Conditional Distribution

Le Wang

Motivation and Basic Definition

Motivation

We are interested in whether or not the relationship exists. But more important, we are interested in predictions.

Given a value of X, what will Y be?

We will discuss the discrete case first, which is completely nonparametric and model-free.

Classification Problems: A Numerical Example

ID	X	Y
1	1	0
2	1	0
3	1	0
4	2	1
5	2	0
6	2	1
7	2	1

Questions: What are your predictions of Y when $X = 1, 2$, respectively?

Definition

Definition. Conditional Distribution is a probability distribution for a **sub-population**.

That is, a conditional probability distribution describes the probability that a randomly selected person from a sub-population has the one characteristic of interest.

$$\Pr[Y|X]$$

How to calculate conditional distribution?

How to calculate conditional distribution?

In a sample

$$\Pr[Y = y \mid X = x] = \frac{\sum \mathbb{I}[Y = y, X = x]}{\sum \mathbb{I}[X = x]}$$

How to calculate conditional distribution?

In a sample

$$
\begin{aligned}
\Pr[Y = y \mid X = x] &= \frac{\sum \mathbb{I}[Y = y, X = x]}{\sum \mathbb{I}[X = x]} \\
&= \frac{N \cdot \frac{1}{N} \sum \mathbb{I}[Y = y, X = x]}{N \cdot \frac{1}{N} \sum \mathbb{I}[X = x]}
\end{aligned}
$$

How to calculate conditional distribution?

In a sample

$$
\begin{aligned}
\Pr[Y = y \mid X = x] &= \frac{\sum \mathbb{I}[Y = y, X = x]}{\sum \mathbb{I}[X = x]} \\
&= \frac{N \cdot \frac{1}{N} \sum \mathbb{I}[Y = y, X = x]}{N \cdot \frac{1}{N} \sum \mathbb{I}[X = x]} \\
&= \frac{\frac{1}{N} \sum \mathbb{I}[Y = y, X = x]}{\frac{1}{N} \sum \mathbb{I}[X = x]}
\end{aligned}
$$

How to calculate conditional distribution?

In a sample

$$
\begin{aligned}
\Pr[Y = y \mid X = x] &= \frac{\sum \mathbb{I}[Y = y, X = x]}{\sum \mathbb{I}[X = x]} \\
&= \frac{N \cdot \frac{1}{N} \sum \mathbb{I}[Y = y, X = x]}{N \cdot \frac{1}{N} \sum \mathbb{I}[X = x]} \\
&= \frac{\frac{1}{N} \sum \mathbb{I}[Y = y, X = x]}{\frac{1}{N} \sum \mathbb{I}[X = x]} \\
&= \frac{\Pr[Y = y, X = x]}{\Pr[X = x]}
\end{aligned}
$$

Joint, Margianl, and Conditional Dists

$$\Pr[Y|X] = \frac{\Pr[Y, X]}{\Pr[X]}$$

$$p(y|x) = \frac{p(x, y)}{p(x)}$$

Joint, Margianl, and Conditional Dists

$$\Pr[Y|X] = \frac{\Pr[Y,X]}{\Pr[X]}$$

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$$p(x,y) = p(y \mid x)p(x)$$

Note: We will also use the later equality a lot later.

Definition

Conditional distribution of Y given X is nothing but the distribution of Y for the subsample when $X = x$.

What is the probability of having $Y = 1$ when $X = 1$?

Table 2: Joint Prob

X\Y	0	1
0	0.3	0.4
1	0.2	0.1

joint and conditional Distributions

Table 3: Joint Prob

X\Y	0	1	$p(x)$
1	0.2	0.1	0.3
1	0.2/0.3	0.1/0.3	

Extensions to Continuous Variable

Definition (Greene, Appendix B.8)

$$f(y \mid x) = \frac{f(x, y)}{f(x)}$$

Extensions to Continuous Variable

Definition (Greene, Appendix B.8)

$$f(y \mid x) = \frac{f(x, y)}{f(x)}$$

$$f(x, y) = f(y \mid x) f(x)$$

Extensions to More than Two Variables

Discrete Variables:

$$\Pr[X \& Y | Z] = \frac{\Pr[X \& Y \& Z]}{\Pr[Z]}$$

$$\Pr[Y | X, Z] = \frac{\Pr[X \& Y \& Z]}{\Pr[X \& Z]}$$

Continuous Variables:

$$f(x, y | z) = \frac{f(x, y, z)}{f(z)}$$

$$f(y | x, z) = \frac{f(x, y, z)}{f(x, z)}$$

Extensions to More than Two Variables

$$f(\text{outcome} \mid \text{predictors}) = \frac{f(\text{outcome,predictors})}{f(\text{predictors})}$$

