Sample Questions for Module 2

- Module 2 is about sampling distributions.
- Main aim is to be familiar with the t, χ^2, and F-distributions, and be able to derive CI based on a pivotal quantity.
- Important results: sampling distribution of the sample mean and the sample variance (can use these without proof in the exam)

Sample Question 4

Consider two independent samples of size n_1 and n_2:

$$Y_{i1} \sim i.i.d. N(\mu_1, \sigma_1^2), \qquad i = 1, 2, \ldots, n_1$$
$$Y_{j2} \sim i.i.d. N(\mu_2, \sigma_2^2), \qquad j = 1, 2, \ldots, n_2$$

and writing $\boxed{\sigma_2^2 = k\sigma_1^2,}$ where $k > 0$ is a constant.
Let $\bar{Y}_1$ and $\bar{Y}_2$ be their samples means and S_1^2 and S_2^2 be their sample variances.

a) Show that $Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sigma_1\sqrt{\frac{1}{n_1} + \frac{k}{n_2}}}$ has a standard normal distribution.

b) Show that $W = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2/k}{\sigma_1^2}$ has a χ^2-distribution.

c) Show that $T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{k}{n_2}}}$has a t-distribution,

   where $s_p^2 = \frac{\sigma_1^2}{n_1 + n_2 - 2} W$.

d) Using the results of part c), construct a $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$.

Sample Question 4 Solution

(a) $\bar{Y}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$ independent of $\bar{Y}_2 \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$.

By the property of linear combination of independent normal random variables,

$\bar{Y}_1 - \bar{Y}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$.

$$Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - E(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{var(\bar{Y}_1 - \bar{Y}_2)}} \sim N(0,1)$$

$$= \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

(Write $\sigma_2^2 = k\sigma_1^2$)

$$= \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sigma_1\sqrt{\frac{1}{n_1} + \frac{k}{n_2}}}$$

$\therefore Z \sim N(0,1)$.

Sample Question 4 Solution

(b) We know $U_1 = \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$ independent of

$$U_2 = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}.$$

$$W = U_1 + U_2 \sim \chi^2_{n_1+n_2-2}$$

$$= \frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{(n_2-1)S_2^2}{\sigma_2^2}$$

$$= \frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{(n_2-1)S_2^2}{k\sigma_1^2}$$

$$= \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2\left(\frac{1}{k}\right)}{\sigma_1^2} \sim \chi^2_{n_1+n_2-2}$$

Sample Question 4 Solution

(c) From parts a and c, we have $Z \sim N(0,1)$ independent of $W \sim \chi^2_{n_1+n_2-2}$.

By the definition of the t-distribution, we have

$$T = \frac{Z}{\sqrt{\frac{W}{n_1+n_2-2}}} \sim t_{n_1+n_2-2}$$

$$= \frac{\dfrac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{\sigma_1\sqrt{\frac{1}{n_1}+\frac{k}{n_2}}}}{\sqrt{\frac{W}{n_1+n_2-2}}}$$

$$= \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\boxed{\sigma_1^2}\left(\frac{1}{n_1}+\frac{k}{n_2}\right)\boxed{\left(\frac{W}{n_1+n_2-2}\right)}}}$$

S_p^2

$$= \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{k}{n_2}\right)}}$$

$$= \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{k}{n_2}}}$$

$$\sim t_{n_1+n_2-2}$$

Sample Question 4 Solution

d) $$1-\alpha = P\left(-t_{n_1+n_2-2,\frac{\alpha}{2}} < T < t_{n_1+n_2-2,\frac{\alpha}{2}}\right)$$

$$= P\left(-t_{n_1+n_2-2,\frac{\alpha}{2}} < \frac{(\bar{Y}_1-\bar{Y}_2)-(\mu_2-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{k}{n_2}}} < t_{n_1+n_2-2,\frac{\alpha}{2}}\right)$$

$$= P\left((\bar{Y}_1-\bar{Y}_2) - t_{n_1+n_2-2,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{k}{n_2}} < \mu_1-\mu_2 < (\bar{Y}_1-\bar{Y}_2) + t_{n_1+n_2-2,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{k}{n_2}}\right).$$

