Bios 661: 1 – 5; Bios 673: 1 – 5.

1. C&B 9.3
2. C&B 9.4
3. C&B 9.17
4. Let $X_1, \ldots, X_n$ be a random sample from a distribution with probability density function
$$f(x|\theta) = \left(\frac{a}{\theta}\right)\left(\frac{x}{\theta}\right)^{a-1}, \quad 0 < x < \theta,$$
where $a \geq 1$ is known and $\theta > 0$ is unknown.

   (a) Construct a confidence interval for θ with coverage probability $1-\alpha$ by using the cumulative distribution function of the largest order statistic $X_{(n)}$.

   Solution: The cumulative density function of X is
$$F(x|\theta) = \int_0^x \left(\frac{a}{\theta}\right)\left(\frac{y}{\theta}\right)^{a-1} dy = \left(\frac{x}{\theta}\right)^a, \quad 0 < x < \theta.$$
The cumulative density function of $X_{(n)}$ is
$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = \{P(X_1 \leq x)\}^n = (x/\theta)^{na}.$$
Since $F_{X_{(n)}}(x)$ is an monotone decreasing function of θ, one can use the cumulative distribution function as the pivotal quantity, i.e.,
$$\begin{aligned} 1-\alpha &= P(\alpha_1 \leq F_{X_{(n)}}(x|\theta) \leq 1-\alpha_2) \\ &= P(L(x_{(n)}) \leq \theta \leq U(x_{(n)})), \end{aligned}$$
where
$$1-\alpha_2 = \left(\frac{x_{(n)}}{L(x_{(n)})}\right)^{na},$$
and
$$\alpha_1 = \left(\frac{x_{(n)}}{U(x_{(n)})}\right)^{na}.$$

   (b) Show that $(X_{(n)}/\theta)^{na}$ is a pivotal quantity and derive the $1-\alpha$ confidence interval using the quantity.

Solution: Let $Y = (X_{(n)}/\theta)^{na}$. One can derive

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P((X_{(n)}/\theta)^{na} \leq y) \\ &= P(X_{(n)} \leq \theta y^{1/(na)}) \\ &= (\theta y^{1/(na)}/\theta)^{na} \\ &= y, \end{aligned}$$

which is independent of θ. Therefore, $(X_{(n)}/\theta)^{na}$ is a pivotal quantity.
Since $Y = (X_{(n)}/\theta)^{na}$ follows a uniform distribution. One can use the same approach in (a) to find the confidence interval.

(c) Compare the intervals in (a) and (b) and comment on which one would you prefer if they are different.

Solution: The intervals are the same.

5. [**2014 final exam**] The exponential distribution is often used to model survival times. This problem develops a simple model for comparing survival times in two groups of patients. Let $X_1, \ldots, X_m$ be a random sample from an exponential distribution with pdf
$$f(x|\mu_1) = \frac{1}{\mu_1} e^{-x/\mu_1}, \quad x > 0, \quad \mu_1 > 0,$$
and let $Y_1, \ldots, Y_n$ be a random sample from an exponential distribution with pdf
$$f(y|\mu_2) = \frac{1}{\mu_2} e^{-y/\mu_2}, \quad y > 0, \quad \mu_2 > 0.$$
Assume that X and Y are independent. Define $\psi = \mu_2/\mu_1$, and let $\bar{X} = m^{-1} \sum_{i=1}^m X_i$ and $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$ be the sample means.

(a) Show that the **exact** likelihood ratio test statistic for the hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 \neq 0$ is
$$\lambda(\boldsymbol{x}, \boldsymbol{y}) = \frac{(m+n)^{m+n}}{m^m n^n} w^m (1-w)^n,$$
where $w = m/(m + nr)$ and $r = \bar{y}/\bar{x}$.

