Math 40 Probability and Statistical Inference

Winter 2021

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Lecture 9-1: Conditional Density (3.3)

Conditional Density (3.3)

Doing data science is related to estimating a conditional density. That is, by knowing about a random variable Y, we want to make a better (in some sense) estimation of X. This process can be done by knowing the conditional density of X by Y

$$p(X|Y)$$

Conditional Density (3.3)

Example 3.19 The exponential distribution is memoryless. Let random variable X have an exponential distribution. Prove that $Pr(X > s + t|X > s) = Pr(X > t)$, where s and t are positive numbers.

Conditional Density (3.3)

Example 3.19 The exponential distribution is memoryless. Let random variable X have an exponential distribution. Prove that $Pr(X > s + t|X > s) = Pr(X > t)$, where s and t are positive numbers.

Solution

$$Pr(X > s+t|X > s) = \frac{Pr(X > s + t \text{ and } X > s)}{Pr(X > s)} = \frac{Pr(X > s + t)}{Pr(X > s)}$$

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = Pr(X > t)$$

Conditional Density (3.3)

Example 3.22 Bayes formula. The joint distribution of (X, Y) is defined through the conditional density of Y given X and marginal density of X. Find the conditional density of X given Y.

$$f_{X|Y=y}(x) = \frac{f_{Y|X=x}(y)f_X(x)}{f_Y(y)}$$

$$= \frac{f_{Y|X=x}(y)f_X(x)}{\int f_{Y|X=x}(y)f_X(x)dx}$$

Conditional mean and variance (3.3.1)

A regression describes a relationship between two random variables, X and Y. The conditional mean of Y given X, $E[Y|X]$ is related to a regression function, say $\mu(x)$, (under some condition) in describing Y given X.

- The conditional mean of Y given X

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

- The conditional variance of Y given X

$$Var(Y|X = x) = E[(Y - \mu(x))^2 | X = x)$$

$$= E[Y^2|X = x) - E^2[Y|X = x]$$

Note If X and Y are independent, there are simple structures for the conditional mean and variance (they are constants; check Theorem 3.24).

Conditional mean and variance (3.3.1)

Example 3.26 *Regression is an optimal predictor* Show that the conditional mean/regression, $E[Y|X = x] = \mu(x)$, is the optimal predictor of Y given $X = x$.
The minimum is attained when $c = \mu(x) = E[Y|X = x]$.

Conditional mean and variance (3.3.1)

Example 3.26 *Regression is an optimal predictor* Show that the conditional mean/regression, $E[Y|X=x]=\mu(x)$, is the optimal predictor of Y given $X=x$.

Solution For an unknown value c depending on x (thus c is in fact a function of x), we have

$$E[(Y-c)^2|X=x]=E[(Y-\mu(x)+\mu(x)-c)^2|X=x]$$

$$=E[(Y-\mu(x))^2|X=x]+2E[(Y-\mu(x))(\mu(x)-c)]+(\mu(x)-c)^2$$

$$=E[(Y-\mu(x))^2|X=x]+(\mu(x)-c)^2$$

The minimum is attained when $c=\mu(x)=E[Y|X=x]$.

Conditional mean and variance (3.3.1)

Example 3.26

- What is the meaning of this example?
- It says $c = E[Y|X = x]$ minimizes $E[(Y - c)^2|X = x]$.
- If we interpret c as a predictor of Y given X, $Y - c$ is the error of c.
- $E[(Y - c)^2|X = x]$ is the expected square of the error given X. Thus, the conditional mean $E[Y|X = x]$ has the minimum expected squared error.

Conditional mean and variance (3.3.1)

Here is another question. What is your predictor c that minimizes the expected squared error $E[(Y-c)^2]$? (there is no conditioning). Surprisingly, your predictor c must be $E[Y|X]$.

The left hand side is minimized if the inner term of the right hand side is minimized for each X.

Conditional mean and variance (3.3.1)

Here is another question. What is your predictor c that minimizes the expected squared error $E[(Y-c)^2]$? (there is no conditioning). Surprisingly, your predictor c must be $E[Y|X]$.

