Lecture 36 - 11/18

Wednesday, November 16, 2022 1:07 PM

Review:

- Let X and Y be random variables with finite second moments:

  the correlation: $E[XY]$

  the covariance: $Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$

  the correlation coefficient: $\rho_{XY} = \dfrac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \dfrac{Cov(X,Y)}{\sigma_X \sigma_Y}$

- Scaling and linearity of Covariance:

  (i) $Cov(X+Y, U+V) = Cov(X,U) + Cov(X,V) + Cov(Y,U) + Cov(Y,V)$

  (ii) $Cov(aX+c, cY+d) = ac\,Cov(X,Y)$

- Important def. & relations.

  ① uncorrelated, negatively correlated, positively correlated

  ② $Var(X) = Cov(X,X)$. In particular if X and Y are uncorrelated

  $$Var(X+Y) = Cov(X+Y, X+Y) = Cov(X,X) + Cov(Y,Y) = Var(X) + Var(Y)$$

  ③ If X & Y are independent, they are uncorrelated. The revers does not hold in general.

  (Note: If X & Y are jointly guassian & uncorrelated, they are independent.)

Today:

① Covariance, correlation coefficient & standardized version of random variables

② Sample mean & variance of a data set

③ jointly guassian random variables

---

① Covariance, correlation coefficient & standardized version of random variables

Notice that $E[XY]$ and $Cov(X,Y)$ have units, i.e., their units is product of the units of X and Y.

e.g.: X in km, Y in km then $Cov(X,Y)$ and $E[XY]$ are in $(km)^2$

In particular, changing unit (e.g. from km to m) changes the value of $E[XY]$ & $Cov(X,Y)$. One solution is to use the standardized version of X and Y:

$$Cov\left(\frac{X-E[X]}{\sigma_X}, \frac{Y-E[Y]}{\sigma_Y}\right) = Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \rho_{XY}$$

$$E\left[\left(\frac{X-E[X]}{\sigma_X}\right)\left(\frac{Y-E[Y]}{\sigma_Y}\right)\right] = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \rho_{XY}$$

In particular, $\rho_{X,Y}$ is a number & does not have unit. Notice that $\rho_{aX+b,\,cY+d} = \rho_{X,Y}$ for any $a,c>0$.

The question is, what values can $\rho_{X,Y}$ take.

Proposition: Schwartz's inequality

For two random var. X & Y, we have:

$$|E[XY]| \leqslant \sqrt{E[X^2]\,E[Y^2]}$$

Moreover, if $E[X^2] \neq 0$, equality holds if & only if $P\{Y = cX\} = 1$ for some c.

proof: Let $\lambda = E[XY]/E[X^2]$. We have

$$0 \leqslant E[(Y-\lambda X)^2] = E[Y^2] + \lambda^2 E[X^2] - 2\lambda E[XY] = E[Y^2] - \frac{E[XY]^2}{E[X^2]}$$

Notice that equality means $E[(Y-\lambda X)^2] = 0$ which is same as $P(Y - \lambda X = 0)$ since $(Y-\lambda X)^2 \geq 0$.

Cor. 4.8.4. For two random variables X and Y.

$$|Cov(X,Y)| \leqslant \sqrt{Var(X)\,Var(Y)}$$

Furthermore, if $Var(X) \neq 0$ then equality holds if & only if $Y = aX + b$ for some a & b.

Consequently, if Var(X) & Var(Y) are not zero, $\rho_{X,Y}$ is well-defined &

- $|\rho_{X,Y}| \leq 1$
- $\rho_{X,Y} = 1$ if & only if $Y = aX + b$ for some $a>0$ & b.
- $\rho_{X,Y} = -1$ if & only if $Y = aX + b$ for some $a<0$ & b.

proof: Using Cauchy-Schwartz for $W = X - E[X]$ & $Z = Y - E[Y]$, we have

$$|Cov(X,Y)| = |E[WZ]| \leqslant \sqrt{E[W^2]E[Z^2]} = \sqrt{Var(X)\,Var(Y)}$$

Hence, $\rho_{X,Y} = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \in [-1,1]$ since $|\rho_{X,Y}| \leqslant 1$.

- If $\rho_{X,Y} = 1$ then $P(Z = aY) = 1$, i.e., $P\big(Y - E[Y] = a(X - E[X])\big) = 1$. Which can be written as $P(Y = aX + b) = 1$. Since $\rho_{X,Y} = 1$, we have $Cov(W,Z) > 0$ which implies $a>0$.
- Similarly, if $\rho_{X,Y} = -1$, we have $P(Y = aX + b) = 1$ for some $a<0$.

Intuitively speaking:

- $\rho_{X,Y} > 0$ means X & Y both tends to be larger than average or both to be smaller than average

- $\rho_{X,Y} < 0$ means X & Y tends to be opposite, i.e., X larger than $E[X]$ tends to indicate Y is smaller than $E[Y]$

@ Sample mean & variance of a data set

Example 4.8.7 Suppose $X_1, \ldots, X_n$ are independent and identically distributed random variables, with mean μ and variance σ^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 μ and σ^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 μ? (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 σ^2?

Solution:

(a) given data points $X_1, \ldots, X_k$ 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 μ.

(b) Notice that

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

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

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

where (a) follows by independece of X_ks, 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 μ, i.e., we are using sample mean instead of the actual mean in our estimation.

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

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

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

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

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

where follows by indepence of X_is:

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

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

---

③ jointly guassion random variables

Def. We say random variables X and Y are jointly guassion random variables if any linear combination of X and Y are guassion random variables, i.e., for any $a,b \in \mathbb{R}$, $aX+bY$ is a guassion random variables.

joint distribution of guassion random variables. Suppose that X, Y are non-degenerate, i.e., they are not linearly related. Then their joint distribution is given by:

$$f_{X,Y}(u,v) = \frac{1}{2\pi\,\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(\frac{\left(\frac{u-\mu_X}{\sigma_X}\right)^2 + \left(\frac{v-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{u-\mu_X}{\sigma_X}\right)\left(\frac{v-\mu_Y}{\sigma_Y}\right)}{2(1-\rho^2)}\right)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_X}\exp\left(\frac{(X-\mu_X)^2}{2\sigma_X^2(1-\rho^2)}\right)\cdot\frac{1}{\sqrt{2\pi}\,\sigma_Y}\exp\left(\frac{(Y-\mu_Y)^2}{2\sigma_Y^2(1-\rho^2)}\right)\cdot\frac{1}{\sqrt{1-\rho^2}}\exp\left(\frac{-\rho\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)}{(1-\rho)^2}\right)$$

where

$$\mu_X = E[X], \quad \sigma_X^2 = Var(X)$$

$$\mu_Y = E[Y], \quad \sigma_Y^2 = Var(Y)$$

$$\rho = \rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X\,\sigma_Y}$$

Important remarks:

① setting $a=0$ and $b=1$, implies that $aX+bY=Y$ is $\mathcal{N}(\mu_Y,\sigma_Y^2)$. Similarly, X is $\mathcal{N}(\mu_X,\sigma_X^2)$.

② If $\rho=0$ then $f_{X,Y}(u,v) = f_X(u)f_Y(v)$. Hence uncorrelated jointly guassion random variables are independent.

③ Let $Z = aX+bY$. Then X and Z are jointly guassion.

④ If X and Y are guassion, it does not mean that they are jointly guassion.

4 If X and Y are guassian, it does not mean that they are jointly guassian.