➢ Example. If a sample of n balls is drawn without replacement from a box containing R red balls, W white balls, and $N-R-W$ blue balls. Let

X = # of red balls in the sample,

Y = # of white balls in the sample,

($n-X-Y$: # of blue balls)

then, the joint pmf of (X, Y) is

$$p_{X,Y}(x,y) = \frac{\binom{R}{x}\binom{W}{y}\binom{N-R-W}{n-x-y}}{\binom{N}{n}},$$

($x \in \{0,1,\ldots,R\}$, $y \in \{0,1,\ldots,W\}$, $0 \le x+y \le n$)

(generalization of hypergeometric ← without or with replacement (similar to the generalization from binomial to multinomial))

Find $E_Y(Y)$.

Sol. Because $Y|X=x \sim$ hypergeometric$(n-x, N-R, W)$,

$g(x) \equiv E_{Y|X}(Y|X=x) = (n-x)[W/(N-R)]$. (LNp.5-42)

Because $X \sim$ hypergeometric$(n, N, R) \Rightarrow E_X(X) = n(R/N)$, and then

$$\begin{aligned} E_Y(Y) &= E_X[E_{Y|X}(Y|X)] = E_X[g(X)] \\ &= E_X\left[(n-X)\frac{W}{N-R}\right] = \frac{W}{N-R}[n - E_X(X)] \\ &= \frac{W}{N-R}\left(n - n\frac{R}{N}\right) = n\frac{W}{N}. \end{aligned}$$

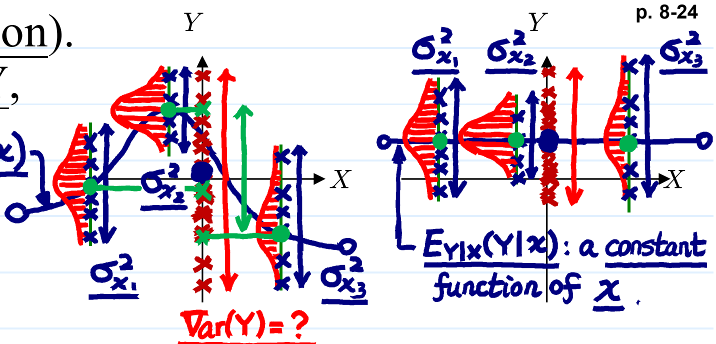
($E_Y E_{X|Y} = E_{X,Y}$)

Note that $Y \sim$ hypergeometric$(n, N, W) \Rightarrow E_Y(Y) = n(W/N)$.

⦿ Theorem (Variance Decomposition). For two random vectors $\mathbf{X}$ and $\mathbf{Y}$,

$$\begin{aligned} Var_{\mathbf{Y}}(Y_i) &= Var_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})] \\ &+ E_{\mathbf{X}}[Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]. \end{aligned}$$

("between group" variation; "within group" variation)



→ The concept behind the Thm leads to "Analysis of Variance (ANOVA)" in statistics.

Proof. $Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{x}) = E_{\mathbf{Y}|\mathbf{X}}(Y_i^2|\mathbf{x}) - [E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{x})]^2$,

(Note: $Var(Z) = E(Z-\mu)^2 = E(Z^2) - [E(Z)]^2$)

and,
$$\begin{aligned} E_{\mathbf{X}}[Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})] &= E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i^2|\mathbf{X})] - E_{\mathbf{X}}\{[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]^2\}. \end{aligned}$$

Also, $Var_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]$ ($\equiv g(\mathbf{X}) = Z$, cf. same)
$$= E_{\mathbf{X}}\{[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]^2\} - \{E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]\}^2.$$

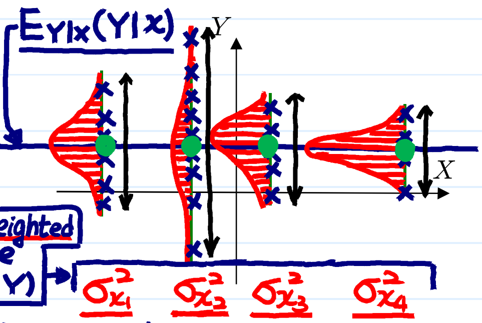
Now, (Law of Total Expectation (LNp.8-22))
$$\begin{aligned} Var_{\mathbf{Y}}(Y_i) &= E_{\mathbf{Y}}(Y_i^2) - [E_{\mathbf{Y}}(Y_i)]^2 \\ &= E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i^2|\mathbf{X})] - \{E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]\}^2 \\ &= E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i^2|\mathbf{X})] - E_{\mathbf{X}}\{[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]^2\} \\ &\quad + E_{\mathbf{X}}\{[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]^2\} - \{E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]\}^2 \\ &= E_{\mathbf{X}}[Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})] + Var_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]. \end{aligned}$$

➢ Corollary.

- $Var_{\mathbf{Y}}(Y_i) \geq E_{\mathbf{X}}[Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]$ and the equality holds if and only if

$$E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})=E_{\mathbf{Y}}(Y_i)$$

with probability one.



