STATS 7107 Statistical Modelling and Inference
Tutorial 3 (Week 6)

Trimester 3, 2023

1. Suppose $Y_1, \ldots, Y_n$ is a random sample of size n from a population that follows a gamma distribution with parameters $\alpha = 2$ and $\lambda = 1/\beta$. Then $E[Y_i] = 2\beta$ and $\mathrm{var}(Y_i) = 2\beta^2$ for $i = 1, \ldots, n$.

   (a) Using moment generating functions, show that $X = \dfrac{2}{\beta}\displaystyle\sum_{i=1}^{n} Y_i$ is a pivotal quantity and $X \sim \chi^2_{4n}$.
   Hint: Recall that if $Y \sim Gamma(\alpha = 2, \lambda = 2\beta)$, then $M_Y(t) = (1 - \beta t)^{-2}$, $t < 1/\beta$.

   Solution:
   Observe that since $M_{Y_i}(t) = (1 - \beta t)^{-2}$, we have

$$
\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= E\left[\exp\left\{t\frac{2}{\beta}\sum_{i=1}^{n} Y_i\right\}\right] \\
&= E\left[\exp\left\{t\frac{2}{\beta}Y_1\right\}\right] \ldots E\left[\exp\left\{t\frac{2}{\beta}Y_n\right\}\right], \\
&\quad \text{by independence of } Y_i \\
&= M_{Y_1}\left(\frac{2}{\beta}t\right) \ldots M_{Y_n}\left(\frac{2}{\beta}t\right) \\
&= \prod_{i=1}^{n} M_{Y_i}\left(\frac{2}{\beta}t\right) \\
&= \prod_{i=1}^{n} \left(1 - \beta\frac{2t}{\beta}\right)^{-2} \\
&= (1 - 2t)^{-2n},\ t < 1/2
\end{aligned}
$$

   Then, by uniqueness of MGFs, $X \sim \chi^2_{4n}$. Since the distribution of X does not depend on β, X is a pivotal quantity.

   (b) Use the pivotal quantity X to derive a symmetric 95% confidence interval for β.

   Solution:
   We want to find L and U such that $P(L \leq \beta \leq U) = 0.95$. For a

symmetric CI, we need $P(\beta \le L) = P(U \le \beta) = 0.025$.

$$
\begin{aligned}
0.95 &= P(L \le \beta \le U) \\
&= P\left(\frac{1}{U} \le \frac{1}{\beta} \le \frac{1}{L}\right) \\
&= P\left(\frac{2\sum_{i=1}^n Y_i}{U} \le \frac{2}{\beta}\sum_{i=1}^n Y_i \le \frac{2\sum_{i=1}^n Y_i}{L}\right) \\
&= P\left(\frac{2\sum_{i=1}^n Y_i}{U} \le X \le \frac{2\sum_{i=1}^n Y_i}{L}\right) \\
&= P\left(\chi^2_{4n}(0.975) \le X \le \chi^2_{4n}(0.025)\right), \text{ as } X \sim \chi^2_{4n}
\end{aligned}
$$

Thus

$$
2\sum_{i=1}^n Y_i/L = \chi^2_{4n}(0.025) \text{ and } 2\sum_{i=1}^n Y_i/U = \chi^2_{4n}(0.975),
$$

so

$$
2\sum_{i=1}^n Y_i/\chi^2_{4n}(0.025) = L \text{ and } 2\sum_{i=1}^n Y_i/\chi^2_{4n}(0.975) = U,
$$

and hence the symmetric 95% CI for β is

$$
\left(2\sum_{i=1}^n Y_i/\chi^2_{4n}(0.025),\ 2\sum_{i=1}^n Y_i/\chi^2_{4n}(0.975)\right).
$$

(c) Suppose, for a sample size $n = 5$, a sample mean $\bar{y} = 5.39$ is obtained. Use the results of part (b) to calculate the symmetric 95% CI for β.

Solution:
Observe that $\sum_{i=1}^n Y_i = n\bar{y} = 5 \times 5.39 = 26.95$.
The critical values are $\chi^2_{20}(0.025) = 34.1696$ and $\chi^2_{20}(0.975) = 9.5908$ (using `qchisq(.975,20)` and `qchisq(0.025,20)` in R).
Then the 95% CI is

$$
\left(\frac{2 \times 26.95}{34.1696}, \frac{2 \times 26.95}{9.5908}\right) \approx (1.5774,\ 5.6200).
$$

2. Consider independent random variables $Y_{i,j} \sim N(\mu_i, \sigma^2)$, $i = 1, 2$, $j = 1, \ldots, n_i$, and let $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$.

(a) Show that S_p^2 is an unbiased estimator for σ^2.

