10-751 Assignment 5

Due: Tuesday, November 12

Problem 1. Suppose that we independent samples $X_1, X_2, \ldots, X_n$, where $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

1. Show that $\overline{X}_n = \frac{1}{n}\sum_i X_i$ is an unbiased estimator of μ.
2. Show that the estimator
$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$
is a biased estimator of σ^2 and that the estimator
$$\hat{\sigma}_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$
is an unbiased estimator of σ^2.
3. Show that $\hat{\sigma}_n^2$ has lower mean squared error than $\hat{\sigma}_{n-1}^2$, thus illustrating that biased estimators may be more "efficient" than unbiased estimators.

Solution. For the first problem, let $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$. Then $E\overline{X_n} = \frac{1}{n}\sum EX_i = \mu$. Thus $\overline{X_n}$ is an unbiased estimator of μ.

For the second problem, by the independence of the X_i's, we have

$$\mathrm{var}(\overline{X_n}) = \frac{1}{n^2}\sum_i \mathrm{var}(X_i) = \frac{\sigma^2}{n}.$$

Also,

$$\begin{aligned} E(X_i - \mu)(\overline{X_n} - \mu) &= E(X_i - \mu)\left(\frac{1}{n}\sum_j (X_j - \mu)\right) \\ &= \frac{1}{n}E(X_i - \mu)^2 + \frac{1}{n}\sum_{j \neq i} E(X_i - \mu)(X_j - \mu) \\ &= \frac{\sigma^2}{n}, \end{aligned}$$

since by independence, $E(X_i - \mu)(X_j - \mu) = 0$.

Letting

$$W = \sum_i (X_i - \overline{X_n})^2 = \sum_i \left((X_i - \mu) - (\overline{X_n} - \mu)\right)^2,$$

we have

$$\begin{aligned}
EW &= \sum_i E(X_i-\mu)^2 - 2\sum_i E(X_i-\mu)(\overline{X_n}-\mu) + nE(\overline{X_n}-\mu)^2 \\
&= n\sigma^2 - 2n\frac{\sigma^2}{n} + n\frac{\sigma^2}{n} \\
&= (n-1)\sigma^2
\end{aligned}$$

Thus,

$$S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{W}{n}$$

has $ES_n^2 = \frac{n-1}{n}\sigma^2$, and it is therefore a biased estimator of σ^2. However,

$$S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{W}{n-1}$$

has expected value σ^2 and is therefore an unbiased estimator of σ^2.

For the third problem,

$$\begin{aligned}
E(S_{n-1}^2-\sigma^2)^2 - E(S_n^2-\sigma^2)^2 &= E(S_{n-1}^4 - 2S_{n-1}^2\sigma^2 + \sigma^4 - S_n^4 + 2S_n^2\sigma^2 - \sigma^4) \\
&= E(S_{n-1}^4 - 2S_{n-1}^2\sigma^2 - \frac{(n-1)^2}{n^2}S_{n-1}^4 + \frac{2(n-1)}{n}S_n^2\sigma^2 \\
&= E(\frac{2n-1}{n^2}S_{n-1}^4 - \frac{2}{n}S_{n-1}^2\sigma^2) \\
&= \frac{2n-1}{n^2}E(S_{n-1}^4) - \frac{2}{n}\sigma^2
\end{aligned}$$

Now, using the fact that

$$\mathrm{Var}(S_{n-1}^2) \geq \frac{2\sigma^4}{n}$$

we conclude that

$$\begin{aligned}
E(S_{n-1}^4) &= E^2(S_{n-1}^2) + \mathrm{Var}(S_{n-1}^2) \\
&= \sigma^4 + \mathrm{Var}(S_{n-1}^2) \\
&\geq \sigma^4 + \frac{2\sigma^4}{n}
\end{aligned}$$

Plugging into the above, we get

$$\begin{aligned}
E(S_{n-1}^2-\sigma^2)^2 - E(S_n^2-\sigma^2)^2 &\geq \frac{2n-1}{n^2}(\sigma^4 + \frac{2\sigma^4}{n}) - \frac{2}{n}\sigma^4 \\
&= \frac{3n-2}{n^3}\sigma^4 \\
&> 0
\end{aligned}$$

Note: this problem was a little more involved that anticipated. It's not necessary to get the entire derivation in part 3.

Problem 2. Suppose that $X_1, \ldots, X_n$ are independent with $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Let $\hat{\mu}$ and $\hat{\sigma}$ denote the maximum likelihood estimates of μ and σ. With $n = 17$, find a value of k such that $\Pr(\hat{\mu} > \mu + k\hat{\sigma}) = 0.95$. (You will need to use a table of probabilities for the t distribution, or calculate these numerically.)