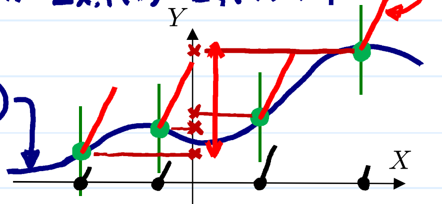
$Var_X[E_{Y|X}(Y|X)]=0 \Rightarrow g(x)=E_{Y|X}(Y|x)$ is a constant over $\mathcal{X}$ almost surely, and the constant is $E_X[g(X)]=E_XE_{Y|X}(Y|X)=E_{X,Y}(Y)=E_Y(Y)=\mu_Y$ ($E_{Y|X}(Y|X) \equiv g(X)$)

$Y=g(X)$, $Var_Y(Y) = Var_X(g(X))$

- $Var_{\mathbf{Y}}(Y_i) \geq Var_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]$ and the equality hold if and only if

$h(X)=Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})=0$ ($\Rightarrow Y_i=E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})$) ($g(X) \equiv E_{Y|X}(Y|X)$)

with probability one. ← $E_X[Var_{Y|X}(Y|X)]=0$, $\equiv h(X) \geq 0$



$\sigma^2_{x_1}=\sigma^2_{x_2}=\sigma^2_{x_3}=\sigma^2_{x_4}$ $= 0$

❖ **Reading**: textbook, Sec 7.5

Conditional Expectation and Prediction

- Problem formulation: predicting the value of a r.v. Y (random) on the basis of the observed value of a r.v. X ← fixed

  strong relation or association between X & Y ⇒ good prediction

  ➢ Data: X and Y (example?)

X	身高	雨量	地點
Y	體重	米產量	房價

➢ Statistical modeling: assigning (X, Y) a (known) joint distribution (cdf $F(x, y)$, pdf $f(x, y)$, or pmf $p(x, y)$) ← If unknown, a statistical problem

➢ Objective: predicting Y by using a function of X, i.e., (how the distribution can help?)

When $X=x$, use $g(x)$ to predict value of Y → $g(X)$ ← predictor ← e.g., for example in LNp.8-21, g transform year into cm

➢ Predictor: considering the following three groups of g's

(i) $G_1 = \{g(x) : g(x) = c,$ where $c \in \mathbb{R}\}$ ← not use the information of X

(ii) $G_2 = \{g(x) : g(x) = a + bx,$ where $a, b \in \mathbb{R}\}$

(iii) $G_3 = \{g(x) : g$ is an arbitrary function$\}$

Note. $G_1 \subset G_2 \subset G_3$. ← best in G_2 is no worse than best in G_1; best in G_3 : : : : best in G_2

➢ Question: Within each group, what is the "best" predictor? → how to choose c in G_1? : : : a,b in G_2? : : : g in G_3?

➢ Criterion: minimizing mean square error

meaning? → $\text{MSE} \equiv E_{X,Y}\{[Y - g(X)]^2\}$

Why taking expectation over X? · true value · error · predicted value

Theorem (best *constant* predictor under MSE).

G1

$$E_{X,Y}(Y-c)^2 = E_Y(Y-c)^2 \geq E_Y[Y-E_Y(Y)]^2 = Var_Y(Y)$$

(a constant function of x) (minimum)

The equality holds if and only if $c = E_Y(Y)$. (only need to know μ_Y)

Proof. $h(Y)=(Y-c)^2$: a function of Y

$E_{X,Y}[h(Y)] = E_Y[h(Y)]$

$\| \; E_Y E_{X|Y}[h(Y)|Y]$ (LNp.8-22) ($h(Y)$)

Thm in LNp.5-19

$$E_Y(Y-c)^2 = Var_Y(Y) + (\mu_Y - c)^2 \geq Var_Y(Y)$$

(cf. LNp.8-24)

12/12

Theorem (best predictor under MSE).

G3

$$E_{X,Y}[Y-g(X)]^2 \geq E_{X,Y}[Y-E_{Y|X}(Y|X)]^2 = E_X[Var_{Y|X}(Y|X)]$$

(LNp.8-24 minimum) (*)

The equality holds if and only if $g(x) = E_{Y|X}(Y|x)$. (a function of X only) (cf. (LNp.8-21))

Proof. $E_{X,Y}[Y-g(X)]^2$

$$= E_{X,Y}\{[Y-E_{Y|X}(Y|X)] + [E_{Y|X}(Y|X) - g(X)]\}^2$$

$$= E_{X,Y}[Y-E_{Y|X}(Y|X)]^2 + E_X[E_{Y|X}(Y|X)-g(X)]^2$$

$$+ 2\cdot E_{X,Y}\{[Y-E_{Y|X}(Y|X)][E_{Y|X}(Y|X)-g(X)]\}$$

($E_{X,Y}$) (a function of X only) (= 0)

last "="

$$= E_{X,Y}[Y-E_{Y|X}(Y|X)]^2 + E_X[E_{Y|X}(Y|X)-g(X)]^2$$

$$\geq E_{X,Y}[Y-E_{Y|X}(Y|X)]^2$$

(= 0 iff $g(x)=E_{Y|X}(Y|X)$)