Stat 512
Emerson, Fall 15
Homework #6
November 17 2015

Written solutions to the homework problems are due on Wednesday, November 25, 2015 at the beginning of class.

The homework problems are divided into "regular" and "more involved" problems. In order to facilitate multiple graders, you should hand in these categories of problems separately. That is, hand in one paper that contains only the "regular" problems, and another paper that contains only the "more involved" problems.
As noted on the syllabus and discussed during the first class, copying of homework solutions is not allowed and, when detected, will be investigated as an infraction of the academy integrity policy of the University of Washington. While it is permissible to discuss problems with other students, TAs, or the instructor in order to learn how to solve a problem, your written solutions must be prepared without directly referencing any notes or solutions derived from other students or sources found on the internet.

REGULAR PROBLEMS

1. Show that an asymptotically unbiased sequence of estimators need not be consistent. *(Hint: Consider independent random variables $X_1, X_2, \ldots,$ with $X_i \sim \mathcal{N}(\mu, 2i)$ and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be an estimator of μ.)*

2. Show that a consistent sequence of estimators for θ need not be asymptotically unbiased.
(Hint: Consider a sequence of estimators $\{T_n\}_{n=1}^{\infty}$ in which latent mixing random variable $Y_n \sim \mathcal{B}(1, p_n)$ has $(T_n|Y_n = 0) \sim \mathcal{N}(\theta, \frac{\sigma^2}{n})$ and $(T_n|Y_n = 1) \sim \mathcal{N}(n^2, 1)$.)

3. Show that for a sequence of distribution functions $T_n \sim F_{T_n}$ and a limiting distribution function $T \sim F_T$ such that
$$\forall x \in (-\infty, \infty),\ F_{T_n}(x) \to F_T(x),$$
we can have the "asymptotic variance" $Var(T) \neq \lim_{n\to\infty} Var(T_n)$.

4. Let $\vec{X}$ be a random vector and $T(\vec{X})$ be an estimator of θ with bias function $b(T, \theta)$ show that mean squared error satisfies
$$MST(T, \theta) = Var(T(\vec{X})|\theta) + b^2(T, \theta).$$

5. Let $X_1, X_2, \ldots$ be a random sample of independent, identically distributed variables with $X_i \sim (\mu, \sigma^2, \omega, \gamma)$, where μ is the first moment, and σ^2, ω, and $\gamma < \infty$ are 2nd, 3rd, and 4th central moments, respectively. Define nonparametric estimators of μ and σ^2

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$
$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

When using a statistic $T(\vec{X})$ as an estimator of some parameter θ, we define

$$\text{Bias function} \quad b(T(\vec{X}), \theta) = E[T(\vec{X})] - \theta$$
$$\text{Mean squared error function} \quad MSE(T(\vec{X}), \theta) = E[(T(\vec{X}) - \theta)^2]$$

(a) Find the bias function $b_n(\overline{X}_n, \mu)$ of $\overline{X}_n$ as an estimator of μ. Is $\overline{X}_n$ unbiased for μ? Is it asymptotically unbiased for μ?

(b) Prove or disprove: $\overline{X}_n$ is median unbiased for μ in the general case.

(c) Find the mean squared error function $MSE_n(\overline{X}_n, \mu)$ of $\overline{X}_n$ as an estimator of μ.

(d) Find the bias function $b_n(\hat{\sigma}_n^2, \sigma^2)$ of $\hat{\sigma}_n^2$ as an estimator of σ^2. Is $\hat{\sigma}_n^2$ unbiased for σ^2? Is it asymptotically unbiased for μ?

(e) Find $Var(\hat{\sigma}_n^2)$.
(Hint: You may find the following results about "quadratic forms" useful:

- *Theorem: For random vector $\vec{Y} = (Y_1, \ldots, Yn)$ with totally independent components and $Y_i \sim (\mu_i, \sigma^2, \omega, \gamma < \infty)$, define "quadratic form" $Q = \vec{Y}^T \mathbf{A} \vec{Y}$ for some n by n matrix $\mathbf{A}$ with $diag(\mathbf{A}) = \vec{a}$. Then*

$$Var(Q) = (\gamma - 3\sigma^4)\vec{a}^T\vec{a} + 2\sigma^4 tr(\mathbf{AA}) + 4\sigma^2\vec{\mu}^T\mathbf{AA}\vec{\mu} + 4\sigma^2\vec{\mu}^T A\vec{a},$$

  with $\vec{\mu} = (\mu_1, \ldots, \mu_n)^T$ and $tr(\mathbf{A})$ being the trace of matrix $\mathbf{A}$.