Alternative Definitions (linking to means)

$$Pr[Y = y] = \mathbb{E}[\mathbb{I}(Y = y)]$$

$$Pr[Y \leq y] = \mathbb{E}[\mathbb{I}(Y \leq y)]$$

Alternative Definitions (linking to means)

$$Pr[Y = y \mid X] = \mathbb{E}[\mathbb{I}(Y = y)]$$

$$Pr[Y \leq y] = \mathbb{E}[\mathbb{I}(Y \leq y)]$$

Alternative Definitions (linking to means)

$$Pr[Y = y \mid X] = \mathbb{E}[\mathbb{I}(Y = y) \mid X]$$

$$Pr[Y \leq y] = \mathbb{E}[\mathbb{I}(Y \leq y)]$$

Alternative Definitions (linking to means)

$$Pr[Y = y \mid X] = \mathbb{E}[\mathbb{I}(Y = y) \mid X]$$

$$Pr[Y \leq y \mid X] = \mathbb{E}[\mathbb{I}(Y \leq y)]$$

Alternative Definitions (linking to means)

$$Pr[Y = y \mid X] = \mathbb{E}[\mathbb{I}(Y = y) \mid X]$$

$$Pr[Y \leq y \mid X] = \mathbb{E}[\mathbb{I}(Y \leq y) \mid X]$$

Applications

- Transition Matrix and Income Mobility

$$\Pr[Y_t|Y_{t-1}] = \frac{\Pr[Y_t, Y_{t-1}]}{\Pr[Y_{t-1}]}$$

- Predictions for discrete variables (also called **classification problem** in machine learning)

1. Whether or not an email is a spam
2. Whether or not an individual is an Asian.

- Partial Effects: Estimation of the impact of X on Y

Special Conditional Distribution

1. Multivariate Normal Distribution
2. Truncated Distribution.

Special Conditional Distribution (I)

Greene, Appendix B. on Multivariate Normal Distribution

Let

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right)$$

where Σ_{XX} is positive definite. Then,

Special Conditional Distribution (I)

Greene, Appendix B. on Multivariate Normal Distribution

Let

$$\left(\begin{array}{c} \mathbf{X} \\ \mathbf{Y} \end{array} \right) \sim \mathcal{N}\left(\left(\begin{array}{c} \mu_X \\ \mu_Y \end{array} \right), \left(\begin{array}{cc} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{array} \right) \right)$$

where Σ_{XX} is positive definite. Then,

$$Y|X \sim \mathcal{N}(\mu_{Y|X}, \Sigma_{Y|X})$$

Special conditional Distributions (II): Truncated Distribution

Chapter 19

Suppose that we only observe individuals who receive wage offers greater than L will enter the labor force. What is the density function for wages among the sample of workers?

Theorem 19.1

$$\begin{aligned} f^*(y) &= f(y|Y > L) \\ &= \frac{f(y)}{\Pr[Y > L]} \\ &= \frac{f(y)}{[1 - F*(L)]} \end{aligned}$$

Intuition: The original density function is no longer a proper density function since it does not integrate to one. Then, how can we reflect this?

We inflate the density by $\frac{1}{\Pr[Y>L]}$!

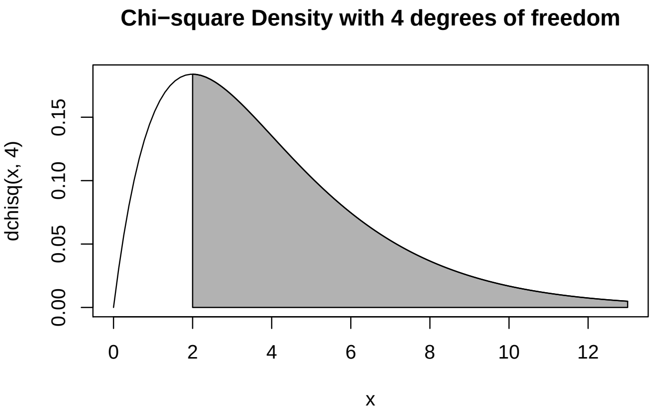
In your **homework**: you will be asked to show the result above holds. Below is some hint.

In your **homework**: you will be asked to show the result above holds. Below is some hint.

