EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Philosophische Fakultät
Seminar für Sprachwissenschaft

Mathematical Models: Statistics

Peter Hendrix

Expectation

- What is expectation?
  - The expected value of a random variable
  - Definition:

$$\mathrm{E}[X] = \mathrm{mean}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Expectation

```
# Define grades
grades = c(6,7,8,6,7,9,7,8,5,7,7,8,5,9,4,6,10,6,7,8)

# Define sum grades
sum_grades = sum(grades)
sum_grades

# [1] 140

# Define length grades
length_grades = length(grades)
length_grades

# [1] 20
```


Expectation

```
# Calculate mean grade
mean_grade = 1/length_grades * sum_grades
mean_grade

# [1] 7
```

Expectation

- What is expectation?
  - The expected value of a random variable
  - Definition:

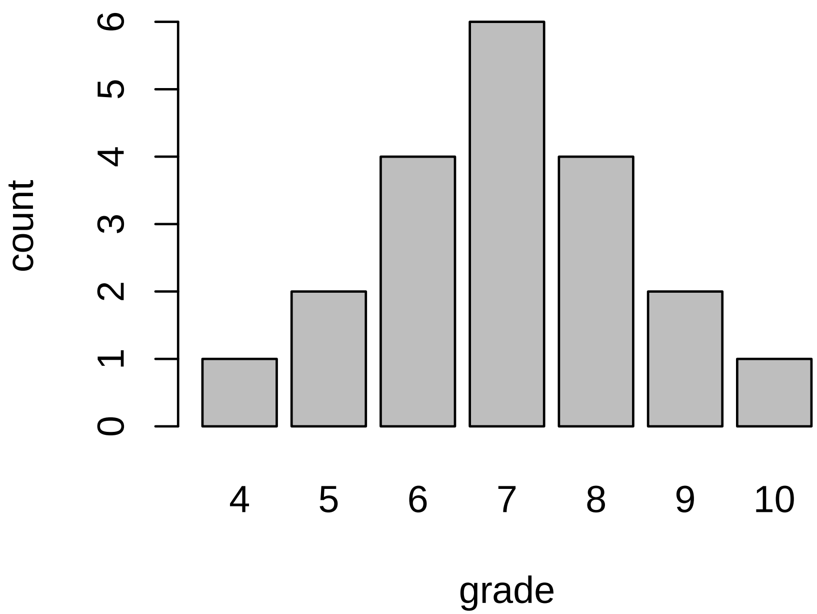
$$
\begin{aligned}
\mathrm{E}[X] &= \mathrm{mean}(X) \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i \\
&= \sum_{x} x * \Pr(X = x)
\end{aligned}
$$

Expectation

```
# Make a table of the grades
tab = table(grades)
tab

# grades
#  4  5  6  7  8  9 10
#  1  2  4  6  4  2  1
```

Expectation

```r
# Show a histogram
par(mar=c(4,4,2,4))
barplot(tab,xlab="grade",ylab="count")
```



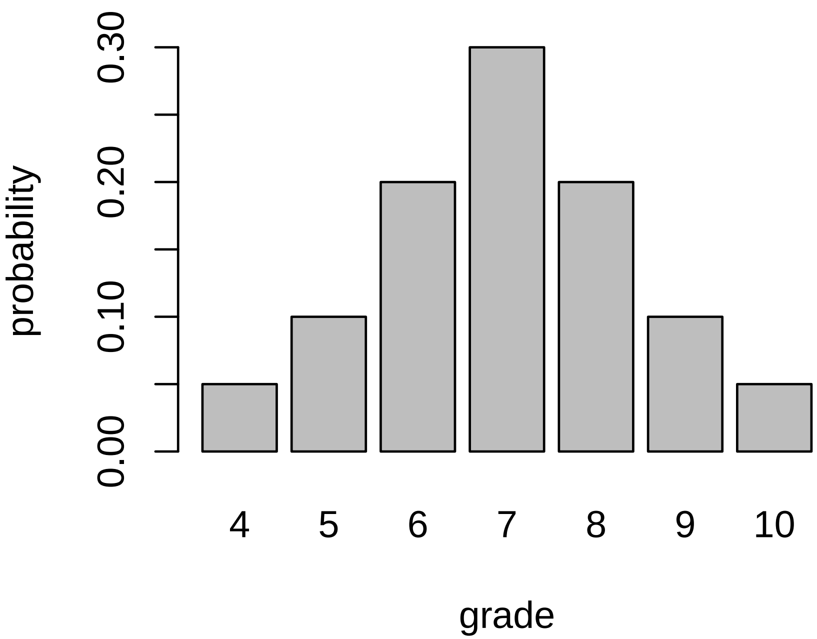
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Expectation

```
# Get the probabilities of the grades
prob_grades = tab/length_grades
prob_grades

# grades
#    4    5    6    7    8    9   10
# 0.05 0.10 0.20 0.30 0.20 0.10 0.05

sum(prob_grades)

# [1] 1
```

