Chapter 5 Mathematical Expectation for Multivariate Case

- ► Goldberger, Ch. 4
- ► Wackerly et al. chapter 5

  Yale note chapter 9

- ► Expectation for Functions of Several Random Variable

Product Moment

Linear Function

Conditional Expectation

Conditional Variance

Independence

1. Expectation for Functions of Several Random Variable

- Let $\left(X_1, X_2, \cdots, X_n\right) \sim f(x_1, x_2, \cdots, x_n)$ be joint p.d.f.

Let $Y = g\left(X_1, X_2, \cdots, X_n\right)$.

Then, $E\left[Y\right] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \cdots, x_n) f(x_1, x_2, \cdots, x_n) dx_1 \cdots dx_n$.

- For discrete random variable, $E\left[Y\right] = \sum_{x_1} \cdots \sum_{x_n} g(x_1, x_2, \cdots, x_n) f(x_1, x_2, \cdots, x_n)$.

(Special case) <u>Bivariate Case</u>

Let $Z = g\left(X, Y\right)$, $E\left[Z\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dxdy$.

(Note) If Z is a function of only one random variable, $Z = g(\mathrm{X})$,

$E[Z] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x) f(x, y) dxdy = \int_{-\infty}^{\infty} g(x) f_X(x) dx$; marginal expectation.

2. Product Moment

(Definition) The rth and sth *uncentered product moment*(*product moment about origin*) is

$$\mu'_{r,s} = E\left[X^r Y^s\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r y^s f(x,y)\,dxdy, \qquad \text{for } r,s = 0,1,\cdots$$

(Example) $\mu'_{1,0} = \mu_X$, $\mu'_{0,1} = \mu_Y$.

(Definition) The rth and sth *centered product moment*(*product moment about mean*) is

$$\mu_{r,s} = E\left[(X-\mu_X)^r (Y-\mu_Y)^s\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)^r (y-\mu_Y)^s f(x,y)\,dxdy$$

(Definition) ***Covariance***($\mu_{1,1}$; σ_{XY}, $Cov(X,Y)$, $C(X,Y)$)

$\sigma_{XY} = E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]$; measures of co-movement between X and Y.

☞ $E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right] = E\left[XY\right] - E\left[X\right]E\left[Y\right]$.

- Interpreting $E\left[\cdot\right]$ as the average in infinitely many trials, the covariance between two random variables is a measure of "average" co-movement.

► $\sigma_{XY} > 0$, if X and Y jointly tend to be above their means on average.

$\Rightarrow$ positive relationship.

► $\sigma_{XY} < 0$, if X and Y move in opposite directions relative to their means.

$\Rightarrow$ negative relationship.

► $\sigma_{XY} = 0$, if X and Y do not move together on average relative to their means.

(Some comments)

ⓐ σ_{XY} measures association, NOT causation.

ⓑ σ_{XY} is only a linear measure.

• To measure the strength of co-variation, it is useful to normalize σ_{XY} and define the *correlation coefficient*.

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}, \qquad -1 \le \rho_{XY} \le 1.$$

► If $\sigma_{XY} = 0, \ \text{then} \ \rho_{XY} = 0$.

(Remarks)

① If X and Y are independent, $Cov(X,Y)=0$ **and** $\rho_{XY}=0$.

- Since X and Y are independent, $E(XY)=E(X)E(Y)$.

(Proof)

► However, $Cov(X,Y)=0$ does not imply independence,

because $Cov(X,Y)$ only measures the linear relationship.

► There could be a nonlinear relationship even if $Cov(X,Y)=0$.

- All these results can be extended to more than two variables.

► Independence of $X_1, X_2, \cdots, X_n$: $E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n)$.

② $Cov\left(aX,bY\right)=ab\cdot Cov(X,Y)$ but $\rho_{aX,bY}=\pm\rho_{XY}$.

If $a\cdot b>0$이면 $\rho_{aX,bY}=\rho_{XY}$.

If $a\cdot b<0$이면 $\rho_{aX,bY}=-\rho_{XY}$.

(Proof)

③ $\left|\rho_{XY}\right| \leq 1$: Cauchy-Schwartz inequality.

(Proof)

3. Expectation of Linear Functions of Two Random Variables

① (<u>Linear function</u>)

Suppose $Z = a + bX + cY$, where a,b,c are constants.

Then,

$$E[Z] = a + bE[X] + cE[Y]$$
$$V[Z] = b^2V[X] + c^2V[Y] + 2bc \cdot Cov[X,Y].$$

► If X and Y are independent, then $Cov(X,Y)$ drops out.

► In general, $V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2\sum_{i \neq j} a_i a_j Cov(X_i, X_j)$.