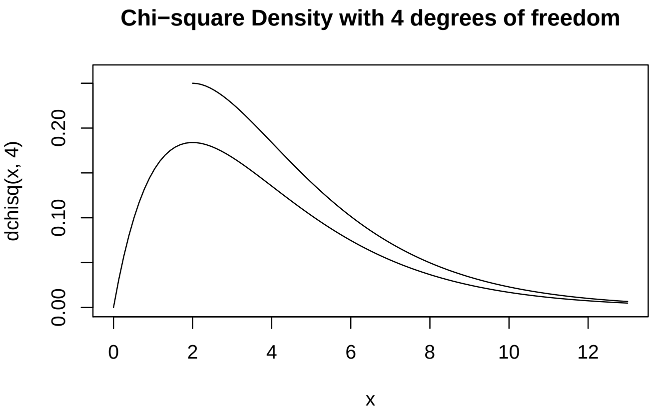
$$f(y \mid y > L) = \frac{d}{dy} F(y \mid Y > L)$$

In your **homework**: you will be asked to show the result above holds. Below is some hint.

$$
\begin{aligned}
f(y \mid y > L) &= \frac{d}{dy} F(y \mid Y > L) \\
&= \frac{d}{dy} \frac{\Pr[Y \leq y, Y > L]}{\Pr[Y > L]}
\end{aligned}
$$



Chi−square Density with 4 degrees of freedom
dchisq(x, 4)
0.00
0.05
0.10
0.15
0
2
4
6
8
10
12
x


Chi−square Density with 4 degrees of freedom
dchisq(x, 4)
0.00
0.10
0.20
0
2
4
6
8
10
12
x

Using the definition, we can easily derive the expectation of the trancated variable later:

$$\mathbb{E}[y \mid y > L] = \int_L^{\infty} yf(y \mid y > L)dy$$

Question: Which one is bigger, $\mathbb{E}[y]$ or $\mathbb{E}[y \mid y > L]$?

A special truncated distribution: **Truncated Normal**

Suppose that $Y \sim \mathcal{N}(\mu, \sigma^2)$. Then the following results hold

$$\Pr[Y > L] = 1 - \Phi(\frac{L - \mu}{\sigma}) = 1 - \Phi(\alpha)$$

A special truncated distribution: **Truncated Normal**

Suppose that $Y \sim \mathcal{N}(\mu, \sigma^2)$. Then the following results hold

$$\begin{aligned}
\Pr[Y > L] &= 1 - \Phi(\frac{L - \mu}{\sigma}) = 1 - \Phi(\alpha) \\
f(y|Y > L) &= \frac{f(y)}{1 - \Phi(\alpha)} \\
&= \frac{\frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})}{1 - \Phi(\alpha)}
\end{aligned}$$

where $\alpha = \frac{L-\mu}{\sigma}$, $\phi(\cdot)$ ($\Phi(\cdot)$) is the **standard** normal density (distribution) function.

Later we will use this density function to derive moments of moments of the truncated normal variables.

Conditional Distribution and Related Concepts and Results

Conditional Distribution and Related Concepts and Results

1. Conditional distribution and **indepdence**
2. Conditional distribution and **law of total probability** (discrete and continuous cases)
3. Conditional distribution and **Bayes' Rule**
4. Conditional distribution and **Skorohod Representation**

Conditional Distribution and Independence

$$p(x, y) = p(x)p(y)$$

Conditional Distribution and Independence

$$p(x, y) = p(x)p(y)$$

$$p(y|x) = \frac{p(y, x)}{p(x)}$$

Conditional Distribution and Independence

$$p(x, y) = p(x)p(y)$$

$$p(y|x) = \frac{p(y, x)}{p(x)}$$
$$= \frac{p(x)p(y)}{p(x)}$$

Conditional Distribution and Independence

$$p(x, y) = p(x)p(y)$$

$$\begin{aligned} p(y|x) &= \frac{p(y, x)}{p(x)} \\ &= \frac{p(x)p(y)}{p(x)} \\ &= p(y) \end{aligned}$$

Independence implies that conditional distribution is **marginal** distribution!

Intuitively, NO predictive power at all! as it should be for independent variables!

Equilvalent Definitions of Dependence

For $Y = 0, 1$ and $X = 0, 1$, there are also alternative definitions of independence. These concepts are sometimes called **risk differences**, **risk ratio**, and **odds ratios**, respectively.

$$\Pr[Y = 1] = \Pr[Y = 1 | X = 1] = \Pr[Y = 1 | X = 0]$$

Equilvalent Definitions of Dependence

For $Y = 0, 1$ and $X = 0, 1$, there are also alternative definitions of independence. These concepts are sometimes called **risk differences**, **risk ratio**, and **odds ratios**, respectively.

$$\Pr[Y = 1] = \Pr[Y = 1 | X = 1] = \Pr[Y = 1 | X = 0]$$

$$\frac{\Pr[Y = 1 | X = 1]}{\Pr[Y = 1 | X = 0]} = 1$$

Equilvalent Definitions of Dependence

For $Y = 0, 1$ and $X = 0, 1$, there are also alternative definitions of independence. These concepts are sometimes called **risk differences**, **risk ratio**, and **odds ratios**, respectively.

