SNU MAE   Intro. Eng. Prob   note 11

Conditional Variance

Def Conditional variance of X given that $Y=y$

$$Var(X|Y) = E\left[(X-E[X|Y])^2 \mid Y\right]$$

analogous to the usual definition of variance, but all expectations are conditional on the fact that Y is known

Recall $var(X) = E[X^2] - (E[X])^2$

$$var(X|Y) = E[X^2|Y] - (E[X|Y])^2$$

$$E\left[Var(X|Y)\right] = \underbrace{E\left[E[X^2|Y]\right]}_{=E[X^2]} - E\left[(E[X|Y])^2\right] \qquad (1)$$

$$Var\left(E[X|Y]\right) = E\left[(E[X|Y])^2\right] - \underbrace{\left(E\left[E[X|Y]\right]\right)^2}_{=(E[X])^2} \qquad (2)$$

(1) & (2) $\Rightarrow$ $$\boxed{Var(X) = E\left[Var(X|Y)\right] + Var\left(E[X|Y]\right)} \qquad (3)$$

Ex4n $X_1, X_2, \ldots$: a sequence of iid r.v.

N: nonnegative integer-valued r.v., indept of X_i, $i \geq 1$.

$Var\left(\sum_{i=1}^{N} X_i\right) = ?$

Sol Condition on N and use (3)

$$E\left[\sum_{i=1}^{N} X_i \mid N\right] = N\,E[X]$$ ~ just the sum of the individual means

$$Var\left(\sum_{i=1}^{N} X_i \mid N\right) = N\,Var[X]$$

$$Var\left(\sum_{i=1}^{N} X_i\right) = E\left[N\ Var[X]\right] + Var\left(N\,E[X]\right)$$
$$= E[N]\ Var[X] + \left(E[X]\right)^2 Var(N).$$

//

Conditional Expectation & Prediction

Based on the observed value of a r.v. X, predict the value of a second r.v. Y.

Let $g(X)$ denote the predictor:

If X is observed to equal x,
then $g(x)$ is our prediction for the value of Y.

Want g so that $g(X) \approx Y$.

↳ One possible criterion: minimize $E\left[(Y-g(X))^2\right]$.

$$\Rightarrow \underline{g(X) = E[Y|X]}.$$

Why? → Prop.

Prop $$\boxed{E\left[(Y-g(X))^2\right] \geq E\left[(Y-E[Y|X])^2\right]} \quad (4)$$

Pf $$E\left[(Y-g(X))^2|X\right] = E\left[(Y-E[Y|X]+E[Y|X]-g(X))^2|X\right]$$
$$= E\left[(Y-E[Y|X])^2|X\right] + E\left[(E[Y|X]-g(X))^2|X\right]$$
$$+2\,\underbrace{E\left[(Y-E[Y|X])(E[Y|X]-g(X))|X\right]}_{\bigstar}$$

$E[Y|X]-g(X)$: ftn of X.

↳ given X, it can be treated as a const.

$$\bigstar = \left(E[Y|X]-g(X)\right) E\left[Y-E[Y|X]\,|X\right]$$
$$= \quad \text{"} \quad \left(E[Y|X]-E[Y|X]\right) = 0.$$

$$\therefore\ E\left[(Y-g(X))^2|X\right] \geq E\left[(Y-E[Y|X])^2|X\right].$$

↳ take expectations. → (4) //

The best estimate (minimum mean. square error)

$$E[S|R=r] = \frac{1}{1+\sigma^2}\mu + \frac{\sigma^2}{1+\sigma^2} r$$

= weighted avrg of a priori exp value (μ) and the value received (r).

relative weights : 1 to σ^2

1 ← conditional variance of the received signal when S is sent

σ^2 ← variance of the signal to be sent.

