Stat 135, Fall 2006 A. Adhikari Midterm Review of General Results

We have studied two main aspects of inference: estimation and the testing of hypotheses.

ESTIMATION

I. Estimating the mean of a finite population. The population consists of N numbers with mean μ and standard deviation σ . In case the population is divided into two categories, the numbers are 0's and 1's, the mean is the proportion p of 1's, and the SD is \sqrt{pq} . 7.1, 7.2.

Random sampling (with or without replacement) results in observations X_1, X_2, \ldots, X_n . If the sampling is done with replacement, the X_i 's are i.i.d. If it is done without replacement, the sample is called a **simple random sample**. Its elements all have the same distribution, and indeed are exchangeable, but they are not independent. **7.3**.

The sample average X is an **unbiased estimate** of μ , whether the sampling is done with replacement or without. The **standard error** of \overline{X} is σ/\sqrt{n} if the sampling is done with replacement. This standard error must be multiplied by the **finite population correction** factor if the sampling is done without replacement. 7.3.1, 7.3.2.

If the sampling is done with replacement, the CLT implies that for large n the **sampling distribution** of \overline{X} will be approximately normal, no matter what the distribution of the population. This can be used to construct **confidence intervals** for μ . The same holds in the case of simple random sampling, provided n is large but small compared to N. **7.3.3**.

If σ is unknown and n is large, the estimates $\hat{\sigma}$ or S can be used in place of σ in the calculation of the confidence interval. This is an example of **bootstrapping**. The estimate S^2 is unbiased for σ^2 . **7.3.2**. When the sample size is large, $\hat{\sigma}$ and S are almost equal.

If σ is unknown and and you are sampling without replacement a large number relative to the population, then you have to be careful about corrections. The table at the end of 7.3.2 (page 214) has a summary, but you don't have to memorize them for the midterm.

We discussed what "confidence" means, what confidence intervals can and cannot be used for, and how to adjust what we know about the population mean in order to estimate a population total. **7.3.2**, **7.3.3**.

II. Estimating parameters of an underlying distribution. Now the model is that we have n i.i.d. observations from a distribution which has parameters. There are two main techniques for estimating these parameters. 8.3.

IIA. The method of moments (8.4). The population moments are computed from the underlying distribution, and are functions of the parameters. The sample moments are averages of powers of the sample. The sample moments are unbiased estimates of the corresponding population moments, and converge in probability to the corresponding population moments.

To compute method of moments estimates,

(i) Calculate the population moments in terms of the parameters.

(ii) Re-write the results of (i) so that the parameters are expressed in terms of the population moments.

(iii) The method of moments estimate of a parameter is its expression in (ii) with the population moments replaced by the corresponding sample moments.

If the distribution of the method of moments estimate is known or can be approximated, it may be possible to construct confidence intervals for the parameter.

Tools. Population moments can be computed directly from the population distribution, or by using **moment generating functions**, **4.5.** [You are not expected to work with m.g.f.s on the midterm.] If the expectation and variance of the method of moments estimates are not easy to compute directly, they can be approximated. The δ -method is one way to do this **4.6**.

IIB. Maximum likelihood (8.5.). The **likelihood** of the data is the joint density, or the joint probability function in the discrete case, of the data. The data may be i.i.d. from a distribution, or they may be dependent observations (e.g. multinomial), or they may be functions of i.i.d. variables (e.g. in one of your homework problems, instead of i.i.d. X_1 , X_2, \ldots , the observation was the minimum of the X_i 's).

To compute the maximum likelihood estimate of a parameter, treat the data as fixed and maximize the likelihood as a function of the parameter. The maximizing value of the parameter is the estimate. How you maximize the likelihood depends on the complexity of the likelihood function. Don't compute the log right away - first look at the likelihood itself and see if it's easy to maximize directly (e.g. in a couple of your homework problems the likelihood was monotone). If not, and if it's a product, take the log and now stare at the log likelihood. Perhaps that's easy to maximize directly. If not, take the derivative, set equal to 0, etc.

If you have n i.i.d. observations from an underlying density (or probability function) that is smooth and well-behaved, the maximum likelihood estimate of the parameter θ has nice properties when the sample size is large. It is consistent, it is asymptotically unbiased, its approximate variance is $1/nI(\theta)$ where $I(\theta)$ is the Fisher information, and it is asymptotically normal. All this can be used to construct approximate confidence intervals for θ . 8.5.2, 8.5.3.

The Cramer-Rao bound says that the variance of every unbiased estimate of θ based on X_1, X_2, \ldots, X_n is at least as large as $1/nI(\theta)$. Just read the statement of Theorem A in 8.7. So the MLE is **asympotically efficient.** And for a fixed n, if the MLE is unbiased and has a variance equal to $1/nI(\theta)$, then no other unbiased estimator can beat it. (However, there may be a biased estimator with smaller variance ...!)

TESTING

The set-up: X_1, X_2, \ldots, X_n i.i.d. from some distribution which has an unknown parameter (maybe more than one). There is a claim about the value of an unknown parameter. How to evaluate the claim?

Terminology. Null and alternative hypotheses, which may be simple or composite. Composite hypotheses are often either one sided or two sided. A test has errors of Type I and Type II. It also has a significance level, a rejection region whose boundary is often specified by critical values, and an observed significance level which is also called the *P*-value. It has power against each fixed value of the alternative. 9.1, 9.2.