$$\Pr[Y = 1] = \Pr[Y = 1|X = 1] = \Pr[Y = 1|X = 0]$$

$$\frac{\Pr[Y = 1|X = 1]}{\Pr[Y = 1|X = 0]} = 1$$

$$\frac{\Pr[Y = 1|X = 1]}{\Pr[Y = 1|X = 0]} \cdot \frac{\Pr[Y = 0|X = 0]}{\Pr[Y = 0|X = 1]} = 1$$

Conditional Distribution and Related Concepts and Results

1. Conditional distribution and **indepdence**
2. Conditional distribution and **law of total probability** (discrete and continuous cases)

Conditional Distribution and Law of Total Probability

- Construct from **Joint Distribution**

$$\Pr[Y = y] = p(y) = \sum p(x_i, y)$$

Conditional Distribution and Law of Total Probability

- Construct from **Joint Distribution**

$$\Pr[Y = y] = p(y) = \sum p(x_i, y)$$

- Construct from **conditional distribution** (a weighted sum of conditional probabilities)

$$\begin{aligned}\Pr[Y = y] &= p(y) \\ &= \sum p(x_i, y)\end{aligned}$$

Conditional Distribution and Law of Total Probability

- Construct from **Joint Distribution**

$$\Pr[Y = y] = p(y) = \sum p(x_i, y)$$

- Construct from **conditional distribution** (a weighted sum of conditional probabilities)

$$\begin{aligned}\Pr[Y = y] &= p(y) \\ &= \sum p(y|x_i) \cdot p(x_i)\end{aligned}$$

Conditional Distribution and Law of Total Probability

- Construct from **Joint Distribution**

$$\Pr[Y = y] = p(y) = \sum p(x_i, y)$$

- Construct from **conditional distribution** (a weighted sum of conditional probabilities)

$$\begin{aligned}
\Pr[Y = y] &= p(y) \\
&= \sum p(y|x_i) \cdot p(x_i) \\
&= \Pr[Y = y|X = x_1] \cdot \Pr[X = x_1] + \\
&\Pr[Y = y|X = x_2] \cdot \Pr[X = x_2] \\
&+ \cdots + \Pr[Y = y|X = x_n] \cdot \Pr[X = x_n]
\end{aligned}$$

$$p(y) = \sum p(y \mid x_i) \cdot p(x_i)$$
$$p(x) = \sum p(x \mid y_i) \cdot p(y_i)$$

We will use the latter equality as well in showing general Bayes' Rule.

Law of Total Probability

Continuous Variables

$$f(y) = \int f(y|x)f(x)dx$$

It is easy to show that the law of total probability is satisfied based on the definition.

$$\begin{aligned} f(y) &= \int f(x, y)dx \\ &= \int f(y|x)f(x)dx \end{aligned}$$

$$f(y) = \int f(y|x)f(x)dx$$

This expression is very useful, e.g., for thinking how the distribution of y is determined. It consists of two parts

1. $f(y \mid x)$: how y is linked to x (for example, the wage determination process linking education to wages)
2. $f(x)$: the distribution of x (e.g., education)

Law of Total Probability (CDF version)

We can also simlarly show that the following holds

$$F(y) = \int F(y|x)f(x)dx$$

This result is particularly useful when we would like to use conditional distribution (conditioanl quantile function) to recover marginal distribution.

Intuition: The percentage of values smaller than y is equal to the weighted average of the percentage of values smaller than y in every subgroup with the weight being the probability of the subgroup.

Law of Total Probability (Proof)

$$F(y) = \int_{-\infty}^{y} f(t)dt$$

Law of Total Probability (Proof)

$$
\begin{aligned}
F(y) &= \int_{-\infty}^{y} f(t)dt \\
&= \int_{-\infty}^{y} \int f(t|x)f(x)dxdt
\end{aligned}
$$

Law of Total Probability (Proof)

$$
\begin{aligned}
F(y) &= \int_{-\infty}^{y} f(t)dt \\
&= \int_{-\infty}^{y} \int f(t|x)f(x)dxdt \\
&= \int \int_{-\infty}^{y} f(t|x)dtf(x)dx
\end{aligned}
$$

Law of Total Probability (Proof)

$$\begin{aligned}
F(y) &= \int_{-\infty}^{y} f(t)dt \\
&= \int_{-\infty}^{y} \int f(t|x)f(x)dxdt \\
&= \int \int_{-\infty}^{y} f(t|x)dtf(x)dx \\
&= \int F(y|x)f(x)dx
\end{aligned}$$

We can also simlarly show that the following holds

$$F(y) = \sum F(y|x)p(x)$$

Application 1: Partial Identification of the Distribution in the Presence of Sample Selection

Bounding the wage distribution in the sample selection:

We do not know the wages for women who do not work. In other words, we can only observe wages for those who do work ($S = 1$) and have the knowledge of conditional distribution of wages

$$F(y|S = 1)$$

Question is: What is $F(y)$?