(proof)

(Example) Consider the random variables, X, Y, Z with

$$\mu_X = 2, \quad \mu_Y = -3, \quad \mu_Z = 4,$$
$$\sigma_X^2 = 1, \quad \sigma_Y^2 = 5, \quad \sigma_Z^2 = 2,$$
$$\sigma_{XY} = -2, \quad \sigma_{XZ} = -1, \quad \sigma_{YZ} = 1.$$

Find the mean and variance of $W = 3X - Y + 2Z$.

ⓐ $E(W) = 17$.

ⓑ $V(W) = 18$.

② (Covariance and variance)

$$Cov\left(X,X\right)=V\left(X\right)$$

③ (<u>Pair of linear function</u>)

Suppose $Z_1 = a_1 + b_1 X + c_1 Y,\ Z_2 = a_2 + b_2 X + c_2 Y$, where $a_1, b_1, c_1, a_2, b_2, c_2$ are constants.

Then, $Cov(Z_1, Z_2) = b_1 b_2 V\left(X\right) + c_1 c_2 V\left(Y\right) + \left(b_1 c_2 + b_2 c_1\right) Cov(X, Y)$.

(Proof)

4. Conditional Expectation

◎ Let the random vector (X, Y) have joint pdf $f(x,y) = h_2(y|x) f_X(x)$, and let $Z = g(X,Y)$ be a function of (X, Y). Then, the <u>*conditional expectation*</u> of Z given $X = x$ is

$$E\left[Z|x\right] = \int_{-\infty}^{\infty} g\left(x,y\right) h_2\left(y|x\right) dy.$$

• In particular, for $Z = g(X,Y) = Y$, we obtain <u>*conditional mean*</u> of Y at given x:

$$\mu_{Y|x} = E\left[Y|x\right] = \int_{-\infty}^{\infty} y h_2\left(y|x\right) dy.$$

• Similarly, the <u>*conditional variance*</u> of Y given x is:

$$\sigma_{Y|x}^2 = E\left[\left(Y - \mu_{Y|x}\right)^2 \middle| x\right] = E\left[Y^2|x\right] - \left(\mu_{Y|x}\right)^2.$$

(Note) $E\left[Z|X\right]=E\left[g(X,Y)|x\right]$ is only a function of x (given value).

(Example) Recall the joint pdf $f(x,y)=\begin{cases}\frac{2}{3}(x+2y) & 0<x<1,\ 0<y<1\\ 0 & elsewhere\end{cases}$.

Find the conditional mean and conditional variance of X given $Y=1/2$.

(Interpretation) Income-age profile and expected income of immigrant.

(<u>Some useful rules</u>: $E\left(Z \middle| x\right)$)

① For $Z = g(X)$, $E\left(g(X) \middle| x\right) = g(x)$.

② For $Z = g(X)Y$, $E\left(g(X)Y \middle| x\right) = g(x)E\left(Y \middle| x\right)$.

③ For $Z = a + bX + cY$, $E\left(Z \middle| x\right) = a + bx + cE\left(Y \middle| x\right)$.

④ For $Z = \left(Y - \mu_{Y|x}\right)$, $E\left(\left(Y - \mu_{Y|x}\right) \middle| x\right) = \mu_{Y|x} - \mu_{Y|x} = 0$.

⑤ For $Z = \left(Y - \mu_Y\right)$, $E\left(\left(Y - \mu_Y\right) \middle| x\right) = \mu_{Y|x} - \mu_Y$.

⑥ For $Z = \left(Y - \mu_Y\right)^2$, $E\left(Z \middle| x\right) = \sigma^2_{Y|x} + \left(\mu_{Y|x} - \mu_Y\right)^2$.

(1) Law of Iterated Expectation

(Theorem) The (marginal) expectation of $Z = g(X,Y)$ is the expectation of its conditional expectations: $E(Z) = E_X\left[E\left(Z|x\right)\right]$.

(Proof)

(Special case ①) $E(Y) = E_X\left[E\left(Y|x\right)\right]$.

(Example) Suppose a point X is chosen from $U(0,1)$. After the value $X=x$ has been observed $(0<x<1)$, a point Y is chosen from $U(x,1)$. Determine the value of $E(Y)$.

(Solution)

(Motivation) LS estimator: when $(\mathrm{X},\ \varepsilon)$~stochastic, find $\mathrm{E}\left[(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\varepsilon\right]$.

(Special case ②) $E(XY) = E\left(X \cdot \mu_{Y|x}\right)$.

(Special case ③) $C(X,Y) = C\left(X, \mu_{Y|x}\right)$.

(2) <u>Analysis of Variance</u> (삭제?)

