STATS 2107
Statistical Modelling and Inference II
Tutorial 3
Solutions

Sharon Lee, Matt Ryan

Semester 2 2022

1. Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n from a gamma-distributed population with parameters $\alpha = 2$ and $\lambda = 1/\beta$, that is, with mean 2β and variance $2\beta^2$.
   (a) Use the method of moment generating functions to show that $X = \frac{2}{\beta}\sum_{i=1}^{n} Y_i$ is a pivotal quantity and has a χ^2_{4n} distribution. **Recall that if $Y \sim$ Gamma$(\alpha = 2, \lambda = 1/\beta)$, then $M_Y(t) = (1-\beta t)^{-2}$.**

Solutions:
Observe that if $Y \sim \text{Gamma}(\alpha = 2, \lambda = 1/\beta)$, then its moment generating function (MGF) is given by $M_Y(t) = (1-\beta t)^{-2}$, for $t < \frac{1}{\beta}$.

$$
\begin{aligned}
M_X(t) &= E\left[e^{tX}\right] \\
&= E\left[e^{t\left(\frac{2}{\beta}\sum_{i=1}^{n} Y_i\right)}\right] \\
&= E\left[e^{t\left(\frac{2}{\beta}Y_1\right)}\right] E\left[e^{t\left(\frac{2}{\beta}Y_2\right)}\right] \ldots E\left[e^{t\left(\frac{2}{\beta}Y_n\right)}\right] \quad \{\text{by independence}\} \\
&= \prod_{i=1}^{n} M_{Y_i}\left(\frac{2}{\beta}t\right) \\
&= \prod_{i=1}^{n} \left[1 - \beta\left(\frac{2t}{\beta}\right)\right]^{-2} \\
&= (1-2t)^{-2n}, \quad t < \frac{1}{2}
\end{aligned}
$$

The above is the MGF of χ^2_{4n}. Hence,

$$X \sim \chi^2_{4n}.$$

Also, the distribution of X does not depend on β. Hence X is a pivotal quantity.

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

Solutions:
We want to solve for L and U such that $P(L \leq \beta \leq U) = 0.95$. For symmetric confidence interval, we have $P(\beta \leq L) = 0.025$ and $P(\beta \geq U) = 0.025$.

$$
\begin{aligned}
0.95 &= P(L \leq \beta \leq U) \\
&= P\left(\frac{1}{U} \leq \frac{1}{\beta} \leq \frac{1}{L}\right) \\
&= P\left(\frac{2\sum_{i=1}^{n} Y_i}{U} \leq \frac{2}{\beta}\sum_{i=1}^{n} Y_i \leq \frac{2\sum_{i=1}^{n} Y_i}{L}\right) \\
&= P\left(\frac{2\sum_{i=1}^{n} Y_i}{U} \leq X \leq \frac{2\sum_{i=1}^{n} Y_i}{L}\right)
\end{aligned}
$$

As $X \sim \chi^2_{4n}$, and for symmetric confidence interval, we have $P(\chi^2_{4n,0.975} \leq X \leq \chi^2_{4n,0.025}) = 0.95$.

This implies

$$\frac{2\sum_{i=1}^{n} Y_i}{L} = \chi^2_{4n,0.025} \quad \text{and} \quad \frac{2\sum_{i=1}^{n} Y_i}{U} = \chi^2_{4n,0.975}.$$

Hence,

$$L = \frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{4n,0.025}} \quad \text{and} \quad U = \frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{4n,0.975}}.$$

It follows that the 95% symmetric confidence interval (CI) for β is

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

(c) If a sample of size $n = 5$ yields $\bar{y} = 5.39$, use the results from part (b) to give a 95% symmetric confidence interval for β.

Solutions:
Observe that $\sum_{i=1}^{n} Y_i = n\bar{y} = 5(5.39) = 26.95$. The critical values in the CI are $\chi^2_{20,0.025} = 34.1696$ and $\chi^2_{20,0.975} = 9.5908$.

They can be obtained from `R` using `qchisq(0.975,20)` and `qchisq(0.025,20)`, respectively.
It follows that the required CI is

$$\left(\frac{2(26.95)}{34.1696}, \frac{2(26.95)}{9.5908}\right) \approx (1.577, 5.620).$$

---

2. Consider the independent random variables

$$Y_{ij},\ i = 1, 2;\ j = 1, 2, \ldots, n_i$$

with $Y_{ij} \sim N(\mu_i, \sigma^2)$. Let

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

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

Solutions:
Recall $\mathrm{E}(S_1^2) = \mathrm{E}(S_2^2) = \sigma^2$. Hence,

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

(b) Prove that

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

Solutions:
Note that S_p^2 can be rewritten as

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

Recall

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2_{n_1-1} \quad \text{and} \quad \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2_{n_2-1} \text{ independently.}$$

Hence,

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} + \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}, \quad \{\text{using Tutorial 2 Question 1}\}$$

Therefore,

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

(c) Show that

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

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

Solutions:

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

Multiply by -1, thus swaping the inequalities, then swap endpoints to give

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

---

3. Suppose we have independent random variables

$$Y_{ij},\ i=1,2,3;\ j=1,2,\dots,n_i$$

with $Y_{ij}\sim N(\mu_i,\sigma^2)$.We would like to do inference on a linear combination of the mean $\theta=a_1\mu_1+a_2\mu_2+a_3\mu_3$. 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 $Y_{i1},Y_{i2},\dots,Y_{in_i}$.

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

Solutions:
By independence of Y_{ij},

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

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

Solutions:
Since $\hat{\theta}$ is a linear combination of normal random variables Y_{ij}, by Lemma 1, 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 $Y_{i1},Y_{i2},\dots,Y_{in_i}$. State the distribution of

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

Solutions:

Observe that $\dfrac{(n_i-1)S_i^2}{\sigma^2} \sim \chi^2_{n_i-1}$ independently for $i = 1,2,3$. It follows that their sum is a $\chi^2_{n_1+n_2+n_3-3}$ random variable. So

$$\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2}+\frac{(n_3-1)S_3^2}{\sigma^2} = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2+(n_3-1)S_3^2}{\sigma^2} = W \sim \chi^2_{n_1+n_2+n_3-3}.$$

Furthermore, we have $T = \frac{Z}{\sqrt{W/(n_1+n_2+n_3-3)}} \sim t_{n_1+n_2+n_3-3}$, where $Z \sim N(0,1)$. Hence,

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

In summary, we have

$$W \sim \chi^2_{n_1+n_2+n_3-3} \quad \text{and} \quad T \sim t_{n_1+n_2+n_3-3}$$

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

Solutions:

$$\text{CI} = \hat\theta \pm t_{n_1+n2+n_3-3,\alpha/2} S_p\sqrt{\frac{a_1^2}{n_1}+\frac{a_2^2}{n_2}+\frac{a_3^2}{n_3}}$$

(e) Using the results from part (c), develop a hypothesis test for testing $H_0 : \theta = \theta_0$ vs $H_a : \theta \neq \theta_0$.

Solutions:

- Hypotheses: $H_0 : \theta = \theta_0$ vs $H_a : \theta \neq \theta_0$
- test statistic: $t = \dfrac{\hat\theta-\theta_0}{S_p\sqrt{\frac{a_1^2}{n_1}+\frac{a_2^2}{n_2}+\frac{a_3^2}{n_3}}}$
- (Under H_0, we have $t \sim t_{n_1+n_2+n_3-3}$.)
- critical region: $|t| \geq t_{n_1+n_2+n_3-3,\alpha/2}$.