Bounds on the Wage Distribution in the presence of Sample Selection

$$F(y) = F(y|S=1)\Pr[S=1] + F(y|S=0)\Pr[S=0]$$

Bounds on the Wage Distribution in the presence of Sample Selection

$$F(y) = F(y|S=1)\Pr[S=1] + F(y|S=0)\Pr[S=0]$$

- Upper Bound: $F(y|S=0) = 1 \implies$

$$F^{UB}(y) = F(y|S=1)\Pr[S=1] + \Pr[S=0]$$

Bounds on the Wage Distribution in the presence of Sample Selection

$$F(y) = F(y|S=1)\Pr[S=1] + F(y|S=0)\Pr[S=0]$$

- Upper Bound: $F(y|S=0) = 1 \implies$

$$F^{UB}(y) = F(y|S=1)\Pr[S=1] + \Pr[S=0]$$

- Lower Bound: $F(y|S=0) = 1 \implies$

$$F^{LB}(y) = F(y|S=1)\Pr[S=1]$$

Such case is also called **worst-case bounds**

Bounds on the Wage Distribution in the presence of Sample Selection

Let's draw a graph of these bounds

Application 2: Decomposition of Wage Distribution between Groups

$F^1(y)$ and $F^0(y)$ are wage distributions for men and women.

The Gender Gap is defined as

$$F^1(y) - F^0(y) = \underbrace{[F^1(y) - F^c(y)]}_{\text{Structural effects}} + \underbrace{[F^c(y) - F^0(y)]}_{\text{Composition effects}}$$

Question: How to define the counterfactual so that it reflects the wage distribution for individuals in group 1 under the wage structure for group 0 but holding fixed the distribution of characteristics, X?

Application 2: Decomposition of Wage Distribution between Groups

How do wages, y, relate to the characteristics, x? Through the conitional distribution!

$$F(y) = \int F(y|x)f(x)dx$$

Application 2: Decomposition of Wage Distribution between Groups

How do wages, y, relate to the characteristics, x? Through the conitional distribution!

$$F(y) = \int F(y|x)f(x)dx$$

$$F^c(y) = \int F^0(y|x)f^1(x)dx$$

Conditional Distribution and Related Concepts and Results

1. Conditional distribution and **indepdence**
2. Conditional distribution and **law of total probability** (discrete and continuous cases)
3. Conditional distribution and **Bayes' Rule**

(General) Baye's Rule

$$p(y|x) = \frac{p(y, x)}{p(x)}$$

(General) Baye's Rule

$$p(y|x) = \frac{p(y,x)}{p(x)}$$
$$= \frac{p(x|y) \cdot p(y)}{p(x)}$$

(General) Baye's Rule

$$
\begin{aligned}
p(y|x) &= \frac{p(y,x)}{p(x)} \\
&= \frac{p(x|y) \cdot p(y)}{p(x)} \\
&= \frac{p(x|y) \cdot p(y)}{\sum_j p(x|y_j) \cdot p(y_j)}
\end{aligned}
$$

The last equality comes from the **law of total probability**.

We can similarly write the rule for the continuous case (simply by replacing p.m.f with p.d.f and the summation sign with integral):

$$f(y|x) = \frac{f(x|y) \cdot f(y)}{f(x)} = \frac{f(x|y) \cdot f(y)}{\int f(x|y) \cdot f(y)dy}$$

(General) Baye's Rule (Skip)

Proof:

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

$$f(x|y) = \frac{f(x,y)}{f(y)} \implies f(x,y) = f(x|y) \cdot f(y)$$

$$\begin{aligned} f(y|x) &= \frac{f(x,y)}{f(x)} \\ &= \frac{f(x|y) \cdot f(y)}{f(x)} \end{aligned}$$

Bayes' Rule (Mixed/Hybrid Variables)

$$f_{X|D=1}(\cdot) = \frac{\Pr[D = 1|X] \cdot f(x)}{\Pr[D = 1]}$$

Such result is useful for analyzing the distribution by various types (e.g., compliers in the local average treatment effects literature)

Example: The distribution of, say, education (X) given that someone participated in a job training program ($D = 1$)

With Bayes' Rule and law of total probability, we can derive the counterfactual distribution!

