Lecture 38 - 11/30

Saturday, December 03, 2022 3:44 PM

Review:

① Sample mean & variance of data set

Given n independent & identically distributed random variables $X_1, X_2, X_3, \dots, X_n$, the sample mean & variance are unbiased estimator and are given as follows:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad , \quad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

② Minimum mean square error constant estimator

Knowing the distribution of a random variable Y, the best constant estimator of Y that minimizes the mean square error is $\delta^* = E[Y]$. Mean square error of this estimator is $E[(Y-\delta^*)^2] = \sigma_Y^2$

③ Minimum mean square error unconstrained estimator

Knowing the joint pdf of X and Y, and observing the value of X, the best unconstrained estimator of Y given X that minimizes the mean square error is $g^*(X) = E[Y|X]$. In particular, for any u in the support of f_X, we have

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

The mean square error of $g^*(X)$ is $E[(Y-g^*(Y))^2] = E[Y^2] - E[(E[Y|X])^2]$

Today: ① Minimum mean squre linear estimator

② Law of large numbers

③ Central limit theorem

① Minimum mean squre linear estimator

Suppose that the joint distribution of X and Y is known. We observe the value of X and we want to estimate Y. We know the best estimator that minimizes mean square error is $g^*(X) = E[Y|X]$. However, in practice calculating $E[Y|X]$ might be infeasible or hard to implement. A popular choice for estimator is linear estimator $g(X) = aX+b$. Our goal is to find a and b that minimizes $E[(Y-(aX+b))^2]$. We will do this in two steps:

Step 1. Fixing $a>0$, find the best choice for b.

Step 1. Fixing $a>0$, find the best choice for b.

Let $Z = Y - aX$. We have $E[(Y-aX-b)^2] = E[(Z-b)^2]$. Notice that the best constant estimator of Z is $\delta^* = E[Z]$. Hence, the best choice of $b \in \mathbb{R}$, is $b^* = E[Z] = \mu_Y - a\mu_X$ where $\mu_Y = E[Y]$ and $\mu_X = E[X]$.

Step 2. Find $a \in \mathbb{R}$ that minimizes $E[(Y-\mu_X - a(X-\mu_X))^2]$.

Notice that $E[(Y-\mu_X - a(X-\mu_X))^2] = Var(Y-aX) = var(Y) + a^2 Var(X) - 2a\,Cov(X,Y)$.

By the first order condition, $a \in \mathbb{R}$ that minimizes above should satisfy,

$$2a\,Var(X) - 2\,cov(X,Y) = 0 \quad \Rightarrow a^* = \frac{Cov(X,Y)}{Var(X)} = \frac{\sigma_Y}{\sigma_X}\rho_{X,Y}$$

where $\rho_{X,Y}$ is correlation coefficient.

$\Rightarrow$ The best linear estimator of Y give X is $L^*(X) = a^*X + b^* = \mu_Y + \sigma_Y \rho_{X,Y}\left(\frac{X-\mu_X}{\sigma_X}\right)$

Then mean square error of $L^*(X)$ is

$$E[(Y-L^*(X))^2] = Var(Y-a^*X) = \sigma_Y^2 + (a^*)^2\sigma_X^2 - 2a^* Cov(X,Y)$$
$$= \sigma_Y^2 + \sigma_Y^2\rho_{X,Y}^2 - 2\sigma_Y^2\rho_{X,Y}^2 = \sigma_Y^2(1-\rho_{X,Y}^2)$$

Important Remarks:

① We also use $\hat{E}[Y|X]$ to denote best linear estimator of Y given X:

$$\hat{E}[Y|X] = L^*(X) = \mu_Y + \sigma_Y \rho_{X,Y}\left(\frac{X-\mu_X}{\sigma_X}\right)$$

$\hat{E}[Y|X]$ is called the wide sense conditional expectation of Y given X.

① The correlation coefficient is an integral part of the best linear estimator.

Notice that if instead of estimating Y using a linear function of X, we estimate standardized version of Y $\left(\tilde{Y} = \frac{Y-\mu_Y}{\sigma_Y}\right)$ using a linear function of standardized version of X $\left(\tilde{X} = \frac{X-\mu_X}{\sigma_X}\right)$, the best linear estimator is

$$\hat{E}[\tilde{Y}|\tilde{X}] = \rho_{\tilde{X},\tilde{Y}}\tilde{X} = \rho_{X,Y}\tilde{X} = \rho_{X,Y}\left(\frac{X-\mu_X}{\sigma_X}\right)$$

② The class of constant estimators is a subset of class of linear estimators, which itself is a subset of class of unconstrained estimators. Hence, we have the following relation between the best estimators we have studied so far.

