Exercise 5

1. (i) Let $b_i = \alpha a_i + \beta$, $i = 1, \ldots, n$, and $\bar{a} = \frac{1}{n}\sum_{i=1}^n a_i$, $\bar{b} = \frac{1}{n}\sum_{i=1}^n b_i$. Show that $\bar{b} = \alpha\bar{a} + \beta$, and

$$\sum_{i=1}^n (b_i - \bar{b})^2 = \alpha^2 \sum_{i=1}^n (a_i - \bar{a})^2.$$

(Note that the sample variance of $\{b_i\}$ is independent of β.)

(ii) Let $\bar{a} = \frac{1}{n}\sum_{i=1}^n a_i$ and $\bar{b} = \frac{1}{n}\sum_{i=1}^n b_i$. Show that

$$\sum_{i=1}^n (a_i - c)(b_i - d) = \sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b}) + n(\bar{a} - c)(\bar{b} - d)$$

for any constants c and d, and

$$\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b}) = \sum_{i=1}^n a_i b_i - n\bar{a}\bar{b}$$

2. Let $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_m\}$ be two independent samples, where $X_1, \ldots, X_n$ are independent observations from $N(\mu_1, \sigma^2)$, and $Y_1, \ldots, Y_m$ are independent observations from $N(\mu_2, \sigma^2)$. Define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \quad s_x^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2,$$

$$\bar{Y} = \frac{1}{m}\sum_{i=1}^m Y_i, \quad s_y^2 = \frac{1}{m-1}\sum_{i=1}^m (Y_i - \bar{Y})^2.$$

Find the sampling distributions of (i) $\bar{X} - \bar{Y}$, (ii) $T \equiv (n-1)s_x^2 + (m-1)s_y^2$, and (iii) $\frac{\bar{X}-\bar{Y}}{\sqrt{T}}\sqrt{\frac{nm(n+m-2)}{m+n}}$ when $\mu_1 = \mu_2$.

3. A finite population has N_a individuals of type A and N_b individuals of type B. The overall number $N_a + N_b$ is equal to 100, but N_a is unknown. If a sample of 5 individuals is taken without replacement, write down the likelihood function for N_a when the observed sequence is (i) A, B, A, B, A, or (ii) A, A, B, B, A. Compare them with the likelihood function when the precise sequence of the outcomes is not available, but it is known that three elements are A and two others are B, and comment on your findings.

4. Given a random sample $Y_1, ..., Y_n$ from the following distributions, find a simple sufficient statistic using the factorisation criterion.

   (a) Poisson with mean λ

   (b) Normal with mean 0 and variance σ^2

   (c) Geometric

   (d) Uniform on (θ_1, θ_2) (Hint: $f_Y(y; \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1} I_{(\theta_1, \infty)}(y) I_{(-\infty, \theta_2)}(y)$.)

   (e) density $f_Y(y; \theta) = \exp(-(y - \theta))$ for $y > \theta$ (Use indicator functions, and take care with the factorisation)

Exercise 6

1. Let $Y_1, \cdots, Y_n$ be a random sample from a Poisson distribution with mean $\theta > 0$ unknown.
   (a) Show that the sum $Y \equiv Y_1 + \cdots + Y_n$ is a sufficient statistic. Find its mean and variance. What is the distribution of Y?
   (b) Obtain the MLE for θ.
   (c) Suppose now that only the first m ($m < n$) observations of the sample are known explicitly, while for the other $n - m$ only their sum, Z say, is known, determine the MLE of θ.
   (d) Compare the answers to (b) and (c), and comment.
2. Find the maximum likelihood estimator of λ given a random sample from the gamma distribution with density function
$$f(x) = \frac{1}{\Gamma(r)} \exp(-\lambda x) x^{r-1} \lambda^r.$$
   Assume that r is known.
3. Find the maximum likelihood estimator for θ from a random sample from the population with density function
$$f(y;\theta) = \frac{2y}{\theta^2} \quad 0 < y \leq \theta, \theta > 0.$$
   Do not use calculus. Draw a picture of the likelihood.
4. Suppose that we have a random sample of size n from the logistic distribution with density function
$$f(y;\mu) = \frac{\exp(y-\mu)}{[1+\exp(y-\mu)]^2}.$$
   Find the likelihood equation for μ, and write the iteration equations for the Fisher's score method. (Hint: write quantities as much as possible in terms of the density function and the distribution function for the logistic.)

