Chapter 4 part 3 §4.4 Conditional Expectation and Prediction

In Exam 2, we have

Y\X	1	2	3	4	
3	0.10	0.05	0.02	0.02	0.19
4	0.05	0.20	0.05	0.02	0.32
5	0.02	0.05	0.20	0.04	0.31
6	0.02	0.02	0.04	0.10	0.18
	0.19	0.32	0.31	0.18	1.00

From joint frequency functions $p(x,y) = P(X=x, Y=y)$, one can obtain marginal frequency function $p_X(x)$ and $p_Y(y)$.

e.g. $P(X=2) = p(2,3) + p(2,4) + p(2,5) + p(2,6) = 0.32.$

summary:

X	1	2	3	4
$p(x)$	0.19	0.32	0.31	0.18

Also, one can obtain conditional dist. of X, given $Y=5$.

i.e. $P(X=1 | Y=5) = \frac{p(1,5)}{p(Y=5)} = \frac{0.02}{0.31} = \frac{2}{31} = 0.065.$

$P(X=2 | Y=5) = P(X=2, Y=5)/P(Y=5) = \frac{0.05}{0.31} = 5/31 = 0.161.$

Summary.

$X \mid Y=5$	1	2	3	4
$p(x\|5)$	0.065	0.161	0.645	0.129

$X|Y=5$ could be considered as an ordinary r.v. with its frequency function. Thus, one can consider its mean — conditional mean (in this case)

e.g.

$$E(X|Y=5) = 1\cdot(0.065) + 2\cdot(0.161) + 3\cdot(0.645) + 4\cdot(0.129) = 2.838.$$

compare $E(X) = 1\cdot(0.19) + 2\cdot(0.32) + 3\cdot(0.31) + 4\cdot(0.18) = 2.48.$

Definition: $E(X|Y=y) = \sum_i x_i \cdot p_{X|Y}(x_i|y).$

$$E(Y|X=x) = \sum_j y_j \cdot p_{Y|X}(y_j|x).$$

rewrite:

$X\|Y=y$	x_1	x_2	x_3	$\cdots\cdots$	x_n	$\cdots\cdots$
$p_{X\|Y}(x\|y)$	$p(x_1\|y)$	$p(x_2\|y)$	$p(x_3\|y)$	$\cdots\cdots$	$p(x_n\|y)$	$\cdots\cdots$

then
$$E(X|Y=y) = x_1\cdot p(x_1|y) + x_2\cdot p(x_2|y) + x_3\cdot p(x_3|y) + \cdots + x_n\cdot p(x_n|y) + \cdots$$
$$= \sum_{i=1}^{\infty} x_i\cdot p(x_i|y)$$
$$= \sum_{i=1}^{\infty} x_i\cdot \frac{P(X=x_i, Y=y)}{P(Y=y)}.$$

For continuous r.v.s $(X,Y)\sim f(x,y)$.
$$X|Y=y \sim f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \qquad f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx.$$

thus $E(X|Y=y)=\int_{-\infty}^{\infty} x\cdot f_{X|Y}(x|y)\,dx.$ $\left[\begin{array}{l}\text{compare } X\sim f(x),\\ E(X)=\int_{-\infty}^{\infty} x\cdot f(x)\,dx\end{array}\right]$

Similarly, $E(Y|X=x)=\int_{-\infty}^{\infty} y\cdot f_{Y|X}(y|x)\,dy.$

further more, $E[h(Y)|X=x]=\int_{-\infty}^{\infty} h(y)\cdot f_{Y|X}(y|x)\,dy$

$\left\{\begin{array}{l}\text{compare } Y\sim f_Y(y),\\ E[h(Y)]=\int_{-\infty}^{\infty} h(y)\cdot f_Y(y)\,dy\end{array}\right\}$

Expl A.