Expectation

```r
# Show a histogram
par(mar=c(4,4,2,4))
barplot(prob_grades,xlab="grade",ylab="probability")
```



Expectation

```r
# Get the unique grades
values = as.numeric(names(tab))
values

# [1]  4  5  6  7  8  9 10

# Get the mean grade
mean_grade = sum(values * prob_grades)
mean_grade

# [1] 7
```

Expectation

- What is expectation?
  - The function $f_X(x)$ that assigns the probabilities Pr(X = x) is known as the probability mass function of X
  - Definition:

$$
\begin{aligned}
\mathrm{E}[X] &= \mathrm{mean}(X) \\
&= \sum_x x * \Pr(X = x) \\
&= \sum_x x * f_X(x)
\end{aligned}
$$

Spread

- To what extent do individual values differ from the expected value?
- Different potential definitions

Spread

```
# Grades
grades = c(6,7,8,6,7,9,7,8,5,7,7,8,5,9,4,6,10,6,7,8)

# Mean grade
mean_grade = mean(grades)
mean_grade

# [1] 7
```

Spread

```
# Grades
grades

#  [1]  6  7  8  6  7  9  7  8  5  7  7  8  5  9  4  6 10
# [18]  6  7  8

differences = grades - mean_grade
differences

#  [1] -1  0  1 -1  0  2  0  1 -2  0  0  1 -2  2 -3 -1  3
# [18] -1  0  1

mean(differences)

# [1] 0
```

Spread

```
# Grades
grades

#  [1]  6  7  8  6  7  9  7  8  5  7  7  8  5  9  4  6 10
# [18]  6  7  8

differences = abs(grades - mean_grade)
differences

#  [1] 1 0 1 1 0 2 0 1 2 0 0 1 2 2 3 1 3 1 0 1

mean(differences)

# [1] 1.1
```

Spread

```
# Grades
grades

#  [1]  6  7  8  6  7  9  7  8  5  7  7  8  5  9  4  6 10
# [18]  6  7  8

differences = (grades - mean_grade)^2
differences

#  [1] 1 0 1 1 0 4 0 1 4 0 0 1 4 4 9 1 9 1 0 1

mean(differences)

# [1] 2.1
```

Variance

```r
# Grades
variance = mean(differences)
variance

# [1] 2.1

sdev = sqrt(variance)
sdev

# [1] 1.449138
```

Variance

- Variance is a measure of spread
    - Definition:

$$\begin{aligned}\text{Var}[X] &= \textit{mean}((X-\mu)^2)\\ &= \text{E}[(X-\text{E}[X])^2]\end{aligned}$$

Note: $\mu = \text{mean}(X)$

Variance

- Variance is a measure of spread
  - Definition:

$$
\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}[(X - \mathrm{E}[X])^2] \\
&= \mathrm{E}[(X^2 - 2 * X * \mathrm{E}[X] + \mathrm{E}[X]^2)] \\
&= \mathrm{E}[X^2] - \mathrm{E}[2 * X * \mathrm{E}[X]] + \mathrm{E}[\mathrm{E}[X]^2] \\
&= \mathrm{E}[X^2] - 2 * \mathrm{E}[X] * \mathrm{E}[X] + \mathrm{E}[X]^2 \\
&= \mathrm{E}[X^2] - 2 * \mathrm{E}[X]^2 + \mathrm{E}[X]^2 \\
&= \mathrm{E}[X^2] - \mathrm{E}[X]^2
\end{aligned}
$$

Variance

```
# Grades
grades = c(6,7,8,6,7,9,7,8,5,7,7,8,5,9,4,6,10,6,7,8)

# Calculate variance
variance = mean(grades^2) - mean(grades)^2
variance

# [1] 2.1
```

Law of the unconscious statistician (LOTUS)

- What is LOTUS?
  - LOTUS provides $\mathrm{E}[Y]$ given that $Y = g(X)$
  - Definition:

