# Lecture 37 - 11/28

Monday, November 28, 2022 9:26 AM

**Today:**

① Sample mean & variance of data set

② Minimum mean square error constant estimator

③ Minimum mean square error unconstrained estimator

---

① Sample mean & variance of a data set

Suppose that we have a device that fills bags of rice. The device is not accurate, and amount of rice in each bag is random. More specifically, suppose that the amount of rice in ith bag is given by $X_i$. Suppose that $X_1, \dots, X_n$ are independent & identically distributed. We measure the amount of rice in the first n bags, and our goal is to estimate $E[X_i]$ and $Var(X_i)$ (so we can label the bags before selling them, i.e., using "$E[X_i] \pm \sqrt{Var(X_i)}$" as the label.

---

**Example 4.8.7** Suppose $X_1, \dots, X_n$ are independent and identically distributed random variables, with mean $\mu$ and variance $\sigma^2$. It might be that the mean and variance are unknown, and that the distribution is not even known to be a particular type, so maximum likelihood estimation is not appropriate. In this case it is reasonable to estimate $\mu$ and $\sigma^2$ by the *sample mean* and *sample variance* defined as follows:

$$\widehat{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \widehat{\sigma^2} = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \widehat{X})^2.$$

Note the perhaps unexpected appearance of $n-1$ in the sample variance. Of course, we should have $n \geq 2$ to estimate the variance (assuming we don't know the mean) so it is not surprising that the formula is not defined if $n = 1$. Recall that an estimator is called *unbiased* if the mean of the estimator is equal to the parameter that is being estimated. (a) Is the sample mean an unbiased estimator of $\mu$? (b) Find the mean square error, $E[(\mu - \widehat{X})^2]$, for estimation of the mean by the sample mean. (c) Is the sample variance an unbiased estimator of $\sigma^2$?

**Solution:**

(a) given data points $X_1, \dots, X_k$ our estimation is $\hat{X} = \frac{1}{n}\sum_{k=1}^{n} X_i$. We have

$$E[\hat{X}] = E\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] = \frac{1}{n}\sum_{k=1}^{n} E[X_k] = \mu$$

Hence, $\hat{X}$ is an unbiased estimation of $\mu$.

(b) Notice that

$$E\left[(\mu - \hat{X})^2\right] = E\left[\left(\mu - \frac{1}{n}\sum_{k=1}^{n} X_k\right)^2\right]$$

$$= \frac{1}{n^2} E\left[\left(\sum_{k=1}^{n}(\mu - X_k)\right)^2\right]$$

$$\overset{(i)}{=} \frac{1}{n^2}\sum_{k=1}^{n} E\left[(\mu - X_k)^2\right] = \frac{\sigma^2}{n}$$

$$\overset{(i)}{=} \frac{1}{n^2} \sum_{k=1}^{n} E\left[(\mu - X_k)^2\right] = \frac{\sigma^2}{n}$$

and (i) follows by independece of $X_k$s, i.e, for two independent random variables X and Y, $Var(X+Y) = Var(X) + Var(Y)$. Hence, the mean square error goes to zero with rate $\frac{1}{n}$.

(c) Notice that using a single data point, we cannot say anything about variance. Hence $\frac{1}{n-1}$ makes sense. The main reason, however, is that in the definition of $\hat{\sigma}^2$, we plugged in $\hat{X}$ instead of $\mu$, i.e., we are using sample mean instead of the actual mean in our estimation.

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n-1} \sum_{k=1}^{n} (X_k - \hat{X})^2\right] = E\left[\frac{1}{n-1} \sum_{k=1}^{n} \left(X_k - \frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{k=1}^{n} \left(\frac{n-1}{n} X_k - \frac{1}{n}\sum_{\substack{i=1 \\ i\neq k}}^{n} X_i\right)^2\right]$$

$$\overset{(i)}{=} \frac{1}{n-1} \sum_{k=1}^{n} Var\left(\frac{n-1}{n} X_k - \frac{1}{n}\sum_{\substack{i=1 \\ i\neq k}}^{n} X_i\right)$$

$$\overset{(ii)}{=} \frac{1}{n-1} \sum_{k=1}^{n} \left(\left(\frac{n-1}{n}\right)^2 \sigma^2 + \frac{(n-1)}{n^2}\sigma^2\right) = \sigma^2$$

where (i) follows by the fact that

$$E\left(\frac{n-1}{n} X_i - \frac{1}{n}\sum_{\substack{k=1 \\ k\neq i}}^{n} X_k\right) = \frac{n-1}{n}\mu - \frac{1}{n}\cdot(n-1)\mu = 0,$$

where (ii) follows by independece of $X_i$s:

$$Var\left(\frac{n-1}{n} X_k + \sum_{\substack{i=1 \\ i\neq k}}^{n} \frac{X_i}{n}\right) = Var\left(\frac{n-1}{n} X_k\right) + \sum_{\substack{i=1 \\ i\neq k}}^{n} Var\left(\frac{X_i}{n}\right) = \frac{(n-1)^2}{n^2}\sigma^2 + (n-1)\cdot\frac{1}{n^2}\sigma^2 = \frac{n-1}{n}\sigma^2$$

Hence, $\hat{\sigma}^2$ is an unbiased estimation of $\sigma^2$.