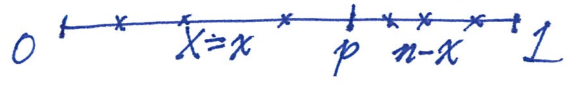
Consider a Poisson process on $[0,1]$ with mean λ, and let N be the number of points in $[0,1]$. For $p<1$, let X be the number of points in $[0,p]$. Find the conditional distribution and conditional mean of X given $N=n$.

Ans: try to find

$X\|N=n$	0	1	2	$\cdots\cdots$	n
$p(x\|n)$	?	?	?		?

$$P(X=x|N=n)=\frac{P(X=x, N=n)}{P(N=n)}, \qquad P(N=n)=\frac{\lambda^n}{n!}e^{-\lambda} \text{ poisson.}$$

$$P(X=x,N=n)=P(X=x)P(N=n|X=x)=\frac{(\lambda p)^x e^{-\lambda p}}{x!}\cdot\frac{[\lambda(1-p)]^{n-x}\cdot e^{-\lambda(1-p)}}{(n-x)!}$$



i.e. $X \sim P(\lambda p)$, from poisson process assumption.

also $P(N=n|X=x)$ is the prob. that there are $n-x$ points over $[p, 1]$.

which is Poisson with parameter $(1-p)\lambda$.

thus, $P(X=x, N=n) = \dfrac{\lambda^x \cdot p^x \cdot \cancel{e^{-\lambda p}}}{x!} \cdot \dfrac{\lambda^{n-x} \cdot (1-p)^{n-x}\, e^{-\lambda} \cdot \cancel{e^{\lambda p}}}{(n-x)!}$

$$= \lambda^n e^{-\lambda} \cdot \frac{p^x \cdot (1-p)^{n-x}}{x!\,(n-x)!} \qquad \begin{array}{l} x=0,1,2,\cdots,n \\ n=0,1,2,\cdots \end{array}$$

therefore, $P(X=x|N=n) = \dfrac{\lambda^n e^{-\lambda} \cdot p^x \cdot (1-p)^{n-x}}{x!\,(n-x)!} \Big/ \dfrac{\lambda^n \cdot e^{-\lambda}}{n!} = \dfrac{n!}{x!\,(n-x)!} \cdot p^x (1-p)^{n-x}$

i.e. $P(X=x|N=n) = \binom{n}{x} p^x (1-p)^{n-x}$, i.e., $X|_{N=n} \sim B(n, p)$.

finally,

$X/N=n$	0	$\cdots$	k	$\cdots$	n
$p(x\|n)$	$\binom{n}{0} p^0 (1-p)^{n-0}$	$\cdots$	$\binom{n}{k} p^k (1-p)^{n-k}$		$\binom{n}{n} p^n (1-p)^{n-n}$

Last, $E(X|N=n) = n \cdot p$.

Expl. Suppose $(X, Y) \sim f(x,y) = \begin{cases} \frac{1}{y} e^{-y} \cdot e^{-x/y}; & x>0, y>0 \\ 0 \quad ; & \text{else} \end{cases}$

find $E(X|Y=y)$, $y>0$.

Ans: $E(X|Y=y) = \int_{-\infty}^{\infty} x \cdot f(x|y)\,dx$, $\quad f(x|y) = \dfrac{f(x,y)}{f_Y(y)}$.

$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{\infty} \frac{1}{y} e^{-y} e^{-\frac{x}{y}}\,dx = e^{-y} \int_0^{\infty} \frac{1}{y} e^{-\frac{x}{y}}\,dx$

$t=\frac{x}{y}$ $= e^{-y} \int_0^{\infty} e^{-t}\,dt = e^{-y}$, $\quad f(x|y) = \frac{1}{y} e^{-x/y}$, $x>0$.