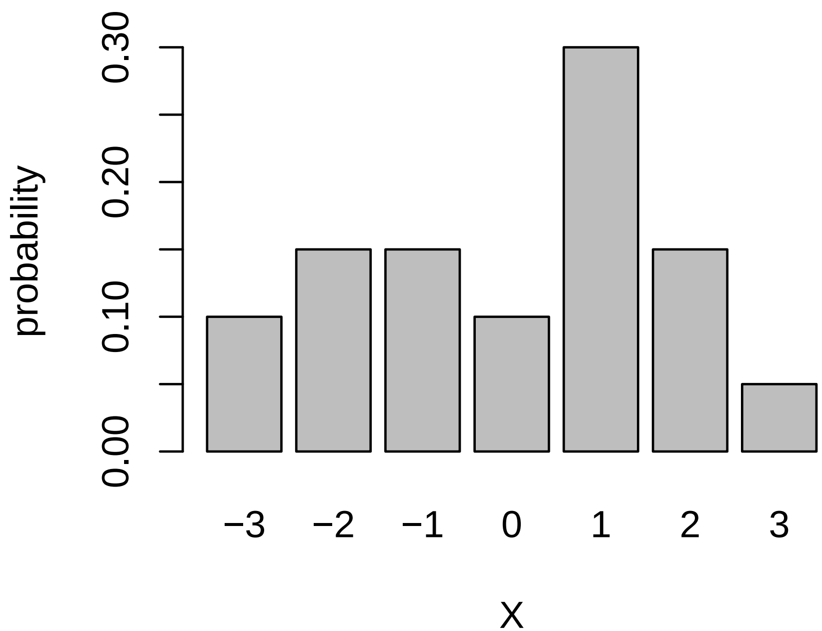
$$\mathrm{E}[Y] = \sum_{x} g(x) * f_X(x)$$

LOTUS

```r
# Define X
X = c(-1,2,1,2,-2,1,3,1,-3,1,0,-3,-1,1,-1,-2,1,-2,2,0)

# Make a table of X
tab = table(X)

# Get values of X
valuesX = as.numeric(names(tab))

# Get probability mass function of X
probX = tab/length(X)
```

LOTUS

```
# Show a histogram
par(mar=c(4,4,2,4))
barplot(probX,xlab="X",ylab="probability")
```



LOTUS

```
# Calculate E[X]
expX = sum(valuesX*probX)
expX

# [1] -2.775558e-17

round(expX,10)

# [1] 0
```

LOTUS: example

- We now define $g(X)$:

$$Y = g(X) = X^2$$

- Applying LOTUS gives:

$$\begin{aligned} \mathrm{E}[Y] &= \sum_x g(x) * f_X(x) \\ &= \sum_x x^2 * f_X(x) \end{aligned}$$

LOTUS: example

```
# Define Y
valuesX

# [1] -3 -2 -1  0  1  2  3

valuesY = valuesX^2
valuesY

# [1] 9 4 1 0 1 4 9

# Calculate expected value of Y
expY = sum(valuesY * probX)
expY

# [1] 3
```

Law of the unconscious statistician (LOTUS)

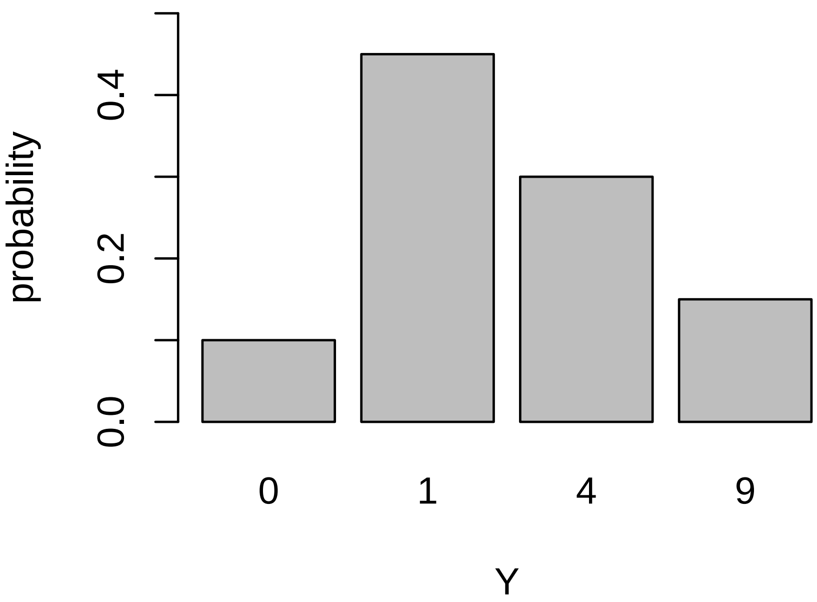
- We calculated $\mathrm{E}[Y]$ without knowing the probability mass function of Y
- Without LOTUS, we would first have to calculate this probability mass function $g_Y(y)$

Without LOTUS

```
# Grades
Y = X^2

# Make a table of Y
tab = table(Y)

# Get values of Y
valuesY = as.numeric(names(tab))

# Get probability mass function of Y
probY = tab/length(Y)
```

Without LOTUS

```
# Show a histogram
par(mar=c(4,4,2,4))
barplot(probY,xlab="Y",ylab="probability",ylim=c(0,0.5))
```



