STAT 269 – Winter 2009

Midterm Test 1

1. **Method of moments, maximum likelihood, best unbiased estimators.** Let $x_1, ..., x_n$ be an independent sample from a population with probability density

$$f(x|\theta) = \frac{1}{\theta} x^{\frac{1}{\theta}-1}, \quad 0 \le x \le 1,$$

depending on an unknown parameter $\theta > 0$.

a) Find a method of moments estimator $\bar{\theta}$ of the unknown parameter θ.

b) Find the maximum likelihood estimator $\hat{\theta}$ of θ.

c) Show that $\hat{\theta}$ is the best unbiased estimator of θ.

d) Give a $100(1-\alpha)\%$ confidence interval for the unknown parameter θ, based on the maximum likelihood estimator $\hat{\theta}$, where α is any small number.

Solution: a) Since

$$\mu = \int_0^1 x f(x|\theta)\, dx = \frac{1}{\theta} \int_0^1 x^{\frac{1}{\theta}}\, dx = \frac{1}{\theta} \left. \frac{x^{\frac{1}{\theta}+1}}{\frac{1}{\theta}+1} \right|_0^1 = \frac{1}{\theta} \frac{1}{\frac{1}{\theta}+1} = \frac{1}{1+\theta},$$

we can express θ in terms of μ as

$$\theta = \frac{1}{\mu} - 1.$$

Thus the method of moments estimator of θ is

$$\bar{\theta} = \frac{1}{\bar{x}} - 1.$$

b) Following the steps of the what-to-do list, we find

$$\ln f(x|\theta) = -\ln\theta + \left(\frac{1}{\theta} - 1\right) \ln x;$$

$$\frac{\partial}{\partial\theta} \ln f(x|\theta) = -\frac{1}{\theta} - \frac{1}{\theta^2} \ln x.$$

From *expected score is aero*, it follows that

$$E(\ln X) = -\theta. \tag{1}$$

The *maximum likelihood equation*

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(x|\theta) = \sum_{i=1}^{n} \left(-\frac{1}{\theta} - \frac{1}{\theta^2} \ln x_i\right) = 0,$$

the maximum likelihood equation is

$$\hat{\theta} = -\frac{1}{n}\sum_{i=1}^{n} \ln x_i. \tag{2}$$

By (1), this is an unbiased estimator of θ.

The second score is

$$\frac{\partial^2}{\partial\theta^2} \ln f(x|\theta) = \frac{1}{\theta^2} + \frac{2}{\theta^3} \ln x.$$

Using (1) again, we find the Fisher information:

$$I(\theta) = E\left(-\frac{\partial^2}{\partial\theta^2} \ln f(x|\theta)\right) = -\left(\frac{1}{\theta^2} + \frac{2}{\theta^3} E(\ln X)\right) = -\left(\frac{1}{\theta^2} - \frac{2}{\theta^2}\right) = \frac{1}{\theta^2}. \tag{3}$$

By the alternative formula for the Fisher information,

$$\frac{1}{\theta^2} = I(\theta) = \mathbf{E}\left(\frac{\partial}{\partial\theta} \ln f(X|\theta)\right)^2 = E\left(-\frac{1}{\theta} - \frac{1}{\theta^2} \ln X\right)^2 = \frac{E(\ln X + \theta)^2}{\theta^4} = \frac{V(\ln X)}{\theta^4},$$

i.e.

$$V(\ln X) = \theta^2. \tag{4}$$

c) From (2) and (4),

$$V(\hat{\theta}) = \frac{V(\ln X)}{n} = \frac{\theta^2}{n} \equiv \frac{1}{nI(\theta)}.$$

Thus $\hat{\theta}$ is the best unbiased estimator of θ.

d) Using (3), a $100(1-\alpha)\%$ confidence interval for θ is given by

$$\hat{\theta} \pm \frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta})}} = \hat{\theta} \pm \frac{z_{\alpha/2}\hat{\theta}}{\sqrt{n}} = \hat{\theta}\left(1 \pm \frac{z_{\alpha/2}}{\sqrt{n}}\right).$$

2. **Confidence intervals for the differences in means, standard deviations.** The following data represents oxygen consumption in millimeters^3 per hour by fish (trout) in a rapidly flowing river (sample 1) and in slow moving waters (sample 2).

High	105	108	86	103	103	107	124	105
Low	89	92	84	97	103	107	111	97

a) Give a 95% confidence interval for the difference between means.

b) Is the null hypothesis about equality of the means acceptable, at the significance level of 5%?

c) Give a 90% confidence interval for the ratio of standard deviations.

d) Is the null hypothesis about equality of the standard deviations acceptable, at the significance level of 10%?