$\frac{x}{y}=t \quad dx = y\,dt$

thus $E(X|Y=y) = \int_0^{\infty} x \cdot \frac{1}{y} e^{-\frac{x}{y}}\,dx = \int_0^{\infty} \frac{x}{y} e^{-\frac{x}{y}}\,dx = \int_0^{\infty} t e^{-t} \cdot y\,dt$

$= y\int_0^\infty t e^{-t} dt = y \cdot \Gamma(2) = y.$ $\left\{ \text{recall } \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx,\ \Gamma(n+1) = n! \right\}$

i.e., $E(X|Y=y) = y.$ e.g. $E(X|Y=10) = 10.$

Consider $E(Y|X=x)$ again. Conditional expection of Y given $X=x$. So it is a well-defined function of X, hence $E(Y|X)$ is a random variable. values are $E[Y|X=x]$ with p.m.f $p(x)$ or p.d.f. $f(x)$. { just like $g(X)$. Values are $g(x)$ with $p(x)$ or $f(x)$.

$$E[g(X)] = \begin{cases} \sum_i g(x_i) p(x_i) \\ \int g(x) f(x) dx. \end{cases}$$
}

In Expl A, we find $E(X|N=n) = np.$

thus $E(X|N) = Np$ is a r.v. which is a function of N.

We can find $E[E(X|N)]$, $Var[E(X|N)]$ from above relationship, $E[E(X|N)] = E[Np] = pE[N] = p \cdot \lambda.$ (*)

$$Var[E(X|N)] = Var[Np] = p^2 \, Var[N] = p^2 \cdot \lambda.$$

Theorem A $E(Y) = E[E(Y|X)].$

proof: $E(Y|X)$ is a r.v, which is a function of X. $\left\{ E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx \right\}$

$$\text{thus } E[E(Y|X)] = \int_{-\infty}^{\infty} E(Y|X=x) \cdot f(x) dx$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y f(y|x) dy \right] f(x) dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(y|x) \cdot f(x) dy\, dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y \cdot \frac{f(x,y)}{f(x)} \cdot f(x) dy\, dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y) dx\, dy = \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x,y) dx\, dy$$

$$= \int_{-\infty}^{\infty} y \cdot f_Y(y) dy = E(Y).$$

e.g. $E[E(X|N)] \stackrel{\text{Thm}}{=} E(X) = p \cdot \lambda,$ Since $X \sim \text{poisson}(p\lambda)$ same as that in (*).

law of total expectation: $E(Y)$ is calculated by weighting the conditional expectations appropriately and summing or integrating.

Expl D. Random Sums. $T=\sum_{i=1}^{N} X_i$, N is a r.v. $E(N)<\infty$, X_i are r.v.s, $E(X_i)$ same.

X_i and N are independent.

e.g. 1. Insurance company; N claims, $X_1, X_2, \cdots$ amounts of claims.

e.g. 2. N is # of customers entering store, X_i, amount of money spent for i-th.

e.g. 3. N is # of jobs in a single server queue, X_i, service time for i-th job. T is the time to serve all the jobs in the queue.

From Thm A, $E(T)=E[E(T|N)]$.

Since $E(T|N=n)=E\left(\sum_{i=1}^{n} X_i \middle| N=n\right) \overset{X_i \perp\!\!\!\perp N}{=} E\left(\sum_{i=1}^{n} X_i\right) = n\cdot E(X)$,

thus, $E(T|N)=N\cdot E(X)$. finally $E[E(T|N)]=E[N\cdot E(X)]=E(X)\cdot E(N)$.

i.e., $E(T)=E(X)\cdot E(N)$. average value of N times average amount of time to complete a job.

Expl. previous Expl. $(X,Y)\sim f(x,y)=\begin{cases}\frac{1}{y}e^{-y}e^{-x/y}; & x>0,\ y>0\\ 0; & \text{else.}\end{cases}$

one finds $E(X|Y=y)=y$. find $E(X|Y)$, $E[E(X|Y)]$.