$\therefore$ The required CI for $\mu_1-\mu_2$ is

$$(\bar{Y}_1-\bar{Y}_2) \pm t_{n_1+n_2-2,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}+\frac{k}{n_2}}$$

Sample Question 5

Suppose $Y_1, Y_2, \dots, Y_n$ is a sample of size n from a $\mathrm{Beta}(\theta, 1)$. Consider the quantity $X = -2\theta \sum_{i=1}^{n} \log Y_i$.

a) Find the distribution of X and show that X is a pivotal quantity.

b) Use the above pivotal quantity to construct a symmetric $100(1-\alpha)\%$ confidence interval for θ.

c) Suppose we have observed the following values for Y:

0.72, 0.48, 0.80, 0.45, 0.67

Using the results of part b), give a 95% confidence interval for θ.

Sample Question 5 Solution

$Y_1, Y_2, \ldots, Y_n \sim \text{iid Beta}(\theta, 1)$

$f_{Y_i}(y) = \theta y^{\theta-1}$ for $0 < y < 1$

$X = -2\theta \sum_{i=1}^{n} \log Y_i$

(a) Let $X_i = -2\theta \log Y_i$, then $Y_i = e^{-\frac{X_i}{2\theta}}$. (inverse transformation)

$$\frac{dY_i}{dX_i} = -\frac{1}{2\theta} e^{-\frac{X_i}{2\theta}}$$

$$f_{X_i}(x) = \cancel{\theta} \left(e^{-\frac{x}{\cancel{2\theta}}} \right)^{\cancel{\theta-1}} \left| -\frac{1}{2\cancel{\theta}} e^{\cancel{-\frac{x}{2\theta}}} \right|$$ (transformation formula)

$$= \frac{1}{2} e^{-\frac{x}{2}} \quad \text{for } x > 0$$

$\therefore \quad X_i \sim \text{iid Exp}\left(\frac{1}{2}\right)$

$\therefore \quad X = \sum_{i=1}^{n} X_i \sim \text{Gamma}\left(n, \frac{1}{2}\right)$ or χ^2_{2n}.

Sample Question 5 Solution

Since X is a function θ and the sample $Y_1, Y_2, \dots, Y_n$, θ is the only unknown parameter in X, and the distribution of X does not depend on θ, so X is a pivotal quantity of θ.

(b) $$1-\alpha = P\left(\chi^2_{2n,1-\frac{\alpha}{2}} < X < \chi^2_{2n,\frac{\alpha}{2}} \right)$$

$$= P\left(\chi^2_{2n,1-\frac{\alpha}{2}} < -2\theta \underbrace{\boxed{\sum_{i=1}^{n} \log Y_i}}_{\text{less than } 0} < \chi^2_{2n,\frac{\alpha}{2}} \right)$$

$$= P\left(\frac{\chi^2_{2n,1-\frac{\alpha}{2}}}{-2\sum_{i=1}^{n} \log Y_i} < \theta < \frac{\chi^2_{2n,\frac{\alpha}{2}}}{-2\sum_{i=1}^{n} \log Y_i} \right)$$

$$\therefore \quad CI = \left(\frac{\chi^2_{2n,1-\frac{\alpha}{2}}}{-2\sum_{i=1}^{n} \log Y_i}, \frac{\chi^2_{2n,\frac{\alpha}{2}}}{-2\sum_{i=1}^{n} \log Y_i} \right)$$

Sample Question 5 Solution

c) $-2\sum_{i=1}^{n}\log Y_i = -2\left(\log 0.72 + \log 0.48 + \log 0.80 + \log 0.45 + \log 0.67\right)$

$= 4.9692$

$n = 5$,

$\chi^2_{10,0.025} = 20.4832$, qchisq(0.975, 10)

$\chi^2_{10,0.975} = 3.2470$, qchisq(0.025, 10)

$$CI = \left(\frac{3.2470}{4.9692}, \frac{20.4832}{4.9692}\right)$$

$$\approx (0.653, 4.122)$$