General facts. These apply to likelihoods based on well-behaved density or probability functions.

A. If both the hypotheses are simple, and if there is a test of level α that rejects the null when the **likelihood ratio** is small, then the Neyman-Pearson Lemma says that the likelihood ratio test is at least as powerful as any other test of the same or lower level. 9.2.

B. The generalized likelihood ratio is used to test composite hypotheses. As before, the null hypothesis is rejected when the ratio is small. This can be used to set up rejection regions to achieve a specified level. 9.4.

C. There is an obvious duality between confidence intervals and tests. Thus you may be able to examine a confidence interval to decide whether or not a particular null hypothesis will be rejected. **9.3**.

D. The **power** of a test against a particular value of the alternative is the chance of rejecting the null hypothesis when that particular value of the alternative happens to be the truth. To calculate power you have to first figure out the rejection region (this will be based on the significance level and the null distribution of the natural statistic) and then calculate the probability of the rejection region under the particular value of the alternative. We have discussed the rough shape of the power function of certain standard tests.

E. A test is **uniformly most powerful** against a composite hypothesis if for each fixed value of the alternative it is the most powerful test of its level.

Applications. All the tests described below are likelihood ratio tests. We have covered z and t tests for the mean of a population, and for the difference between the means of two populations.

Note. Remember the duality between confidence intervals and tests. In the situations below, if you can test hypotheses about a parameter then you should be able to construct confidence intervals for the parameter.

Tests for the population mean μ . Your sample is i.i.d. from some distribution with mean μ and variance σ^2 . An important special case is that of the dichotomous variable, where $\mu = p$ and $\sigma^2 = pq$. The natural statistic is the sample mean \bar{X} which has expectation

 μ and standard error σ/\sqrt{n} .

Large n. By the CLT, the distribution of \bar{X} is roughly normal no matter what the underlying distribution of the population. Also, if σ is unknown it can be estimated by either $\hat{\sigma}$ or S because the two will be almost equal. The normal curve can be used to set up rejection regions and compute approximate p-values. This is called a "one-sample z-test". The **power** of this test against a particular μ in the alternative can be be computed by using the normal curve.

This applies also when the sample is a large SRS from a finite population, provided the sample size is small relative to the population size. In such a case the sample is essentially i.i.d.

Small n. Now assumptions about the underlying distribution become important.

(i) If the population distribution is normal and σ is known, then the distribution of X is normal with mean μ and standard error σ/\sqrt{n} , so the z-test works but as an exact test this time, not an approximation.

(ii) If the population distribution is normal with an unknown σ , then $(\bar{X} - \mu)/(S/\sqrt{n})$ has the t distribution with n-1 degrees of freedom. This can be used in place of the normal curve and is called the t-test. There is no approximation here either.

(iii) In the dichotomous case the exact distribution of the number of successes is binomial (or hypergeometric, for simple random sampling). This can be used to get exact *p*-values etc. Again, no approximation.

(iv) If the population distribution is neither normal nor dichotomous you may be able to look at the likelihood ratio directly and come up with a rejection region. You may have to use facts about the distribution in question, such as "sums of independent Poissons are Poisson."

(v) If all else fails, use the bootstrap. Resample, and use the observed distribution of sample means instead of the normal, t, etc. All your p-values etc will most definitely be approximations.

Tests for the difference between two population means. If you have independent samples from the two populations, then just use the fact that $\bar{X} - \bar{Y}$ has expectation $\mu_X - \mu_Y$ and standard error $\sqrt{\sigma_X^2/n_X + \sigma_Y^2/n_Y}$.

The discussion is much as before, with the CLT etc. kicking in to make the large sample case very easy. That is called the "two-sample *z*-test". In some situations, however, there is the new element of **pooling**.

(i) Large n, dichotomous case: If you are testing for the equality of two population proportions, then you should use the natural **pooled** estimate of this common proportion when you estimate standard errors under the null hypothesis.

(ii) If you have two small independent samples from normal populations with the same unknown variance, there use the **pooled** estimate of the common variance, and use

the *t*-test with $n_X + n_Y - 2$ degrees of freedom.

The "paired" case. If you have one set of individuals and you want to compare their mean responses to two different treatments (e.g. you want to compare the mean pre-treatment blood-pressure and the mean post-treatment blood-pressure of patients in a population) then your observations will be pairs (X_i, Y_i) where both elements of the pair are measurements on the same individual. So if you think of your data as two samples X_1, X_2, \ldots, X_n and Y_1, Y_2, \ldots, Y_n , then the two samples will be dependent then so will \bar{X} and \bar{Y} . Therefore you can no longer use the formula $\sigma_X^2/n_X + \sigma_Y^2/n_Y$ for the variance of $\bar{X} - \bar{Y}$. You have to account for the dependence. There are two main ways of doing this.

The first way is to compute the individual differences $X_i - Y_i$ (thus combining the two samples into one), and then do a one-sample test for the mean of the differences. This SE calculation in this method is equivalent to including the covariance term in the calculation of $Var(\bar{X} - \bar{Y})$.

The second way of dealing with this is the **sign test.** This just computes the proportion of positive differences $X_i - Y_i$, and compares that to 1/2 using the binomial/normal distributions.