14.30 Statistics - Fall 2003
Exam #3 Solutions

Prepared by Eric Moos

1. True, False, or Uncertain:

   (a) False - Consistency does not necessarily imply unbiasedness. Consider the following estimator for σ^2:
   $$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$
   Since this estimator uses $\frac{1}{n}$ instead of $\frac{1}{n-1}$, as S^2 does, and since S^2 is unbiased, the estimator $\widehat{\sigma^2}$ is biased. But it is consistent, as a simple LLN argument proves.

   (b) False - An interval estimator is a random interval containing the true parameter value with some probability, called the confidence level. The random interval is typically constructed to achieve some particular confidence level.

   (c) False - The power function is defined as
   $$\pi\left(\theta|\delta\right) \equiv P\left(\text{rejecting } H_0|\theta \in \Omega\right)$$
   The parameter θ does not necessarily need to be in the null hypothesis. Also, α_θ and β_θ cannot sum to 1 because they are defined on different intervals. α_θ is defined only on the null hypothesis Ω_0, while β_θ is defined only on the alternative hypothesis Ω_1.

   (d) False - The standard error is the standard deviation of of an estimator.

2. Four estimators from a $N\left(\theta_1, \theta_2\right)$ population:

   (a) Expected values:
   $$\begin{aligned} E\left[T_2\right] &= E\left[\frac{1}{n-3}\sum_{i=1}^{n} X_i\right] = \frac{1}{n-3}\sum_{i=1}^{n} E\left[X_i\right] \\ &= \frac{1}{n-3}\sum_{i=1}^{n}\theta_1 = \frac{n}{n-3}\theta_1 \end{aligned}$$
   Therefore, T_2 is biased.
   $$\begin{aligned} E\left[T_3\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n/2} X_i\right] = \frac{1}{n}\sum_{i=1}^{n/2} E\left[X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n/2}\theta_1 = \frac{1}{n}\left(\frac{n}{2}\right)\theta_1 = \frac{\theta_1}{2} \end{aligned}$$
   T_3 is also biased.

   (b) Consistency: We rely on the Law of Large Numbers to prove consistency.
   $$T_2 = \frac{1}{n-3}\sum_{i=1}^{n} X_i = \frac{n}{n-3}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \rightarrow 1 * \theta_1 = \theta_1$$
   Therefore, T_2 is consistent.
   $$T_3 = \frac{1}{n}\sum_{i=1}^{n/2} X_i = \frac{1}{2}\left(\frac{1}{n/2}\sum_{i=1}^{n/2} X_i\right) \rightarrow \frac{1}{2}\theta_1$$

Therefore, T_3 is inconsistent.

$$T_4 = \frac{1}{n^2}\sum_{i=1}^{n} X_i = \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \to 0 * \theta_1 = 0$$

Therefore, T_4 is inconsistent.

(c) Variancs of estimators:

$$Var\left[T_1\right] = Var\left[\overline{X}\right] = \frac{\theta_2}{n}$$

and

$$\begin{aligned} Var\left[T_3\right] &= Var\left[\frac{1}{n}\sum_{i=1}^{n/2} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n/2} Var\left[X_i\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n/2}\theta_2 = \frac{1}{n^2}\left(\frac{n}{2}\theta_2\right) = \frac{\theta_2}{2n} \end{aligned}$$

Therefore, $Var\left[T_3\right] < Var\left[T_1\right]$.

(d) Actual data:

$$\begin{aligned} \overline{x} &= \frac{1}{3}\left(1+5+3\right) = \frac{9}{3} = 3 \\ s^2 &= \frac{1}{3-1}\left[\left(1-3\right)^2 + \left(5-3\right)^2 + \left(3-3\right)^2\right] \\ &= \frac{1}{2}\left(8\right) = 4 \\ Var\left[\overline{X}\right] &= \frac{\theta_2}{3} \end{aligned}$$

For method of moments estimation,

$$\begin{aligned} \overline{x} &= E\left[X\right] = \theta_1 \\ \widehat{\theta_1} &= \overline{x} \\ \frac{1}{n}\sum_{i=1}^{n} x^2 &= E\left[X^2\right] = \theta_2 + \theta_1^2 \\ \widehat{\theta_2} &= \frac{1}{n}\sum_{i=1}^{n} x^2 - \widehat{\theta_1}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} x^2 - \overline{x}^2 \\ &= 8/3 \end{aligned}$$