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

How do wages, y, relate to the characteristics, x? Through the conitional distribution!

$$F(y) = \int F(y|x)f(x)dx$$

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

How do wages, y, relate to the characteristics, x? Through the conitional distribution!

$$F(y) = \int F(y|x)f(x)dx$$

$$F^c(y) = \int F^0(y|x)f^1(x)dx$$

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

How do actually recover the counterfactual distribution, $F^c(y)$ and all the characteristics of the distribution? Reweighting!

$$F^c(y) = \mathbb{E}[w \cdot \mathbb{I}[Y \leq y] \mid D = 0]$$

where $w = \frac{p(x)}{1-p(x)} \frac{1-p}{p}$. Here, $p = \Pr[D = 1]$ and $p(x) = \Pr[D = 1|X]$

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

$$F^c(y) = \int F^0(y|x)f^1(x)dx$$

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

$$
\begin{aligned}
F^c(y) &= \int F^0(y|x) f^1(x) dx \\
&= \int F^0(y|x) \frac{f^1(x)}{f^0(x)} f^0(x) dx
\end{aligned}
$$

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

$$
\begin{aligned}
F^c(y) &= \int F^0(y|x) f^1(x) dx \\
&= \int F^0(y|x) \frac{f^1(x)}{f^0(x)} f^0(x) dx \\
&= \int F^0(y|x) \frac{f_{X|D=1}(x)}{f_{X|D=0}(x)} f_{X|d=0}(x) dx
\end{aligned}
$$

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

$$
\begin{aligned}
F^c(y) &= \int F^0(y|x) f^1(x) dx \\
&= \int F^0(y|x) \frac{f^1(x)}{f^0(x)} f^0(x) dx \\
&= \int F^0(y|x) \frac{f_{X|D=1}(x)}{f_{X|D=0}(x)} f_{X|d=0}(x) dx \\
&= \int F^0(y|x) \frac{\frac{\Pr[D=1|X] \cdot f(x)}{\Pr[D=1]}}{\frac{1-\Pr[D=1|X] \cdot f(x)}{1-\Pr[d=1]}} f_{X|D=0}(x) dx
\end{aligned}
$$

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

$$
\begin{aligned}
F^c(y) &= \int F^0(y|x) f^1(x) dx \\
&= \int F^0(y|x) \frac{f^1(x)}{f^0(x)} f^0(x) dx \\
&= \int F^0(y|x) \frac{f_{X|D=1}(x)}{f_{X|D=0}(x)} f_{X|d=0}(x) dx \\
&= \int F^0(y|x) \frac{\frac{\Pr[D=1|X]\cdot f(x)}{\Pr[D=1]}}{\frac{1-\Pr[D=1|X]\cdot f(x)}{1-\Pr[d=1]}} f_{X|D=0}(x) dx \\
&= \int F^0(y|x) \frac{\frac{p(x)}{p}}{\frac{1-p(x)}{1-p}} f_{X|D=0}(x) dx
\end{aligned}
$$

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

$$
\begin{aligned}
F^c(y) &= \int F^0(y|x)f^1(x)dx \\
&= \int F^0(y|x)\frac{f^1(x)}{f^0(x)}f^0(x)dx \\
&= \int F^0(y|x)\frac{f_{X|D=1}(x)}{f_{X|D=0}(x)}f_{X|d=0}(x)dx \\
&= \int F^0(y|x)\frac{\frac{\Pr[D=1|X]\cdot f(x)}{\Pr[D=1]}}{\frac{1-\Pr[D=1|X]\cdot f(x)}{1-\Pr[d=1]}}f_{X|D=0}(x)dx \\
&= \int F^0(y|x)\frac{\frac{p(x)}{p}}{\frac{1-p(x)}{1-p}}f_{X|D=0}(x)dx \\
&= \int F^0(y|x)\frac{p(x)}{1-p(x)}\cdot\frac{1-p}{p}f_{X|D=0}(x)dx
\end{aligned}
$$

Application 2: Decomposition of Wage Distribution between Groups (Skip in Class)

$$
\begin{aligned}
F^c(y) &= \int F^0(y|x) f^1(x) dx \\
&= \int F^0(y|x) \frac{f^1(x)}{f^0(x)} f^0(x) dx \\
&= \int F^0(y|x) \frac{f_{X|D=1}(x)}{f_{X|D=0}(x)} f_{X|d=0}(x) dx \\
&= \int F^0(y|x) \frac{\frac{\Pr[D=1|X] \cdot f(x)}{\Pr[D=1]}}{\frac{1-\Pr[D=1|X] \cdot f(x)}{1-\Pr[d=1]}} f_{X|D=0}(x) dx \\
&= \int F^0(y|x) \frac{\frac{p(x)}{p}}{\frac{1-p(x)}{1-p}} f_{X|D=0}(x) dx \\
&= \int F^0(y|x) \frac{p(x)}{1-p(x)} \cdot \frac{1-p}{p} f_{X|D=0}(x) dx \\
&= \mathbb{E}[w \cdot \mathbb{I}[Y \leq y] | D=0]
\end{aligned}
$$

