Joint probability distribution
Lecture: 4

Hamid Raza
Assistant Professor in Economics
raza@business.aau.dk

Aalborg University

Statistics - Statistik

Outline

Joint probability distribution:

- In today's lecture, we look at the probabilities that several random variables of interest simultaneously take specific values
- At this point we will concentrate on two random variables, but the concepts apply to more than two

Joint probability distribution:

Let X and Y be a pair of discrete random variables. Their joint probability distribution expresses the probability that X and Y simultaneously takes some specific values

- Recall lecture 2, where we presented the probability of the intersection of bivariate events by $P(A \cap B)$.
- Here we use random variables and denote the joint probability distribution by $P(x, y)$, such that:

$$P(x, y) = P(X = x \cap Y = y)$$

Joint probability distribution:

Example:

For example, employees of a company are asked their membership status of a volleyball club in Aalborg as well as their gender:

```r
volley <- matrix(c(30, 20, 60, 30, 20, 20), ncol = 2, byrow = TRUE)
colnames(volley) <- c("Male", "Female")
rownames(volley) <- c("Current", "Former", "Never")
volley1 <- as.table(volley)
prop.table(volley)
```

```
##           Male Female
## Current 0.1667 0.1111
## Former  0.3333 0.1667
## Never   0.1111 0.1111
```

Let's say $X = (male, female)$ and $Y = (current, former, never)$

0.16 expresses the probability that X and Y simultaneously takes some specific values, which in this example are $X = male$, and $Y = current\ member$

Outline

Marginal probability distribution:

Marginal probability distribution:

Let X and Y be a pair of jointly distributed random variables. In this context the probability distribution of the random variable X is called its marginal probability distribution and is obtained by summing the joint probabilities over all possible values—that is,

$$P(x) = \sum_{y} P(x, y)$$

Similarly, the marginal probability distribution of the random variable Y is as follows:

$$P(y) = \sum_{x} P(x, y)$$

Example: In our example, the marginal probabilities are as follows:

	Male	Female	$P(y)$
Current	0.17	0.11	0.28
Former	0.33	0.17	0.5
Never	0.11	0.11	0.22
$P(x)$	0.61	0.39	1

Marginal probability distribution:

Marginal probability distribution in R:

Example: The marginal probabilities are as follows:

	Male	Female	$P(y)$
Current	0.17	0.11	0.28
Former	0.33	0.17	0.50
Never	0.11	0.11	0.22
$P(x)$	0.61	0.39	1

We can easily calculate this in R by simply typing:

```
addmargins(prop.table(volley))
```

```
##         Male Female  Sum
## Current 0.17   0.11 0.28
## Former  0.33   0.17 0.50
## Never   0.11   0.11 0.22
## Sum     0.61   0.39 1.00
```

Properties of joint probability distribution:

Properties of joint probability distribution:

Let X and Y be discrete random variables with joint probability distribution, $P(x, y)$, Then:

1. $0 \leq P(x, y) \leq 1$
2. the sum of the joint probabilities $P(x, y)$ over all possible pairs of values must be 1

Outline

Conditional Probability Distribution

Conditional Probability Distribution:

Let X and Y be a pair of jointly distributed discrete random variables

The conditional probability distribution of the random variable Y, given that the random variable X takes the value x, expresses the probability that Y takes the value y, as a function of y, when the value x is fixed for X

This is denoted $P(y|x)$, and so, by the definition of conditional probability, is as follows:

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

Similarly, the conditional probability distribution of X, given $Y = y$, is as follows:

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

Conditional Probability Distribution

Conditional Probability Distribution (Example):

Reconsider the table:

	Male	Female	$P(y)$
Current	0.17	0.11	0.28
Former	0.33	0.17	0.50
Never	0.11	0.11	0.22
$P(x)$	0.61	0.39	1

For example, what is the probability that an employee who is a male is also a former club member?

$$P(former|male) = \frac{P(former, male)}{P(male)} = \frac{0.33}{0.61} = 0.55$$

What is the probability that an employee who is a female is also a current club member?

