Statistics 245: Midterm

1. Consider i.i.d. observations $X_1, ..., X_n \sim N(\mu, 1)$.

   (a) Compute $\mathbb{E}(X_1|\bar{X})$.

   (b) Compute $\mathbb{E}\left(\frac{X_1+X_2}{2}\middle|\bar{X}\right)$.

   (c) Discuss your finding (hint: $\bar{X}$ is a sufficient statistic).

   Solution: It is easy to see that

   $$\mathbb{E}(X_1) = \mathbb{E}(\bar{X}) = \mu.$$

   We also have

   $$\mathsf{Var}(X_1) = 1, \quad \mathsf{Var}(\bar{X}) = n^{-1}, \quad \mathsf{Cov}(X_1, \bar{X}) = n^{-1}.$$

   Therefore,

   $$\begin{pmatrix} X_1 \\ \bar{X} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & n^{-1} \\ n^{-1} & n^{-1} \end{pmatrix}\right).$$

   This implies $(X_1|\bar{X}) \sim N(\bar{X}, 1-n^{-1})$. Therefore, $\mathbb{E}(X_1|\bar{X}) = \bar{X}$. The same argument also implies $\mathbb{E}(X_2|\bar{X}) = \bar{X}$. Therefore, $\mathbb{E}\left(\frac{X_1+X_2}{2}\middle|\bar{X}\right) = \frac{1}{2}\mathbb{E}(X_1|\bar{X}) + \frac{1}{2}\mathbb{E}(X_2|\bar{X}) = \bar{X}$. Since $\bar{X}$ is a sufficient statistic, by conditioning an estimator on $\bar{X}$, we obtain a new estimator that is a function of sufficient statistic. In other words, by conditioning on the sufficient statistic, we use more information.

2. Consider independent observations $X_1, ..., X_n \sim N(\mu, 1)$ and $X_{n+1}, ..., X_{2n} \sim N(\mu, 2)$.

   (a) Compute the mean squared error $\mathbb{E}(\bar{X} - \mu)^2$, where $\bar{X} = \frac{1}{2n}\sum_{i=1}^{2n} X_i$.

   (b) Find the MLE $\hat{\mu}$, and compute $\mathbb{E}(\hat{\mu} - \mu)^2$.

   (c) Construct a 95% confidence interval using the MLE.

   Solution: Note that

   $$\mathbb{E}(\bar{X} - \mu)^2 = \mathsf{Var}(\bar{X}) = \frac{1}{4n^2}\sum_{i=1}^{2n}\mathsf{Var}(X_i) = \frac{n+2n}{4n^2} = \frac{3}{4n}.$$

   The likelihood function is proportional to

   $$\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{1}{4}\sum_{i=n+1}^{2n}(X_i - \mu)^2\right).$$

   Thus, maximizing the likelihood function is equivalent to minimizing

   $$\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^2 + \frac{1}{4}\sum_{i=n+1}^{2n}(X_i - \mu)^2.$$

Set the derivative to zero, and we get the solution

$$\hat{\mu} = \frac{2\sum_{i=1}^{n} X_i + \sum_{i=n+1}^{2n} X_i}{3n}.$$

It is obvious that $\mathbb{E}(\hat{\mu}) = \mu$. This implies

$$\mathbb{E}(\hat{\mu} - \mu)^2 = \mathsf{Var}(\hat{\mu}) = \frac{4}{9n^2}\sum_{i=1}^{n} \mathsf{Var}(X_i) + \frac{1}{9n^2}\sum_{i=n+1}^{2n} \mathsf{Var}(X_i) = \frac{2}{3n}.$$

The fact that $\hat{\mu}$ is a linear combination of independent normal random variables implies $\hat{\mu} \sim N\left(\mu, \frac{2}{3n}\right)$, or equivalently

$$\frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{2/3}} \sim N(0, 1).$$

Thus, a 95% confidence interval is

$$\left[\hat{\mu} - \frac{\sqrt{2/3}}{\sqrt{n}} \times 1.96, \hat{\mu} + \frac{\sqrt{2/3}}{\sqrt{n}} \times 1.96\right].$$

3. Consider i.i.d. observations $X_1, ..., X_n \sim N(0, \sigma^2)$.

   (a) Find the MLE of σ^2, denoted by $\hat{\sigma}^2$.

   (b) Find the asymptotic distribution that $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ converges to.

   (c) Find a variance stabilization transformation so that $\sqrt{n}(g(\hat{\sigma}^2) - g(\sigma^2))$ converges to a distribution that does not depend on σ^2.

   (d) Find a testing procedure using $g(\hat{\sigma}^2)$ for $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$ with asymptotic Type-I error 0.05.

Solution: Set the derivative of the log-likelihood function to zero, and we get

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

By CLT,

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \rightsquigarrow N(0, 2\sigma^4).$$

The VST can be found by setting $|g'(\sigma^2)|^2\sigma^4$ as a constant. One solution is $g(t) = \log t$. This gives

$$\sqrt{n}(g(\hat{\sigma}^2) - g(\sigma^2)) \rightsquigarrow N(0, 2).$$

Under the null,

$$\sqrt{n}(g(\hat{\sigma}^2) - g(\sigma_0^2)) \rightsquigarrow N(0, 2).$$

Therefore, one can reject the null when

$$|\sqrt{n}(g(\hat{\sigma}^2) - g(\sigma_0^2))|/\sqrt{2} > 1.96.$$

4. Consider density functions f and g on $\mathbb{R}$ that satisfy

$$\int_{-\infty}^{\infty} xf(x)dx = 4, \quad \int_{-\infty}^{\infty} xg(x)dx = 6.$$

Find the value of

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x-t)g(t)dtdx.$$

Solution: Consider random variables $X \sim f$ and $Y \sim g$ and are independent of each other. Then, $X + Y \sim h$, where $h(x) = \int_{-\infty}^{\infty} f(x-t)g(t)dt$. This implies

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x-t)g(t)dtdx = \mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 10.$$