L.2: Review of Probability 2

Econometrics 1: ver. 2024 Fall Semester

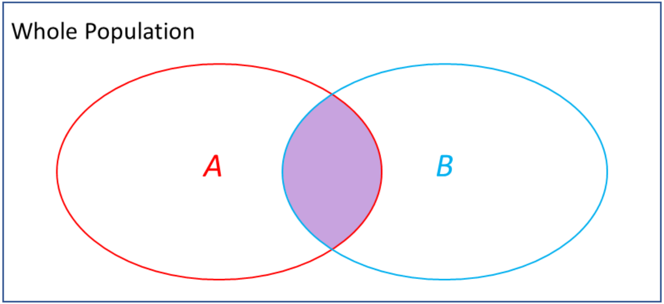
Naoki Awaya

Multiple Random Variables

Conditional Probability and Joint Probability

- Consider two events, A and B. For example,
  - A: a randomly selected person has a part-time job.
  - B: a randomly selected person is an undergraduate student.
- **Conditional probability** $\Pr(A|B)$: the probability of selecting a person who has a part-time job given that she is an undergrad student.
  $=$ the proportion of part-time workers among undergraduate students.
- **Joint probability** $\Pr(A, B)$: the probability of selecting a person who has a part-time job and is an undergraduate student.
  $=$ the proportion of part-time workers who are undergraduate students among the whole population.

Conditional Probability and Joint Probability



$$\Pr(A|B) = \frac{\Pr(A,B)}{\Pr(B)} = \frac{\quad}{B}$$

$$\Pr(A,B) = \frac{\Pr(A,B)}{\Pr(\text{Whole Population})} = \frac{\quad}{\boxed{\text{Whole Pop.}}}$$

Conditional Probability and Joint Probability

Conditional Distribution Function

For a pair of random variables (X, Y) and constants (x, y), the conditional probability of $\{Y \leq y\}$ given $X = x$ is called the conditional distribution function, and it is denoted as $F_{Y|X}(y|X = x)$:

$$F_{Y|X}(y|X = x) = \Pr(Y \leq y|X = x)$$

- For example, $Y =$ annual inc in mill JPY, $X =$ education in years.
  - $F_{Y|X}(3|X = 9) = \Pr($anninc $\leq$ 3 mill JPY | jr high graduate$)$.
  - $F_{Y|X}(3|X = 16) = \Pr($anninc $\leq$ 3 mill JPY | college graduate$)$.
  - Normally, $F_{Y|X}(3|X = 9) \geq F_{Y|X}(3|X = 16)$
    $\Rightarrow$ the "condition" affects the probability distribution.

Conditional Probability and Joint Probability

Conditional Density Function

The derivative of $F_{Y|X}(y|X=x)$ with respect to y is called the conditional density function, and it is denoted as $f_{Y|X}(y|X=x)$:

$$f_{Y|X}(y|X=x) = \frac{\partial F_{Y|X}(y|X=x)}{\partial y}.$$

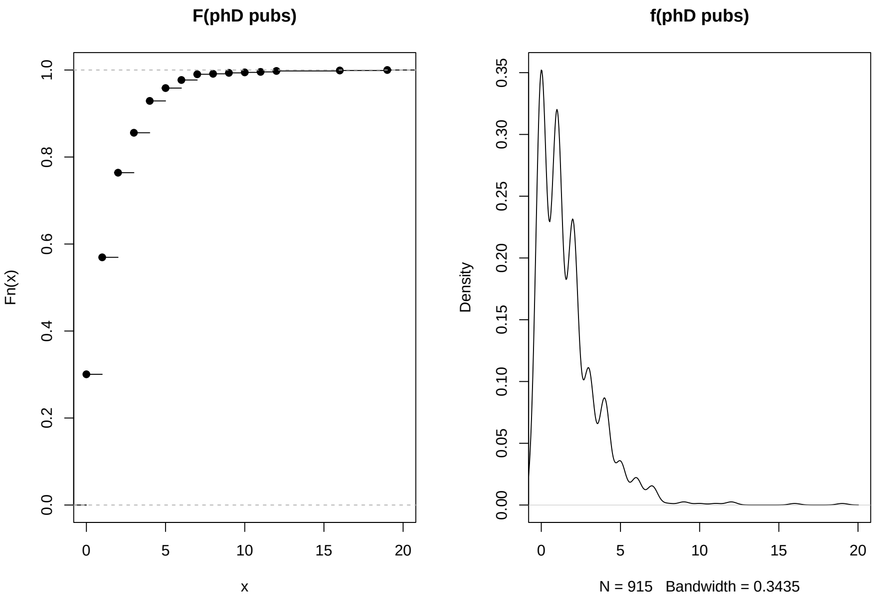
- This is just the PDF of Y for a sub-population satisfying $X=x$.
- For example,
  - $f_{Y|X}(3|X=9)$ = how {anninc = 3 mill JPY} is likely to occur for junior high graduates.
  - $f_{Y|X}(3|X=16)$ = how {anninc = 3 mill JPY} is likely to occur for college graduates.

