Midterm Solutions
MATH50011 Statistical Modelling 1

1. This question asked about *randomised response* procedures. These are often used in surveys where individuals are to be asked sensitive or stigmatizing questions.

(a) We have

$$\begin{aligned} P(X_i = 1) &= P(\text{spinner points to A})\, P(\text{individual } i \text{ has A}) \\ &\quad + P(\text{spinner points to not A})\, P(\text{individual } i \text{ has not A}) \\ &= \theta\pi + (1-\theta)(1-\pi) \\ &= \pi(2\theta - 1) + (1-\theta) \end{aligned}$$

(2 marks)

(b) The direct approach is to write a likelihood in terms of π. Since $X_i \sim \text{Bernoulli}(p)$ with $p = P(X_i = 1)$, we have

$$\begin{aligned} L(\pi) &= \prod_{i=1}^{n} [\pi(2\theta-1) + (1-\theta)]^{x_i} [1 - \pi(2\theta-1) - (1-\theta)]^{1-x_i} \\ &= [\pi(2\theta-1) + (1-\theta)]^{s} [1 - \pi(2\theta-1) - (1-\theta)]^{n-s} \end{aligned}$$

where $s = \sum_{i=1}^{n} x_i$.
The log-likelihood $\ell = \log L$ is then

$$\ell(\pi) = s \log[\pi(2\theta-1) + (1-\theta)] + (n-s)\log[1 - \pi(2\theta-1) - (1-\theta)].$$

(3 marks)

Differentiating this, we find that

$$\ell'(\pi) = \frac{s(2\theta-1)}{\pi(2\theta-1) + (1-\theta)} + \frac{(n-s)(2\theta-1)}{1 - \pi(2\theta-1) - (1-\theta)} \equiv 0$$

(2 marks)

which implies that the MLE is

$$\hat{\pi} = \frac{s - n(1-\theta)}{n(2\theta-1)} = \frac{\frac{s}{n} - (1-\theta)}{2\theta - 1}.$$

(2 marks)

Alternatively, one can note immediately that the MLE for $p = P(X_i = 1)$ is $\hat{p} = s/n$ using the likelihood for a random sample from Bernoulli(p). Then, by part (a) we have $\pi = [p - (1-\theta)]/(2\theta - 1)$ so that by invariance of the MLE we again have

$$\hat{\pi} = \frac{s - n(1-\theta)}{n(2\theta-1)} = \frac{\frac{s}{n} - (1-\theta)}{2\theta - 1}.$$

(c) First, we find that

$$E(\hat{\pi}) = \frac{\frac{E(S)}{n} - (1-\theta)}{2\theta-1} = \frac{\pi(2\theta-1) + (1-\theta) - (1-\theta)}{2\theta-1} = \frac{\pi(2\theta-1)}{2\theta-1} = \pi$$

and hence $bias(\hat{\pi}) = \pi - \pi = 0$. (2 marks)
For the variance, we find by direct calculation

$$Var(\hat{\pi}) = \frac{\frac{Var(X_i)}{n}}{(2\theta-1)^2} = \frac{p(1-p)}{n(2\theta-1)^2} = \frac{[\pi\theta + (1-\theta)(1-\pi)][\pi(1-\theta) + \theta(1-\pi)]}{n(2\theta-1)^2}$$

which would be the same expression as evaluating

$$\frac{1}{-E\left[\ell''(\pi)\right]} = \frac{1}{I_n(\pi)}$$

based on asymptotic normality of the MLE. (2 marks)

(d) No. By part (c), the variance of $\hat{\pi}$ equals the *Cramer-Rao lower bound* and $\hat{\pi}$ is itself unbiased. (2 marks)

2. Note that the variance of $\bar{X} - \bar{Y}$ depends on

$$Var(X_i - Y_i) = (2\sigma^2 - 2\tau) < 2\sigma^2. \quad \text{(1 mark)}$$

However, we have $S_X^2 + S_Y^2 \to_p 2\sigma^2$. (1 mark)

By Slutsky's theorem and the central limit theorem, we have under H_0 that

$$T = \frac{\sqrt{2\sigma^2}}{\sqrt{S_X^2 + S_Y^2}} \frac{\sqrt{2\sigma^2 - 2\tau}}{\sqrt{2\sigma^2}} \sqrt{n} \frac{\bar{X} - \bar{Y}}{\sqrt{2\sigma^2 - 2\tau}} \to_d N(0, \gamma^2)$$

where the asymptotic variance is

$$\gamma^2 = 1 - \frac{\tau}{\sigma^2} < 1. \quad \text{(2 marks)}$$

This implies that as $n \to \infty$, the type I error rate satisfies

$$P(|T| > z_{0.975}) \to 2(1 - \Phi(z_{0.975}/\gamma)) < 2(1 - \Phi(z_{0.975})) = 0.05$$

Thus, *the type I error rate is less than the nominal* $\alpha = 0.05$. (1 mark)

(Total 20 marks)