- when the joint distribution of X & Y is not completely known, or when the calculation of $E[Y|X=x]$ is mathematically intractable?
  - ↳ If mean/var of X & Y, and the correlation of X&Y are known, we can determine the <u>best linear predictor</u> of Y w.r.t X.

    i.e. choose a & b so as to minimize $E[(Y-(a+bX))^2]$.

$$E[(Y-(a+bX))^2] = E[Y^2] - 2aE[Y] - 2bE[XY] + a^2 + 2ab\,E[X] + b^2E[X^2].$$

$$\hookrightarrow \begin{cases} \frac{\partial}{\partial a}(\ \cdots\ ) = -2E[Y] + 2a + 2bE[X] = 0 \\ \frac{\partial}{\partial b}(\ \cdots\ ) = -2E[XY] + 2a\,E[X] + 2bE[X^2] = 0 \end{cases}$$

Letting ρ = correlation(X,Y), $\sigma_x^2 = Var(X)$, $\sigma_y^2 = Var(Y)$,

$$\begin{cases} b = \dfrac{E[XY] - E[X]E[Y]}{E[X^2] - (E[X])^2} = \dfrac{Cov(X,Y)}{\sigma_x^2} = \rho\dfrac{\sigma_y}{\sigma_x} \\ a = E[Y] - bE[X] = E[Y] - \dfrac{\rho\sigma_y}{\sigma_x}E[X] \end{cases}$$

$\Rightarrow$ The best lin predictor Y w.r.t X is

$$\mu_y + \frac{\rho\sigma_y}{\sigma_x}(X-\mu_x).$$

the mean sq err of this predictor

$$= E\left[\left(Y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(X-\mu_x)\right)^2\right]$$

$$= E\left[(Y-\mu_y)^2\right] + \rho^2\frac{\sigma_y^2}{\sigma_x^2}E\left[(X-\mu_x)^2\right] - 2\rho\frac{\sigma_y}{\sigma_x}E\left[(Y-\mu_y)(X-\mu_x)\right]$$

$$= \sigma_y^2 + \rho^2\sigma_y^2 - 2\rho^2\sigma_y^2$$

$$= \sigma_y^2(1-\rho^2) \quad \approx 0 \quad \text{as } \rho \approx \pm 1. \ //$$

Moment Generating Ftns

Def $M(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x) & : X \text{ discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\,dx & : X \text{ conti} \end{cases}$

moment gen. ftn. of the r.v. X (defined for all real values of t)

all the moments of X can be obtained by $M'(t)$, $M''(t)$, $\cdots$ at $t=0$.

Ex $M'(t) = \frac{d}{dt}E[e^{tX}]$ — assume diff & E interchangeable.

i.e. $\begin{cases} \frac{d}{dt}\left[\sum_x e^{tx}p(x)\right] = \sum_x \frac{d}{dt}\left[e^{tx}p(x)\right] \\ \frac{d}{dt}\left[\int e^{tx}f(x)\,dx\right] = \int \frac{d}{dt}\left[e^{tx}f(x)\right]dx \end{cases}$

$$= E\left[\frac{d}{dt}(e^{tX})\right]$$

$$= E[Xe^{tX}]$$

$$M''(t) = \frac{d}{dt}E[Xe^{tX}]$$

$$= E\left[\frac{d}{dt}(Xe^{tX})\right]$$

$$= E[X^2e^{tX}]$$

$$\vdots$$

$$M^{(n)}(t) = E[X^n e^{tX}]$$

$M'(0) = E[X]$

$M''(0) = E[X^2]$

$\vdots$

$M^{(n)}(0) = E[X^n]. \qquad n \geq 1$

Ex 6d [Normal distribution] $Z \sim N(0,1)$

$$M_Z(t) = E[e^{tZ}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{(x-t)^2}{2} + \frac{t^2}{2} \right\} dx$$

$$= e^{\frac{t^2}{2}}.$$

$\leftarrow$ moment gen ftn of a unit normal r.v.