We already derived that $E\left(\left(Y-\mu_y\right)^2 \middle| x\right)=\sigma_{Y|x}^2+\left(\mu_{Y|x}-\mu_Y\right)^2$. (refer note p17, ⑤)

By law of iterated expectation,

$$E\left[\left(Y-\mu_Y\right)^2\right]=E_X\left[\sigma_{Y|x}^2\right]+E_X\left[\left(\mu_{Y|x}-\mu_Y\right)^2\right]$$

$$\Rightarrow V\left(Y\right)=E_X\left[\sigma_{Y|x}^2\right]+V_X\left[\mu_{Y|x}\right]$$

► $E_X\left[\sigma_{Y|x}^2\right]$: residual variance.

► $V_X\left[\mu_{Y|x}\right]$: regression variance.

(Example) $X : R\,\&\,D \rightarrow Y : Patents$.

We are interested in $E(patents \mid R\,\&\,D)$.

$V(Patents) =$ ⓐ due to variation of patents at each level of R&D.

ⓑ due to variation of mean patents as R&D varies.

5. Conditional Expectation Function(C.E.F.)

(Note) $E\left(Y|x\right)=\int_{-\infty}^{\infty} y h_2\left(y|x\right)dy$: conditional mean at $X=x$.

As we change x, we get $E\left(Y|X\right)=\mu_{Y|X}$: <u>*C.E.F.*</u>(<u>*(population) regression function*</u>).

6. Prediction 2 (Conditional Prediction)

(1) Best Predictor

- $\left(X, Y\right) \sim f(x, y)$: *known*

A single draw is made. You are told the value of X that was drawn, and asked to predict the value of Y, using any function of X, $g(X)$.

What is your ***best predictor***, in the sense of minimizing $E\left(v^2\right)$ where $v = Y - g(X)$?

(Example) 신입생 선발: 내신성적⇒CGPA

(Answer) $g^*(X) = E\left(Y \middle| X\right)$.

(Proof)

(<u>Properties of C.E.F.</u>)

Let $\varepsilon = Y - E\left(Y \middle| X\right)$.

① $E\left(\varepsilon \middle| X\right) = 0 \;\Rightarrow\; E(\varepsilon) = 0$.

② $V\left(\varepsilon \middle| X\right) = V\left(Y \middle| X\right) (= \sigma^2_{Y|X})$.

(Proof)

③ $Cov(X, \varepsilon) = 0$

(Proof)

④ $Cov\left(h(X), \varepsilon\right) = 0$ for all function $h(\cdot)$.

(2) Best Linear Predictor

- Suppose your choice is limited to linear function of X, $g(X) = a + bX$.

The *best linear predictor*, which minimizes $E\left(v^2\right)$ where $v = Y - (a + bX)$, is:

$$h^*(X) = \alpha + \beta X,\ \beta = \frac{\sigma_{XY}}{\sigma_X^2},\ \alpha = \mu_Y - \beta\mu_X$$

► *Linear projection* of Y on X.

(Proof)

7. Independence

- X and Y are *statistically independent* iff $f(x,y) = f_X(x) f_Y(y)$ for all (x,y) of (X,Y) in sample space.

(Example) $f(x,y)=\frac{2}{3}(2y+x), \qquad 0<x,y<1$

$\Rightarrow$ <u>NOT</u> independent

(Example) $f(x,y)=\begin{cases} e^{-(x+y)} & x,y>0 \\ 0 & otherwise \end{cases}$

$\Rightarrow$ Independent

• If $X_1, X_2, \cdots, X_n$ have joint p.d.f. $f(x_1, x_2, \cdots, x_n)$, then $X_1, X_2, \cdots, X_n$ are <u>(*mutually*)</u> <u>*statistically independent*</u> iff $f(x_1, x_2,, x_n) = f_1(x_1)\cdots f_n(x_n)$.

• <u>Random Sample</u>: Suppose we repeat an experiment n times and let $X_1, X_2, \cdots, X_n$ be random variables representing these n trials.

► $X_1, X_2, \cdots, X_n$ is random sample if they are <u>*independent and identically distributed*</u>(*i.i.d.*).

►

$$
\begin{aligned}
P\left(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\right) &= f(x_1, x_2, \cdots, x_n) \\
&= f_1(x_1) f_2(x_2) \cdots f_n(x_n); \text{ } \textit{independent} \\
&= f(x_1) f(x_2) \cdots f(x_n); \text{ } \textit{identically distributed} \\
&= \prod_{i=1}^{n} f(x_i)
\end{aligned}
$$

• If X and Y are independent, then $h(X)$ and $g(Y)$ are also independent for any function $h(\cdot)$ & $g(\cdot)$.

• If X and Y are uncorrelated, then $h(X)$ and $g(Y)$ are uncorrelated for <u>linear</u> function of $h(\cdot)$ & $g(\cdot)$.

► If X and Y are uncorrelated, $Cov(X,Y) = E(XY) - E(X)E(Y) = 0$

$\Rightarrow$ $E(XY) = E(X)E(Y)$.

► If X and Y are independent, $E\left(X^r Y^s\right) = E\left(X^r\right)E\left(Y^s\right)$ for all r, s.