Lsn26

Clark

Admin

Let $X \sim Bin(n, p)$. Let $\hat{\sigma}^2 = n\hat{p}(1 - \hat{p})$ where $\hat{p} = \frac{X}{n}$ show that $\hat{\sigma}^2$ is a biased estimator of $\sigma^2 = np(1 - p)$
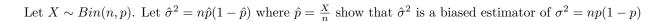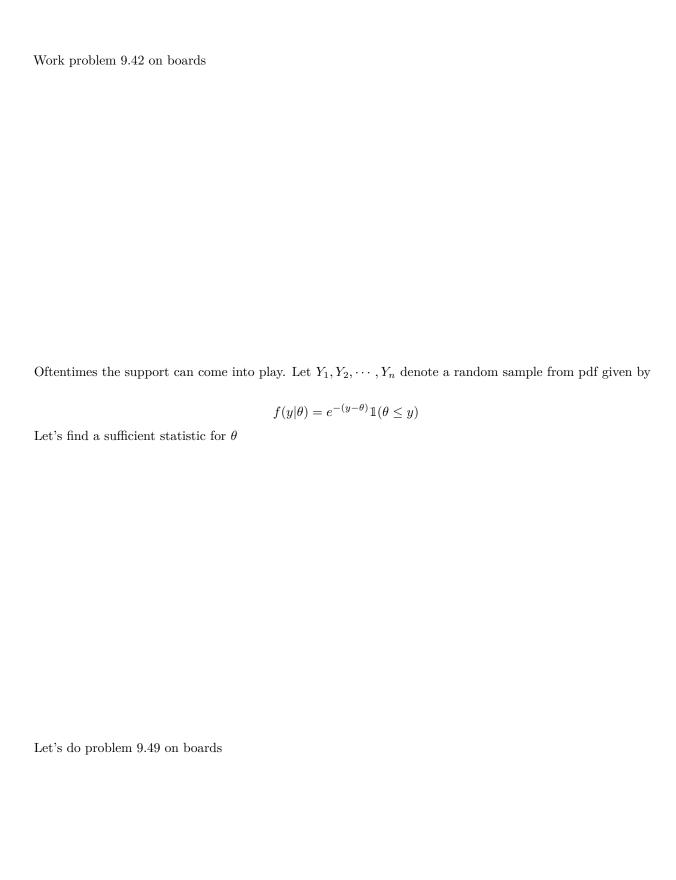
Propose a function of $\hat{\sigma}^2$ that is unbiased

One of the points of statistics is to efficiently summarize our data. For instance, if $X_i \sim N(\mu, \sigma)$ for $i = 1, \cdots, n$ and we want to say something about the distribution our data come from, we don't really need to know all of our data points, we only really need to know $\bar{X}$ and S^2. From there we can infer the most likely values of μ and σ to some level of certaintly. The notion of *sufficiency* means that by using, say, $\bar{X}$ and S^2, we aren't actually losing any information. If we aren't losing any information, then we say that our estimators are *sufficient statistics*. Formally, a statistic, $\hat{\theta}$ is sufficient for θ if the conditional distribution of our data, $X_1, \cdots, X_n$ given our sufficient statistic does not depend on θ.

The classic example, let $X_i \sim Bern(p)$ and let's let $\hat{p} = \sum_{i=1}^{n} X_i$, we can appeal directly to the definition of sufficiency to prove $\hat{p}$ is sufficient for p.

Using the definition is all well and good if we have a candidate for a sufficient statistic. However, it's often much easier to use what is called the *factorization theorem*. Though in order to talk about the factorization theorem we need to talk about a *likelihood function* first.

The likelihood of θ given our data is written as:

Note that if our data is discrete, then our likelihood is a probability. However if our data are continuous we have no guarantees that our likelihood is bounded by 1, so it cannot be a probability. With the idea of a likelihood we can use theorem 9.4, the *factorization theorem*, to find a sufficient statistic.

As an example, consider $Y_1, \cdots, Y_n \sim Po(\lambda)$ and we'll use the factorization theorem to find a sufficient statistic.

Work problem 9.42 on boards

Oftentimes the support can come into play. Let $Y_1, Y_2, \cdots, Y_n$ denote a random sample from pdf given by

$$f(y|\theta) = e^{-(y-\theta)}\mathbb{1}(\theta \leq y)$$

Let's find a sufficient statistic for θ

Let's do problem 9.49 on boards

Next lesson, COL Watts will talk about MVUE, which is the unbiased estimator that has the lowest variance. Recall that when we compare efficiency of two unbiased estimators we are looking at the ratio of the variances. A MVUE is the unbiased estimator that is guaranteed to have the lowest variance of ALL the unbiased estimators. Though the Rao-Blackwell theorem is a bit clunky, the biggest take away is that if we use the factorization theorem to find a sufficient statistic, we actually have what is called a *minimally sufficient statistic* and if we have an unbiased estimator that is a function of the minimally sufficient statistic, then we are **guaranteed** to have the MVUE.

whew. . . .