$$P(current|female) = \frac{P(current, female)}{P(female)} = \frac{0.11}{0.39} = 0.29$$

Conditional Probability Distribution

Conditional Probability Distribution (Example in R):

R can calculate all the conditional probabilities for us:

```
prop.table(volley, margin = 2)
```

```
##         Male Female
## Current 0.27   0.29
## Former  0.55   0.43
## Never   0.18   0.29
```

Assuming $X = (male, female)$, and $Y = (current, former, never)$. The above command calculates all conditional probabilities when X is given, e.g., our earlier example said: given that we have a male, what is the probability that he is a former member. The answer is 0.55

```
prop.table(volley, margin = 1)
```

```
##         Male Female
## Current 0.60   0.40
## Former  0.67   0.33
## Never   0.50   0.50
```

Assuming $X = (male, female)$, and $Y = (current, former, never)$. The above command calculates all conditional probabilities when Y is given.

Conditional mean and Variance

Conditional mean and Variance:

Mean: The conditional expectation of Y given $X = x$:

$$\mu_{Y|X} = E[Y|X] = \sum_{y} (y|x)P(y|x)$$

Variance: The conditional variance of Y given $X = x$:

$$\sigma^2_{Y|X} = Var[Y|X] = \sum_{y} ((y - \mu_{Y|X})^2|x)P(y|x)$$

Assume $X = (x = 1, x = 2)$ and $Y = (y = 1, y = 2, y = 3)$. We have the following joint probability distribution:

	X		
Y	1	2	P(y)
1	0.17	0.11	0.28
2	0.33	0.17	0.50
3	0.11	0.11	0.22
$P(x)$	0.61	0.39	1

What is the mean and variance of Y, when $X = 2$?

Conditional mean and Variance

Conditional mean and Variance (cont):

- **Solution**
- **Conditional Mean:** We need to calculate the conditional mean of Y ($E[Y|x=2]$), given $x=2$

$$E[Y|x=2] = \sum_y (y|x=2)P(y|x=2) = 1\frac{(0.11)}{0.39} + 2\frac{(0.17)}{0.39} + 3\frac{(0.11)}{0.39} = 2$$

- **Conditional Variance:** We now calculate the conditional variance of Y, given $x=2$

$$\sigma^2_{Y|x=2} = \sum_y ((y-2)^2|x=2)P(y|x=2)$$
$$= (1-2)^2\frac{0.11}{0.39} + (2-2)^2\frac{0.17}{0.39} + (3-2)^2\frac{0.11}{0.39} = 0.56$$

Conditional mean and Variance (cont):

Solution in R:

- First create the matrix of probabilities and convert it into a table in R:

```
data <- matrix(c(0.17, 0.11,
                 0.33, 0.17,
                 0.11, 0.11),
                 ncol = 2, byrow = TRUE)
# Define column names
colnames(data) <- c("1", "2")
# Define row names
rownames(data) <- c("1", "2", "3")
data <- as.table(data)
data

##      1    2
## 1 0.17 0.11
## 2 0.33 0.17
## 3 0.11 0.11
```

- Now calculate marginal probabilities of X and Y:

```
table <- addmargins(prop.table(data))
table

##        1    2  Sum
## 1   0.17 0.11 0.28
## 2   0.33 0.17 0.50
## 3   0.11 0.11 0.22
## Sum 0.61 0.39 1.00
```

Conditional mean and Variance (cont):

Solution in R:

- Create the vectors that you are going to need:
  - What are the possible values taken by Y?

```
y=c(1,2,3)
```

  - What are the conditional probabilities of Y when $X = 2$?

```
py2=c(0.11, 0.17, 0.11)
```

  - What is the marginal probability of Y when $X = 2$? [0.39]
  - So, we can calculate the conditional mean of Y when X=2 as follows:

```
mean_y= sum(y*py2/0.39)
mean_y

## [1] 2
```

Conditional mean and Variance (cont):

Solution in R:

