Math Statistics
Weekly HW 6

Ozaner Hansha

October 23, 2020

Question 1

Problem: Consider the distribution $\text{Ber}(\theta)$. For a sample of n i.i.d. variables $X_i \sim \text{Ber}(\theta)$, show that the following is a sufficient statistic of θ:

$$\hat{\theta} = \sum_{i=1}^{n} X_i$$

Solution: Note the following:

$$\begin{aligned}
p_X(\mathbf{x};\theta) &= \prod_{i=1}^{n} p_{X_i}(x_i;\theta) && \text{(independent sample)} \\
&= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} && (X_i \sim \text{Ber}(\theta)) \\
&= p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)} && \text{(sum of exponents)} \\
&= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} && \text{(linearity of summation)} \\
&= p^{\hat{\theta}(\mathbf{x})}(1-p)^{n-\hat{\theta}(\mathbf{x})} && \text{(def. of } \hat{\theta}\text{)} \\
&= \underbrace{1}_{h(\mathbf{x})} \cdot \underbrace{p^{\hat{\theta}(\mathbf{x})}(1-p)^{n-\hat{\theta}(\mathbf{x})}}_{g(\hat{\theta}(\mathbf{x});\theta)}
\end{aligned}$$

And so by the factorization theorem, we have that $\hat{\theta}$ is a sufficient statistic of θ.

Question 2

Consider two distributions with the same mean μ but with different variances σ_1^2, σ_2^2. Suppose we take independent samples from each distribution of sizes n_1, n_2, and that these samples have sample means $\bar{X}_1$, $\bar{X}_2$.

Part a: Show that for any $\omega \in \mathbb{R}$, the statistic $\hat{\theta}_\omega = \omega\bar{X}_1 + (1-\omega)\bar{X}_2$ is an unbiased estimator of μ.

Solution: Let us compute the expected value of $\hat{\theta} - \omega$:

$$\begin{aligned}
E[\hat{\theta}_\omega] &= E[\omega\bar{X}_1 + (1-\omega)\bar{X}_2] && \text{(def. of } \hat{\theta}_\omega\text{)} \\
&= \omega E[\bar{X}_1] + (1-\omega)E[\bar{X}_2] && \text{(linearity of expectation)} \\
&= \omega\mu + (1-\omega)\mu && \text{(mean of sample mean)} \\
&= \omega\mu + \mu - \omega\mu \\
&= \mu
\end{aligned}$$

And so we have that, for any real ω, the mean of our estimator $\hat{\theta}_\omega$ is the mean of the population. Thus our estimator is unbiased.

Part b: Give the variance of the estimator $\hat{\theta}_\omega$.

Solution: The variance of $\hat{\theta}_\omega$ is given by:

$$\begin{aligned}
\operatorname{Var}(\hat{\theta}_\omega) &= \operatorname{Var}(\omega \bar{X}_1 + (1-\omega)\bar{X}_2) && \text{(def. of } \hat{\theta}_\omega\text{)} \\
&= \operatorname{Var}(\omega \bar{X}_1) + \operatorname{Var}((1-\omega)\bar{X}_2) && \text{(variance of independent RVs)} \\
&= \omega^2 \operatorname{Var}(\bar{X}_1) + (1-\omega)^2 \operatorname{Var}(\bar{X}_2) && \text{(variance of multiple of RV)} \\
&= \omega^2 \frac{\sigma_1^2}{n_1} + (1-\omega)^2 \frac{\sigma_2^2}{n_2} && \text{(variance of sample mean)}
\end{aligned}$$

Part c: Show that $\operatorname{Var}(\hat{\theta}_\omega)$ is minimized when:

$$\omega = \frac{n_1 \sigma_2^2}{n_2 \sigma_1^2 + n_1 \sigma_2^2}$$

Solution: Our goal is to compute the following:

$$\begin{aligned}
\arg\min_{\omega \in \mathbb{R}} \operatorname{Var}(\hat{\theta}_\omega) &= \arg\min_{\omega \in \mathbb{R}} \omega^2 \frac{\sigma_1^2}{n_1} + (1-\omega)^2 \frac{\sigma_2^2}{n_2} && \text{(part b)} \\
&= \arg\min_{\omega \in \mathbb{R}} \underbrace{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}_{a} \omega^2 + \underbrace{\left(-\frac{2\sigma_2^2}{n_2}\right)}_{b} \omega + \underbrace{\frac{\sigma^2}{n_2}}_{c}
\end{aligned}$$

Note that the expression we are trying to minimize is a quadratic polynomial in ω. Recall that all quadratics in one variable have a single extremum, and that extremum is a minimum if $a > 0$ and a maximum if $a < 0$.

In our case, $a = \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$ is always positive, unless both σ_1^2 and σ_2^2 were zero (i.e. both RVs the distributions correspond to are just constants). And so the ω that minimizes estimator's variance is given by the zero of the above quadratic:

$$\begin{aligned}
\arg\min_{\omega \in \mathbb{R}} \operatorname{Var}(\hat{\theta}_\omega) &= \arg\min_{\omega \in \mathbb{R}} \underbrace{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}_{a} \omega^2 + \underbrace{\left(-\frac{2\sigma_2^2}{n_2}\right)}_{b} \omega + \underbrace{\frac{\sigma^2}{n_2}}_{c} \\
&= -\frac{b}{2a} && \text{(zero of a quadratic)} \\
&= -\frac{-\frac{2\sigma_2^2}{n_2}}{2\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)} \\
&= \frac{\sigma_2^2}{n_2\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)} \\
&= \frac{\sigma_2^2}{\frac{n_2 \sigma_1^2}{n_1} + \sigma_2^2} \\
&= \frac{n_1 \sigma_2^2}{n_2 \sigma_1^2 + n_1 \sigma_2^2}
\end{aligned}$$

And so we are done.