Stat 426 : Homework 3.

Moulinath Banerjee

March 8, 2004

**Announcement:** For purposes of the homework, you can cite any results in the handouts or the text-book or any others proved in class, without proof. The homework carries a total of 62 points. The maximum possible score is 60 points.

- 1. Consider the standard estimator of $\sigma^2$ based on $X_1, X_2, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$. This is

$$s^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$ 

We will show that $s^2$ is consistent for $\sigma^2$ - i.e. $s^2$ converges in probability to $\sigma^2$.

To this end, define $Y_i = X_i - \mu$ and show that

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n \bar{Y}^2.$$ 

Thus,

$$s^2 = \frac{n}{n - 1} \left( \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 \right).$$ 

How are the $Y_i$’s distributed? Now, use the Weak Law of Large Numbers to show that $s^2$ converges in probability to $\sigma^2$. (8 points)

- 2. (a) **Two sample problems and confidence intervals:** Let $X_1, X_2, \ldots, X_n$ be a sample from a $N(\mu_X, \sigma_X^2)$ distribution and $Y_1, Y_2, \ldots, Y_m$ be a sample from a $N(\mu_Y, \sigma_Y^2)$ distribution and let the $X_i$’s be independent of the $Y_j$’s. For example, you can think of the $X_i$’s as a random sample of SAT scores for one year and the $Y_j$’s as a random sample of SAT scores from a different year.

(i) Suppose first that $\sigma_X^2$ and $\sigma_Y^2$ are known. How would you construct a confidence interval of a fixed level $1 - \alpha$ for the difference of the population means, i.e. $\mu_X - \mu_Y$?

**Hint:** What is the distribution of $\bar{X} - \bar{Y}$? Can you construct a pivot out of this?
(ii) Now, suppose that \( \sigma_X^2 \) and \( \sigma_Y^2 \) are unknown. Let

\[
s_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2
\]

be the usual unbiased estimator of the variance \( \sigma_X^2 \) and let

\[
s_Y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \overline{Y})^2
\]

be the usual unbiased estimator of \( \sigma_Y^2 \). How would you find a confidence interval for the ratio of variances \( \sigma_X^2 / \sigma_Y^2 \) using \( F \)-distribution tables?

(iii) Now let \( \sigma_X^2 = \sigma_Y^2 = \sigma^2 \) and suppose that \( \sigma^2 \) is unknown. What is the distribution of

\[
(n - 1) \sigma_X^2 + (m - 1) \sigma_Y^2
\]

How would you use this result to find a confidence interval for \( \sigma^2 \)?

(iv) What is the distribution of \( \overline{X} - \overline{Y} \) when we have the set-up in (iii)? How would you find a confidence interval for \( \mu_X - \mu_Y \) in this case? Also suggest a level \( \alpha \) test for testing \( H_0 : \mu_X = \mu_Y = 0 \).

(v) **Behrens-Fisher Problem:** Finally consider the most general situation: \( X_1, X_2, \ldots, X_n \) are i.i.d. \( N(\mu_X, \sigma_X^2) \) and \( Y_1, Y_2, \ldots, Y_m \) are i.i.d. \( N(\mu_Y, \sigma_Y^2) \). Both \( \sigma_X^2 \) and \( \sigma_Y^2 \) are unknown; we seek to find a confidence interval for \( \mu_X - \mu_Y \). Consider the following quantity:

\[
D = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{s_X^2/n + s_Y^2/m}}.
\]

The exact distribution of \( D \) is difficult to obtain. Argue that as the sample sizes \( n \) and \( m \) both become large the above quantity behaves approximately like a \( N(0,1) \) random variable; therefore it becomes an approximate pivot.

Use this fact to get an approximate level \( 1 - \alpha \) confidence interval for \( \mu_X - \mu_Y \). \( (4 + 4 + 4 + 7 + 8 = 27 \text{ points}) \)

• 3. Let \( U \) and \( V \) be i.i.d. \( N(0,1) \) random variables. Let \( X = U \) and \( Y = U/V \). Use the change of variable theorem to compute the joint density of \( (X,Y) \). Hence deduce the marginal density of \( Y = U/V \). Do you recognize this as something you have seen before?

Next, let \( G = U/|V| \). Show that \( G \) follows a \( t \) distribution on 1 degree of freedom. **Without using the formula for the \( t \) density** show that \( G \) and \( Y \) have the same distribution. \( (6 + 6 = 12 \text{ points}) \)
• 4. If $X$ and $Y$ are independent exponential random variables with the same parameter $\lambda$, then show that $X/Y$ follows an $F$ distribution and identify the degrees of freedom. (8 points)

• 5. Suppose that $X$ follows Exponential $\lambda$. How can you use $\Psi(\lambda, X) = \lambda X$ to construct a 95% confidence interval for $\lambda$? (7)