STAT 513 hw 5

1. Let $X_{11}, \dots, X_{1n_1}$ and $X_{21}, \dots, X_{2n_2}$ be independent random samples from the $\text{Normal}(\mu_1, \sigma^2)$ and $\text{Normal}(\mu_1, \sigma^2)$ distributions, respectively, where μ_1, μ_2, and σ^2 are unknown. Suppose it is of interest to test the hypotheses H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$.

   (a) Give the likelihood function $L(\mu_1, \mu_2, \sigma^2; X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2})$ for $X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}$. *Hint: It is the product of the likelihood functions of the two samples.*

   (b) Give the log-likelihood function $\ell(\mu_1, \mu_2, \sigma^2; X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2})$.

   (c) Find the maximum likelihood estimators $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\sigma}^2$ of μ_1, μ_2, and σ^2, respectively; that is, find

   $$(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2) = \text{argmax}_{\mu_1, \mu_2, \sigma^2} L(\mu_1, \mu_2, \sigma^2; X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}).$$

   Hint: Use calculus methods on the log-likelihood function.

   (d) Under H_0, we have $\mu_1 = \mu_2 = \mu$, say, where μ denotes the common mean. Let $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ be

   $$(\hat{\mu}_0, \hat{\sigma}_0^2) = \text{argmax}_{\mu, \sigma^2} L(\mu, \mu, \sigma^2; X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}).$$

   Find expressions for $\hat{\mu}_0$ and $\hat{\sigma}_0^2$.

   (e) Show that the likelihood ratio

   $$\text{LR}(X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}) = \frac{L(\hat{\mu}_0, \hat{\mu}_0, \hat{\sigma}_0^2; X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2})}{L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2; X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2})}$$

   can be simplified to

   $$\text{LR}(X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}) = \left[\frac{\sum_{i=1}^{n_1}(X_{1i} - \hat{\mu}_1)^2 + \sum_{j=1}^{n_2}(X_{2j} - \hat{\mu}_2)^2}{\sum_{i=1}^{n_1}(X_{1i} - \hat{\mu}_0)^2 + \sum_{j=1}^{n_2}(X_{2j} - \hat{\mu}_0)^2} \right]^{(n_1+n_2)/2}.$$

   (f) Show that for any $c \in [0, 1]$, there exists a c_1 such that the likelihood ratio test

   $$\text{Reject } H_0 \text{ iff } \text{LR}(X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}) < c$$

   is equivalent to the test

   $$\frac{|\bar{X}_1 - \bar{X}_2|}{S_{\text{pooled}}\sqrt{1/n_1 + 1/n_2}} > c_1.$$

   Note: Please just attempt this question. It is quite tricky. You will get points for trying.

   (g) Provide the value c_1 such that the test in the previous part has size α for any $\alpha \in (0, 1)$.

2. Let $X_1, \dots, X_n$ be a random sample from the $\text{Gamma}(\alpha, \beta)$ distribution with density

   $$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta) \mathbb{1}(x > 0),$$

   where α and β are unknown.

(a) Give the likelihood function $L(\alpha, \beta; X_1, \dots, X_n)$ for the sample $X_1, \dots, X_n$.

(b) Give the log-likelihood function $\ell(\alpha, \beta; X_1, \dots, X_n)$ for the sample $X_1, \dots, X_n$.

(c) For any $\alpha \geq 0$, let $\hat{\beta}(\alpha)$ be the value of β which maximizes $L(\alpha, \beta; X_1, \dots, X_n)$. Get an expression for $\hat{\beta}(\alpha)$.

(d) Consider testing the hypotheses H_0: $\alpha = \alpha_0$ versus H_1: $\alpha \neq \alpha_0$ and let $\hat{\alpha}$ be the maximum likelihood estimator for α. Then the likelihood ratio is given by

$$\begin{aligned}\mathrm{LR}(X_1, \dots, X_n) &= \frac{\sup_{\{\alpha,\beta:\alpha=\alpha_0,\beta\geq 0\}} L(\alpha, \beta; X_1, \dots, X_n)}{\sup_{\{\alpha,\beta:\alpha\geq 0,\beta\geq 0\}} L(\alpha, \beta; X_1, \dots, X_n)} \\ &= \frac{L(\alpha_0, \hat{\beta}(\alpha_0); X_1, \dots, X_n)}{L(\hat{\alpha}, \hat{\beta}(\hat{\alpha}); X_1, \dots, X_n)}.\end{aligned}$$

Show that $-2\log\mathrm{LR}(X_1, \dots, X_n)$ can be simplified to

$$\begin{aligned}&-2\log\mathrm{LR}(X_1, \dots, X_n) \\ &\quad = -2\left[n\log\left(\frac{\Gamma(\hat{\alpha})}{\Gamma(\alpha_0)}\right) + n(\hat{\alpha} - \alpha_0)\left(\log\bar{X}_n - n^{-1}\sum_{i=1}^{n}\log X_i + 1\right) + n\alpha_0\log\alpha_0 - n\hat{\alpha}\log\hat{\alpha}\right].\end{aligned}$$

(e) The following R code stores in the vector `X` the survival times of several guinea pigs from the point in time at which they were infected with virulent tubercle bacilli and computes on these data the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ for the $\mathrm{Gammma}(\alpha, \beta)$ distribution. The data are taken from Bjerkedal (1960).

```
X <-   c(12, 15, 22, 24, 24, 32, 32, 33, 34, 38, 38, 43, 44, 48, 52,
        53, 54, 54, 55, 56, 57, 58, 58, 59, 60, 60, 60, 60, 61, 62,
        63, 65, 65, 67, 68, 70, 70, 72, 73, 75, 76, 76, 81, 83, 84,
        85, 87, 91, 95, 96, 98, 99, 109, 110, 121, 127, 129, 131,
        143, 146, 146, 175, 175, 211, 233, 258, 258, 263, 297, 341, 341, 376)

library(MASS) # pull in library of functions including the fitdistr() function
fitdistr(X,"gamma") # gives alpha.hat and 1/beta.hat
```

Compute the value of $-2\log\mathrm{LR}(X_1, \dots, X_n)$ for these data when testing the hypotheses H_0: $\alpha = 1$ versus $H_0 : \alpha \neq 1$.

(f) Report the p-value for testing the hypotheses in the previous question, using the asymptotic distribution of $-2\log\mathrm{LR}(X_1, \dots, X_n)$ under the null hypothesis.

(g) Consider testing H_0: $\alpha = \alpha_0$ versus H_1: $\alpha \neq \alpha_0$ using the guinea pig data. Find an interval such that you fail to reject H_0 at the 0.01 significance level for all α_0 in the interval. *Hint: Compute $-2\log\mathrm{LR}(X_1, \dots, X_n)$ over many values of α_0 and find those values of α_0 (search, say, between $1/2$ and 4) for which $-2\log\mathrm{LR}(X_1, \dots, X_n) < \chi^2_{1,0.01}$.*

(h) Give an interpretation of this interval.

(i) Based on these results, do you think it would be reasonable to model these data using the $\mathrm{Exponential}(\beta)$ distribution?

References

Bjerkedal, T. (1960). Acquisition of Resistance in Guinea Pigs infected with Different Doses of Virulent Tubercle Bacilli. *American Journal of Hygiene*, 72(1), 130-48.