Solution: The joint pdf of $\boldsymbol{X} = (X_1, \ldots, X_m)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ are
$$L(\mu_1, \mu_2|\boldsymbol{x}, \boldsymbol{y}) = L(\mu_1|\boldsymbol{x})L(\mu_2|\boldsymbol{y}) = \mu_1^{-m} e^{-m\bar{x}/\mu_1} \mu_2^{-n} e^{-n\bar{y}/\mu_2}.$$

Under the null hypothesis $\mu_1 = \mu_2 = \mu_0$, the maximization is over the function

$$L(\mu_0|\boldsymbol{x}, \boldsymbol{y}) = \mu_0^{-m} e^{-m\bar{x}/\mu_0} \mu_0^{-n} e^{-n\bar{y}/\mu_0}.$$

One can obtain the MLE of μ_0 as

$$\hat{\mu}_0 = (m\bar{x} + n\bar{y})/(m + n).$$

Under unrestricted space, the MLE of μ_1 and μ_2 can be solved by maximizing individual likelihood functions due to independence between $\boldsymbol{X}$ and $\boldsymbol{Y}$. One obtains $\hat{\mu}_1 = \bar{x}$ and $\hat{\mu}_2 = \bar{y}$. We hence can have the likelihood ratio statistic as

$$\begin{aligned}
\lambda(\boldsymbol{x}, \boldsymbol{y}) &= \frac{\sup_{H_0} L(\mu_1, \mu_2|\boldsymbol{x}, \boldsymbol{y})}{\sup_{H_0 \bigcup H_1} L(\mu_1, \mu_2|\boldsymbol{x}, \boldsymbol{y})} \\
&= \frac{L(\hat{\mu}_0|\boldsymbol{x}, \boldsymbol{y})}{L(\hat{\mu}_1, \hat{\mu}_2|\boldsymbol{x}, \boldsymbol{y})} = \frac{(m+n)^{m+n}(m\bar{x})^m(n\bar{y})^n}{m^m n^n (m\bar{x} + n\bar{y})^{m+n}} \\
&= \frac{(m+n)^{m+n}}{m^m n^n} w^m (1-w)^n,
\end{aligned}$$

where $w = m\bar{x}/(m\bar{x} + n\bar{y}) = m/(m + nr)$ and $r = \bar{y}/\bar{x}$.

(b) Demonstrate that the rejection region $\{(\boldsymbol{x}, \boldsymbol{y}); \lambda(\boldsymbol{x}, \boldsymbol{y}) < c\}$ is equivalent to $\{r; r < c_1^*\} \cup \{r; r > c_2^*\}$. That means one may reject the null hypothesis by observing either $r < c_1^*$ or $r > c_2^*$. Given a type-I error rate α, find c_1^* and c_2^* using the fact that $\mu_1 \bar{Y}/\mu_2 \bar{X}$ follows $F_{2n,2m}$, which is F distribution with degree of freedoms $2n$ and $2m$.

Solution: Here $r > 0$, $w \in (0, 1)$, and $\lambda(\boldsymbol{x}, \boldsymbol{y})$ is unimodal and concave in w. That means

$$\lambda(\boldsymbol{x}, \boldsymbol{y}) < c \Leftrightarrow \{w < c_1\} \cup \{w > c_2\}.$$

Further, since w is monotone decreasing in r, we can have the critical region written as

$$\{r < c_1^*\} \cup \{r > c_2^*\}.$$

Since $\mu_1 \bar{Y}/\mu_2 \bar{X}$ follows $F_{2n,2m}$, one may choose $c_1^* = \psi F_{2n,2m,\alpha/2}$ and $c_2^* = \psi F_{2n,2m,1-\alpha/2}$, where $F_{2n,2m,\alpha}$ is the $(1-\alpha)$th quantile of $F_{2n,2m}$. In this case, under the null hypothesis, one have $\psi = 1$.

(c) Explain why $\psi \bar{X}/\bar{Y}$ is a pivotal quantity. Use that pivot to derive an exact 95% confidence interval for ψ.