Conditional Distribution and Related Concepts and Results

1. Conditional distribution and **indepdence**
2. Conditional distribution and **law of total probability** (discrete and continuous cases)
3. Conditional distribution and **Bayes' Rule**
4. Conditional distribution and **Skorohod Representation**

Skorohod Representation

Theorem Conditioning on X, for a random variable Y, there exists $U \sim \mathcal{U}(0,1)$ (the standard uniform distribution) such that

$$Y = F^{-1}(U|X) = m(X,U)$$

holds almost surely. This $m(X,U)$ is the quantile function.

Skorohod Representation

Proof. Let's show it for the simple case where Y is continuous. In other words, $F(y|x)$ is strictly increasing, and so is its inverse function. Define U such that

$$F(Y|X) = U$$

Skorohod Representation

Proof. Let's show it for the simple case where Y is continuous. In other words, $F(y|x)$ is strictly increasing, and so is its inverse function. Define U such that

$$F(Y|X) = U$$

And let y be the τ^{th} quantile of the conditional distribution

$$F(y|X) = \tau$$

Skorohod Representation

Proof. Let's show it for the simple case where Y is continuous. In other words, $F(y|x)$ is strictly increasing, and so is its inverse function. Define U such that

$$F(Y|X) = U$$

And let y be the τ^{th} quantile of the conditional distribution

$$F(y|X) = \tau$$

$$\Pr[U \leq \tau|X] = \Pr[F^{-1}(U|X) \leq F^{-1}(\tau|X)|X]$$

Skorohod Representation

Proof. Let's show it for the simple case where Y is continuous. In other words, $F(y|x)$ is strictly increasing, and so is its inverse function. Define U such that

$$F(Y|X) = U$$

And let y be the τ^{th} quantile of the conditional distribution

$$F(y|X) = \tau$$

$$\begin{aligned}\Pr[U \leq \tau|X] &= \Pr[F^{-1}(U|X) \leq F^{-1}(\tau|X)|X] \\ &= \Pr[Y \leq y|X]\end{aligned}$$

Skorohod Representation

Proof. Let's show it for the simple case where Y is continuous. In other words, $F(y|x)$ is strictly increasing, and so is its inverse function. Define U such that

$$F(Y|X) = U$$

And let y be the τ^{th} quantile of the conditional distribution

$$F(y|X) = \tau$$

$$\begin{aligned}
\Pr[U \leq \tau|X] &= \Pr[F^{-1}(U|X) \leq F^{-1}(\tau|X)|X] \\
&= \Pr[Y \leq y|X] \\
&= F(y|X)
\end{aligned}$$

Skorohod Representation

Proof. Let's show it for the simple case where Y is continuous. In other words, $F(y|x)$ is strictly increasing, and so is its inverse function. Define U such that

$$F(Y|X) = U$$

And let y be the τ^{th} quantile of the conditional distribution

$$F(y|X) = \tau$$

$$\begin{aligned}
\Pr[U \leq \tau|X] &= \Pr[F^{-1}(U|X) \leq F^{-1}(\tau|X)|X] \\
&= \Pr[Y \leq y|X] \\
&= F(y|X) \\
&= \tau
\end{aligned}$$

$\implies U|X \sim \mathcal{U}(0,1)$.

Question: Are U and X dependent?

Question: Are U and X dependent?

Answer: Yes. You can show the following is true.

Question: Are U and X dependent?

Answer: Yes. You can show the following is true.

$$\Pr[U \leq \tau \mid X] = \Pr[U \leq \tau]$$

Hint: You can use the CDF version of the law of total probability.

Conditional Distribution and Conditional Independence

Conditional Distribution and Independence

$$f(y|x) = f(y)$$

This result immediately follows from the fact that independence implies that $f(x, y) = f(x)f(y)$

