Problem 1 Let $T \sim \mathcal{N}(\mu, \sigma^2)$ with σ^2 known. Inference for the unknown μ proceeds in two stages. First, a two-sided test of the hypothesis that $\mu = 0$ is conducted. If this test is significant at level 2α, then a confidence interval $\mathcal{I}_{1-2\alpha}$ with limits $T \pm \sigma z_{1-\alpha}$ is computed. Show that if $\mu \geq 0$ then the coverage of this interval is

$$\frac{1 - \alpha - \Phi\{\max(z_\alpha, z_{1-\alpha} - \mu/\sigma)\}}{\Phi(z_\alpha + \mu/\sigma) + \Phi(z_\alpha - \mu/\sigma)},$$

and discuss how this depends on $\theta = \mu/\sigma$.

Problem 2 Randomisation for selective inference on θ based on $T \sim \mathcal{N}(\theta, 1)$ and an independent $W \sim \mathcal{N}(0, 1)$ results in $U = T + pW$ and $V = T - W/p$.

(a) Find the distribution of U, and that of T conditional on $U = u$.

(b) Show that the expected information for θ can be decomposed as

$$\imath(\theta) = \mathrm{E}_U\left\{-\frac{\mathrm{d}^2 \log f(U;\theta)}{\mathrm{d}\theta^2}\right\} + \mathrm{E}_U\left[\mathrm{E}_{T\mid U}\left\{-\frac{\mathrm{d}^2 \log f(T \mid U;\theta)}{\mathrm{d}\theta^2}\right\}\right],$$

and find these terms in the situation above. Hence deduce that randomisation can be viewed as partitioning the overall information between the two phases of inference.

Problem 3 To investigate how randomisation for dealing with selection might be applied to Poisson variables, suppose that Y has a Poisson distribution with mean θ, and let $X = (X_1, \ldots, X_K)$ be a multinomial variable with denominator y and probability vector $(p_1, \ldots, p_K)$.

(a) Given that the joint moment-generating function (MGF) of X given that $Y = y$ is

$$\mathrm{E}\left\{\exp\left(\sum_{k=1}^K t_k X_k\right) \,\middle|\, Y = y\right\} = \left(\sum_{k=1}^K p_k e^{t_k}\right)^y, \quad t_1, \ldots, t_K \in \mathbb{R},$$

show that after marginalisation over Y the components of X are independent Poisson variables with means $p_1\theta, \ldots, p_K\theta$. (Recall that a Poisson variable with mean λ has MGF $\exp\{\lambda(e^t - 1)\}$.)

(b) Discuss the use of randomisation for selective inference when the data consist of independent Poisson variables $Y_1, \ldots, Y_n$ with means $\theta_1, \ldots, \theta_n$.

Problem 4 The *Simes procedure* for simultaneously testing null hypotheses $H_1, \ldots, H_m$ with respective P-values $P_1, \ldots, P_m$ ordered as $P_{(1)} \leq \cdots \leq P_{(m)}$ rejects the global null hypothesis $H_0 = H_1 \cap \cdots \cap H_m$ if $P_{(k)} \leq k\alpha/m$ for at least one k.

(a) If $U_{(1)} \leq \cdots \leq U_{(n)}$ are the order statistics of $n > 1$ independent $U(0, 1)$ variables, show that the conditional distribution of $U_{(1)}/u_n \leq \cdots \leq U_{(n-1)}/u_n$ given that $U_{(n)} = u_n$ is that of the order statistics of $n - 1$ independent $U(0, 1)$ variables.

(b) Let $\alpha \in [0, 1]$ and define $A_m(\alpha) = \mathrm{P}(P_{(k)} > k\alpha/m, k = 1, \ldots, m)$. Show by induction that under H_0 and when the P_k are independent, $A_m(\alpha) = 1 - \alpha$ for any m, and deduce that the familywise error rate of the Simes procedure is exactly α. For the induction, condition on the value of $P_{(m+1)}$.

(c) Which of the Simes and Bonferroni procedures is preferable?