3. Maximum likelihood estimation:

(a) In order to calculate the MLE of the mean, we need to calculate the mean of the distribution.

$$\begin{aligned} E\left[Y\right] &= \int_0^1 y\theta y^{\theta-1} dy = \int_0^1 \theta y^{\theta} dy \\ &= \frac{\theta}{\theta+1} y^{\theta+1}\Big|_0^1 = \frac{\theta}{\theta+1} \end{aligned}$$

Calculating the MLE of $\frac{\theta}{\theta+1}$ is simplfied by the invariance property of MLE's; we need only calculate the MLE of θ, $\widehat{\theta}$, and then use $\frac{\widehat{\theta}}{\widehat{\theta}+1}$ as the MLE. To calculate the MLE, we need the joint pdf of the sample, now known as the likelihood function; it will be convenient to use the natural logarithm of the likelihood function.

$$\begin{aligned}\ln f(\mathbf{y}|\theta) &= \ln\prod_{i=1}^{n} f(y_i|\theta) = \sum_{i=1}^{n}\ln f(y_i|\theta) \\ &= \sum_{i=1}^{n}\ln\left(\theta y_i^{\theta-1}\right) = \sum_{i=1}^{n}\left[\ln\theta + (\theta-1)\ln y_i\right] \\ &= n\ln\theta + (\theta-1)\sum_{i=1}^{n}\ln y_i\end{aligned}$$

The first-order condition of the maximization problem is the following:

$$\begin{aligned}\frac{n}{\theta} + \sum_{i=1}^{n}\ln y_i &= 0 \\ \widehat{\theta} &= \frac{-n}{\sum_{i=1}^{n}\ln y_i}\end{aligned}$$

Therefore,

$$\begin{aligned}\widehat{\mu} &= \frac{\widehat{\theta}}{\widehat{\theta}+1} = \frac{\frac{-n}{\sum_{i=1}^{n}\ln y_i}}{1+\frac{-n}{\sum_{i=1}^{n}\ln y_i}} \\ &= \frac{-n}{\sum_{i=1}^{n}\ln y_i - n}\end{aligned}$$

(b) Considering the limited information constraint, the probability that we observe 0.4 is the following:

$$\begin{aligned}P(Y \le 0.4) &= \int_{0}^{0.4}\theta y^{\theta-1}dy \\ &= y^{\theta}\Big|_{0}^{0.4} = (0.4)^{\theta}\end{aligned}$$

Letting m be the number of observations that are not equal to 0.4 (hence, $n-m$ observations are 0.4), the new likelihood function is the following:

$$\begin{aligned}f(\mathbf{y}|\theta) &= \prod_{i=1}^{n}\ln f(y_i|\theta) = (0.4)^{(n-m)\theta}\prod_{i=1}^{m}\theta y_i^{\theta-1} \\ &= (0.4)^{(n-m)\theta}\,\theta^{n}\left(\prod_{i=1}^{m}y_i\right)^{\theta-1}\end{aligned}$$

with the following log-likelihood function:

$$\ln f(\mathbf{y}|\theta) = (n-m)\,\theta\ln(0.4) + m\ln\theta + (\theta-1)\sum_{i=1}^{m}\ln y_i$$

where The first-order condition for this new, limited information MLE problem is the following:

$$\begin{aligned}(n-m)\ln(0.4)+\frac{m}{\theta}+\sum_{i=1}^{m}\ln y_i &= 0\\ \widehat{\theta} &= \frac{-m}{(n-m)\ln(0.4)+\sum\limits_{i=1}^{m}\ln y_i}\end{aligned}$$

4. Given the distributions of the X sample and Y sample, we can conclude that

$$\overline{X}\sim N\left(\mu_X,\frac{\sigma_X^2}{n_X}\right)$$

and

$$\overline{Y}\sim N\left(\mu_Y,\frac{\sigma_Y^2}{nY}\right)$$

Therefore,

$$\begin{aligned}\overline{X}+\overline{Y} &\sim N\left(\mu_X+\mu_Y,\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}\right)\\ \frac{\overline{X}+\overline{Y}-\mu_X+\mu_Y}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}} &\sim N(0,1)\end{aligned}$$

since the samples are independent. In order to construct the confidence interval, we do the following:

$$\begin{aligned}95\% &= P\left(-1.96\leq\frac{\overline{X}+\overline{Y}-\mu_X-\mu_Y}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}}\leq 1.96\right)\\ &= P\left(-1.96\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}\leq\overline{X}+\overline{Y}-\mu_X-\mu_Y\leq 1.96\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}\right)\\ & P\left(\overline{X}+\overline{Y}-1.96\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}\leq\mu_X+\mu_Y\leq\overline{X}+\overline{Y}+1.96\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}\right)\end{aligned}$$