- *Consider a quadratic form using $\mathbf{A} = \mathbf{I}_n - \frac{1}{n}\vec{1}_n\vec{1}_n^T$ where $\mathbf{I}_n$ is the n by n identity matrix and $\vec{1}_n$ is an n dimensional vector of 1's.)*

(f) Find the mean squared error function $MSE_n(\hat{\sigma}_n^2, \sigma^2)$ of $\hat{\sigma}_n^2$ as an estimator of σ^2.

(g) Find the bias function $b_n(s_n^2, \sigma^2)$ of s_n^2 as an estimator of σ^2. Is s_n^2 unbiased for σ^2? Is it asymptotically unbiased for σ^2?

(h) Find $Var(s_n^2)$

(i) Find the mean squared error function $MSE_n(s_n^2, \sigma^2)$ of s_n^2 as an estimator of σ^2.

(j) Which of the above estimators for σ^2 has smaller MSE?

(k) Show that $s_n = \sqrt{s_n^2}$ is biased for σ.

6. Let $X_1, X_2, \ldots$ be a random sample of independent, identically distributed variables with $X_i \sim \mathcal{U}(0, \theta)$. Define estimator $\hat{\theta}_n = \max(X_1, X_2, \ldots, X_n)$.

   (a) Show that sequence of estimators $\{\hat{\theta}_n\}_{n=1}^{\infty}$ are asymptotically unbiased and consistent estimators of θ.

   (b) Let $\tilde{\theta}_n = a_n\hat{\theta}_n$ be an unbiased estimator of θ. Find a_n, and show that sequence of estimators $\{\tilde{\theta}_n\}_{n=1}^{\infty}$ are asymptotically unbiased and consistent estimators of θ.

MORE INVOLVED PROBLEMS

7. Let $X_1, X_2, \ldots$ be a random sample of independent, identically distributed variables with $X_i \sim (\mu, \sigma^2 < \infty)$. Chebyshev's inequality provides that for any random variable Y with finite variance and any $\delta > 0$

$$Pr\left(\,|Y - E[Y]\,| \geq \delta\right) \leq \frac{Var(Y)}{\delta^2}.$$

   Use Chebyshev's inequality to prove $\overline{X}_n$ is a consistent estimator of μ.

8. Let $X_1, X_2, \ldots$ be a random sample of independent, identically distributed variables with $\{T_n(\vec{X})\}_{n=1}^{\infty}$ a sequence of asymptotically unbiased estimators of θ with $Var(T_n) < \infty, \forall n$ and $\lim_{n\to\infty} Var(T_n) = 0$. Show that the sequence of estimators are consistent for θ.

9. Let $X_1, X_2, \ldots$ be a random sample of independent, identically distributed variables with $X_i \sim \mathcal{B}(1, \theta)$. We are interested in parametric estimators of $Var(X_i) = g(\theta) = \theta(1 - \theta)$.

   (a) Show that $T_n(\vec{X}_n) = \overline{X}_n(1 - \overline{X}_n)$ is asymptotically unbiased and consistent for $Var(X_i) = g(\theta) = \theta(1 - \theta)$.

   (b) Find an estimator $T_n^*(\vec{X}_n) = a_nT_n(\vec{X}_n)$ that is unbiased and consistent for $g(\theta)$. How does $MSE(T_n^*, g(\theta))$ compare to $MSE(T_n, g(\theta))$?

   (Hint: As derived in class notes, the kurtosis of the Bernoulli distribution is $p(1 - p)(1 - 3p + 3p^2) < \infty$.)

10. Let $X_1, X_2, \ldots$ be a random sample of independent, identically distributed variables with $X_i \sim \mathcal{P}(\theta)$. We are interested in parametric estimators of $Var(X_i) = g(\theta)$.

(a) Find a parametric estimator $T_n(\vec{X}_n)$ that is unbiased and consistent for $g(\theta)$.

(b) How does the MSE of $T_n(\vec{X}_n)$ compare to the MSE of s_n^2 as found in problem 5?

(Hint:The kurtosis of the Poisson distribution is $3\lambda^2 + \lambda < \infty$*.)*

11. Consider again Chebyshev's inequality.

(a) Show that Chebyshevs inequality can be extremely conservative for $\delta > 1$ by finding a nondegenerate distribution (i.e., with a positive variance) that has 100% of its data within 1 standard deviation of the mean.

(b) Show that Chebyshevs inequality is not always conservative, because for each $\delta \geq 1$ there is a distribution that meets the bound exactly. That is, for arbitrary $\delta \geq 1$, find a distribution function F_δ such that exactly proportion $\frac{1}{\delta^2}$ of the data lie at least δ standard deviations away from its mean.

(Hint: Consider a trinomial distribution that takes on values -1 with probability p, 1 with probability p, and 0 with probability 1-2p.)