Solution. Since $\hat{\mu} = \overline{X}_n$ and $\hat{\sigma}^2 = S_n^2/n$, it follows that

$$
\begin{aligned}
\Pr(\hat{\mu} > \mu + k\hat{\sigma}) &= \Pr\left(\frac{\overline{X}_n - \mu}{\hat{\sigma}} > k\right) \\
&= \Pr(T > k\sqrt{n-1})
\end{aligned}
$$

Since T has a t distribution with $n-1$ degrees of freedom and $n = 17$, we must choose k so that $\Pr(T > 4k) = 0.95$. From the table for the t distribution we find $\Pr(T < 1.746) = 0.95$. Thus, $\Pr(T > -1.746) = 0.95$ and it follows that $4k = -1.746$, or $k = -0.4365$.

Problem 3. Let $X_1, \ldots, X_n$ be independent with $X_i \sim \text{Uniform}(0, \theta)$, and let $\hat{\theta}_n = \max_i X_i$.

1. Compute the bias of $\hat{\theta}_n$.
2. Compute the standard error of $\hat{\theta}_n$.
3. Computer the mean squared error of $\hat{\theta}_n$.

Solution. For $0 < y < \theta$, the cdf of $\max_{1 \le i \le n} X_i$ is

$$
\begin{aligned}
F(x) &= \Pr(X_1 \le x, \ldots, X_n \le x) \\
&= \left(\frac{x}{\theta}\right)^n
\end{aligned}
$$

Thus the density is given by

$$
f(x) = \frac{nx^{n-1}}{\theta}
$$

and the mean is

$$
\begin{aligned}
E_\theta(\max_{1 \le i \le n} X_i) &= \int_0^\theta x \frac{nx^{n-1}}{\theta}\, dx \\
&= \frac{n}{n+1}\theta
\end{aligned}
$$

Therefore the bias is

$$
\begin{aligned}
\text{bias}(\max_{1 \le i \le n} X_i) &= \frac{n}{n+1}\theta - \theta \\
&= -\frac{1}{n+1}\theta
\end{aligned}
$$

Using the density of $\max X_i$ we can also easily compute

$$\begin{aligned}
\mathrm{se}(\hat{\theta}_n) &= \sqrt{E\left(\hat{\theta}_n - \frac{n}{n+1}\theta\right)^2} \\
&= \sqrt{\int_0^\theta \left(x - \frac{n}{n+1}\theta\right)^2 \frac{nx^{n-1}}{\theta}\,dx} \\
&= \sqrt{\theta^{n+1}\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)}
\end{aligned}$$

and the the mean squared error is given by

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{bias}(\hat{\theta})^2 + \mathrm{var}(\hat{\theta})$$

Problem 4. Suppose that $X_1, \ldots, X_{n+1}$ form a sample from a normal distribution, and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $T_n = \left(\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2\right)^{1/2}$. Determine the value of a constant c such that $c(X_{n+1} - \overline{X}_n)/T_n$ will have a t distribution. (Remember that if Z is standard normal and Y is a χ^2 distribution with n degrees of freedom, $Z/\sqrt{\frac{1}{n}Y}$ has a t distribution with n degrees of freedom.)

Solution. The difference $X_{n+1} - \overline{X}_n$ has a normal distribution with mean 0 and variance $(1+\frac{1}{n})\sigma^2$. Thus, the distribution of

$$\left(\frac{n}{n+1}\right)^{1/2} \frac{X_{n+1} - \overline{X}_n}{\sigma}$$

is standard normal. Also, nT_n/σ^2 has an independent χ^2 distribution with $n-1$ degrees of freedom. Thus, the following ratio will have a t distribution with $n-1$ degrees of freedom:

$$\frac{\left(\frac{n}{n+1}\right)^{1/2} \frac{X_{n+1} - \overline{X}_n}{\sigma}}{\left(\frac{nT_n^2}{(n-1)\sigma^2}\right)^{1/2}} = \left(\frac{n-1}{n+1}\right)^{1/2} \frac{X_{n+1} - \overline{X}_n}{T_n}$$

Therefore, we can should $c = \left(\frac{n-1}{n+1}\right)^{1/2}$.

Problem 5. Using Matlab, generate $n = 20$ samples from a Poisson(λ_0) distribution, where $\lambda_0 = 1$. Perform the Wald test with $\alpha = 0.05$. Repeat many times and count how many times the null hypothesis $\lambda = \lambda_0$ is rejected. How close is the type I error rate to 0.05? Give a description of your results and include your Matlab code with your assignment.