Ans: From $E(X|Y=y)=y$, obtain $E(X|Y)=Y$.

thus $E[E(X|Y)]=E(Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f(x,y)\,dx\,dy = \int_0^{\infty}\int_0^{\infty} y\cdot\frac{1}{y}e^{-y}e^{-\frac{x}{y}}\,dx\,dy$

$=\int_0^{\infty} e^{-y}\int_0^{\infty} e^{-\frac{x}{y}}\,dx\,dy = \int_0^{\infty} e^{-y}\cdot y\,dy = 1.$

From Thm A. $E[E(X|Y)]=E(X)$.

verify: $E(X)=\int_0^{\infty}\int_0^{\infty} x\cdot\frac{1}{y}e^{-y}e^{-\frac{x}{y}}\,dx\,dy = \int_0^{\infty} e^{-y}\int_0^{\infty}\frac{x}{y}e^{-\frac{x}{y}}\,dx\,dy = \int_0^{\infty} e^{-y}\cdot y\,dy = 1.$

Theorem B. $Var(Y) = Var[E(Y|X)] + E[Var(Y|X)]$.

first. $Var(Y|X)$ is a function of X, its value $Var(Y|X=x) = E(Y^2|X=x) - [E(Y|X=x)]^2$.

thus $Var(Y|X) = E(Y^2|X) - [E(Y|X)]^2$.

Now,
$$E[Var(Y|X)] = E[E(Y^2|X)] - E\{[E(Y|X)]^2\}$$
$$= E(Y^2) - E\{[E(Y|X)]^2\}$$

also
$$Var[E(Y|X)] = E\{[E(Y|X)]^2\} - \{E[E(Y|X)]\}^2$$
$$= E\{[E(Y|X)]^2\} - [E(Y)]^2$$

refer $Var[g(X)] = E\{[g(X)]^2\} - \{E[g(X)]\}^2$

Now, $RHS = Var[E(Y|X)] + E[Var(Y|X)] = E(Y^2) - [E(Y)]^2 = Var(Y)$.

Expl. E. Random Sums $T = \sum_{i=1}^{N} X_i$, assume $E(X_i)$, $Var(X_i)$, $\exists$. same. $Var(N) < \infty$.

then
$$Var(T) = Var[E(T|N)] + E[Var(T|N)]$$
$$= Var[N \cdot E(X)] + E[N \cdot Var(X)]$$
$$= [E(X)]^2 \cdot Var(N) + Var(X) \cdot E(N).$$

$\left\{ \begin{array}{l} E(T|N) = N \cdot E(X) \\ Var(T|N=n) \\ = Var(\sum_{i=1}^{n} X_i | N=n) \\ = n \cdot Var(X) \end{array} \right\}$

§4.4.2 Prediction

predicting the value of one random variable from another r.v.

e.g.1 (X, Y), joint dist. of heights of fathers & sons.

we wish to predict the value of Y (son's) from the value of X (father's).

e.g.2. For a whole forest, (X, Y), X diameter, Y, volume of a tree. have some joint distribution. one attempts to predict Y (volume) based on diameter X.

Consider simple case: predicting Y by a constant c.

criteria, MSE = "mean squared error". $MSE = E[(Y-c)^2]$

$$MSE = E[(Y-E(Y)+E(Y)-c)^2] = E\{[Y-E(Y)]^2 + 2(Y-E(Y))(E(Y)-c) + [E(Y)-c]^2\}$$

$$= E\{[Y-E(Y)]^2\} + 2\cdot[E(Y)-c]\cdot \underbrace{E\{[Y-E(Y)]\}}_{E(Y)-E(Y)=0} + [E(Y)-c]^2$$

$$= Var(Y) + [E(Y)-c]^2.$$

find c to minimize MSE, $Var(Y)$ doesn't dependent on c. thus $c = E(Y)$.