5. Hypothesis testing:

   (a) Let us consider a simple two-sided test of the following hypotheses:

   $$\begin{aligned}H_0 &: \mu=0\\ H_1 &: \mu\neq 0\end{aligned}$$

   For a simple two-sided test, we reject H_0 if

   $$|\overline{x}|>k$$

   (Typically, our decision rule will be something of the form "reject H_0 if $\overline{x}>k_1$ or if $\overline{x}<k_2$." But since the normal and $t-$distributions, which we will be using, are symmetric, we know that $k_1=k_2=k$. Since the distribution is unknown, we must hope that some limiting result helps us. Fortunately, the Central Limit Theorem assures us that $\overline{X}$ is asymptotically normal. A sample size of 100 should be (hopefully) sufficiently large to assure normality.

   $$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

Although we do not know σ^2, this is no problem. We can estimate it using s^2. Although this might suggest to you that we must use the $t-$distribution, we still use the normal distribution.

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$$

If n is large enough to let us use the CLT and assume normality, then n is large enough that estimation of σ^2 is qualitatively inconsequential.

$$s^2 \approx \sigma^2$$

Notice that the $t-$distribution table only includes degrees of freedom less than 39 at which point it skips to infinity, which corresponds exactly with the normal distribution. The cut-off value k is chosen to satisfy the size requirement, $\alpha = 5\%$.

$$\begin{aligned}
0.5 &= \alpha \equiv P\left(\left|\overline{X}\right| > k|\mu = 0\right) \\
&= 1 - P\left(\left|\overline{X}\right| \leq k|\mu = 0\right) \\
&= 1 - P\left(-k \leq \overline{X} \leq k|\mu = 0\right) \\
95\% &= P\left(-\sqrt{n}\frac{k}{s} \leq \sqrt{n}\frac{\overline{X}}{s} \leq \sqrt{n}\frac{k}{s}|\mu = 0\right) \\
&= P\left(-\sqrt{n}\frac{k}{s} \leq Z \leq \sqrt{n}\frac{k}{s}|\mu = 0\right) \\
\sqrt{n}\frac{k}{s} &= 1.96 \\
k &= \frac{1.96s}{\sqrt{n}} = \frac{1.96\,(10)}{10} = 1.96
\end{aligned}$$

The decision rule is, thus, to reject H_0 if $|\overline{x}| > 1.96$.

(b) We had our choice in part a as to whether we used the normal or $t-$distributions. In general, the $t-$distribution is "fatter" than the normal distribution. The fact that we use the estimated s^2 instead of a known σ^2 tends to add more uncertainty (i.e., variance). For our "large" sample, this distinction is minor, so we can just assume normality. Using the $t-$distribution is a conservative approach, we are less likely to reject the null if we use the $t-$distribution rather than the normal due to the added uncertainty introduced through estimating σ^2.

(c) Because our estimate value of $\overline{x}$, 3, is in the rejection region, we can reject the null hypothesis; there is enough statistical evidence to reject H_0. In saying that there is "enough statistical evidence" to reject H_0, we mean that the observed sample is sufficiently improbable under the assumption that the null hypothesis is true that we should not conclude that the null hypothesis is correct.

(d) The p-value is the minimum level of significance that would have implied rejecting the null hypothesis, given the particular realization of the random sample. This occurs when the $\overline{x}$ we observe is exactly the cut-off value for the rejection region. If the cut-off is larger, then we would not reject H_0 under the given sample; if the cut-off is smaller then the significance level is not minimized. Therefore, the p-value can be calculated according to the following:

$$\begin{aligned}
p - value &= P\left(\left|\overline{X}\right| > \overline{x} = 3|\mu = 0\right) \\
&= 1 - P\left(-3 < \overline{X} < 3|\mu = 0\right) \\
&= 1 - P\left(-\sqrt{n}\frac{3}{s} < Z < \sqrt{n}\frac{3}{s}|\mu = 0\right) \\
&= 1 - P\left(-3 < Z < 3|\mu = 0\right) \\
&= 1 - 0.9974 \\
&= 0.26\%
\end{aligned}$$

Our p-value is quite small.