Exercise 7

1. Suppose $X_1, \cdots, X_n$ are a random sample from Cauchy distribution with the density function
$$f(x, \theta) = \frac{1}{\pi}\{1 + (x - \theta)^2\}^{-1}.$$
Find the equation of which the MLE of θ is the solution. Write down the Newton-Raphson algorithm for calculating the MLE numerically. How would you choose an initial value for the algorithm?

2. Let $X_1, \cdots, X_n$ be a random sample from an exponential distribution with the density function $f(x, \lambda) = \lambda e^{-\lambda x} I(x \geq 0)$, where $\lambda > 0$ is an unknown parameter. Suppose the last $n - m$ $(m < n)$ observations are *censored* in the sense that we only known $X_j \in [u_j, v_j]$ for some given $v_j > u_j > 0$ while X_j itself is unknown, $j = m+1, \cdots, n$. Outline an EM algorithm for estimating the parameter $\lambda > 0$.

3. Let $X_1, \cdots, X_n$ be a random sample from Bernoulli distribution, i.e. $P(X_1 = 1) = p = 1 - P(X_1 = 0)$, where $p \in (0, 1)$ is unknown. Let $\theta = p^2$.
   (a) Find the Cramér-Rao lower bound for the variance of unbiased estimators for θ.
   (b) Find the MLE $\widehat{\theta}$ for the parameter θ.
   (c) Show that $E(\widehat{\theta}) \neq \theta$.
   (d) Outline a bootstrap procedure for estimating the bias of $\widehat{\theta}$.

4. Given a random sample of n observations from the N(0,θ) distribution, construct a most powerful test of the hypothesis $H_0 : \theta = 1$ against $H_1 : \theta = 2$. Find the power of the test of the size 0.05 when $n = 10$.

5. Let $(X_1, \cdots, X_n)$ be a random sample from $N(0, \sigma^2)$. Find the UMPT for testing hypotheses $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$, where $\sigma_0^2 > 0$ is a given constant.

Exercise 8

1. Suppose that $X_1, ..., X_n$ and $Y_1, ..., Y_n$ are two independent random samples from two exponential distributions with mean μ_1 and μ_2 respectively. Find the likelihood ratio test for $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. Specify the asymptotic distribution of the test statistic under H_0.

2. A survey of the use a particular product was conducted in four areas, and a random sample of 200 potential users was interviewed in each area. In area i, for $i = 1, 2, 3, 4$, x_i of the 200 said that they used the product. Construct a likelihood ratio test to test whether the proportion of the population using the product is the same in each of the four areas. Carry out the test at 5% level for the case $x_1 = 76, x_2 = 53, x_3 = 59$ and $x_4 = 48$.

3. A random sample $X_1, \cdots, X_n$ of size n is selected from a normal distribution with known mean μ and unknown variance σ^2. Two possible confidence intervals for σ^2 are shown below, where a_1, a_2, b_1 and b_2 are constants.

$$(a_1^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,\ a_2^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2), \qquad (b_1^{-1}\sum_{i=1}^{n}(X_i - \mu)^2,\ b_2^{-1}\sum_{i=1}^{n}(X_i - \mu)^2).$$

   For the case $n = 10$, find values of these constants which give intervals with confidence level 0.90. Compare the expected lengths of these intervals. Comment on your findings.

4. Let $X_1, \cdots, X_n$ be a random sample from the uniform distribution on the interval $[0, \theta]$ $(\theta > 0)$. Find a confidence interval for θ.