1. According to Hardy-Weinberg law, if gene frequencies are in equilibrium, the chance of observing genotypes AA, Aa, aa in a population equals $(1-\theta)^2$, $2\theta(1-\theta)$, and θ^2, respectively. Let X_1, X_2, and X_3 denote the counts observed for blood type M, MN, and N, respectively, out of sample size n. Assuming Hardy–Weinberg law holds, one may assume X_1, X_2 and X_3 follow a multinomial distribution with probability density function
$$f(x_1, x_2, x_3|\theta) = \frac{n!}{x_1!x_2!x_3!}(1-\theta)^{2x_1}\{2\theta(1-\theta)\}^{x_2}\theta^{2x_3},$$
and $E(X_1) = n(1-\theta)^2$, $E(X_2) = 2n\theta(1-\theta)$, and $E(X_3) = n\theta^2$.

   (a) Find the maximum likelihood estimator of θ and comment on if it is an unbiased.

   Solution: The MLE is $\hat{\theta} = (X_2 + 2X_3)/2n$, and
$$\begin{aligned} E(\hat{\theta}) &= (2E(X_2) + E(X_3))/2n \\ &= (2n\theta(1-\theta) + 2n\theta^2)/2n \\ &= \theta, \end{aligned}$$
   which shows $\hat{\theta}$ is an unbiased estimator of θ.

   (b) Find the Cramér–Rao lower bound (CRLB) for any unbiased estimator of θ.

   Solution: The denominator of the CRLB is
$$\begin{aligned} -E\left(\frac{\partial^2}{\partial\theta^2}\log f(x_1, x_2, x_3|\theta)\right) &= 2E(X_1)\frac{1}{(1-\theta)^2} + E(X_2)\frac{1-2\theta+2\theta^2}{\theta^2(1-\theta)^2} + 2E(X_3)\frac{1}{\theta^2} \\ &= 2n + 2n\frac{1-2\theta+2\theta^2}{\theta(1-\theta)} + 2n \\ &= \frac{2n}{\theta(1-\theta)}. \end{aligned}$$
   Hence the CRLB equals $\theta(1-\theta)/(2n)$.

   (c) Naively, one may use an unbiased estimator X_3/n to estimate θ^2. Show that
$$\sqrt{n}(X_3/n - \theta^2) \to_d N(0, \theta^2(1-\theta^2)),$$
   and find σ^2 such that
$$\sqrt{n}(\sqrt{X_3/n} - \theta) \to_d N(0, \sigma^2).$$

Compare σ^2/n to the CRLB in (b) and comment on which one is smaller.

Solution: By Central Limit Theorem, one can have

$$\sqrt{n}(X_3/n - \theta^2) \to_d N(0, \theta^2(1-\theta^2)),$$

since X_3 can be written as a summation of n i.i.d. Bernoulli indicators with mean θ^2 and variance $\theta^2(1-\theta^2)$. Then, by the delta method, one can have

$$\sqrt{n}(g(X_3/n) - g(\theta^2)) \to_d N(0, \{g'(\theta^2)\}^2\theta^2(1-\theta^2)),$$

where $g(x) = \sqrt{x}$ with $g'(x) = (1/2)x^{-1/2}$. Accordingly, one has

$$\sqrt{n}(\sqrt{X_3/n} - \theta) \to_d N(0, (1-\theta^2)/4).$$

If we compare the limiting variance of $\sqrt{X_3/n}$, which is $(1-\theta^2)/4$, to n times of CRLB, which is $\theta(1-\theta)/2$, one can see that the limiting variance of $\sqrt{X_3/n}$ is always larger than n times of CRLB.

2. Let T_1 and T_2 be sufficient statistics for θ, and suppose that U be an unbiased estimator of θ. Let

$$V_1 = E(U|T_1)$$
$$V_2 = E(V_1|T_2).$$

(a) Show that both V_1 and V_2 are unbiased estimators of θ.

Solution: One can have $E(V_1) = EE(U|T_1) = E(U) = \theta$ and $E(V_2) = EE(V_1|T_2) = E(V_1) = \theta$.

(b) Show that $\text{Var}(V_2) \leq \text{Var}(V_1)$.

