parameter whose distribution is known and fully specified. Thus a pivotal quantity $Q(\boldsymbol{X}, \theta)$ must follow a distribution which is independent of the value of θ, i.e. has the same distribution for all values of θ.

For example, in N(μ, σ^2), the t statistic $(\bar{X} - \mu)/(\mathrm{S}/\sqrt{n})$ is a pivotal quantity because its distribution does not depend on μ (or σ^2).

For a pivotal quantity Q, it is possible to find constants a and b such that $P(a \le Q(\boldsymbol{X}, \theta) \le b)$ is a chosen "confidence" value. For example, for the t statistic as above,

$$P\left(-a \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \le a\right) = P\left(-a \le t_{n-1} \le a\right).$$

The chosen probability for this statement is 0.95, or other suitable value depending on the problem, and a is chosen accordingly. The inequality in the bracket is then rewritten in terms of the parameter; thus, for the t example, rewriting in terms of μ gives confidence limits $\bar{x} \pm t_{n-1}\dfrac{s}{\sqrt{n}}$ for the true μ.

(b) We will need $P\left(X \le x\right) = \int_0^x \theta u^{\theta-1} du = x^\theta$.

(i) Let $Y = -\theta \log X$. Then $P\left(Y \le y\right) = P\left(-\theta \log X \le y\right) = P\left(X \ge e^{-y/\theta}\right) = 1 - e^{-y}$.

This is independent of θ, so Y is a pivotal quantity. [It has the exponential distribution with mean 1.]

(ii) The lower and upper 5% points of such an exponential distribution are –log 0.95 and –log 0.05, i.e. 0.051 and 2.996.

Hence $P\left(\dfrac{0.051}{-\log X} < \theta < \dfrac{2.996}{-\log X}\right) = 0.90$, so that the required 90% confidence interval for θ is $\left(\dfrac{0.051}{-\log x}, \dfrac{2.996}{-\log x}\right)$.

(iii) The corresponding interval for $\dfrac{1}{\theta}$ is $\left(\dfrac{-\log x}{2.996}, \dfrac{-\log x}{0.051}\right)$. For $x = 0.5$, the value of –log x is 0.6931, so the interval is (0.231, 13.59).

Graduate Diploma, Statistical Theory & Methods, Paper II, 2006. Question 8

In a discussion like this, credit is gained for all relevant points made, for clarity and depth of explanation, and for practical as well as theoretical considerations. The notes below cover a selection of points that could be made. **The solution continues on the next page.**

Suppose a parameter θ is to be estimated in a statistical distribution on the basis of a sample of data from which suitable statistics can be obtained – for example, the sample mean to use as an estimate of central location.

An estimator is *unbiased* if the expectation (mean) of its sampling distribution is equal to the parameter being estimated. This means that the estimator "gets the right answer *on average*" – but *individual* values of it could be a long way away from the true value of the parameter being estimated. Unbiasedness has some intuitive appeal, but is not a particularly important criterion for an estimator to possess; biased estimators are often used, particularly if the bias is (in some sense) small.

An estimator is *consistent* if the probability of it differing from the parameter being estimated by more than ε, a very small quantity, approaches 0 as the sample size $\to \infty$. It is however easier to use a criterion based on variance: if the variance of the sampling distribution $\to 0$ as the sample size $\to \infty$, the estimator is *consistent*. [Some care is needed in using this criterion for biased estimators, in case the estimator is "homing in" on the wrong place. Provided any bias itself $\to 0$ as the sample size $\to \infty$, the criterion is satisfactory.] Consistency is a very important criterion for an estimator to possess; in most situations an inconsistent estimator would not be entertained, as it "gets the wrong answer in large samples".

As an example, the usual estimator of μ in N(μ, σ^2) is $\bar{X}$, and we have the standard results $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \sigma^2/n$. So $\bar{X}$ is both unbiased and consistent.

As another example, an estimator of σ^2 is $\frac{1}{n}\Sigma(X_i - \bar{X})^2$. This is not unbiased: standard results give that divisor $n - 1$ is required for unbiasedness, whereas the expectation of this estimator (divisor n) is $[(n-1)/n]\sigma^2$. But it is consistent.

For unbiased estimators, the precision with which θ is estimated is measured by the variance of the estimator, though if the underlying sampling distribution is not symmetrical the variance is less useful. Mean square error (MSE) is a useful combination of bias and variance that can be used in a similar way for biased estimators. If W is an estimator of θ, the MSE of W is defined as $E([W - \theta]^2)$, the average squared difference between W and θ. This can be written as

$$E\left([W-\theta]^2\right) = \text{Var}(W) + \left\{E(W) - \theta\right\}^2 = \text{Var}(W) + \left\{\text{Bias of } W\right\}^2.$$

Thus small MSE indicates small combined variance and (squared) bias.

Estimators are often compared by their *efficiency* – the reciprocal of the ratio of their

variances if they are unbiased or their MSEs if biased. Other things being equal, the more efficient estimator of a pair would be preferred.

As an example, the sample median, M say, is also an unbiased and consistent estimator of μ in N(μ, σ^2). Its variance can be shown to be $\pi\sigma^2/(2n)$. So its efficiency relative to the sample mean is

$$\frac{\frac{\sigma^2}{n}}{\frac{\pi\sigma^2}{2n}} = \frac{2}{\pi}$$ – i.e. it is less efficient than the sample mean.

Provided certain "regularity conditions" are satisfied, it can be shown that there is a minimum below which the variance of an unbiased estimator cannot fall. This is the Cramér-Rao lower bound. If we can find an unbiased estimator whose variance actually attains this bound, it is likely to be very useful – it is a "best unbiased estimator".

As an example, it can be shown that the Cramér-Rao lower bound for unbiased estimators of μ in N(μ, σ^2) is σ^2/n. But we know that this is the variance of the sample mean $\bar{X}$, so this is the "best" estimator on this criterion.

It is also useful to consider sufficient statistics as the basis for best unbiased estimators (the Rao-Blackwell approach).

Maximum likelihood estimators often have very good asymptotic (i.e. large-sample) properties (e.g. consistency, asymptotic unbiasedness, small asymptotic variance, asymptotic underlying Normality as a basis for confidence intervals), and this (as well as their intuitive appeal) makes them desirable. But their properties can be far from optimal in small samples. Some MLEs are very good in small samples, but unless the small-sample behaviour is actually examined an MLE should not necessarily be automatically used in preference to others. This consideration arises, for example, in estimating variances.

In decision theoretic analysis, the concept of admissibility can be useful, together with risk and loss functions.