Solution:
Recall that $E[S_1^2] = E[S_2^2] = \sigma^2$. Hence

$$\begin{aligned}
E[S_p^2 &= E\left[\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right] \\
&= \frac{n_1-1}{n_1+n_2-2}E[S_1^2] + \frac{n_2-1}{n_1+n_2-2}E[S_2^2] \\
&= \frac{n_1-1}{n_1+n_2-2}\sigma^2 + \frac{n_2-1}{n_1+n_2-2}\sigma^2 \\
&= \sigma^2.
\end{aligned}$$

Thus, S_p^2 is an unbiased estimator for σ^2.

(b) Show that $\dfrac{(n_1+n_2-2)S_p^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$.

Solution:
Observe that

$$\frac{(n_1+n_2-2)S_p^2}{\sigma^2} = \frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2}.$$

Recall that

$$\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2_{n_1-1}$$

and

$$\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_2-1},$$

independently. Then

$$\frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

(see Tutorial 2, question 1). Hence,

$$\frac{(n_1+n_2-2)S_p^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}.$$

(c) Show that

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

is a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ if σ^2 is not known.

Solution:

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

3. Consider independent random variables $Y_{i,j} \sim N(\mu_i, \sigma^2)$, $i = 1, 2, 3$, $j = 1, \ldots, n_i$. Define $\theta = a_1\mu_1 + a_2\mu_2 + a_3\mu_3$, and suppose we wish to perform inference for θ. An intuitive estimator for θ is $\hat{\theta} = a_1\bar{Y}_1 + a_2\bar{Y}_2 + a_3\bar{Y}_3$, where $\bar{Y}_i$ is the sample mean of the ith sample.

   (a) Find the standard error of $\hat{\theta}$.

   Solution:
   By independence of $Y_{i,j}$

$$SE(\hat{\theta}) = \sqrt{\operatorname{var}(\hat{\theta})}$$
$$= \sqrt{a_1^2 \operatorname{var}(\bar{Y}_1) + a_2^2 \operatorname{var}(\bar{Y}_2) + a_3^2 \operatorname{var}(\bar{Y}_3)}$$
$$= \sigma\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}}.$$

   (b) Find the distribution of $\hat{\theta}$.

   Solution:
   As $\hat{\theta}$ is a linear combination of normally distributed random variables $Y_{i,j}$, we have

$$\hat{\theta} \sim N\left(\theta, \sigma^2\left(\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}\right)\right).$$

(c) A pooled estimator for σ^2 is

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2 + (n_3-1)S_3^2}{n_1+n_2+n_3-3},$$

where S_i^2 is the sample variance of the ith sample. State the distribution of

$$W = \frac{(n_1+n_2+n_3-3)S_p^2}{\sigma^2} \text{ and } T = \frac{\hat{\theta}-\theta}{S_p\sqrt{\frac{a_1^2}{n_1}+\frac{a_2^2}{n_2}+\frac{a_3^2}{n_3}}}.$$

Solution:
As $\frac{(n_i-1)S_i^2}{\sigma^2} \sim \chi^2_{n_i-1}$ independently for $i=1,2,3$, it follows that their sum must be a $\chi^2_{n_1+n_2+n_3-3}$ distributed random variable. Thus $W \sim \chi^2_{n_1+n_2+n_3-3}$.
Further, we know that $T = \frac{Z}{\sqrt{W/(n_1+n_2+n_3-3)}} \sim t_{n_1+n_2+n_3-3}$, where $Z \sim N(0,1)$ and $W \sim \chi^2_{n_1+n_2+n_3-3}$. Then

$$\begin{aligned}
T &= \frac{Z}{\sqrt{W/(n_1+n_2+n_3-3)}} \\
&= \frac{\hat{\theta}-\theta}{\sigma\sqrt{a_1^2/n_1+a_2^2/n_2+a_3^2/n_3}} \Bigg/ \sqrt{\frac{W}{n_1+n_2+n_3-3}} \\
&= \frac{\hat{\theta}-\theta}{\sigma\sqrt{a_1^2/n_1+a_2^2/n_2+a_3^2/n_3}} \Bigg/ \sqrt{\frac{(n_1+n_2+n_3-3)S_p^2}{(n_1+n_2+n_3-3)\sigma^2}} \\
&= \frac{\hat{\theta}-\theta}{S_p\sqrt{\frac{a_1^2}{n_1}+\frac{a_2^2}{n_2}+\frac{a_3^2}{n_3}}} \sim t_{n_1+n_2+n_3-3}.
\end{aligned}$$

(d) Using the results from part (c), give a $100(1-\alpha)\%$ confidence interval for θ.

Solution:
$CI = \hat{\theta} \pm t_{n_1+n_2+n_3-3}(\alpha/2) S_p\sqrt{\frac{a_1^2}{n_1}+\frac{a_2^2}{n_2}+\frac{a_3^2}{n_3}}$.