200B HW#6 solution

7.10 Supplementrary Exercises

22. The sample variance is $s^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$. The joint pmf is

$$f_n(\mathbf{x}|\theta) = \prod_{i=1}^n \frac{\theta^{X_i} e^{-\theta}}{X_i!} = \frac{\theta^{\sum_{i=1}^n X_i} e^{-n\theta}}{\prod_{i=1}^n X_i!}.$$

$T = \sum_{i=1}^n X_i$ is a sufficient statistic by factorization theorem with $u(\mathbf{x}) = \frac{1}{\prod_{i=1}^n X_i!}$ and $v(T,\theta) = \theta^T e^{-n\theta}$. Since s^2 is not a function of T, it is not admissible. (There exists another estimator $\delta_0(\mathbf{X}) = E(s^2|T)$, by Rao-Blackwell theorem, which is an improvement over s^2 in terms of the MSE.)

8.7 Unbiased Estimators

3. We know that the sample variance $\frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$ is an unbiased estimator of $\mathrm{var}(X)$ and $\frac{\sum_{i=1}^n X_i^2}{n}$ is an unbiased estimator of $E(X^2)$. So by $[E(x)]^2 = E(X^2) - \mathrm{var}(X)$, $\frac{\sum_{i=1}^n X_i^2}{n} - \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$ is an unbiased estimator of $[E(x)]^2$.

4. The mean of the geometric distribution is $E(X) = \frac{1-p}{p} = \frac{1}{p} - 1$. Therefore, $E(X+1) = \frac{1}{p}$ and $\delta(X) = X + 1$ is an unbiased estimator of $\frac{1}{p}$.

5. Following the hint, we have

$$\sum_{x=0}^{\infty} \frac{\delta(x)\lambda^x}{x!} = e^{2\lambda}.$$

Taylor expansion of the right hand side is $e^{2\lambda} = \sum_{x=0}^{\infty} \frac{(2\lambda)^x}{x!}$. Then comparing the expansion with left hand side we have $\delta(X) = 2^X$, which implies 2^X is an unbiased estimator of

e^λ.

6. $\frac{\sum_{i=1}^n (X_i - \bar{X})}{\sigma^2}$ follows a $\mathcal{X}^2$ distribution with degree of freedom $n-1$ and its mean and variance are $n-1$, $2(n-1)$ respectively. Then we immediately get $E(\hat{\sigma}_0^2) = \frac{n-1}{n}\sigma^2$, $\text{var}(\hat{\sigma}_0^2) = \frac{2(n-1)}{n^2}\sigma^4$, $E(\hat{\sigma}_1^2) = \sigma^2$ and $E(\hat{\sigma}_1^2) = \frac{2}{n-1}\sigma^4$. By theorem in the lecture notes page 71,

$$MSE(\hat{\sigma}_0^2) = [E(\hat{\sigma}_0^2) - \sigma^2]^2 + \text{var}(\hat{\sigma}_0^2) = \frac{2n-1}{n^2}\sigma^4,$$
$$MSE(\hat{\sigma}_1^2) = [E(\hat{\sigma}_1^2) - \sigma^2]^2 + \text{var}(\hat{\sigma}_1^2) = \frac{2}{n-1}\sigma^4.$$

Then by $(2n-1)(n-1) = 2n^2 - 3n + 1 < 2n^2$, we have $\frac{2n-1}{n^2} < \frac{2}{n-1}$ and therefore MSE of $\hat{\sigma}_0^2$ is smaller.

12. (a) Let X denote the value of the characteristic for a person chosen at random from the total population, and let A_i denote the event that the person belongs to stratum $i(i = (1, \ldots, k))$. Then

$$\mu = E(X) = \sum_{i=1}^k E(X|A_i)P(A_i) = \sum_{i=1}^k p_i\mu_i.$$

Also,

$$E(\hat{\mu}) = \sum_{i=1}^k p_i E(\bar{X}_i) = \sum_{i=1}^k p_i\mu_i = \mu.$$

(b) Since the samples are taken independently of each other, the variables $\bar{X}_1, \ldots, \bar{X}_k$ are independent. Therefore,

$$\text{var}(\hat{\mu}) = \sum_{i=1}^k p_i^2 \text{var}(\bar{X}_i) = \sum_{i=1}^k \frac{p_i^2\sigma_i^2}{n_i}.$$