Conditional Independence

$$x \perp y | z$$

if and only if the following statements are satisfied

$$f(x, y|z) = f(x|z)f(y|z)$$

$$f(y|x, z) = f(y|z)$$

Note that these results extend to many variables

$$x \perp (y_1, y_2, y_3, \ldots, y_k) \mid z$$

if and only if **one** of the following **equivalent** statements are satisfied

$$f(x, y_1, y_2, y_3, \ldots, y_k|z) = f(x|z)f(y_1, y_2, y_3, \ldots, y_k|z)$$

Note that these results extend to many variables

$$x \perp (y_1, y_2, y_3, \ldots, y_k) \mid z$$

if and only if **one** of the following **equivalent** statements are satisfied

$$f(x, y_1, y_2, y_3, \ldots, y_k|z) = f(x|z)f(y_1, y_2, y_3, \ldots, y_k|z)$$

$$f(y_1, y_2, y_3, \ldots, y_k \mid x, z) = f(y_1, y_2, y_3, \ldots, y_k \mid z)$$

Note that these results extend to many variables

$$x \perp (y_1, y_2, y_3, \ldots, y_k) \mid z$$

if and only if **one** of the following **equivalent** statements are satisfied

$$f(x, y_1, y_2, y_3, \ldots, y_k|z) = f(x|z)f(y_1, y_2, y_3, \ldots, y_k|z)$$

$$f(y_1, y_2, y_3, \ldots, y_k \mid x, z) = f(y_1, y_2, y_3, \ldots, y_k \mid z)$$

$$f(x \mid y_1, y_2, y_3, \ldots, y_k, z) = f(x \mid z)$$

Applications in the treatment effects

Potential Outcome Framework

Consider the impacts of job training program on wages, then for each state (whether or not one participates in the program, $D = 0, 1$), there is a potential wage offer

$$Y(0) \text{ if } D = 0$$
$$Y(1) \text{ if } D = 1$$

We never observe both, but only one of them. The observed outcome can be written as

Applications in the treatment effects

Potential Outcome Framework

Consider the impacts of job training program on wages, then for each state (whether or not one participates in the program, $D = 0, 1$), there is a potential wage offer

$$Y(0) \text{ if } D = 0$$
$$Y(1) \text{ if } D = 1$$

We never observe both, but only one of them. The observed outcome can be written as

$$Y = D \cdot Y(1) + (1 - D) \cdot Y(0)$$

Application 1: Strong Ignorability Assumption

Variants of this assumption is employed to identify models such as OLS and Propensity Score Matching

$$(Y(0), Y(1)) \perp D \mid X$$

Application 2: Marginal, Conditional, and Joint Independence

$$(Y(0), Y(1)) \perp D$$

Does marginal independence imply the joint depedence?

Application 2: Marginal, Conditional, and Joint Independence

$$(Y(0), Y(1)) \perp D$$

Does marginal independence imply the joint depedence?

Answer No!

What do we need?

The following result states one of the ways (with two conditions) to ensure joint independence:

$$Y(1) \perp D \mid Y(0)$$

and

$$Y(0) \perp D$$

In other words, these two conditions imply that

$$(Y(1), Y(0)) \perp D$$

Application 2: Marginal, Conditional, and Joint Independence

Proof: **Note I** use lower case for a specific value

$$\Pr[d|y_1, y_0] = \ldots \text{(you fill in here)} = \Pr[d]$$

Application 2: Marginal, Conditional, and Joint Independence

Proof: **Note I** use lower case for a specific value

$$\Pr[d|y_1, y_0] = \ldots \text{(you fill in here)} = \Pr[d]$$

Or, you can start with joint distribution (then by factoring out $f(y_0))

$$\begin{aligned} f(y_1, y_0, d) &= \ldots \text{(you fill in here)} \\ &= \ldots \text{(you fill in here)} \\ &= f(d)f(y_1, y_0) \end{aligned}$$

This is left as a **homework** question.

Why care about marginals?

Why can we just assume joint dependence to estiamte the effects? Consider the following model:

$$D = \mathbb{I}[Y(1) \geq Y(0)]$$

Joint indepedence rules out selection based on personal gains, $y_1 - y_0 \geq 0$. Such case is economically uninteresting.

Conditional Indpendence (Some Basic Properties)

Angrist (1997), Conditional Indpedence in Sample Selection Models, *Economics Letters*

Lemma Let R_1, R_2, R_3 and R_4 be random variables defined on a common probability space with joint proability measure. Then the following are equivalent.

- $R_1 \perp R_2 | R_3$ and $R_1 \perp R_4 | (R_2, R_3)$
- $R_1 \perp (R_2, R_4) | R_3$
- $R_1 \perp R_4 | R_3$ and $R_1 \perp R_2 | (R_3, R_4)$

Note that

- $R_1 \perp R_2 | R_3$ and $R_1 \perp R_4 | (R_2, R_3)$
- $R_1 \perp (R_2, R_4) | R_3$

implies that (the unconditional one is just omitting R_3), which exactly what we show earlier!

- $R_1 \perp R_2$ and $R_1 \perp R_4 | R_2$
- $R_1 \perp (R_2, R_4)$