Without LOTUS

```
# Show the probability mass function of Y
probY

# Y
#    0    1    4    9
# 0.10 0.45 0.30 0.15

# Calculate E[X]
expY = sum(valuesY*probY)
expY

# [1] 3
```

LOTUS

- Proof of LOTUS:

$$
\begin{aligned}
\mathrm{E}[Y] &= \sum_{x} g(x) * f_X(x) \\
&= \sum_{y} \sum_{x:g(x)=y} g(x) * f_X(x) \\
&= \sum_{y} \sum_{x:g(x)=y} y * f_X(x) \\
&= \sum_{y} y * \sum_{x:g(x)=y} f_X(x) \\
&= \sum_{y} y * f_Y(y)
\end{aligned}
$$

Linearity of expectation

- Expectation is linear:

$$\mathrm{E}[a * X + b] = a * E[X] + b$$

- Special case of LOTUS where $g(x)$ is linear

Linearity of expectation

```
# Grades
grades = c(6,7,8,6,7,9,7,8,5,7,7,8,5,9,4,6,10,6,7,8)

# Mean
mean(grades)

# [1] 7
```

Linearity of expectation

```
# Show old grades
grades

#  [1]  6  7  8  6  7  9  7  8  5  7  7  8  5  9  4  6 10
# [18]  6  7  8

# Calculate new grades
new_grades = 0.9 * grades + 1

# Show new grades
new_grades

#  [1]  6.4  7.3  8.2  6.4  7.3  9.1  7.3  8.2  5.5  7.3
# [11]  7.3  8.2  5.5  9.1  4.6  6.4 10.0  6.4  7.3  8.2
```

Linearity of expectation

```
# E[X]
mean(grades)

# [1] 7

# E[a * X + b]
mean(new_grades)

# [1] 7.3

# Show that E[a * X + b] = a * E[X] + b
0.9 * mean(grades) + 1

# [1] 7.3
```

Linearity of expectation

- Proof:

$$\begin{aligned}
\mathrm{E}[a * X + b] &= \sum_x (a * x + b) * f_X(x) \\
&= \sum_x (a * x) * f_X(x) + \sum_x b * f_X(x) \\
&= a * \sum_x x * f_X(x) + b * \sum_x * f_X(x) \\
&= a * \sum_x x * f_X(x) + b \\
&= a * \mathrm{E}[x] + b
\end{aligned}$$

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Non-linearity of variance

- Variance is not linear:

$$\text{Var}[a * X + b] \neq a * Var[X](+b)$$

- What is $\text{Var}[a * X + b]$ then?

Non-linearity of variance

```
# Var[X]
varX = mean((grades- mean(grades))^2)
varX

# [1] 2.1

# Var[a * X + b]
var_newX = mean((new_grades- mean(new_grades))^2)
var_newX

# [1] 1.701
```

Non-linearity of variance

```r
# Var[X]
varX

# [1] 2.1

# Var[a * X + b]
var_newX

# [1] 1.701

# Var[a * X + b] = a^2 * Var[x]
0.9^2 * varX

# [1] 1.701
```

Non-linearity of variance

- Proof:

$$
\begin{aligned}
\text{Var}[a * X + b] &= \text{E}[((a * x + b) - E[a * X + b])^2] \\
&= \text{E}[((a * x + b) - (a * E[X] + b))^2] \\
&= \text{E}[(a * x - a * E[X])^2] \\
&= \text{E}[(a * (x - E[X])^2] \\
&= a^2 * \text{E}[(x - E[X])^2] \\
&= a^2 * \text{Var}[X]
\end{aligned}
$$

Joint distributions

- Probability mass functions for multiple variables:

$$f_{XY}(x, y)$$

where:

$$\begin{aligned} \Pr(A, B) &= f_{XY}(X = A, Y = B) \\ &= \Pr(A \cap B) \end{aligned}$$

- Example: baby names in the U.S. in 2016
(data from the Social Security Administration)

Joint distributions

```
# Load data
names = read.table("data/us_names.txt",T)
head(names,n=8)

#   Rank NameMale NumMale NameFemale NumFemale
# 1    1     Noah  19,015       Emma    19,414
# 2    2     Liam  18,138     Olivia    19,246
# 3    3  William  15,668        Ava    16,237
# 4    4    Mason  15,192     Sophia    16,070
# 5    5    James  14,776   Isabella    14,722
# 6    6 Benjamin  14,569        Mia    14,366
# 7    7    Jacob  14,416  Charlotte    13,030
# 8    8  Michael  13,998    Abigail    11,699
```