We need to minimize $v = \sum_{i=1}^k \frac{p_i^2\sigma_i^2}{n_i}$ subject to the constraint that $\sum_{i=1}^k = n$. Write $n_k = n - \sum_{i=1}^{k-1} n_i$, then

$$v = \sum_{i=1}^{k-1} \frac{p_i^2\sigma_i^2}{n_i} + \frac{p_k^2\sigma_k^2}{n - \sum_{i=1}^{k-1} n_i}$$
$$\frac{\partial v}{\partial n_i} = -\frac{(p_i^2\sigma_i^2)}{n_i^2} + \frac{(p_k\sigma_k)^2}{n_k^2}, \text{ for } i = 1, \ldots, k-1.$$

Set all these partial derivatives to 0, we get $\frac{n_i}{p_i\sigma_i} = \frac{n_k}{p_k\sigma_k}$ for all $i = 1, \ldots, k-1$. Therefore, $n_i = cp_i\sigma_i$ for some constant c. Plug into $\sum_{i=1}^{k} n_i = n$ we have $c = \frac{n}{\sum_{i=1}^{k} p_i\sigma_i}$ and $n_i = \frac{np_i\sigma_i}{\sum_{i=1}^{k} p_i\sigma_i}$.

This analysis ignores the fact that the values of $n_1, \ldots, n_k$ must be integers. The integers $n_1, \ldots, n_k$ for which v is a minimum would presumably be near the minimizing values of $n_1, \ldots, n_k$ which have just been found.

14. For $0 < y < \theta$, the cumulative distribution function of $Y_n = X_{(n)}$ is

$$F(y|\theta) = P(Y \leq y|\theta) = P(X_1 \leq y, \ldots, X_n \leq y|\theta) = \left(\frac{y}{\theta}\right)^n.$$

Therefore the pdf of Y_n is $f(y|\theta) = \frac{dF(y|\theta)}{dy} = \frac{ny^{n-1}}{\theta^n}$ for $0 < y < \theta$ and 0 otherwise.
It then follows that

$$E_\theta(Y_n) = \int_0^\theta y\frac{ny^{n-1}}{\theta^n} = \frac{n}{n+1}\theta.$$

Hence $\frac{n+1}{n}Y_n$ is an unbiased estimator of θ.

Additional Problems

For these problems, you can use all results presented in theorems and examples in the lecture notes, chapters 1-5.

1. Derive the completeness of $\bar{X}$ in the Bernoulli model using the theorem on complete statistics in the exponential family.

   The pmf of Bernoulli distribution with parameter p is $p^x(1-p)^{1-x} = (1-p)(\frac{p}{1-p})^x = (1-p)e^{x\log\frac{p}{1-p}}$ and it belongs to the exponential family with $a(p) = 1-p$, 1, $k = 1$, $c_1(p) = \log\frac{p}{1-p}$ and $d_1(x) = x$. Then by theorem in page 35 of lecture notes $\sum_{i=1}^{n} X_i$ is complete and so is $\bar{X} = \sum_{i=1}^{n} X_i/n$.

2. Derive the UMVUE for θ for a random sample drawn from the model $f(x|\theta) = \theta e^{-\theta x}$, $\theta > 0$, $x \geq 0$.

It belongs to the exponential family with $a(\theta)=\theta$, $b(x)=1$, $k=1$, $c_1(\theta)=-\theta$ and $d_1(x)=x$. Then by theorem in page 35 of lecture notes $\sum_{i=1}^n X_i$ is a complete sufficient statistic. It means if we find $g(\sum_{i=1}^n X_i)$ a unbiased extimator of θ, it is the UMVUE of θ. By $X_i \sim Gamma(1,\theta)$ and property in page 45 of lecture notes, $\sum_{i=1}^n X_i \sim Gamma(n,\theta)$. Then we have

$$E(\frac{1}{\sum_{i=1}^n X_i}) = \int_0^\infty \frac{1}{x}\frac{\theta^n}{\Gamma(n)}x^{n-1}e^{-\theta x}dx = \frac{\theta\Gamma(n-1)}{\Gamma(n)}\int_0^\infty \frac{\theta^{n-1}}{\Gamma(n-1)}x^{n-2}e^{-\theta x}dx = \frac{\theta}{n-1}.$$

So $\frac{n-1}{\sum_{i=1}^n X_i}$ is the UMVUE of θ.

3. Find the UMVUE for σ^2 in the model $N(\mu,\sigma^2)$, where μ is known.

The pdf is $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ and it belongs to exponential family with $a(\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}$, $b(x)=1$, $k=1$, $c_1(\sigma^2)=-\frac{1}{2\sigma^2}$ and $d_1(x)=(x-\mu)^2$. By theorem in page 35 of lecture notes $\sum_{i=1}^n (X_i-\mu)^2$ is a complete sufficient statistic. We further have $E(\sum_{i=1}^n (X_i-\mu)^2)=n\sigma^2$. So $\frac{\sum_{i=1}^n (X_i-\mu)^2}{n}$ is the UMVUE of σ^2