Solution: Since $V_2 = E(V_1|T_2)$ is a conditional expectation of an unbiased estimator V_1 given a sufficient statistic T_2, Rao-Blackwell Theorem tells us that $\text{Var}(V_2) \leq \text{Var}(V_1)$. Another way to solve this inequality is to express $\text{Var}(V_2)$ as

$$\text{Var}(V_2) = \text{Var}(E(V_1|T_2)) = \text{Var}(V_1) - E(\text{Var}(V_1|T_2)).$$

Since $E(\text{Var}(V_1|T_2)) \geq 0$, the inequality holds.

3. In a certain laboratory experiment, the time X (in milliseconds) for a certain clotting agent to show an observable effect is assumed to have an exponential distribution

$$f(x|\beta) = \frac{1}{\beta}\exp(-x/\beta), \quad x > 0, \quad \beta > 0.$$

It is of interest to make statistical inferences about the unknown parameter $\theta = \beta^2$, which is the variance of X.

(a) Develop an explicit expression for MLE $\hat{\theta}$ of θ.

Solution: The MLE of β is $\hat{\beta} = \bar{X}$. The MLE have an invariance property. For which the MLE of θ is $\hat{\theta} = \bar{X}^2$.

(b) Find the uniformly minimum variance unbiased estimator (UMVUE) $\hat{\theta}^*$ of θ.

Solution: Since $E(\bar{X}^2) = (n+1)\beta^2/n$, one can easily see that an unbiased estimator for θ is $\hat{\theta}^* = n\bar{X}^2/(n+1)$. Further, since $\bar{X}$ is a complete sufficient statistic (from exponential family), one can conclude that $\hat{\theta}^*$ is an UMVUE by Lehmann-Scheffe Theorem.

(c) Comment on whether the variance of $\hat{\theta}^*$ reaches CRLB.

Solution: Since $\theta = \beta^2$, one can let $\theta = \tau(\beta)$, where $\tau(x) = x^2$. Hence, the numerator of CRLB is $\{\tau'(\beta)\}^2 = 4\beta^2$. The denominator otherwise equals

$$\begin{aligned} -E\left(\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta)\right) &= E\left(-n\beta^{-2} + 2\sum_{i=1}^{n} X_i\beta^{-3}\right) \\ &= n\beta^{-2}. \end{aligned}$$

Hence the CRLB is β^4/n. The variance of $\hat{\theta}^*$ is

$$\begin{aligned} \mathrm{Var}\left(\frac{n}{n+1}\bar{X}^2\right) &= \frac{1}{n^2(n+1)^2}\mathrm{Var}(W^2) \\ &= \frac{1}{n^2(n+1)^2}[E(W^4) - \{E(W^2)\}^2] \\ &= \frac{1}{n^2(n+1)^2}\left[\frac{\Gamma(n+4)}{\Gamma(n)}\beta^4 - \frac{\Gamma(n+2)^2}{\Gamma(n)^2}\beta^4\right] \\ &= \frac{4n+6}{n(n+1)}\beta^4, \end{aligned}$$

where $W = \sum_{i=1}^n X_i$. One can see that, when $n \to \infty$, $\text{Var}(\hat{\theta}^*) \approx 4\beta^4/n$, which is larger than the CRLB.

(d) Derive the likelihood ratio test statistic $\lambda(\boldsymbol{x})$ of $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$.

(e) Show that the rejection region $R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\}$ is equivalent to $R^* = \{\boldsymbol{x} : \bar{x} \leq c_1^* \text{ or } \bar{x} \geq c_2^*\}$.

Solution: The likelihood ratio statistic is defined by

$$\begin{aligned}\lambda(\boldsymbol{x}) &= \frac{\sup_{\theta \in \Theta_0} L(\theta|\boldsymbol{x})}{\sup_{\theta \in \Theta} L(\theta|\boldsymbol{x})} = \frac{L(\theta_0|\boldsymbol{x})}{L(\bar{x}|\boldsymbol{x})} \\ &= \left(\frac{\beta_0}{\bar{x}}\right)^{-n} \exp\left(-\sum_{i=1}^n x_i/\beta_0 + n\right),\end{aligned}$$

which is a concave function of $\sum_{i=1}^n x_i$. Hence the rejection region with $\{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\}$ is equivalent to $\{\boldsymbol{x} : \sum_{i=1}^n x_i \leq c_1^* \text{ or } \sum_{i=1}^n x_i \geq c_2^*\}$.

Hint 1: If a random variable W follows Gamma(n, β), then, for $r > -n$,

$$E(W^r) = \frac{\Gamma(n+r)}{\Gamma(n)}\beta^r.$$

Hint 2: A function $g(y) = y^n \exp(-y)$ is a quadratic function of y.