Joint distributions

f_{XY}	boy	girl	f_X
consonant	0.41	0.14	0.55
vowel	0.10	0.35	0.45
f_Y	0.51	0.49	1.00

$$\Pr(X = consonant, Y = boy) = 0.41$$
$$\Pr(X = vowel, Y = girl) = 0.35$$
$$\Pr(X = consonant) = 0.41 + 0.14 = 0.55$$
$$\Pr(Y = boy) = 0.41 + 0.10 = 0.51$$

Joint distributions

- Probability mass function for X:

$$f_X(x) = \sum_y f_{XY}(x, y)$$

- Probability mass function for Y:

$$f_Y(y) = \sum_x f_{XY}(x, y)$$

Expectation of joint distributions

- LOTUS also works for joint distributions
  - We define Z as $g(x, y)$, a function of both X and Y
  - Definition:

$$E[Z] = \sum_y \sum_x g(x, y) * f_{XY}(x, y)$$

Expectation of joint distributions

- Special case: addition
  - Definition:

$$Z = X + Y$$

- What is the expected value of Z, $\mathrm{E}[Z]$?

Expectation of joint distributions

$$E[Z] = \sum_Y \sum_X g(x, y) * f_{XY}(x, y)$$

$$E[Z] = \sum_Y \sum_X (x + y) * f_{XY}(x, y)$$

$$E[Z] = \sum_Y \sum_X x * f_{XY}(x, y) + \sum_Y \sum_X y * f_{XY}(x, y)$$

$$E[Z] = \sum_X x \sum_Y f_{XY}(x, y) + \sum_Y y \sum_X f_{XY}(x, y)$$

$$E[Z] = \sum_X x * f_X(x) + \sum_Y y * f_Y(y)$$

$$E[Z] = \mathrm{E}[X] + \mathrm{E}[Y]$$

Expectation of joint distributions

Expectation of joint distributions

```
# Define the sample space for a 6-sided die
die1 = 1:6

# Define the probability mass function for a 6-sided die
prob_die1 = table(die1)/length(die1)
round(prob_die1,6)

# die1
#        1        2        3        4        5        6
# 0.166667 0.166667 0.166667 0.166667 0.166667 0.166667
```

Expectation of joint distributions

```
# Get the expected value of a 6-sided die
exp_die1 = sum(die1 * prob_die1)
exp_die1

# [1] 3.5
```

Expectation of joint distributions

```
# Define the sample space for an 8-sided die
die2 = 1:8

# Define the probability mass function for an 8-sided die
prob_die2 = table(die2)/length(die2)
round(prob_die2,6)

# die2
#     1     2     3     4     5     6     7     8
# 0.125 0.125 0.125 0.125 0.125 0.125 0.125 0.125
```

Expectation of joint distributions

```
# Get the expected value of a 8-sided die
exp_die2 = sum(die2 * prob_die2)
exp_die2

# [1] 4.5
```

Expectation of joint distributions

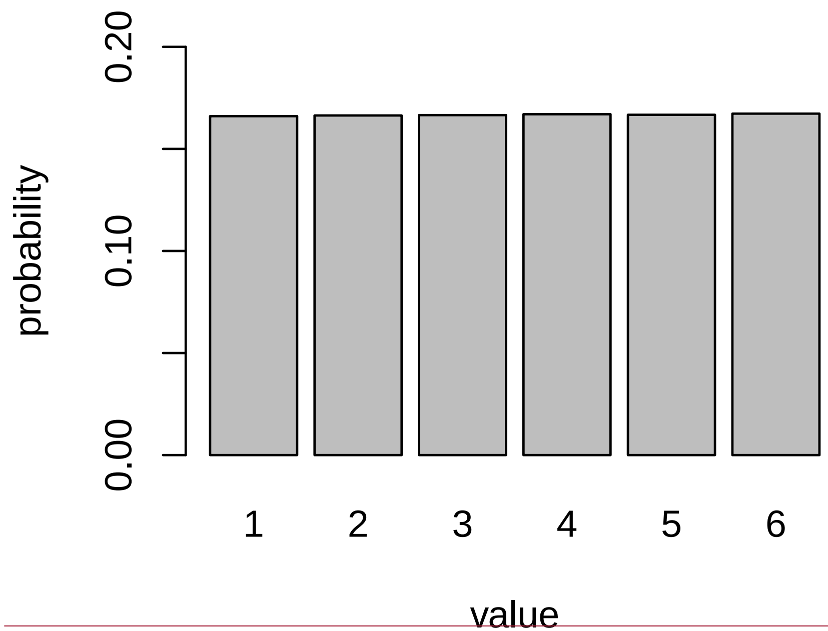
```
# Get the expected value of the sum of both dice
exp_dice = exp_die1 + exp_die2
exp_dice

# [1] 8
```