Why? Theorem 3.27 says $E[Y] = E_X[E[Y|X]]$. The same rule applies to

$$E[(Y-c)^2] = E_X[E[(Y-c)^2|X]]$$

The left hand side is minimized if the inner term of the right hand side is minimized for each X.

Take-home message Throughout the course, we will focus on calculating the conditional expected value of Y, $E[Y|X=x]$ (aka regression function), as the predictor of Y with the minimum expected squared error.

Conditional mean and variance (3.3.1)

Here is another question. What is your predictor c that minimizes the expected squared error $E[(Y-c)^2]$? (there is no conditioning). Surprisingly, your predictor c must be $E[Y|X]$.
Why? Theorem 3.27 says $E[Y] = E_X[E[Y|X]]$. The same rule applies to

$$E[(Y-c)^2] = E_X[E[(Y-c)^2|X]]$$

The left hand side is minimized if the inner term of the right hand side is minimized for each X.
Take-home message Throughout the course, we will focus on calculating the conditional expected value of Y, $E[Y|X=x]$ (aka regression function), as the predictor of Y with the minimum expected squared error.
Note If X and Y are independent, the conditional expected value $E[Y|X=x]$ is a constant. That is, any information of X does not provide any new information for Y.

Conditional mean and variance (3.3.1)

- Regarding the variance, we have the following result (called **Variance Decomposition**)

$$Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$$

  where $Var(E(Y|X))$ is called **explained** variance while $E(Var(Y|X))$ is called **unexplained** (or residual) variance.
- The unexplained variance is related to the error using $E[Y|X]$ as a predictor of Y given X.
- Coefficient of determination (or variation)

$$\rho^2 = \frac{Var(E(Y|X))}{Var(Y)}$$

  represents the proportion of the variation explained by the predictor.

Conditional mean and variance (3.3.1)

Example 3.35 Bivariate Bernoulli distribution. Let Y and X be two Bernoulli random variables, distributed as specified by the 2×2 elementary probability table:

$$Pr(X=0, Y=0) = p_{00}, \quad Pr(X=0, Y=1) = p_{01}$$

$$Pr(X=1, Y=0) = p_{10}, \quad Pr(X=1, Y=1) = p_{11}$$

Find the conditional mean $E[Y|X=x]$ where $x=0$ or 1.

Conditional mean and variance (3.3.1)

Example 3.35 Bivariate Bernoulli distribution. Let Y and X be two Bernoulli random variables, distributed as specified by the 2×2 elementary probability table:

$$Pr(X = 0, Y = 0) = p_{00}, \quad Pr(X = 0, Y = 1) = p_{01}$$

$$Pr(X = 1, Y = 0) = p_{10}, \quad Pr(X = 1, Y = 1) = p_{11}$$

Find the conditional mean $E[Y|X = x]$ where $x = 0$ or 1.

Solution

$$E[Y|X = 1] = \frac{p_{11}}{p_{10} + p_{11}}$$

$$E[Y|X = 0] = \frac{p_{01}}{p_{00} + p_{01}}$$

Conditional mean and variance (3.3.1)

Example 3.35 Let's do some explicit calculations.
Case 1 If

$$Pr(X = 0, Y = 0) = p_{00} = 1/4, \quad Pr(X = 0, Y = 1) = p_{01} = 1/4$$

$$Pr(X = 1, Y = 0) = p_{10} = 1/4, \quad Pr(X = 1, Y = 1) = p_{11} = 1/4,$$

do you believe that knowing the outcome of X is useful to know about Y?

