Lecture 21: Conditional Expectation

Statistics 251

Yi Sun and Zhongjian Wang

Department of Statistics
The University of Chicago

Lecture Outline

Formula for computation

Compute Expectation by Conditioning

Conditional Expectation is the best preditor

Where are we?

Formula for computation

Compute Expectation by Conditioning

Conditional Expectation is the best preditor

Conditional Expectation for Discrete r.v.

Recall that if X and Y are jointly discrete random variables, then the conditional probability mass function of X, given that $Y = y$, is defined, for all y such that $P\{Y = y\} > 0$, by

$$p_{X|Y}(x \mid y) = P\{X = x \mid Y = y\} = \frac{p(x, y)}{p_Y(y)}$$

Conditional Expectation for Discrete r.v.

Recall that if X and Y are jointly discrete random variables, then the conditional probability mass function of X, given that $Y = y$, is defined, for all y such that $P\{Y = y\} > 0$, by

$$p_{X|Y}(x \mid y) = P\{X = x \mid Y = y\} = \frac{p(x, y)}{p_Y(y)}$$

Base on this, we know $p_{X|Y}\{\cdot \mid Y = y\}$ is a distribution function. Now how to define $E[X \mid Y = y]$?

Conditional Expectation for Discrete r.v.

Recall that if X and Y are jointly discrete random variables, then the conditional probability mass function of X, given that $Y = y$, is defined, for all y such that $P\{Y = y\} > 0$, by

$$p_{X|Y}(x \mid y) = P\{X = x \mid Y = y\} = \frac{p(x, y)}{p_Y(y)}$$

Base on this, we know $p_{X|Y}\{\cdot \mid Y = y\}$ is a distribution function. Now how to define $E[X \mid Y = y]$? The conditional expectation of X given that $Y = y$, for all values of y such that $p_Y(y) > 0$, by

$$\begin{aligned} E[X \mid Y = y] &= \sum_x x P\{X = x \mid Y = y\} \\ &= \sum_x x p_{X|Y}(x \mid y) \end{aligned}$$

Conditional Expectation for Continuous r.v.

Similarly, let us recall that if X and Y are jointly continuous with a joint probability density function $f(x, y)$, then the conditional probability density of X, given that $Y = y$, is defined, for all values of y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

Conditional Expectation for Continuous r.v.

Similarly, let us recall that if X and Y are jointly continuous with a joint probability density function $f(x, y)$, then the conditional probability density of X, given that $Y = y$, is defined, for all values of y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

Again, we know $p_{X|Y}\{\cdot \mid Y = y\}$ is a distribution function. Now how to define $E[X \mid Y = y]$?

Conditional Expectation for Continuous r.v.

Similarly, let us recall that if X and Y are jointly continuous with a joint probability density function $f(x, y)$, then the conditional probability density of X, given that $Y = y$, is defined, for all values of y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

Again, we know $p_{X|Y}\{\cdot \mid Y = y\}$ is a distribution function. Now how to define $E[X \mid Y = y]$? The conditional expectation of X, given that $Y = y$, by

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx$$

provided that $f_Y(y) > 0$

Example

Suppose that the joint density of X and Y is given by

$$f(x,y) = \frac{e^{-x/y}e^{-y}}{y} \quad 0 < x < \infty, 0 < y < \infty$$

Compute $E[X \mid Y = y]$.

Where are we?

Formula for computation

Compute Expectation by Conditioning

Conditional Expectation is the best preditor

Formula

$$E[X] = E[E[X \mid Y]]$$

What does it mean?

Formula

$$E[X] = E[E[X \mid Y]]$$

What does it mean?
If Y is a discrete random variable,

$$E[X] = \sum_y E[X \mid Y = y] P\{Y = y\}$$

Formula

$$E[X] = E[E[X \mid Y]]$$

What does it mean?
If Y is a discrete random variable,

$$E[X] = \sum_{y} E[X \mid Y = y] P\{Y = y\}$$

whereas if Y is continuous random variable,

$$E[X] = \int_{-\infty}^{\infty} E[X \mid Y = y] f_Y(y) dy$$

Example: Miner Travel

A miner is trapped in a mine containing 3 doors.

