

## Lecture 22

### Recap:

- Conditionals:

- Let  $B_1, \dots, B_n$  be a partition of  $\Omega$

- ① Discrete on Event:

- $P_{X|B}(k|B) = \frac{P(X=k \cap B)}{P(B)}$

- $P_X(k) = \sum_{i=1}^n P_{X|B_i}(k) \cdot P(B_i)$

- $E(X|B) = \sum_k k P_{X|B}(k)$

- ② Discrete on Discrete

- $P_{X|Y}(x) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$

- $P_X(x) = \sum_y P_{X|Y}(x) \cdot P_Y(y)$

- $E(X|Y=y) = \sum_x x P_{X|Y}(x)$

- $E(X) = \sum_y E(X|Y=y) \cdot P_Y(y) = E(E(X|Y))$

### ③ Jointly Continuous

- $f_{X|Y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
- $f_X(x) = \int_Y f_{X|Y}(y) \cdot f_Y(y) dy$
- $E(X|Y) = \int_X x f_{X|Y}(x) dx$
- $E(X) = \int_Y E(X|Y=y) \cdot f_Y(y) dy = E(E(X|Y))$

## 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:

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

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

or

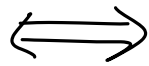
$$\circ \sum_{x_1, \dots, x_n} g(x_1, \dots, x_n) P_{x_1, \dots, x_n | x_{n+1}}(x_1, \dots, x_n)$$

⇒ Properties:

$$\circ \text{Linearity: } E(aX + b | Y) = aE(X | Y) + b$$

## Lemma

- 2 RV  $X, Y$  are independent



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

or

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

- 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

⇒ 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 ⇒ lower is better

Thm

Let  $X, Y$  be 2 RV.

Then  $\forall$  functions  $h$ :

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

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

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

Prf

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

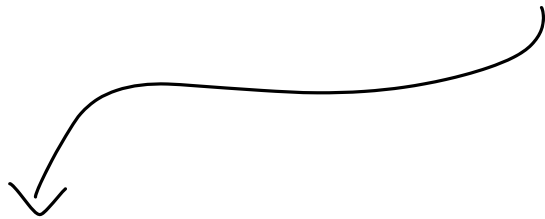
$$= E[(Y - E(Y|X) + E(Y|X) - h(X))^2]$$

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

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

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

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



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

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

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

$\Rightarrow$  So :

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

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

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

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

$$\text{So } E[(Y - E(Y|X))^2] \leq E[(Y - h(X))^2]$$