- Now we can calculate the conditional variance of Y when X=2

```
var_y=sum((y-mean_y)^2*(py2)/0.39)
var_y

## [1] 0.564
```

Statistical independence and probability distributions

Statistical independence and Joint probability:

X and Y are random variables:

$P(x, y)$	Joint probability distribution of X and Y
$P(x)$	Marginal probability distribution function for X
$P(y)$	Marginal probability distribution function for Y

X and Y are independent if:

- $P(x, y) = P(x).P(y)$

Statistical independence and Conditional probability:

For conditional probabilities the independence of X and Y implies:

- $P(x|y) = P(x)$ and $P(y|x) = P(y)$

Outline

Mean, variance and covariance of joint probability dist.

Mean of variables with joint probability dist.

X and Y are random variables with joint probability distribution $P(x, y)$

The expected value of X:

$$E[X] = \sum_x xP(x)$$

Similarly the expected value of Y:

$$E[Y] = \sum_y yP(y)$$

	X			
Y	2	4	6	$P(y)$
1	0.112	0.074	0.224	0.411
2	0.112	0.074	0.074	0.262
3	0.104	0.134	0.086	0.325
$P(x)$	0.329	0.284	0.385	1

Mean of X is $E[X] = \sum_x xP(x) = 2(0.329) + 4(0.284) + 6(0.385) = 4.10$

Mean of Y is $E[Y] = \sum_y yP(y) = 1(0.411) + 2(0.262) + 3(0.325) = 1.92$

Mean of variables with joint probability dist.

Solution in R:

- First create the matrix of probabilities and convert it into a table in R:

```
data <- matrix(c(0.112, 0.074, 0.224,
                 0.112, 0.074, 0.074,
                 0.104, 0.134, 0.086),
            ncol = 3, byrow = TRUE)
# Define column names
colnames(data) <- c("2", "4", "6")
# Define row names
rownames(data) <- c("1", "2", "3")
data <- as.table(data)
prop.table(data)

##        2      4      6
## 1 0.1127 0.0744 0.2254
## 2 0.1127 0.0744 0.0744
## 3 0.1046 0.1348 0.0865
```

- Now calculate marginal probabilities of X and Y:

```
table <- addmargins(prop.table(data))
table

##          2      4      6    Sum
## 1   0.1127 0.0744 0.2254 0.4125
## 2   0.1127 0.0744 0.0744 0.2616
## 3   0.1046 0.1348 0.0865 0.3260
## Sum 0.3300 0.2837 0.3863 1.0000
```

Mean of variables with joint probability dist.

Solution in R:

- Calculate the mean of Y
  - What are the possible values taken by Y?

```
y=c(1,2,3)
```

  - What are the marginal probabilities of Y?

```
py=c(0.44, 0.26, 0.32)
```

  - Now simply calculate $E[Y] = \sum y(P_y)$ in R:

```
mean_y= sum(y*py)
mean_y

## [1] 1.92
```

- **Calculate the mean of X in this example both manually and in R?**

Mean, variance and covariance of joint probability dist.

Mean of variables with joint probability dist.

X and Y are random variables with joint probability distribution $P(x,y)$

The expected value of $g(X,Y)$:

$$E[g(X,Y)] = \sum_{y}\sum_{x} g(x,y)P(x,y)$$

Important properties:

$$E[X+Y] = E[X] + E[Y] = \mu_X + \mu_Y$$

Assume $W = aX + bY + c$:, and a, b and c are some constants, then the mean of W is:

$$E[W] = E[aX + bY + c] = aE[X] + bE[Y] + c = a\mu_X + b\mu_Y + c$$

Statistical independence and Mean of a random variable:

If X and Y are independent variables:

$$E(XY) = E(X)E(Y) = \mu_X\mu_Y$$

Mean, variance and covariance of joint probability dist.