Expectation of joint distributions

```
# Let's verify that E[X+Y] = E[X] + E[Y]
# Set a seed for replication
set.seed(31415)

# Throw both dice a million times
throws_die1 = sample(1:6,1000000,replace=TRUE)
head(throws_die1,n=20)

#  [1] 6 3 1 3 1 3 2 3 6 6 1 5 5 1 5 2 1 6 6 4

throws_die2 = sample(1:8,1000000,replace=TRUE)
head(throws_die2,n=20)

#  [1] 6 8 1 6 6 5 7 6 3 2 6 3 7 7 5 8 8 8 1 2
```

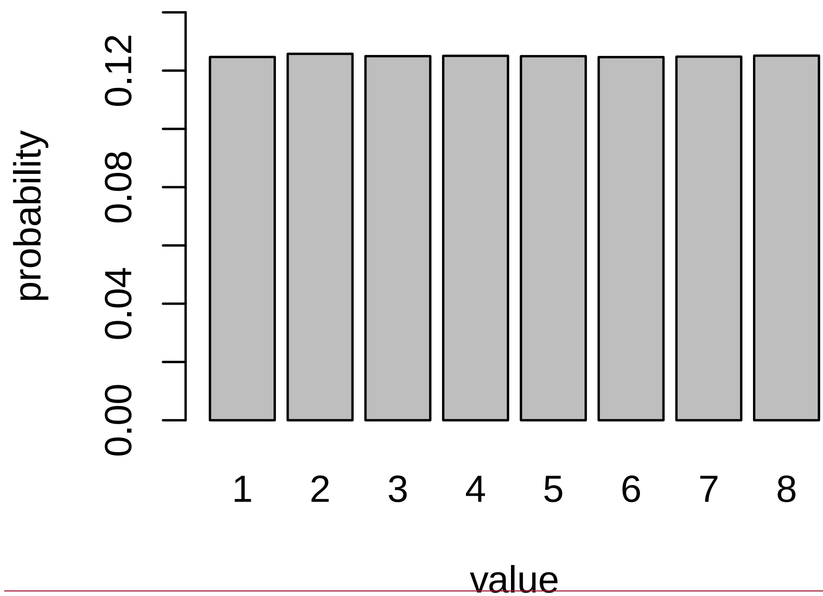
Expectation of joint distributions

```
# Show a histogram for die1
par(mar=c(4,4,2,4))
barplot(table(throws_die1)/1000000,xlab="value",
        ylab="probability",ylim=c(0,0.20))
```



Expectation of joint distributions

```
# Show a histogram for die2
par(mar=c(4,4,2,4))
barplot(table(throws_die2)/1000000,xlab="value",
        ylab="probability",ylim=c(0,0.14))
```



Expectation of joint distributions

```
# Repeat for convenience
head(throws_die1,n=15)

#  [1] 6 3 1 3 1 3 2 3 6 6 1 5 5 1 5

head(throws_die2,n=15)

#  [1] 6 8 1 6 6 5 7 6 3 2 6 3 7 7 5

# Define X + Y: the sum of the dice for each throw
sum_dice = throws_die1 + throws_die2
head(sum_dice,n=15)

#  [1] 12 11  2  9  7  8  9  9  9  8  7  8 12  8 10
```

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Expectation of joint distributions

```
# Repeat analytically derived E[X + Y]
exp_dice = exp_die1 + exp_die2
exp_dice

# [1] 8

# Calculate observed E[X + Y]
mean(sum_dice)

# [1] 8.002453
```

Expectation of joint distributions

- Generalization:

$$E\left[\sum_i X_i\right] = \sum iE\left[X_i\right]$$

Variance of joint distributions

- Variances are not additive

$$\mathrm{Var}[X+Y] \neq \mathrm{Var}[X] + \mathrm{Var}[Y]$$

What is $\mathrm{Var}[X+Y]$?

Variance of joint distributions

$$
\begin{aligned}
\mathrm{Var}[X+Y] &= \mathrm{E}[(X+Y-\mathrm{E}[X+Y])^2] \\
&= \mathrm{E}[(X+Y-\mathrm{E}[X]-\mathrm{E}[Y])^2] \\
&= \mathrm{E}[((X-\mathrm{E}[X])+(Y-\mathrm{E}[Y]))^2] \\
&= \mathrm{E}[(X-\mathrm{E}[X])^2+(Y-\mathrm{E}[Y])^2+ \\
&\qquad +2*(X-\mathrm{E}[X])*(Y-\mathrm{E}[Y])] \\
&= \mathrm{E}[(X-\mathrm{E}[X])^2]+\mathrm{E}[(Y-\mathrm{E}[Y])^2]+ \\
&\qquad +2*\mathrm{E}[(X-\mathrm{E}[X])*(Y-\mathrm{E}[Y])] \\
&= \mathrm{Var}[X]+\mathrm{Var}[Y]+2*\mathrm{Cov}[X,Y]
\end{aligned}
$$

