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

   (ii) $\sum_{i=1}^n(a_i - c)(b_i - d) = \sum_{i=1}^n(a_i - \bar{a} + \bar{a} - c)(b_i - \bar{b} + \bar{b} - d) = \sum_{i=1}^n(a_i - \bar{a})(b_i - \bar{b}) + \{\sum_{i=1}^n(a_i - \bar{a})\}(\bar{b} - d) + (\bar{a} - c)\sum_{i=1}^n(b_i - \bar{b}) + (\bar{a} - c)(\bar{b} - d)$. The first equation follows from the fact that $\sum_{i=1}^n(a_i - \bar{a}) = 0$ and $\sum_{i=1}^n(b_i - \bar{b}) = 0$. The second equation follows from the first equation immediately by letting $c = d = 0$.

2. (i) From the course work, we know that $\bar{X} \sim N(\mu_1, \frac{1}{n}\sigma^2)$ and $\bar{Y} \sim N(\mu_2, \frac{1}{m}\sigma^2)$. Since $\bar{X}$ and $\bar{Y}$ are independent, $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, (\frac{1}{n} + \frac{1}{m})\sigma^2)$.

   (ii) We know that $\sigma^{-2}(n-1)s_x^2 \sim \chi^2(n-1)$ and $\sigma^{-2}(m-1)s_y^2 \sim \chi^2(m-1)$. Note that s_x^2 and s_y^2 are also independent. From the reproductive property of χ^2 distribution, we have $\sigma^{-2}T \sim \chi^2(n+m-2)$.

   (iii) Note that

   $$Z \equiv \frac{\bar{X} - \bar{Y}}{\sqrt{T}}\sqrt{\frac{nm(n+m-2)}{m+n}} = \frac{(\bar{X} - \bar{Y})/\sqrt{\sigma^2(\frac{1}{n} + \frac{1}{m})}}{\sqrt{\sigma^{-2}T/(n+m-2)}}.$$

   On the right hand side of the above expression, the numerator is a standard normal random variable, and the denominator is a square root of χ^2-distributed random variable divided by its degrees of freedom, and the two parts are independent. Hence Z has t-distribution with $(n+m-2)$ degrees of freedom.

3. For outcome (i), the likelihood is

   $$L(N_a) = \frac{N_a}{100}\frac{N_b}{99}\frac{N_a - 1}{98}\frac{N_b - 1}{97}\frac{N_a - 2}{96} \propto N_a(N_a - 1)(N_a - 2)(100 - N_a)(99 - N_a).$$

   Note the probability of outcome (ii) is the same as the probability of outcome (i). Therefore $L(N_a)$ defined above is also the likelihood with outcome (ii).

   The probability for the event of 3 A's and 2 B's with unknown ordering is

   $$\frac{\binom{N_a}{3}\binom{N_b}{2}}{\binom{100}{5}} \propto N_a(N_a - 1)(N_a - 2)(100 - N_a)(99 - N_a),$$

   which is different, as a probability, from those for outcomes (i) and (ii). However the likelihood with this 'combined' outcome is still in principle the same as likelihood based outcome (i) or (ii). According to the likelihood principle, the inference for N_a based all three outcomes should be the same.

4. (a) The joint probability function of the observations is

   $$\prod_{i=1}^n\left[e^{-\lambda}\frac{\lambda^{y_i}}{y_i!}\right] = e^{-n\lambda}\lambda^{\sum y_i} \times \frac{1}{\prod y_i!}.$$

   So we have a factorisation of the joint probability function into

   $$g\left(\sum y_i; \lambda\right) = e^{-n\lambda}\lambda^{\sum y_i}$$

   and $h(\mathbf{y}) = \frac{1}{\prod y_i!}$. It follows that $\sum Y_i$ is a sufficient statistic for λ using the factorisation criterion.

(b) the joint density function is

$$\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}\sigma}e^{-y_i^2/(2\sigma^2)}\right] = \left[\frac{1}{\sqrt{2\pi}\sigma}\right]^n e^{-\sum y_i^2/(2\sigma^2)}.$$

Here we can take $g\left(\sum y_i^2;\sigma^2\right)$ as the joint density and $h(\mathbf{y})$ as identically 1. It follows by the factorisation criterion that $\sum Y_i^2$ is a sufficient statistics for σ^2.

(c) The probability function for the geometric distribution is $f(y_i;\pi) = (1-\pi)\pi^{y_i}, \quad y_i = 0,1,2,\ldots$. The joint probability function is

$$f(\mathbf{y};\pi) = \prod_{i=1}^{n} f(y_i;\pi) = (1-\pi)^n\pi^{\sum y_i}$$

So we can again take $g\left(\sum y_i\right)$ as the joint probability function and $h(\mathbf{y}) = 1$. By the factorisation criterion $\sum Y_i$ is a sufficient statistic for π.

(d) We can write the density function

$$f_{Y_i}(y_i) = \frac{1}{\theta_2-\theta_1}I_{(\theta_1,\infty)}(y_i)I_{(-\infty,\theta_2)}(y_i)$$

where the indicator functions are defined so that for a set A, $I_A(y) = \begin{cases} 0 & y \notin A, \\ 1 & y \in A. \end{cases}$ The joint density function is therefore

$$f_{\mathbf{Y}}(\mathbf{y}) = \left[\frac{1}{\theta_2-\theta_1}\right]^n\left[\prod_{i=1}^{n}I_{(\theta_1,\infty)}(y_i)\right]\left[\prod_{i=1}^{n}I_{(-\infty,\theta_2)}(y_i)\right]$$

which simplifies to

$$\left[\frac{1}{\theta_2-\theta_1}\right]^n I_{(\theta_1,\infty)}(\min_i y_i)I_{(-\infty,\theta_2)}(\max_i y_i)$$

Once again we can take $g(\min_i y_i, \max_i y_i;\theta_1,\theta_2)$ as the joint density function, and $h(\mathbf{y}) = 1$. By the factorisation criterion, $(\min_i Y_i, \max_i Y_i)$ is a sufficient statistic for (θ_1,θ_2).

(e) the joint density function is

$$f_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^{n}\left[e^{-(y_i-\theta)}I_{(\theta,\infty)}(y_i)\right] = e^{n\theta}I_{(\theta,\infty)}(\min_i y_i)e^{-n\bar{y}}.$$

We can take $g(\min_i y_i;\theta) = e^{n\theta}I_{(\theta,\infty)}(\min_i y_i)$ and $h(\mathbf{y}) = e^{-n(\bar{y}}$ to show that $\min_i Y_i$ is a sufficient statistic for θ.