---

@ Minimum mean square error constant estimator

Suppose that Y is a random variable with know distribution. Suppose that we do not observe the realization of Y & we wish to estimate it.

Example. an automated device fills bags of rice which will be sold on the market. The device is not perfect & the amount of rice in each bag is random. We have tested the device & we know the distribution of output of our device, however, we do not observe the exact value in each bag. It seems natural to report the expected value as the however, we do not observe the exact value in each bag. It seems natural to report the expected value as the constant estimator for amount of rice in each bag, but why?

When estimating a quantity, we need a notion of error that quantifies the accuracy of our estimation. One common notion is mean square error, which has nice mathematical properties: given our estimated value is $\delta$, the MSE is defined by

$$\text{MSE ( for estimation of } Y \text{ by a constant } \delta) = E\left[(Y-\delta)^2\right]$$

$$= \int_{-\infty}^{+\infty} (y-\delta)^2 f_Y(y)\,dy \rightsquigarrow \text{ if } Y \text{ is continuous\_type}$$

Notice that $E\left[(Y-\delta)^2\right] = E\left[Y^2\right] + \delta^2 - 2\delta E[Y]$. Using first order condition, we get $\delta = E[Y]$.

Another way to see this, is via the following calculation:

$$E\left[(Y-\delta)^2\right] = E\left[\left(Y - E[Y] + E[Y] - \delta\right)^2\right]$$

$$= Var(Y) + \left(E[Y]-\delta\right)^2 + 2\,E\left[(E[Y]-Y)(E[Y]-\delta)\right]$$

$$= Var(Y) + \left(E[Y]-\delta\right)^2$$

which is minimize if we set $\boxed{\delta = E[Y]}$

---

③ Minimum mean square error linear estimator

Suppose that we wish to estimate $Y$ base on observation $X$. Using $g(X)$ as our estimator, we want to find $g(X)$ that mimizes the mean square error, i.e., minimizes $E\left[(Y-g(X))^2\right]$.

Example: we are buiding a computer, and we want to report temperature of components to the user. However, we only have one temperature sensor that we used to measure temperature of CPU. We know the joint pdf of temperature of CPU and GPU. The question is, given temperature of CPU, what is the best estimate of the temperature of GPU that should be reported to the user.

Suppose that $g^*$ is the minimizer. $g^*(X)$ is called the unconstrained optimal estimator of $Y$ given $X$:

① unconstrained since there is no constraint on $g$

② optimal since by definition $E\left(Y-g^*(X)\right)^2 \leqslant E\left(Y-g(X)\right)^2$ for any function $g$.

Based on previous section, we may guess $g^*$ have something to do with the mean of $Y$. In particular, given $X=u$, the pdf of $Y$ is $f(v|u) = f_{XY}(u,v)$ ... Hence by analysis of previous section, given $X=u$ we have $\delta = \int^{+\infty} v f(v|u)\,dv = E[Y|X=u]$

Based on previous section, we may guess $g^*$ have something to do with the mean of $Y$. In particular, given $X=u$, the pdf of $Y$ is $f_{Y|X}(v|u) = \frac{f_{X,Y}(u,v)}{f_X(u)}$. Hence, by analysis of previous section, given $X=u$ we have $S = \int_{-\infty}^{+\infty} v f_{Y|X}(v|u)\,dv = E[Y|X=u]$.

Hence, a natural guess for $g^*$ is $g^*(X) = E[Y|X]$. We will show this is indeed the case:

$$E\left[(Y-g(X))^2\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (v-g(u))^2 f_{X,Y}(u,v)\,dv\,du$$

$$= \int_{-\infty}^{+\infty} \underbrace{\int_{-\infty}^{+\infty} (v-g(u))^2 f_{Y|X}(v|u)\,dv}\, f_X(u)\,du \quad \text{assuming } f_X(u) \text{ has full support}$$

and the inner part is minimized by setting $g(u) = E[Y|X=u]$. Notice that

$$E\left[(Y-g^*(X))^2\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (v-g^*(u))^2 f_{Y|X}(v|u)\,dv\, f_X(u)\,du$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(v^2 + (g^*(u))^2 - 2v g^*(u)\right) f_{Y|X}(v|u)\,dv\, f_X(u)\,du$$

$$= \int_{-\infty}^{+\infty} v^2 f_Y(v)\,dv + \int_{-\infty}^{+\infty} (g^*(u))^2 f_X(u)\,du - 2\int_{-\infty}^{+\infty} g^*(u) \int_{-\infty}^{+\infty} v f_{Y|X}(v|u)\,dv\, f_X(u)\,du$$

$$= \int_{-\infty}^{+\infty} v^2 f_Y(v)\,dv - \int_{-\infty}^{+\infty} (g^*(u))^2 f_X(u)\,du = E[Y^2] - E\left[(E[Y|X])^2\right]$$

where we have used the fact that $g^*(u) = \int_{-\infty}^{+\infty} v f_{Y|X}(v|u)\,dv$.