Covariance of joint distributions

- What is covariance?
  - Definition:

$$\begin{aligned}\mathrm{Cov}[X, Y] &= \mathrm{E}[(X - \mathrm{E}[X]) * (Y - \mathrm{E}[Y])] \\ &= \frac{1}{n} * \sum_{i=1}^{n} (x_i - \mu_x) * (y_i - \mu_y)\end{aligned}$$

- Covariance is a measure of how much two variables change together

Covariance of joint distributions

- Covariance is zero when X and Y are independent
- In this case (and only in this case!):

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

Covariance of joint distributions

```
# Repeat for convenience
die1

# [1] 1 2 3 4 5 6

exp_die1

# [1] 3.5

# Calculate variance of 6-sided die
var_die1 = mean((die1 - exp_die1)^2)
var_die1

# [1] 2.916667
```

Covariance of joint distributions

```
# Repeat for convenience
die2

# [1] 1 2 3 4 5 6 7 8

exp_die2

# [1] 4.5

# Calculate variance of 8-sided die
var_die2 = mean((die2 - exp_die2)^2)
var_die2

# [1] 5.25
```

Covariance of joint distributions

```
# Given independence of die1 and die2 Var[X + Y] is
# defined as Var[X] + Var[Y]:
var_dice = var_die1 + var_die2
var_dice

# [1] 8.166667

# Verify for our million roll sample
var_sample = mean((sum_dice - mean(sum_dice))^2)
var_sample

# [1] 8.165507
```

Covariance of joint distributions

```
# Cov[X,Y] should be zero
# Cov[X,Y] = E[(X-E[X])*(Y-E[Y])]
cov_xy = mean(
            (throws_die1 - mean(throws_die1)) *
            (throws_die2 - mean(throws_die2))
          )
cov_xy

# [1] -0.0005393453

# Variation in X provides no information about
# variation in Y
```

Conditional probability

- Conditional probability is the probability of $X = x$ given $Y = y$
- Definition:

$$\Pr(X = x|Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$

Conditional probability

f_{XY}	1	2	3	4	f_X
1	0.05	0.05	0.15	0.05	0.30
2	0.00	0.00	0.05	0.10	0.15
3	0.10	0.05	0.10	0.05	0.30
4	0.05	0.00	0.20	0.00	0.25
f_Y	0.20	0.10	0.50	0.20	1.00

Conditional probability

f_{XY}	1	2	3	4	f_X
1	0.05	0.05	0.15	0.05	0.30
2	0.00	0.00	0.05	0.10	0.15
3	0.10	0.05	0.10	0.05	0.30
4	0.05	0.00	0.20	0.00	0.25
f_Y	0.20	0.10	0.50	0.20	1.00

$$\Pr(X = 2|Y = 4) = \frac{\Pr(X=2, Y=4)}{\Pr(Y=4)} = \frac{0.10}{0.20} = 0.50$$

Conditional probability

f_{XY}	1	2	3	4	f_X
1	0.05	0.05	0.15	0.05	0.30
2	0.00	0.00	0.05	0.10	0.15
3	0.10	0.05	0.10	0.05	0.30
4	0.05	0.00	0.20	0.00	0.25
f_Y	0.20	0.10	0.50	0.20	1.00

$$\Pr(Y = 1 | X = 3) = \frac{\Pr(Y=1, X=3)}{\Pr(X=3)} = \frac{0.10}{0.30} = 0.33$$

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Conditional expectation

- What is the conditional expectation $E[X|Y=y]$?
- Definition:

$$\begin{aligned} \mathrm{E}[X|Y=y] &= \sum_x x * \Pr(X=x|Y=y) \\ &= \sum_x x * \frac{\Pr(X=x, Y=y)}{\Pr(Y=y)} \end{aligned}$$

Conditional expectation

f_{XY}	1	2	3	4	f_X
1	0.05	0.05	0.15	0.05	0.30
2	0.00	0.00	0.05	0.10	0.15
3	0.10	0.05	0.10	0.05	0.30
4	0.05	0.00	0.20	0.00	0.25
f_Y	0.20	0.10	0.50	0.20	1.00

$$\begin{aligned} \mathrm{E}(X|Y=1) &= 1 * \frac{0.05}{0.20} + 2 * \frac{0.00}{0.20} + 3 * \frac{0.10}{0.20} + 4 * \frac{0.05}{0.20} \\ &= 2.75 \end{aligned}$$

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Conditional expectation

- How about the expectation of the conditional expectations $E[X|Y]$?