Arbitrary normal r.v. $X = \mu + \sigma Z$

$$M_X(t) = E[e^{t(\mu+\sigma Z)}] = e^{t\mu} E[e^{t\sigma Z}] = e^{t\mu} M_Z(t\sigma)$$

$$= e^{t\mu} e^{\frac{(t\sigma)^2}{2}}$$

$$= \exp\left\{ \frac{\sigma^2 t^2}{2} + \mu t \right\}$$

$\downarrow$

$$\begin{cases} M_X'(t) = (\mu + t\sigma^2) \exp\{ \quad \} \\ M_X''(t) = (\mu + t\sigma^2)^2 \exp\{ \quad \} + \sigma^2 \exp\{ \quad \} \end{cases}$$

$\downarrow$

$$\begin{cases} E[X] = M_X'(0) = \mu \\ E[X^2] = M_X''(0) = \mu^2 + \sigma^2. \end{cases} \Rightarrow \operatorname{var}(X) = E[X^2] - (E[X])^2 = \sigma^2.$$

- moment gen ftn of the sum of indept r.v.s = product of the individual moment gen ftn.

$\because \quad M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} e^{tY}] = E[e^{tX}] E[e^{tY}]$

$= M_X(t) M_Y(t).$ //

Joint Moment Generating Ftns

Def The joint moment gen ftn for any n r.v.s $X_1, \cdots, X_n$:

$$M(t_1, \cdots, t_n) = E\left[e^{t_1 X_1 + \cdots + t_n X_n}\right].$$

$\Downarrow$ individual moment gen ftn

$$M_{X_i}(t) = E\left[e^{tX_i}\right] = M(0, \cdots, 0, \underset{\substack{\uparrow \\ i\text{th}}}{t}, 0, \cdots, 0)$$

Rmk $M(t_1, \cdots, t_n)$ uniquely determines the joint distribution of $X_1, \cdots, X_n$. (pf too difficult).

Fact $X_1, \cdots, X_n$ indept

$\Updownarrow$

$$M(t_1, \cdots, t_n) = M_{X_1}(t_1) \cdots M_{X_n}(t_n).$$

Pf
- If n rvs are indept, then

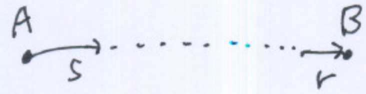
$$\begin{aligned} M(t_1, \cdots, t_n) &= E\left[e^{t_1 X_1} \cdots e^{t_n X_n}\right] \\ &= E\left[e^{t_1 X_1}\right] \cdots E\left[e^{t_n X_n}\right] \quad \text{independence} \\ &= M_{X_1}(t_1) \cdots M_{X_n}(t_n) \qquad (*) \end{aligned}$$

- If $(*)$ holds, then $M(t_1, \cdots, t_n)$ is the joint moment gen ftn of n indept rvs, ith of which has the same distribution as X_i. According to the remark, $M(t_1, \cdots, t_n)$ uniquely determines the joint distribution. $\rightarrow$ indept.

· Rmk [Intuitive argument for (4)]

- $E[(Y-C)^2]$ is minimized at $C=E[Y]$.
- If X is observed to be x, then the prediction problem remains the same with the exception that everything is now conditional on $X=x$.
- The best prediction in this situation is
  conditional expected value of Y given that $X=x$. //

Ex 5b



S: the value of the signal sent at A. $\sim N(\mu, \sigma^2)$

R: the value received at B. $\sim N(s, 1)$

The best estimate of S, given that $R=r$?

Sol.

$$f_{S|R}(s|r) = \frac{f_{S,R}(s,r)}{f_R(r)} = \frac{f_S(s)\, f_{R|S}(r|s)}{f_R(r)}$$

$$= \underset{\uparrow \text{ indept of } s.}{K} e^{-(s-\mu)^2/2\sigma^2}\, e^{-(r-s)^2/2}$$

$$\frac{(s-\mu)^2}{2\sigma^2} + \frac{(r-s)^2}{2} = s^2\left(\frac{1}{2\sigma^2}+\frac{1}{2}\right) - \left(\frac{\mu}{\sigma^2}+r\right)s + C_1$$

$$= \frac{1+\sigma^2}{2\sigma^2}\left[s^2 - 2\left(\frac{\mu+r\sigma^2}{1+\sigma^2}\right)s\right] + C_1$$

$$= \frac{1+\sigma^2}{2\sigma^2}\left(s - \frac{\mu+r\sigma^2}{1+\sigma^2}\right)^2 + C_2$$

$$\therefore\ f_{S|R}(s|r) = C \exp\left[\frac{-\left[s-\frac{\mu+r\sigma^2}{1+\sigma^2}\right]^2}{2\,\frac{\sigma^2}{1+\sigma^2}}\right] \ : \text{normal dist!}$$

$$\begin{cases} E[S\,|\,R=r] = \dfrac{\mu+r\sigma^2}{1+\sigma^2} \\[2mm] Var(S\,|\,R=r) = \dfrac{\sigma^2}{1+\sigma^2} \end{cases}$$