Conditional mean and variance (3.3.1)

Example 3.35 Let's do some explicit calculations.
Case 1 If

$$Pr(X=0, Y=0) = p_{00} = 1/4, \quad Pr(X=0, Y=1) = p_{01} = 1/4$$

$$Pr(X=1, Y=0) = p_{10} = 1/4, \quad Pr(X=1, Y=1) = p_{11} = 1/4,$$

do you believe that knowing the outcome of X is useful to know about Y?

$$E[Y|X=1] = \frac{p_{11}}{p_{10}+p_{11}} = \frac{1}{2}$$

$$[E[Y|X=0] = \frac{p_{01}}{p_{00}+p_{01}} = \frac{1}{2}$$

Conditional mean and variance (3.3.1)

Example 3.35 Let's do some explicit calculations.
Case 2 If

$$Pr(X = 0, Y = 0) = p_{00} = 5/12, \quad Pr(X = 0, Y = 1) = p_{01} = 1/12$$

$$Pr(X = 1, Y = 0) = p_{10} = 1/12, \quad Pr(X = 1, Y = 1) = p_{11} = 5/12,$$

do you believe that knowing the outcome of X is useful to know about Y?

Conditional mean and variance (3.3.1)

Example 3.35 Let's do some explicit calculations.
Case 2 If

$$Pr(X=0,Y=0)=p_{00}=5/12, \quad Pr(X=0,Y=1)=p_{01}=1/12$$

$$Pr(X=1,Y=0)=p_{10}=1/12, \quad Pr(X=1,Y=1)=p_{11}=5/12,$$

do you believe that knowing the outcome of X is useful to know about Y?

$$E[Y|X=1]=\frac{p_{11}}{p_{10}+p_{11}}=\frac{5}{6}$$
$$E[Y|X=0]=\frac{p_{01}}{p_{00}+p_{01}}=\frac{1}{6}$$

Conditional mean and variance (3.3.1)

Example 3.35 For the case 1 and 2, what are the coefficients of variation?

Conditional mean and variance (3.3.1)

Example 3.35 For the case 1 and 2, what are the coefficients of variation?

Case 1: $Var(Y) = 1/4$, $Var(E(Y|X)) = 0$. Thus, $\rho^2 = 0$

Conditional mean and variance (3.3.1)

Example 3.35 For the case 1 and 2, what are the coefficients of variation?
Case 1: $Var(Y) = 1/4$, $Var(E(Y|X)) = 0$. Thus, $\rho^2 = 0$
Case 2: $Var(Y) = 1/4$, $Var(E(Y|X)) = 1/9$. Thus, $\rho^2 = 4/9$.

Mixture distribution and Bayesian statistics (3.3.2)

- Let $f(x)$ be the density of the height of all people at Dartmouth.
- Also, let $f(x|Y = w)$ be the density of the height of all women at Dartmouth while $f(x|Y = m)$ is the density of the height of all men at Dartmouth.
- Thus, the two random variables X and Y are height and gender.
- If we let $p = Pr(Y = w)$, then from the law of total probability, we have

$$f(x) = pf(x|Y = w) + (1 - p)f(x|Y = m),$$

which is a mixture density.

Mixture distribution and Bayesian statistics (3.3.2)

By knowing the height of a person, is it possible to classify the gender of the person?

Mixture distribution and Bayesian statistics (3.3.2)

By knowing the height of a person, is it possible to classify the gender of the person?
Not perfect, but there is a way. It is called **Bayesian classifier**.

$$Pr(Y = w|X = x) = \frac{pPr(X = x|Y = w)}{pPr(X = x|Y = w) + (1 - p)Pr(X = x|Y = m)}$$

Similarly,

$$Pr(Y = m|X = x) = \frac{(1 - p)Pr(X = x|Y = m)}{pPr(X = x|Y = w) + (1 - p)Pr(X = x|Y = m)}$$

Mixture distribution and Bayesian statistics (3.3.2)

Example 3.37 $f(x|Y = w) = \phi(\frac{x-64}{3})$ and $f(x|Y = m) = \phi(\frac{x-70}{4})$. Also $p = 0.5$.
If the height of a person is 68, the probability of being a man is

$$Pr(Y = m|X = 68) = \frac{\frac{1}{2}\phi(\frac{68-70}{4})}{\frac{1}{2}\phi(\frac{68-70}{4}) + \frac{1}{2}\phi(\frac{68-64}{3})} = 0.62$$

If $X = 67.1$, the probability of being a man is 0.5 (check the textbook regarding how to get this value). If $X > 67.1$, we expect that the person is a man.

Homework

Read sections 3.3.3 and 3.3.4.