Now, considerang predicting Y by some function $h(X)$ to minimize

$MSE = E\{[Y-h(X)]^2\}$. $h(X) = ?$ to minimize MSE.

based on $E(Y) = E[E(Y|X)]$, $MSE = E\{E\{[Y-h(X)]^2|X\}\}$

for every $X = x$, $E\{[Y-h(X)]^2|X=x\}$ is minimized when $h(x) = E[Y|X=x]$.

thus $h(X) = E[Y|X]$ minimize MSE.

or $MSE = E\{[Y-E(Y|X)+E(Y|X)-h(X)]^2\}$

$$= E\{[Y-E(Y|X)]^2 + 2\cdot[Y-E(Y|X)]\cdot[E(Y|X)-h(X)] + [E(Y|X)-h(X)]^2\}$$

$$= E\{[Y-E(Y|X)]^2\} + 2E\{[Y-E(Y|X)][E(Y|X)-h(X)]\} + E\{[E(Y|X)-h(X)]^2\}$$

2nd term $= 2\cdot E\{E[[Y-E(Y|X)][E(Y|X)-h(X)]|X]\}$

$$= 2E\{[E(Y|X)-h(X)]\cdot \underbrace{E[[Y-E(Y|X)]|X]}_{0}\}$$

3rd term, take $h(X) = E[Y|X]$, 3rd term $= 0$. which minimize MSE.

hence, $h(X)$ is an optimal prediction. $h(x) = E[Y|X=x]$

$= \int y f(y|x)dy$. But it needs joint. dist. of (X, Y). $f(x,y)$!

Best linear predictor of Y. using $\alpha+\beta X$ to predict Y.

$$MSE = E[(Y-\alpha-\beta X)^2]$$

$\mu_X = E(X)$
$\mu_Y = E(Y)$

$\sigma_X^2 = E(X^2) - \mu_X^2$

$\sigma_Y^2 = E(Y^2) - \mu_Y^2$

$\sigma_{XY} = E(XY) - E(X)E(Y)$
$= E(XY) - \mu_X \mu_Y$

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

$$MSE = E(Y^2+\alpha^2+\beta^2X^2-2\alpha Y-2\beta XY+2\alpha\beta X)$$
$$= (\mu_Y^2+\sigma_Y^2)+\alpha^2+\beta^2(\mu_X^2+\sigma_X^2)-2\cdot\alpha\cdot\mu_Y-2\beta(\sigma_{XY}+\mu_X\mu_Y) + 2\cdot\alpha\cdot\beta\cdot\mu_X$$
$$= \mu_Y^2+\sigma_Y^2+\alpha^2+\mu_X^2\beta^2+\sigma_X^2\beta^2-2\mu_Y\cdot\alpha-2\sigma_{XY}\beta-2\cdot\mu_X\mu_Y\beta+2\mu_X\alpha\beta$$

$$\frac{\partial MSE}{\partial\alpha} = 0+2\alpha-2\mu_Y+2\mu_X\beta \quad (1)$$

$$\frac{\partial MSE}{\partial\beta} = 2\mu_X^2\cdot\beta+2\sigma_X^2\beta-2\sigma_{XY}-2\mu_X\mu_Y+2\mu_X\cdot\alpha \quad (2)$$

from (1) $\alpha = \mu_Y - \mu_X\beta$.

(2). $\cancel{2\mu_X^2\cdot\beta}+2\sigma_X^2\cdot\beta-2\sigma_{XY}-\cancel{2\mu_X\mu_Y}+2\mu_X\cdot(\mu_Y-\mu_X\beta)$

$\cancel{2\mu_X\cdot\mu_Y}-\cancel{2\mu_X^2\cdot\beta}$

$$= 2\sigma_X^2\beta-2\sigma_{XY}=0 \Rightarrow \hat\beta = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\rho\cdot\sigma_X\sigma_Y}{\sigma_X^2} = \rho\cdot\frac{\sigma_Y}{\sigma_X}.$$

$$\hat\alpha = \mu_Y-\mu_X\cdot\hat\beta = \mu_Y-\mu_X\cdot\rho\frac{\sigma_Y}{\sigma_X}.$$

Answer. $\hat\alpha+\hat\beta X = (\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}\cdot\mu_X)+\rho\frac{\sigma_Y}{\sigma_X}\cdot X.$

if $\rho=0$, $\hat\alpha+\hat\beta X = \mu_Y = E(Y)$. when $X \perp\!\!\!\perp Y$.