Multivariate Normal Distribution

$Z_1, \cdots, Z_n$: indept. $\sim N(0,1)$

<u>Def</u> If for some constants a_{ij}, $1 \le i \le m$, $1 \le j \le n$, and μ_i, $1 \le i \le m$,

$$X_1 = a_{11} Z_1 + \cdots + a_{1n} Z_n + \mu_1$$
$$\vdots$$
$$X_i = a_{i1} Z_1 + \cdots + a_{in} Z_n + \mu_i$$
$$\vdots$$
$$X_m = a_{m1} Z_1 + \cdots + a_{mn} Z_n + \mu_m$$

then $X_1, \cdots, X_m$ are said to have a <u>multivariate normal distribution</u>.

<u>Fact</u> The sum of independent normal r.v.s is itself a normal r.v.

$\Rightarrow$ Each X_i is a normal r.v. with

$$E[X_i] = \mu_i$$
$$Var(X_i) = \sum_{j=1}^{n} a_{ij}^2$$

<u>Note</u> $\sum_{i=1}^{m} t_i X_i$ is also normally distributed.

$$E\left[\sum_{i=1}^{m} t_i X_i\right] = \sum_{i=1}^{m} t_i \mu_i \qquad (i)$$

$$Var\left(\sum_{i=1}^{m} t_i X_i\right) = Cov\left(\sum_{i=1}^{m} t_i X_i, \sum_{j=1}^{m} t_j X_j\right)$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{m} t_i t_j \, Cov(X_i, X_j) \qquad (ii)$$

If $Y \sim N(\mu, \sigma^2)$, then

$$E[e^Y] = M_Y(t)\Big|_{t=1} = e^{\frac{\sigma^2}{2} + \mu} \qquad (iii)$$

From (i), (ii), & (iii),

$$M(t_1, \cdots, t_m) = E\left[\exp\{t_1X_1 + \cdots + t_mX_m\}\right]$$

$$= \exp\left\{\sum_{i=1}^{m} t_i\mu_i + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} t_it_j \mathrm{Cov}(X_i, X_j)\right\}$$

//

↑ Joint distribution of $X_1, \cdots, X_m$ is completely determined by $E[X_i]$ and $\mathrm{cov}(X_i, X_j)$. (✱)

Joint Distribution of the Sample Mean & Sample Variances

$X_1, \cdots, X_n$: independent, $X_i \sim N(\mu, \sigma^2)$.

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$: sample mean.

↪ $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

<u>Recall</u>
- $\mathrm{Cov}(\bar{X}, X_i - \bar{X}) = 0$
- $\bar{X}, X_1 - \bar{X}, \cdots, X_n - \bar{X}$ are all lin. combinations of the independent standard normals $\frac{X_i - \mu}{\sigma}$.

$\Rightarrow$ $\bar{X}, X_i - \bar{X}$ has a joint distribution that is multivariate normal.

Let Y be a r.v. indept of X_i,
$Y \sim N(\mu, \frac{\sigma^2}{n})$.

Then $Y, X_i - \bar{X}$, also has a multivariate normal joint distribution. same E and Cov as $\bar{X}, X_i - \bar{X}$.

(✱) $\Rightarrow$ $Y, X_i - \bar{X}$ and $\bar{X}, X_i - \bar{X}$ have the same joint distribution.

$\Rightarrow$ $\bar{X}$ is indept of $X_i - \bar{X}$ $i = 1, \cdots, n$.