Conditional Probability and Joint Probability

```
> library(AER)
> library(tidyverse)
> data(PhDPublications)
> head(PhDPublications, 3)
> F <- ecdf(PhDPublications$articles)
> f <- density(PhDPublications$articles)
> par(mfrow = c(1,2))
> plot(F, xlim = c(0,20), main = "F(phD pubs)")
> plot(f, xlim = c(0,20), main = "f(phD pubs)")
```

- `AER`: Applied Econometrics with R. This includes many econometrics practice datasets.
- `tidyverse`: set of useful packages for data manipulation.
- `ecdf(X)`: compute the (empirical) distribution function of `X`.
- `density(X)`: compute the kernel (i.e., smoothed) density function of `X`.
- `par(mfrow = c(1,2))`: split the graphic window into 1×2 sub-windows.

Conditional Probability and Joint Probability



Conditional Probability and Joint Probability

```
> data1 <- PhDPublications %>% filter(mentor >= 6)
> data2 <- PhDPublications %>% filter(mentor < 6)
```

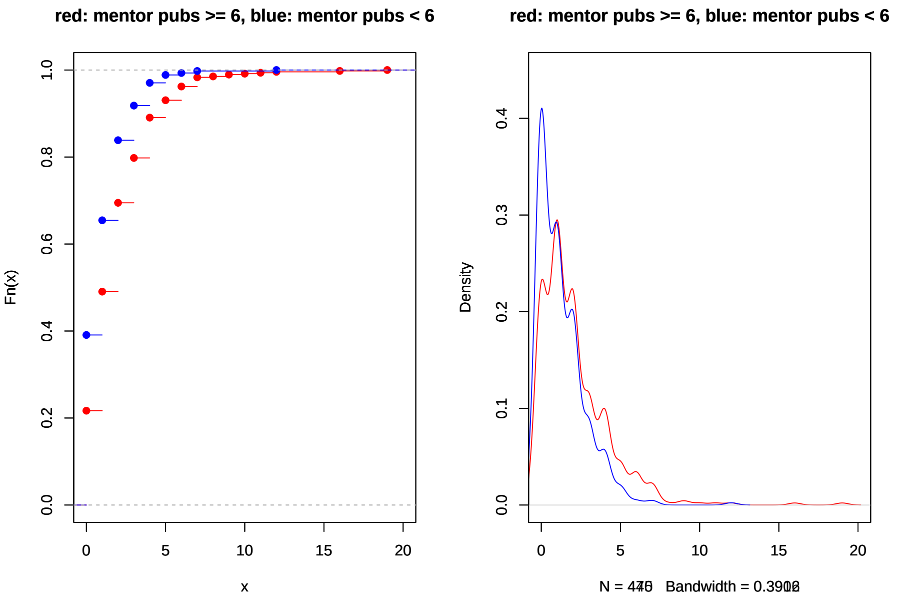
- `%>%`: "pipe" operator. `A %>% B` means "apply `B` to `A`".
    - E.g., `max(x)` = `x %>% max`
- `filter`: select the subset that satisfies the given condition.

Conditional Probability and Joint Probability

```
> F1 <- ecdf(data1$articles)
> F2 <- ecdf(data2$articles)
> f1 <- density(data1$articles)
> f2 <- density(data2$articles)
> par(mfrow = c(1,2))
> fig_title <- "red: mentor pubs >= 6, blue: mentor pubs < 6"
> plot(F1, col = "red", xlim = c(0,20), main = fig_title)
> par(new = T)
> plot(F2, col = "blue", xlim = c(0,20), main = fig_title)
>
> plot(f1, col = "red", xlim = c(0,20), ylim = c(0, 0.45), main = fig_title
> par(new = T)
> plot(f2, col = "blue", xlim = c(0,20), ylim = c(0, 0.45), main = fig_titl
```

- `par(new = T)`: overwrite the current image without erasing it.

Conditional Probability and Joint Probability



Conditional Expectation

Conditional Expectation

Let $f_{Y|X}(y|X=x)$ be the conditional density function of Y given $X=x$.

$$\mathbb{E}(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|X=x)dy$$

is called the conditional expectation of Y given $X=x$.