Solution: Since $\psi\bar{X}/\bar{Y}$ follows $F_{2m,2n}$, an F-distribution with degree of freedoms $2m$ and $2n$ and free of the parameter of interest ψ, one hence can claim that $\psi\bar{X}/\bar{Y}$ is a pivotal quantity. Using the pivotal quantity and its distribution, one can have

$$1-\alpha = P(F_{2m,2n,\alpha_1} < \psi\bar{X}/\bar{Y} < F_{2m,2n,1-\alpha_2}),$$

where $F_{2m,2n,\alpha}$ is the αth quantile of the distribution $F_{2m,2n}$ and $\alpha_1+\alpha_2=\alpha$. Using the equation above, one can easily see

$$1-\alpha = P(F_{2m,2n,\alpha_1}\bar{Y}/\bar{X} < \psi < F_{2m,2n,1-\alpha_2}\bar{Y}/\bar{X}),$$

and the 95% confidence interval for ψ can be

$$\left(\frac{\bar{y}}{\bar{x}}F_{2m,2n,0.25}, \frac{\bar{y}}{\bar{x}}F_{2m,2n,0.975}\right).$$

(d) [**We will discuss this in the review session, no need to return for homework**] Express the critical region of the Wald test for the hypothesis $H_0: \mu_1-\mu_2=0$ against $H_1: \mu_1-\mu_2\neq 0$ given that the type-I error probability is α.

Solution: Under the null, the Wald test statistic by definition is

$$\frac{\bar{X}-\bar{Y}}{\sqrt{\mu_0^2(1/m+1/n)}},$$

since $E(\bar{X}-\bar{Y})=0$ and $Var(\bar{X}-\bar{Y})=\mu_0^2(1/m+1/n)$. However, we did not know what μ_0 is. We need a consistent estimator of μ_0 and plug it into the Wald statistic. We knew that the MLE is a consistent estimator so we can use $\hat{\mu}_0=(m\bar{X}+n\bar{Y})/(m+n)$ to replace μ_0 in the Wald statistic. Therefore, the Wald statistic of practical use is

$$T = \frac{\bar{X}-\bar{Y}}{\hat{\mu}_0\sqrt{(1/m+1/n)}},$$

which follows a standard normal distribution. Hence the critical region is $R=\{x,y;\{T\le z_{\alpha_1}\}\cup\{T\ge z_{1-\alpha_2}\}\}$, where $\alpha_1+\alpha_2=\alpha$.

6. [**Bios 673 class**] Let (X_i, Y_i), $i=1,\dots,n$, be *paired* random variables with respective distributions

$$f_{X_i}(x) = (\theta\phi_i)^{-1}e^{-x_i/(\theta\phi_i)}, \quad x_i>0,$$

and

$$f_{Y_i}(y) = \phi_i^{-1} e^{-y_i/\phi_i}, \quad y_i > 0,$$

where $\phi_i > 0$ is a parameter pertaining to characteristics of the ith pair, and $\theta > 0$ is the parameter reflecting any difference in average values between X and Y. Hence, it is of interest to test if $\theta = 1$ and to indicate if the distribution of X and Y are identical.

(a) Provide an explicit expression for the likelihood function of the random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, assuming that X_i and Y_i are mutually independent for $i = 1, \ldots, n$. Comment on how many parameters one needs to estimate by the method of maximum likelihood estimation.

Solution: Given the independence assumption of X and Y, the likelihood function can be written as

$$\begin{aligned} L(\theta, \phi_1, \ldots, \phi_n | \boldsymbol{x}, \boldsymbol{y}) &= \prod_{i=1}^{n} f_{X_i}(x_i) f_{Y_i}(y_i) \\ &= \prod_{i=1}^{n} (\theta\phi_i)^{-1} e^{-x_i/(\theta\phi_i)} \phi_i^{-1} e^{-y_i/\phi_i} \\ &= \theta^{-n} \prod_{i=1}^{n} \phi_i^{-2} e^{-\sum_{i=1}^{n} x_i/(\theta\phi_i) - \sum_{i=1}^{n} y_i/\phi_i}. \end{aligned}$$

There are $(n+1)$ parameters one needs to estimate.

(b) One points out that the only parameter of real interest is θ. She suggests that an alternative analysis, based just on the n ratios $R_i = X_i/Y_i$, $i = 1, \ldots, n$, should be favored in order to avoid the estimation of $\phi_1, \ldots, \phi_n$. Prove that this approach is feasible by showing

$$f_{R_i}(r_i) = \frac{\theta}{(\theta + r_i)^2}, \quad 0 < r_i < \infty, \quad i = 1, \ldots, n,$$

which is independent of $\phi_1, \ldots, \phi_n$.