$\Rightarrow \bar{X}$ is also indept of the sample variance

$$S^2 \equiv \sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{(n-1)} .$$

Need distribution of S^2.

$$(n-1)S^2 = \sum_{i=1}^{n}\left[(X_i-\mu)^2 + (\mu-\bar{X})^2 + 2(X_i-\mu)(\mu-\bar{X})\right]$$

$$= \sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2 .$$

$\Downarrow$

$$\frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \qquad (*)$$

independent.

sum of the squares of n indept standard normal rvs.
$\Rightarrow$ chi-squared!

moment gen ftn
$(1-2t)^{-1/2}$
(Ex 6i, pg 368)

Recall: the moment gen ftn of the sum of indept rvs is the product of their individual moment gen ftns.

$$(*) \Rightarrow E\left[e^{t(n-1)S^2/\sigma^2}\right](1-2t)^{-1/2} = (1-2t)^{-\frac{n}{2}}$$

$$\Rightarrow E\left[e^{t(n-1)S^2/\sigma^2}\right] = (1-2t)^{-\frac{1}{2}(n-1)}$$

: moment gen ftn of a chi-squared r.v. with $n-1$ dof.

Recall: the moment gen ftn uniquely determines the distribution of the rv.

<u>Prop</u> $X_1, \cdots, X_n$: i.i.d. $\sim N(\mu, \sigma^2)$

$\Rightarrow$ Then the sample mean $\bar{X}$ and the sample variance S^2 are indept.

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$\frac{(n-1)S^2}{\sigma^2}$: chi-squared with $n-1$ dof.

General Def of Expectation (skip?)

There exist rvs that are neither discrete nor conti, and they too may possess E.

ex) X : Bernoulli with parameter $p = \frac{1}{2}$.

Y : $U[0,1]$.

X, Y : indept.

Define $W = \begin{cases} X & \text{if } X = 1 \\ Y & \text{if } X \neq 1. \end{cases}$

- The set of possible values $[0,1]$ is uncountable
  $\Rightarrow$ W is NOT discrete.
- $P\{W = 1\} = \frac{1}{2}$ $\Rightarrow$ W is NOT conti.

How do we define E for an arbitrary rv?

Recall $\int_a^b g(x)\, dx = \lim \sum_{i=1}^n g(x_i)(x_i - x_{i-1})$ for any ftn g.

$\uparrow$

(taken over all $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ as $n \to \infty$ and $\max_{i=1,\cdots,n} (x_i - x_{i-1}) \to 0$.) ※

<u>Def</u> For any distribution ftn F, the Stieltjes integral of the nonnegative ftn g over $[a,b]$

$$\int_a^b g(x)\, dF(x) = \lim \sum_{i=1}^n g(x_i)\left[F(x_i) - F(x_{i-1})\right]$$

$\uparrow$

(...) ※

<u>Def</u> The Stieltjes integral over the whole line

$$\int_{-\infty}^{\infty} g(x)\, dF(x) = \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b g(x)\, dF(x).$$

<u>Def</u> If g is not a nonnegative ftn,

$$g^+(x) = \begin{cases} g(x) & \text{if } g(x) \geq 0 \\ 0 & \end{cases}$$

$$g^-(x) = \begin{cases} 0 & \text{if } g(x) \geq 0 \\ -g(x) & < 0 \end{cases}$$

both nonnegative.

$g = g^+ - g^-$.

$$\Rightarrow \underbrace{\int_{-\infty}^{\infty} g(x)\, dF(x)}_{(*)} = \underbrace{\int_{-\infty}^{\infty} g^+(x)\, dF(x)}_{①} - \underbrace{\int_{-\infty}^{\infty} g^-(x)\, dF(x)}_{②}.$$

We say that (*) exists as long as ① & ② are not both equal to $+\infty$.

<u>Def</u> If X is an arbitrary rv with cumulative distribution F,

$$E[X] = \int_{-\infty}^{\infty} x\, dF(x).$$

→ this matches the previous def of E for conti/discrete rvs.

$$\rightarrow \sum_{i=1}^{n} x_i \underbrace{\left[F(x_i) - F(x_{i-1})\right]}_{\text{prob that } X \text{ will be in } (x_{i-1}, x_i]}.$$

<u>Rmk</u> Stieltjes integrals are mainly of theoretical interest. → compact way of defining / dealing with E.