- This is the expectation of Y for a sub-population satisfying $X=x$.
- For example,
  - $\mathbb{E}(Y|X=9) =$ expected value of anninc for junior high graduates.
  - $\mathbb{E}(Y|X=16) =$ expected value of anninc for college graduates.
  - Normally, $\mathbb{E}(Y|X=9) \leq \mathbb{E}(Y|X=16)$ is expected.

$\mathbb{E}(Y|X=x)$ is the value obtained by plugging x into the conditional expectation function $\mathbb{E}(Y|X=\cdot)$.

Estimation of Conditional Expectation Function

- $\mathbb{E}(Y|X = x)$ can be estimated by the average of Y for the sub-sample satisfying $X = x$.

```
> quant <- quantile(PhDPublications$mentor, c(0.25,0.50,0.75))
> quant
```

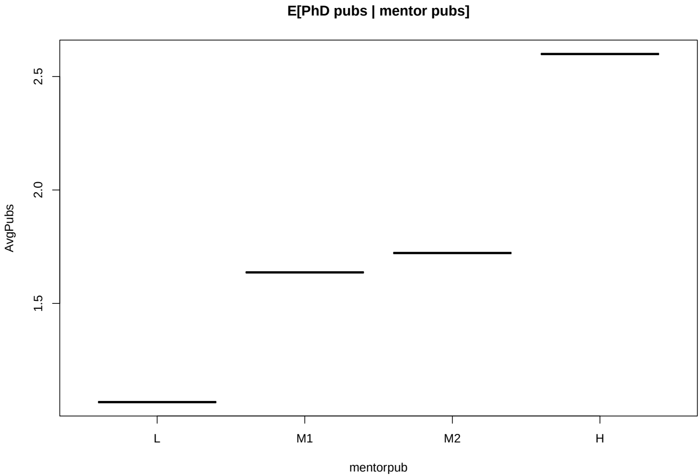
```
## 25% 50% 75%
##   3   6  12
```

```
> pubdata <- PhDPublications %>%
+   mutate(mentorpub = cut(mentor,
+                          breaks = c(-Inf, quant, Inf),
+                          labels = c('L', 'M1', 'M2', 'H')))
> pubdata <- pubdata %>%
+   group_by(mentorpub) %>%
+   summarize(AvgPubs = mean(articles))
```

- `cut`: discretize the variable following the given break points.
- `mutate`: add a new variable

Estimation of Conditional Expectation Function

```
> plot(pubdata, main = "E[PhD pubs | mentor pubs]")
```



Joint Distribution and Joint Density

Joint Distribution Function, Joint Density Function

For random variables (Y, X), the probability of $\{Y \leq y, X \leq x\}$ is called the joint distribution function, and it is denoted as $F_{YX}(y, x)$:

$$F_{YX}(y, x) = \Pr(Y \leq y, X \leq x)$$

In addition, by taking the cross-partial derivative of $F_{YX}(y, x)$, we obtain the joint density function

$$f_{YX}(y, x) = \frac{\partial^2 F_{YX}(y, x)}{\partial y \partial x} \quad \left(F_{YX}(a, b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{YX}(y, x) dx dy \right)$$

Expectation of a product of random variables

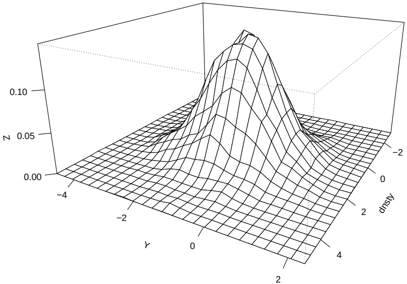
The expectation of the product of random variables Y and X is given by

$$\mathbb{E}(YX) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yx f_{YX}(y, x) dx dy$$

Joint Distribution and Joint Density

Bivariate standard normal distribution with correlation parameter $\rho = 0$:

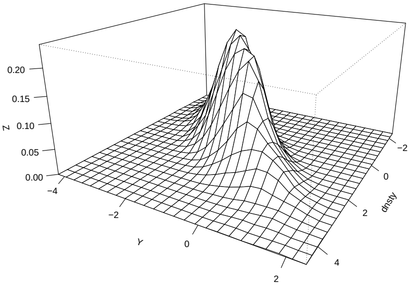
```
> library(MASS)
>
> xy <- mvrnorm(5000, c(1,-1), diag(2))
> x <- xy[,1]; y <- xy[,2]
> dnsty <- kde2d(x, y) # 2-dimensional kernel density
> persp(dnsty, theta=120, phi=20, expand=0.5, ticktype = "detailed")
```