Variance of variables with join probability dist.:

$$Var(X) = \sigma_X^2 = \sum_x (x - \mu_X)^2 P(x)$$

$$Var(Y) = \sigma_Y^2 = \sum_y (y - \mu_Y)^2 P(y)$$

Consider the table with joint probability dist. again:

	X			
Y	2	4	6	$P(y)$
1	0.112	0.074	0.224	0.411
2	0.112	0.074	0.074	0.262
3	0.104	0.134	0.086	0.325
$P(x)$	0.329	0.284	0.385	1

Mean of X is $E[X] = \sum_x xP(x) = 2(0.329) + 4(0.284) + 6(0.385) = 4.10$

Variance of X is
$Var(X) = \sum_x (x - \mu_X)^2 P(x) = (2-4.10)^2(0.329) + (4-4.10)^2(0.284) + (6-4.10)^2(0.385) = 2.84$

Variance of variables with joint probability:

Solution in R:

- Calculate the variance of Y
  - For variance, we need the mean of Y. Earlier, I calculated the mean and stored (see below):

```
mean_y

## [1] 1.92
```

  - I also stored the vector of marginal probabilities (see below):

```
py

## [1] 0.44 0.26 0.32
```

  - I can simply calculate the variance of Y using the formula $Var[y] = \sum(y - \mu_y)^2 P_y$:

```
sum((y-mean_y)^2*py)

## [1] 0.747
```

- **Calculate the variance of X in this example both manually and in R?**

Mean, variance and covariance of joint probability dist.

Variance of variables with join probability dist (cont):

$$Var(X+Y) = \sigma_X^2 + \sigma_Y^2 + 2Cov(X,Y)$$

Assume $W = aX + bY + c$:, and a, b and c are some constants, then the Variance of W is:

$$\sigma_W^2 = Var(aX + bY + c) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X,Y)$$

Variance and statistical independence:

X and Y are statistically independent variable:

Then $Var(X+Y) = Var(X) + Var(Y)$

Mean, variance and covariance of joint probability dist.

Covariance of variables with joint prob dist:

X and Y are random variables with joint probability $P(x, y)$

The covariance between X and Y

$$Cov(X, Y) = E(X - \mu_X)(Y - \mu_Y) = \sum_x \sum_y (x - \mu_X)(y - \mu_Y)P(x, y)$$

Note: $Cov(X, X) = Var(X) = E(X - \mu_X)(X - \mu_X)$

Assume $W = aX + bY$ and a, b are some constants, then the Covariance of X and Y is:

$$Cov(aX, bY) = ab.Cov(X, Y)$$

Also:

$$Cov(a + X, b + Y) = Cov(X, Y)$$

Note: that adding constants to X and Y does not affect the covariance between them

Covariance and statistical independence:

X and Y are statistically independent variable:

Then $Cov(X, Y) = 0$

Outline

Correlation

Correlation with joint prob dist:

X and Y are random variables:

The correlation between X and Y is:

$$\delta = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

The correlation is the covariance divided by the standard deviations of the two random variables

The correlation coefficient, provides a measure of the strength of the linear relationship between two random variables, with the measure being limited to the range from -1 to $+1$

$+1$ means perfect positive correlation, -1 means perfect negative correlation

A correlation closer to 0 means weaker correlation, and 0 means no correlation

Outline

Exercise:

- Assume $X = (x = 1, x = 3, x = 5, x = 7)$ and $Y = (y = 1, y = 2, y = 3, y = 4)$. We have the following joint probability distribution:

	X			
Y	1	3	5	7
1	0.066	0.044	0.132	0.066
2	0.044	0.044	0.033	0.044
3	0.121	0.099	0.084	0.073
4	0.026	0.053	0.042	0.024

- Recreate this table in R
- Calculate the marginal probabilities of X and Y
- Calculate the expected value of X and Y, i.e., $E[X]$ and $E[Y]$
- Given the condition that we have $x = 3$, what is the probability of getting y=2, i.e., Calculate $P(y = 2|x = 3)$?
- Calculate the conditional mean and variance of X when $y = 2$?
- Calculate the variance of X and Y?
- Calculate the covariance between X and Y