Lecture 39 - 12/02

Friday, December 02, 2022 9:30 AM

Review.

① minimum mean square error linear estimator

We have a pair of random variable X and Y. Upon observing X, we want to estimate Y using a linear function $g(X) = aX + b$. The best linear estimator, that minimizes the mean squre error is

$$L^*(X) = a^* X + b^*, \quad a^* = \frac{Cov(Y,X)}{Var(X)}, \quad b^* = \mu_Y - a^* \mu_X$$

We also use use $\hat{E}[Y|X]$ to denote $L^*(X)$:

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

$\hat{E}[Y|X]$ is called wide sense conditional expectation. MSE for $L^*(X)$ equals $\sigma_Y^2 (1 - \rho_{X,Y}^2)$.

② Law of large numbers

Suppose that $X_1, X_2, \dots, X_n$ are uncorrelated, each one has the same mean μ, and their variance is bounded by C. Let $S_n = X_1 + X_2 + \dots + X_n$. Then, for any $\delta > 0$, we have

$$P\left(\left\{ \left| \frac{S_n}{n} - \mu \right| \geq \delta \right\}\right) = P\left(\left\{ w \in \Omega : \left| \frac{S_n(w)}{n} - \mu \right| \geq \delta \right\}\right) \leq \frac{C}{n\delta^2} \xrightarrow{\text{as } n \to \infty} 0$$

③ Central limit theorem (CLT)

Suppose that $X_1, X_2, \dots, X_n$ are independent, identically distributed, each with mean μ and variance σ^2. Let $S_n = X_1 + X_2 + \dots + X_n$. Then, for any c, we have

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

where $\tilde{S}_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{S_n - E[S_n]}{\sqrt{Var(S_n)}}$ is the standardized version of S_n.

Today: ① jointly guassian random variables

---

③ jointly guassian random variables

The mean reason that guassian random variables is common in practice is CLT:

① sum of independent & identically distributed random variables can be approximated by a guassian random variable

② any randomness we observe in practice is supperposition of many "simillar" random events.

Now, suppose that instead of sum of one event, we are interested in two correlated events (V_i, U_i).

Example: Suppose that V_i is temprature of cpu while gaming at day i.

Suppose that U_i is temprature of gpu while gaming at day i.

Example: Suppose that V_i is temprature of cpu while gaming at day i.

Suppose that U_i is temprature of gpu while gaming at day i.

Suppose that $(U_1,V_1),(U_2,V_2),\dots$ are independent and identically distributed. For simplicity, assume $E[U_i]=E[V_i]=0$ and $Var(U_i)=Var(V_i)=1$ for all i. By CLT, for any $c\in\mathbb{R}$

$$\lim_{n\to\infty} P\left(\left\{\frac{U_1+\cdots+U_n}{\sqrt{n}}\leqslant c\right\}\right)=\Phi(c)$$

$$\lim_{n\to\infty} P\left(\left\{\frac{V_1+\cdots+V_n}{\sqrt{n}}\leqslant c\right\}\right)=\Phi(c)$$

How about ther joint distribution, i.e.,

$$\lim_{n\to\infty} P\left(\left\{\frac{U_1+\cdots+U_n}{\sqrt{n}}\leqslant c\right\}\cap\left\{\frac{V_1+\cdots+V_n}{\sqrt{n}}\leqslant c\right\}\right)=?$$

Notice that for any $a,b\in\mathbb{R}$, by CLT we have

$$\lim_{n\to\infty} P\left(\left\{\frac{aU_1+bV_1+aU_2+bV_2+\cdots+aU_n+bV_n}{\sqrt{n(a^2+b^2+2ab\rho_{U,V})}}\leqslant c\right)=\Phi(c)$$

since aU_i+bV_i are independent, $E[aU_i+bV_i]=0$ and $Var(aU_i+bV_i)=a^2+b^2+2ab\,\rho_{U,V}$

where $\rho_{U,V}=\dfrac{Cov(U_1,V_1)}{\sigma_{U_1}\sigma_{V_1}}=cov(U_1,V_1)$.

Hence, if joint distribution of $\left(\frac{U_1+\cdots+U_n}{\sqrt{n}},\frac{V_1+\cdots+V_n}{\sqrt{n}}\right)$ converges to joint distribution of X and Y, then for any $a,b\in\mathbb{R}$, $aX+bY$ must have guassian distribution.

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

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

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

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

where

$$\mu_X=E[X],\quad \sigma_X^2=Var(X)\qquad\qquad \rho=\rho_{X,Y}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

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

This is calle bivariate normal distribution.

Important remark:

If X and Y are independent guassian random variables. then their joint distribution is

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

Hence, they are jointly guassian. In particular, any linear combination of them are guassian. Hence, given $X_1, X_2, \dots, X_n$ independent & identically distributed, $\frac{X_1 + \dots + X_n - nE[X_1]}{\sqrt{n^2 Var(X_1)}}$ is a standard normal random variable.

Def: Suppose W & Z are independent standard normal random variables. Their joint pdf is called the standard bivariate normal pdf:

$$f_{W,Z}(\alpha,\beta) = \frac{1}{2\pi} e^{-\frac{\alpha^2+\beta^2}{2}}$$

Important remarks:

① Suppose that

$$\begin{pmatrix} X \\ Y \end{pmatrix} = A\begin{pmatrix} W \\ Z \end{pmatrix} + \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \quad A = \begin{pmatrix} \sqrt{\frac{\sigma_X^2(1+\rho)}{2}} & -\sqrt{\frac{\sigma_X^2(1-\rho)}{2}} \\ \sqrt{\frac{\sigma_Y^2(1+\rho)}{2}} & \sqrt{\frac{\sigma_Y^2(1-\rho)}{2}} \end{pmatrix}$$

Then $f_{X,Y}$ is given by (**)

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

③ If $\rho=0$ then $f_{X,Y}(u,v) = f_X(u) f_Y(v)$. Hence uncorrelated jointly guassian random variables are independent.

④ Let $Z = aX + bY$. Then X and Z are jointly guassian.

⑤ If X and Y are guassian, it does not mean that they are jointly guassian.

Proposition: Suppose that X & Y have bivariate normal pdf with parameters $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ and ρ.

(a) X has $\mathcal{N}(\mu_X, \sigma_X^2)$ distribution, Y has $\mathcal{N}(\mu_Y, \sigma_Y^2)$ distribution.

(b) For any $a, b \in \mathbb{R}$. $aX+bY$ is a guassion random variable, i.e., X and Y are jointly guassian.

(c) $\rho_{X,Y} = \rho$

(d) if $\rho = 0$, then they are independent

(e) For estimation of Y from X, $L^*(X) = g^*(X)$. Equivalently, $E[Y|X] = \hat{E}[Y|X]$, i.e., the best linear estimator of Y using X, is the best unconstrained estimator.

estimator of Y using X, is the best unconstrained estimator.

(d) The conditional distribution of Y given $X=u$ is $N(\hat{E}[Y|X=u], \sigma_e^2)$ where σ_e^2 is MSE for $\hat{E}[Y|X]$.