- The first door leads to a tunnel that will take him to safety after 3 hours of travel.
- The second door leads to a tunnel that will return him to the mine after 5 hours of travel.
- The third door leads to a tunnel that will return him to the mine after 7 hours.

If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

Example

Suppose that X and Y are independent continuous random variables having densities f_X and f_Y, respectively. Compute $P\{X < Y\}$

Conditional Variance

$$\text{Var}(X) = E[\text{Var}(X \mid Y)] + \text{Var}(E[X \mid Y])$$

Conditional Variance

$$\text{Var}(X) = E[\text{Var}(X \mid Y)] + \text{Var}(E[X \mid Y])$$

Proof can be found on Ross Book and is left as homework.

Where are we?

Formula for computation

Compute Expectation by Conditioning

Conditional Expectation is the best preditor

Derivation

- Assume a random variable X is observed.
- We need to predict the value of a second random variable Y.
- Let $g(X)$ denote the predictor for Y.

Derivation

- Assume a random variable X is observed.
- We need to predict the value of a second random variable Y.
- Let $g(X)$ denote the predictor for Y.
- We would like to show $g(X) = E[Y \mid X]$, is the best possible predictor.

Derivation

- Assume a random variable X is observed.
- We need to predict the value of a second random variable Y.
- Let $g(X)$ denote the predictor for Y.
- We would like to show $g(X) = E[Y \mid X]$, is the best possible predictor.

How to determine a predictor is good (or best)?

Derivation

- Assume a random variable X is observed.
- We need to predict the value of a second random variable Y.
- Let $g(X)$ denote the predictor for Y.
- We would like to show $g(X) = E[Y \mid X]$, is the best possible predictor.

How to determine a predictor is good (or best)?
One possible criterion for closeness is to choose g so as to minimize $E\left[(Y - g(X))^2\right]$.

Proof

In detail, we would like to prove,

$$E\left[(Y - g(X))^2\right] \geq E\left[(Y - E[Y \mid X])^2\right].$$

Proof

In detail, we would like to prove,

$$E\left[(Y - g(X))^2\right] \geq E\left[(Y - E[Y \mid X])^2\right].$$

$$\begin{aligned}
E\left[(Y - g(X))^2 \mid X\right] =& E\left[(Y - E[Y \mid X] + E[Y \mid X] - g(X))^2 \mid X\right] \\
=& E\left[(Y - E[Y \mid X])^2 \mid X\right] \\
& + E\left[(E[Y \mid X] - g(X))^2 \mid X\right] \\
& + 2E[(Y - E[Y \mid X])(E[Y \mid X] - g(X)) \mid X]
\end{aligned}$$

Proof

In detail, we would like to prove,

$$E\left[(Y - g(X))^2\right] \geq E\left[(Y - E[Y \mid X])^2\right].$$

$$\begin{aligned}
E\left[(Y - g(X))^2 \mid X\right] =& E\left[(Y - E[Y \mid X] + E[Y \mid X] - g(X))^2 \mid X\right] \\
=& E\left[(Y - E[Y \mid X])^2 \mid X\right] \\
&+ E\left[(E[Y \mid X] - g(X))^2 \mid X\right] \\
&+ 2E[(Y - E[Y \mid X])(E[Y \mid X] - g(X)) \mid X]
\end{aligned}$$

However, given X, $E[Y \mid X] - g(X)$, being a function of X, can be treated as a constant. Thus,

$$\begin{aligned}
&E[(Y - E[Y \mid X])(E[Y \mid X] - g(X)) \mid X] \\
&\quad = (E[Y \mid X] - g(X))E[Y - E[Y \mid X] \mid X] \\
&\quad = (E[Y \mid X] - g(X))(E[Y \mid X] - E[Y \mid X]) \\
&\quad = 0
\end{aligned}$$