(e) According to the definition of Type II errors,

$$\begin{aligned}\beta(\mu) &= P(\text{accept } H_0|\mu \text{ such that } \mu \neq 0)\\ &= P\left(\left|\overline{X}\right| \leq 1.96|\mu \text{ such that } \mu \neq 0\right)\\ &= P\left(-1.96 \leq \overline{X} \leq 1.96|\mu \text{ such that } \mu \neq 0\right)\\ &= P\left(-\sqrt{n}\frac{1.96-\mu}{s} \leq \sqrt{n}\frac{\overline{X}-\mu}{s} \leq \sqrt{n}\frac{1.96-\mu}{s}|\mu \text{ such that } \mu \neq 0\right)\\ &= P\left(-\sqrt{n}\frac{1.96-\mu}{s} \leq Z \leq \sqrt{n}\frac{1.96-\mu}{s}|\mu \text{ such that } \mu \neq 0\right)\end{aligned}$$

Once again, we simplify by using the normal distribution rather than the $t-$distribution. The Type II error probabilities depend upon μ.

(f) With only a sample of 20 observations and with our assumption of a normal population, we must stick to the $t-$distribution, since s^2 is really just an estimate. Our sample is not large enough to let us abstract and assume normality of the test statistic. Building off of our previous results, our rejection region will still have the form of $|\overline{x}| > k$. In order to choose

$$\begin{aligned}95\% &= P\left(-\sqrt{n}\frac{k}{s} \leq \sqrt{n}\frac{\overline{X}}{s} \leq \sqrt{n}\frac{k}{s}|\mu = 0\right)\\ &= P\left(-\sqrt{n}\frac{k}{s} \leq t_{n-1} \leq \sqrt{n}\frac{k}{s}|\mu = 0\right)\\ &= P\left(-\sqrt{20}\frac{k}{10} \leq t_{19} \leq \sqrt{20}\frac{k}{10}|\mu = 0\right)\\ \sqrt{20}\frac{k}{10} &= 2.093\\ k &= 4.68\end{aligned}$$

Hence, we no longer reject the null hypothesis, as our value for the test statistic does not lie in the rejection region. The smaller sample size has introduced additional uncertainty; we must account for our estimation of σ, using s, which forces us to use the $t-$distribution.

6. Simple hypotheses on a normal distribution:

(a) For a simple one-sided test on the hypotheses

$$\begin{aligned}H_0 &: \mu = 0\\ H_1 &: \mu = 1\end{aligned}$$

we reject the null hypothesis when $\overline{x} > k$. Since $\sigma^2 = 1$ is known, we need not resort to the $t-$distribution. We choose k and n in order to achieve desided size and Type II error probability. The size condition implies the following:

$$\begin{aligned}1\% &= \alpha \equiv P\left(\overline{X} > k|\mu = 0\right)\\ &= P\left(\sqrt{n}\overline{X} > k\sqrt{n}|\mu = 0\right)\\ &= P\left(Z \geq k\sqrt{n}|\mu = 0\right)\\ k\sqrt{n} &\approx 2.33\end{aligned}$$

And the Type II error probability condition implies the following:

$$\begin{aligned}5\% &= P\left(\overline{X} \leq k|\mu = 1\right)\\ &= P\left(\sqrt{n}\left(\overline{X}-1\right) \leq \sqrt{n}(k-1)|\mu = 1\right)\\ &= P\left(Z \leq \sqrt{n}(k-1)|\mu = 1\right)\\ \sqrt{n}(k-1) &= -1.645\end{aligned}$$

This system of 2 equations in 2 unknowns can be solved for k and n as follows:

$$\begin{aligned}
-1.645 &= \sqrt{n}\,(k-1) = k\sqrt{n} - \sqrt{n} \\
&= 2.33 - \sqrt{n} \\
\sqrt{n} &= 3.975 \\
n &= 15.80 \\
k &= \frac{2.33}{\sqrt{n}} = \frac{2.33}{3.975} = 0.586
\end{aligned}$$

(b) This test is optimal! When the population is normal, the Simple 1-Sided test is really just a reduced form of the likelihood ratio test, which the Neyman-Pearson Lemma tells us is optimal.