MSE of $\delta^* = E(Y) \geq$ MSE of $L^*(X) = \mu_Y + \sigma_Y \rho_{X,Y}\left(\frac{X-\mu_X}{\sigma_X}\right)$

$\geq$ MSE of $g^*(X) = E[Y|X]$

Equivalently, we have

$$\sigma_Y^2 \geq \sigma_Y^2(1-\rho_{X,Y}^2) \geq \sigma_Y^2 - E[(E[Y|X])^2]$$

③ If X & Y are independent we have

$$E[Y] = \delta^* = L^*(X) = g^*(X)$$

as we expected (knowing X do not reduces our uncertainty about Y.).

---

②Law of large numbers

Notice that given n samples $X_1,\dots,X_2$ that are independent & identically distributed, the sample average $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is unbiased estimator of μ. The question is: how accurate is our estimator. Intuitively, more samples should result in a more accurate estimator.

From basic calculus, when we say a sequence of numbers $y_1, y_2, \dots$ converges to y, we mean for any $\delta>0$, there is $N>0$ s.t. $|y-y_n|<\delta$ for any $n>N$. However, in the case of sample average $\hat{\mu}_n$, we are dealing with random number.

Consider the set of bad estimates: $\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|>\delta\right\} = \left\{\omega\in\Omega : \left|\frac{X_1(\omega)+X_2(\omega)+\cdots+X_n(\omega)}{n}-\mu\right|>\delta\right\}$

One way to characterize the accuracy of our estimate is to say the probability of observing bad estimates is small. This is exactly what the following Law of large numbers say.

Proposition 4.10.1. (Law of Large Numbers) Suppose that $X_1, X_2, \dots, X_n$ are uncorrelated random variables. they have the same mean $\mu = E[X_1] = \cdots = E[X_n]$, and their variance is less than C, i.e., $Var(X_i)<C$ for all $i\in\{1,2,\dots,n\}$. For any $\delta>0$, we have: $P\left(\left|\frac{S_n}{n}-\mu\right|>\delta\right) \leq \frac{C}{n\delta^2} \xrightarrow[n\to\infty]{as} 0$

where $S_n = X_1+\cdots+X_n$.

Proof: Notice that by chebyshevs inequality:

$$P\left(\left|\frac{S_n}{n}-\mu\right|>\delta\right) \leq \frac{Var\left(\frac{S_n}{n}\right)}{\delta^2} = \frac{Var(S_n)}{n^2\delta^2}$$

Since X_is are uncorrelated, we have

$$Var(S_n) = Var(X_1+\cdots+X_n) = Var(X_1)+\cdots+Var(X_n) \leq nC.$$

Hence,

$$P\left(\left|\frac{S_n}{n}-\mu\right|>\delta\right) \leq \frac{nC}{n^2\delta^2} = \frac{C}{n\delta^2}$$

---

③Central limit theorem

Suppose that we have n random variables $X_1, X_2, \dots, X_n$ which are independent & identically distributed. Law of large numbers

③ Central limit theorem

Suppose that we have n random variables $X_1, X_2, \dots, X_n$ which are independent & identically distributed. Law of large numbers characterizes the accuracy of sample mean estimator. What if we want more, i.e., we want to estimate the distribution of S_n? Notice that $E[S_n] = n\mu$ and $Var(S_n) = n\sigma^2$ where $\mu = E[X_1]$ and $\sigma^2 = Var(X_1)$. Hence, S_n itself is not well-behaved, i.e., its mean & variance blows up. Hence, it makes more sense to study the standardized version of S_n instead, i.e., $\tilde{S}_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$. The central limit theorem characterizes the asymptotic distribution of $\tilde{S}_n$.

Theorem. (Central Limit Theorem) Suppose that $X_1, X_2, \dots, X_n$ are independent & identically distributed random variables. Let $\mu = E[X_1]$, $\sigma^2 = Var(X_1)$. Set $S_n = X_1 + \dots + X_n$. We have

$$\lim_{n\to\infty} P(\tilde{S}_n \leq c) = \lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq c\right) = \Phi(c)$$

Remarks.

① We can use Guassion approximation for sum of independent & identically distributed random variables.

② CLT is called a universal law since it has nothing to do with distribution of X_1 other than existence of μ & σ^2.