Joint Distribution and Joint Density

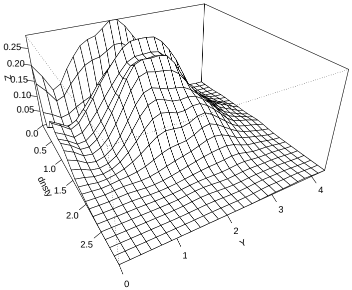
Bivariate standard normal distribution with correlation parameter $\rho = 0.8$:

```
> xy <- mvrnorm(5000, c(1,-1), matrix(c(1,0.8,0.8,1),2,2))
> x <- xy[,1]; y <- xy[,2]
> dnsty <- kde2d(x, y) # 2-dimensional kernel density
> persp(dnsty, theta=120, phi=20, expand=0.5, ticktype = "detailed"
```



Joint Distribution and Joint Density

```
> x <- log(PhDPublications$articles + 1)
> y <- log(PhDPublications$mentor + 1)
> dnsty <- kde2d(x, y) # 2-dimensional kernel density
> persp(dnsty, theta=60, phi=30, expand=0.5, ticktype = "detailed")
```



Marginal Distribution and Marginal Density

- Given a joint distribution of random variables, the distribution and density function of each random variable are called the margianl distribution function and marginal density function, respectively.
- The marginal distribution function of Y, $F_Y(y)$, can be derived from the joint distribution function $F_{YX}(y, x)$ in the following manner:

$$F_Y(y) = \Pr(Y \le y) = \Pr(Y \le y, X \le \infty) = F_{YX}(y, \infty)$$

- Namely, for any a,

$$\int_{-\infty}^{a} f_Y(y)dy = \int_{-\infty}^{a}\int_{-\infty}^{\infty} f_{YX}(y, x)dxdy$$

which further implies that

$$f_Y(y) = \int_{\mathcal{X}} f_{YX}(y, x)dx$$

- Obtaining the marginal density by integrating the joint density is called marginalization.

Conditional Density and Joint Density

- The following relationship between the joint probability $\Pr(A, B)$ and the conditional probability $\Pr(A|B)$ is clear from the definitions:

$$\Pr(A|B) = \frac{\Pr(A, B)}{\Pr(B)}$$

- The same relationship holds for the joint density and the conditional density:

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

If the marginal density $f_X(x)$ is equal to zero, the conditional density $f_{Y|X}(y|X = x)$ is not defined.

Covariance and Correlation

Covariance and Correlation

Covariance

For two random variables X and Y, the covariance is defined by

$$\begin{aligned}\mathbb{C}(Y,X) &= \mathbb{E}[\{Y-\mathbb{E}(Y)\}\{X-\mathbb{E}(X)\}] \\ &= \mathbb{E}(YX)-\mathbb{E}(Y)\mathbb{E}(X)\end{aligned}$$

In particular, if either X or Y has mean zero, $\mathbb{C}(Y,X)=\mathbb{E}(YX)$.

$\mathbb{C}(Y,X)$ represents the strength of the correlation between Y and X.

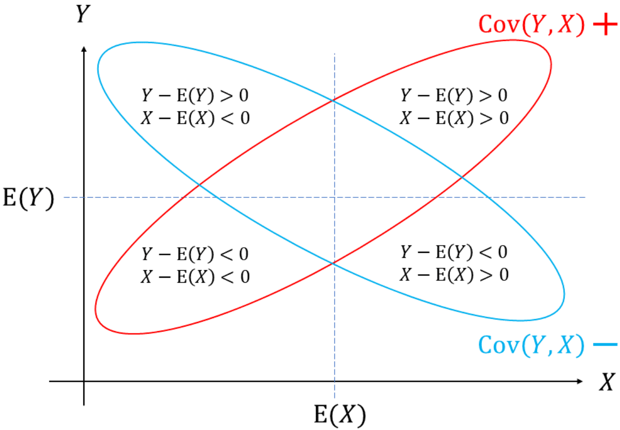
(Pearson's) correlation coefficient

For two random variables Y and X, the correlation coefficient is defined by

$$\text{Cor}(Y,X)=\frac{\mathbb{C}(Y,X)}{\sqrt{\mathbb{V}(Y)}\sqrt{\mathbb{V}(X)}}$$

The Pearson correlation coefficient is a scaled version of the covariance, so that $-1 \leq \text{Cor}(Y,X) \leq 1$.

Covariance and Correlation



$$\mathrm{Cov}(Y,X) = \mathrm{E}[(Y - \mathrm{E}Y)(X - \mathrm{E}X)]$$

Independence

Independence of Random Variables

- For two events $\{Y \leq y\}$ and $\{X \leq x\}$, if

$$\Pr(Y \leq y, X \leq x) = \Pr(Y \leq y) \cdot \Pr(X \leq x)$$

in other words, $F_{YX}(y, x) = F_Y(y) \cdot F_X(x)$

then, these "events" are independent.