Solution: a) Let $x_1, ..., x_8$ be sample 1 *(high data)*, and $y_1, ..., y_8$ – sample 2 *(low data)*. The sample sizes are $n_1 = n_2 = 8$. A $100(1-\alpha)\%$ confidence interval for the difference $\mu_x - \mu_y$ is given by

$$\bar{x} - \bar{y} \pm t_{\frac{\alpha}{2}, n_1+n_2-2} \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}.$$

Here

$$\bar{x} = 105.1, \ \bar{y} = 97.5, \ s_1^2 = 106.1, \ s_2^2 = 84.$$

The pooled unbiased variance estimator is

$$s_p^2 = (s_1^2 + s_2^2)/2 = 95.1.$$

From Table 4,

$$t_{0.025,14} = 2.15.$$

Thus the confidence interval for $\mu_x - \mu_y$ is

$$105.1 - 97.5 \pm 1.15 \cdot \sqrt{\frac{s_p^2}{4}} = [-2.9, 18.1].$$

b) Since the 95% contains 0, the null hypothesis $H_0: \ \mu_x = \mu_y$ can be accepted, at the significance level 5%.

c) Since from the Table 4 $f_{0.05,7,7} = 3.79$, the confidence interval for the ratio of standard deviations σ_1/σ_2 is

$$\left[\sqrt{\frac{s_1^2}{s_2^2 \cdot f_{0.05,7,7}}}, \sqrt{\frac{s_1^2 \cdot f_{0.05,7,7}}{s_2^2}} \right] = \left[\sqrt{\frac{106.1}{84 \cdot 3.79}}, \sqrt{\frac{106.1 \cdot 3.79}{84}} \right] = [0.6, 2.2].$$

d) Since the confidence interval for σ_1/σ_2 contains 1, the null hypothesis $H_0: \ \sigma_1 = \sigma_2$ can be accepted, at the significance level 10%.

3. **Testing two proportions.** A random sample of 100 balls is drawn from box I, and 45 of the balls are black. From box II, a random sample of size 200 is taken, and 115 of these balls are black.

   a) Test the null hypothesis about proportions of black balls in these boxes

$$H_0 : p_1 = p_2,$$

   against the alternative

$$H_1 : p_1 < p_2,$$

   at the (approximate) significance level $\alpha = 0.05$.

   b) Give the (approximate) P-value corresponding to this test. Is there a strong evidence against the null hypothesis?

Solution: a) The test statistic of the text has the form

$$z = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Here $n_1 = 100$ $n_2 = 200$ and

$$\hat{p}_1 = \frac{45}{100} = 0.45, \quad \hat{p}_2 = \frac{115}{200} = 0.575, \quad \hat{p} = \frac{45+115}{100+200} = 0.543.$$

Therefore

$$z = \frac{0.575 - 0.45}{\sqrt{0.543(1-0.543)\left(\frac{1}{100} + \frac{1}{200}\right)}} = 2.05.$$

Since from the Table III,

$$z_{0.05} \approx 1.65,$$

and $z > z_{0.05}$, the null hypothesis H_0 is rejected.

b) From Table III, the approximate P-value is

$$P(Z \geq z) = P(Z \geq 2.05) = 0.02.$$

Thus there is a pretty strong evidence against H_0.

4. **Type I error.** Let $x_1, ..., x_n$ be an independent sample from the normal distribution $\mathcal{N}(\mu, \sigma^2)$, with unknown parameter μ and given σ^2. Suppose a significance level α is chosen. As was explained in the class, the uniformly most powerful level-α test of the null hypothesis

$$H_0 : \ \mu = \mu_0,$$

against the one-sided alternative

$$H_1 : \ \mu > \mu_0,$$

rejects H_0 if and only if

$$\frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} > z_\alpha. \tag{1}$$

Suppose you want to test a more general null hypothesis

$$H_0 : \ \mu \leq \mu_0,$$

against the alternative

$$H_1 : \ \mu > \mu_0.$$

Intuitively, the same test will do the job: reject H_0 if and only if (1) happens.

For a given $\mu \in H_0$ (i.e. $\mu \le \mu_0$), the probability of *type I error* is defined as

$$\alpha(\mu) = P(\text{the test rejects } H_0|\mu).$$

a) Express the probability $\alpha(\mu)$ of type I error in terms of the parameters μ, μ_0, σ, and n. **Hint: standardization**

b) Show that $\alpha(\mu) \le \alpha$, for all $\mu \le \mu_0$.

c) Find $\lim_{\mu \to \mu_0 -} \alpha(\mu)$.

d) Find $\lim_{\mu \to -\infty} \alpha(\mu)$. Interpret your result.

e) Find $\lim_{\sigma \to \infty} \alpha(\mu)$.

f) For $\mu < \mu_0$, find $\lim_{\sigma \to 0} \alpha(\mu)$. Interpret your result.

g) For $\mu < \mu_0$, find $\lim_{n \to \infty} \alpha(\mu)$. Interpret your result.

Solution: a)

$$\alpha(\mu) = P\left(\frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} > z_\alpha \middle| \mu \right) = P\left(\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} + \frac{\mu - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} > z_\alpha \middle| \mu \right) =$$

$$P\left(\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} > \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{n}}} + z_\alpha \middle| \mu \right) = P\left(Z > \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{n}}} + z_\alpha \right)$$

b) Since $\mu_0 - \mu \ge 0$,

$$\alpha(\mu) \le P(Z > z_\alpha) = \alpha.$$

c)

$$\lim_{\mu \to \mu_0 -} \alpha(\mu) = P(Z > z_\alpha) = \alpha.$$

d)

$$\lim_{\mu \to -\infty} \alpha(\mu) = 0.$$

This means that the more the true value μ becomes separated from the hypothetical value μ_0, the less likely type I error is.

e)

$$\lim_{\sigma \to \infty} \alpha(\mu) = P(Z > z_\alpha) = \alpha.$$

f) For $\mu < \mu_0$,

$$\lim_{\sigma \to 0} \alpha(\mu) = 0.$$

Thus, the smaller are errors in observations, the less likely type I error is.

g) For $\mu < \mu_0$,

$$\lim_{n \to \infty} \alpha(\mu) = 0.$$

Thus, the larger is data set, the less likely type I error is.