Lecture 22

Recap:

- **Conditionals:**
  - Let $B_1, \dots, B_n$ be a partition of Ω

① Discrete on Event:

- $P_{X|B}(k|B) = \dfrac{P(X=k \cap B)}{P(B)}$
- $P_X(k) = \sum_{i=1}^{n} P_{X|B_i}(k) \cdot P(B_i)$
- $E(X|B) = \sum_k k \, P_{X|B}(k)$

② Discrete on Discrete

- $P_{X|Y}(x) = \dfrac{P_{X,Y}(x,y)}{P_Y(y)}$
- $P_X(x) = \sum_y P_{X|Y}(x) \cdot P_Y(y)$
- $E(X|Y=y) = \sum_x x \, P_{X|Y}(x)$
- $E(X) = \sum_Y E(X|Y=y) \cdot P_Y(y) = E(E(X|Y))$

③ Jointly Continuous

- $f_{X|Y}(x) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

- $f_X(x) = \int_y f_{X|Y}(x) \cdot f_Y(y)\, dy$

- $E(X|Y) = \int_x x\, f_{X|Y}(x)\, dx$

- $E(X) = \int_y E(X|Y=y) \cdot f_Y(y)\, dy = E(E(X|Y))$

10.3

Multivariate Conditionals

Def

- Let $X_1, \dots, X_{n+1}$ be $n+1$ discrete RV
  The Conditional dist. $(X_1, \dots, X_n) | X_{n+1}$ is:

$$P_{X_1 \dots X_n | X_{n+1}}(k_1, \dots, k_n | x_{n+1}) = \frac{P_{X_1, \dots, X_{n+1}}(k_1, \dots, k_{n+1})}{P_{X_{n+1}}(x_{n+1})}$$

- Let $X_1, \dots, X_{n+1}$ be $n+1$ jointly continuous RV
  The Conditional dist. $(X_1, \dots, X_n) | X_{n+1}$ is:

$$f_{X_1 \dots X_n | X_{n+1}}(x_1, \dots, x_n | x_{n+1}) = \frac{f_{X_1, \dots, X_{n+1}}(x_1, \dots, x_{n+1})}{f_{X_{n+1}}(x_{n+1})}$$

- Expectation:

$E(g(X_1, \dots, X_n) | X_{n+1})$:

- $\underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n \text{ int.}} g(x_1, \dots, x_n) f_{x_1, \dots, x_n | x_{n+1}}(x_1, \dots, x_n | x_{n+1}) \, dx_1 \dots dx_n$

or

- $\sum_{x_1, \dots, x_n} g(x_1, \dots, x_n) P_{x_1, \dots, x_n | x_{n+1}}(x_1, \dots, x_n)$

$\Rightarrow$ Properties:

- Linearity: $E(aX + b | Y) = aE(X|Y) + b$

Lemma

- 2 RV X, Y are independent

$$\Longleftrightarrow$$

1) (Discrete) $P_{X|Y}(x) = P_X(x) \ \forall \ y$ s.t. $p(y) > 0$

<u>or</u>

2) (Continuous) $f_{X|Y}(x) = f_X(x) \ \forall \ y$ s.t. $f_Y(y) > 0$

- If X, Y are independent then

$$E(g(X) | Y) = E(g(X))$$

and

$$E(g(Y) | X) = E(g(Y))$$

Conditional Expectation for Best Predictor

$\Rightarrow$ Goal use X to predict Y with some function $h(X)$.

- Mean-Square-Error:

$$E[(Y-h(X))^2]$$

Metric to measure how good a predictor is $\Rightarrow$ lower is better

Thm

Let X, Y be 2 RV.
Then $\forall$ functions h:

$$E[(Y-h(X))^2] \geq E[(Y-E(Y|X))^2]$$

$\Rightarrow E(Y|X)$ is the best predictor of Y using X

$\Rightarrow$ Equality only occurs for $h(X) = E(X|Y)$

Prf

$$E[(Y-h(X))^2]$$

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

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

$$+ E[(E(Y|X)-h(X))^2]$$

$$\star \quad E[(Y-E(Y|X))(E(Y|X)-h(X))]$$

$$= \iint (y-E(Y|X))(E(Y|X)-h(X)) \cdot f_{X,Y}(x,y)\,dy\,dx$$

$$= \int (E(Y|X)-h(X)) \left[\int (y-E(Y|X)) f_{Y|X}(y)\,dy\right] f_X(x)\,dx$$

$$\int (y-E(Y|X)) f_{Y|X}(y)\,dy$$

$$= \int y\, f_{Y|X}\,dy - E(Y|X) \int f_{Y|X}\,dy$$

$$= E(Y|X) - E(Y|X) = 0$$

$\Rightarrow$ So :

$$E[(y-h(x))^2]$$

$$= E[(y-E(y|x)+E(y|x)-h(x))^2]$$

$$= E[(y-E(y|x))^2] + 2\underbrace{E[(y-E(y|x))(E(y|x)-h(x))]}_{\ast} \rightarrow 0$$

$$+ E[(E(y|x)-h(x))^2]$$

$$= E[(y-E(y|x))^2] + E[(E(y|x)-h(x))^2]$$

So $E[(y-E(y|x))^2] \leq E[(y-h(x))^2]$