Independence of random variables

Let $F_{YX}(y, x)$ be the joint distribution function of (Y, X), and $F_Y(y)$ and $F_X(x)$ be the marginal distribution functions of Y and X, respectively. We say that X and Y are independent if

$$F_{YX}(y, x) = F_Y(y) \cdot F_X(x)$$

holds for any (y, x).

Independence of Random Variables

- When two random variables Y and X are independent, we also have

$$f_{YX}(y,x) = f_Y(y) \cdot f_X(x)$$

because

$$f_{YX}(y,x) = \frac{\partial^2 F_{YX}(y,x)}{\partial y \partial x} = \frac{\partial F_Y(y)}{\partial y} \frac{\partial F_X(x)}{\partial x} = f_Y(y) \cdot f_X(x)$$

Property of independent random variables (1)

$$\begin{aligned} \mathbb{E}(YX) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yx f_Y(y) f_X(x) dx dy \\ &= \int_{-\infty}^{\infty} y f_Y(y) dy \int_{-\infty}^{\infty} x f_X(x) dx = \mathbb{E}(Y)\mathbb{E}(X) \end{aligned}$$

Independence of Random Variables

```
> x <- rnorm(1000,1,1)
> y <- rnorm(1000,1,1)
> mean(x*y); mean(x)*mean(y); cor(x,y)
```

```
## [1] 1.01276
```

```
## [1] 0.9946088
```

```
## [1] 0.01759474
```

```
> x <- rnorm(1000,1,1)
> y <- sin(x)*rnorm(1000,1,1)
> mean(x*y); mean(x)*mean(y); cor(x,y)
```

```
## [1] 0.8661112
```

```
## [1] 0.5146595
```

```
## [1] 0.4017158
```

Independence of Random Variables

- Recalling the definition of $\mathbb{C}(Y, X)$, if Y and X are independent, we obtain

$$\mathbb{C}(Y, X) = 0$$

However, $\mathbb{C}(Y, X) = 0$ does not imply the independence.

$$Y \text{ and } X \text{ are independent} \Rightarrow Y \text{ and } X \text{ are uncorrelated;}$$
$$Y \text{ and } X \text{ are uncorrelated} \nRightarrow Y \text{ and } X \text{ are independent.}$$

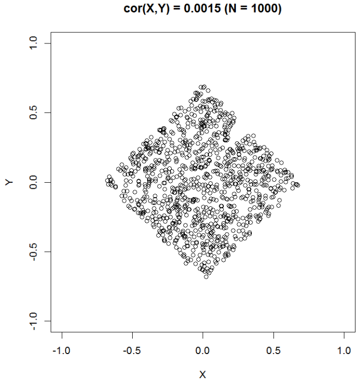
- The independence of Y and X means that they are "irrelevant" to each other.

Independence of Random Variables

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

$$f_{YX}(y,x) = \begin{cases} 1 & \text{if } |y| + |x| \leq \frac{1}{\sqrt{2}} \\ 0 & \text{otherwise} \end{cases}$$

- Scatter plot of X and Y values $(N = 1,000)$



Independence of Random Variables

- As seen from this figure, there is no correlation between X and Y; the sample correlation coefficient is almost zero (0.0015).
- However, they are not independent; for example, when $X = \frac{1}{\sqrt{2}}$, the value of Y is "automatically" determined as $Y = 0$, i.e., the value of Y depends on X, and vice versa.
- As in this example, if Y and X are independent, then the joint support of (Y, X) must have a shape like $\Box$ but not $\Diamond$.

Independence of Random Variables

```r
> x <- rnorm(10000)
> y <- x^2
```

`x` and `y` are clearly dependent but…

```r
> cov(x,y)
```

```
## [1] -0.0513248
```

```r
> cor(x,y)
```

```
## [1] -0.0354668
```

- The above result is due to the fact $\mathbb{E}[X^3] = 0$ when $X \sim N(0, 1)$.

Independence of Random Variables

- If Y and X are independent, the conditional expectation of Y given X coincides with Y's unconditional expectation.
- Note that if Y and X are independent,

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

Thus, for any x, it holds that

Property of independent random variables (2)

$$\mathbb{E}(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|X=x)dy = \int_{-\infty}^{\infty} y f_Y(y)dy = \mathbb{E}(Y)$$

- $Y =$ anninc and $X =$ education, the independence means that education has no effect on annual income.