- Definition:

$$\begin{aligned} \mathrm{E}[\mathrm{E}[X|Y]] &= \sum_y \Pr(Y = y) * \mathrm{E}[X|Y = y] \\ &= \sum_y \Pr(Y = y) * \sum_x x * \Pr(X = x|Y = y) \end{aligned}$$

Conditional expectation

f_{XY}	1	2	3	4	f_X
1	0.05	0.05	0.15	0.05	0.30
2	0.00	0.00	0.05	0.10	0.15
3	0.10	0.05	0.10	0.05	0.30
4	0.05	0.00	0.20	0.00	0.25
f_Y	0.20	0.10	0.50	0.20	1.00

$$
\begin{aligned}
E(X|Y=2) &= 1 * \frac{0.05}{0.10} + 2 * \frac{0.00}{0.10} + 3 * \frac{0.05}{0.10} + 4 * \frac{0.00}{0.10} \\
&= 2.00
\end{aligned}
$$

Conditional expectation

f_{XY}	1	2	3	4	f_X
1	0.05	0.05	0.15	0.05	0.30
2	0.00	0.00	0.05	0.10	0.15
3	0.10	0.05	0.10	0.05	0.30
4	0.05	0.00	0.20	0.00	0.25
f_Y	0.20	0.10	0.50	0.20	1.00

$$\mathrm{E}(X|Y=3) = 1 * \frac{0.15}{0.50} + 2 * \frac{0.05}{0.50} + 3 * \frac{0.10}{0.50} + 4 * \frac{0.20}{0.50}$$
$$= 2.70$$

Conditional expectation

f_{XY}	1	2	3	4	f_X
1	0.05	0.05	0.15	0.05	0.30
2	0.00	0.00	0.05	0.10	0.15
3	0.10	0.05	0.10	0.05	0.30
4	0.05	0.00	0.20	0.00	0.25
f_Y	0.20	0.10	0.50	0.20	1.00

$$
\begin{aligned}
E(X|Y=4) &= 1 * \frac{0.05}{0.20} + 2 * \frac{0.10}{0.20} + 3 * \frac{0.05}{0.20} + 4 * \frac{0.00}{0.20} \\
&= 2.00
\end{aligned}
$$

Conditional expectation

Given

$$\Pr(Y = 1) = 0.20, \mathrm{E}(X|Y = 1) = 2.75$$
$$\Pr(Y = 2) = 0.10, \mathrm{E}(X|Y = 2) = 2.00$$
$$\Pr(Y = 3) = 0.50, \mathrm{E}(X|Y = 3) = 2.70$$
$$\Pr(Y = 4) = 0.20, \mathrm{E}(X|Y = 4) = 2.00$$

we find that

$$\begin{aligned} \mathrm{E}[\mathrm{E}[X|Y]] &= 0.20 * 2.75 + 0.10 * 2.00 + 0.50 * 2.70 + 0.20 * 2.00 \\ &= 2.50 \end{aligned}$$

Conditional expectation

- The expectation of the expectations of the conditional probabilities $\mathrm{E}[X|Y]$ is the expectation of X
- Definition:

$$\mathrm{E}[\mathrm{E}[X|Y]] = \mathrm{E}[X]$$

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Conditional expectation

f_{XY}	1	2	3	4	f_X
1	0.05	0.05	0.15	0.05	0.30
2	0.00	0.00	0.05	0.10	0.15
3	0.10	0.05	0.10	0.05	0.30
4	0.05	0.00	0.20	0.00	0.25
f_Y	0.20	0.10	0.50	0.20	1.00

$$\begin{aligned} \mathrm{E}(X) &= 1 * 0.30 + 2 * 0.15 + 3 * 0.30 + 4 * 0.25 \\ &= 2.50 \end{aligned}$$

Conditional expectation

$$
\begin{aligned}
E[E[X|Y]] &= \sum_y \Pr(Y = y) * \sum_x x * \Pr(X = x|Y = y) \\
&= \sum_y \sum_x x * \Pr(Y = y) * \Pr(X = x|Y = y) \\
&= \sum_y \sum_x x * \Pr(Y = y) * \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)} \\
&= \sum_y \sum_x x * \Pr(X = x, Y = y) \\
&= \sum_x \sum_y x * \Pr(X = x, Y = y) \\
&= \sum_x x \sum_y \Pr(X = x, Y = y)
\end{aligned}
$$

Conditional expectation

$$
\begin{aligned}
\mathrm{E}[\mathrm{E}[X|Y]] &= \sum_x x \sum_y \Pr(X = x, Y = y) \\
&= \sum_x x * \Pr(X = x) \\
&= \mathrm{E}[X]
\end{aligned}
$$

Thank you