Solution: Let $R_i = X_i/Y_i$ and $S_i = Y_i$. One can derive the joint pdf of (R_i, S_i) using the transformation method. First, the inverse function is $X_i = R_i S_i$ and $Y_i = S_i$, which makes the Jacobian as s_i. Then, the joint pdf of (R_i, S_i) can be written as

$$
\begin{aligned}
f_{R_i,S_i}(r_i,s_i) &= f_{X_i,Y_i}(r_is_i,s_i)|J| \\
&= (\theta\phi_i)^{-1}e^{-(r_is_i)/(\theta\phi_i)}\phi_i^{-1}e^{-s_i/\phi_i}s_i \\
&= \frac{s_i}{\theta\phi_i^2}\exp\left(-\frac{r_is_i}{\theta\phi_i}-\frac{s_i}{\phi_i}\right),
\end{aligned}
$$

for $0<r_i<\infty$ and $0<s_i<\infty$. To obtain the marginal pdf of R_i, one can have

$$
\begin{aligned}
f_{R_i}(r_i) &= \frac{1}{\theta\phi_i}\int_0^\infty s_i\exp\left(-\frac{r_is_i}{\theta\phi_i}-\frac{s_i}{\phi_i}\right)ds_i \\
&= \frac{1}{\theta\phi_i}\int_0^\infty s_i\exp\left(-s_i/\frac{\phi_i}{\left(\frac{r_i}{\theta+1}\right)}\right)ds_i \\
&= \frac{1}{\theta\phi_i}\left(\frac{\phi_i}{\frac{r_i}{\theta}+1}\right)^2 \\
&= \frac{\theta}{(\theta+r_i)^2},\quad 0<r_i<\infty.
\end{aligned}
$$

(c) Using mutually independent $R_i=X_i/Y_i$, $i=1,\ldots,n$, test $H_0:\theta=1$ versus $H_1:\theta>1$ by a likelihood-based large sample Wald-type test at the size $\alpha=0.05$.

Solution: With $\boldsymbol{r}=(r_1,\ldots,r_n)$, we have

$$
L(\theta|\boldsymbol{r})=\prod_{i=1}^n\frac{\theta}{(\theta+r_i)^2}=\theta^n\prod_{i=1}^n(\theta+r_i)^{-2}.
$$

The log-likelihood function becomes

$$
\ell(\theta)=n\log\theta-2\sum_{i=1}^n\log(\theta+r_i).
$$

Taking the first derivative, we have

$$
\frac{\partial}{\partial\theta}\ell(\theta)=n\theta^{-1}-2\sum_{i=1}^n(\theta+r_i)^{-1}.
$$

Taking the second derivative, we have

$$
\frac{\partial^2}{\partial\theta^2}\ell(\theta)=-n\theta^{-2}+2\sum_{i=1}^n(\theta+r_i)^{-2}.
$$

The expected information hence is

$$E\left(-\frac{\partial^2}{\partial\theta^2}\ell(\theta)\right) = n\theta^{-2} - 2\sum_{i=1}^{n} E(\theta+R_i)^{-2},$$

where

$$\begin{aligned}
E(\theta+R_i)^{-2} &= \int_0^\infty (\theta+r_i)^{-2}\frac{\theta}{(\theta+r_i)^2}dr_i \\
&= \theta\int_0^\infty (\theta+r_i)^{-4}dr_i \\
&= -\frac{\theta}{3}(\theta+r_i)^{-3}|_0^\infty \\
&= \frac{1}{3\theta^2}.
\end{aligned}$$

Hence, the expected information is $n/(3\theta^2)$. By the large sample property of MLE, one can conclude

$$\sqrt{n}(\hat{\theta}-\theta) \to_d N(0, 3\theta^2),$$

and use a critical region $\{\boldsymbol{r} : \sqrt{n}(\hat{\theta}-1)/\sqrt{3} > 1.96\}$ for testing the null hypothesis $H_0 : \theta = 1$ versus $H_1 : \theta > 1$.