STAT 450/460

Handout 3: Multivariate distributions

Fall 2016

Chapter 5: Multivariate distributions

Up to this point, we have considered a single random variable Y, both discrete or continuous. Often, we are interested in the distribution of a *vector* of random variables, $\{Y_1, Y_2, ..., Y_k\} \in \mathcal{R}^k$. For example, if we have a sample of students, we might have:

- $Y_1 \equiv$ height (marginally normal?)
- $Y_2 \equiv$ sex (marginally binomial?)
- $Y_3 \equiv$ age (marginally normal?)
- $Y_4 \equiv$ income (marginally gamma?)

As we'll see later, "marginally" means when the variable is considered in isolation. In this simple example, $k = 4$. However, are often very interested not only in how variables behanve individually, but how they behave **together**. I.e., we are interested in the *joint* distributions of these random vectors. For example, how does age correlate with income? Multivariate distributions are extremely important in statistics, and critical to fundamental statistical concepts like regression.

Although k can be any dimension, for many of our examples we will consider the bivariate case where $k = 2$.

We'll begin our discussion by considering multivariate discrete distributions.

Multivariate discrete distributions

Definition: Let $\{Y_1, ..., Y_p\} \in \mathcal{R}^p$ be a $p-$vector of discrete random variables. Then the *joint probability function* for $\{Y_1, ..., Y_p\}$ is:

$$p(y_1, y_2, ..., y_p) = P(Y_1 = y_1, Y_2 = y_2, ..., Y_p = y_p); -\infty < y_1 < \infty; ...; -\infty < y_p < \infty$$

Properties:

1. $p(y_1, ..., y_p) \geq 0$ for all $y_1, ..., y_p$.
2. $\sum_{y_1} \sum_{y_2} ... \sum_{y_p} p(y_1, y_2, ..., y_p) = 1$

Marginal and conditional distributions

- **Marginal distribution.** The marginal distribution of Y_i is defined to be:

$$p_i(y_i) = \sum_{y_j : j \neq i} p(y_1, ..., y_p)$$

- **Joint distribution.** Analogously, the joint pmf of Y_i and Y_j is defined to be:

$$p(y_i, y_j) = \sum_{y_k : k \neq i,j} p(y_1, ..., y_p)$$

- **Conditional distribution.** The conditional distribution of $Y_1 | Y_2$ is defined to be:

$$p(y_1|y_2) = P(Y_1 = y_1 | Y_2 = y_2) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)}$$

From here, it follows that:

$$p(y_1, y_2) = p(y_1|y_2)p_2(y_2)$$

- **Conditional expectation.** The conditional expectation of $Y_1 | Y_2 = y_2$ is defined to be:

$$E(Y_1|Y_2 = y_2) = \sum_{all y_1} y_1 p(y_1|y_2)$$

- **Independence.** $Y_1, ..., Y_p$ are independent if and only if:

$$p(y_1, ..., y_p) = \prod_{i=1}^{p} p_i(y_i) \text{ (product of marginal pmfs)}$$

Furthermore, if Y_1 and Y_2 are independent, then:

$$p(y_1|y_2) = p_1(y_1)$$
$$p(y_2|y_1) = p_2(y_2)$$

EXAMPLE

For the bivariate case, we'll use X and Y instead of Y_1 and Y_2. Suppose the joint pmf of X and Y is:

$$p(x,y) = \begin{cases} \frac{1}{30}(x+y) & x = 0,1,2; y = 0,1,2,3 \\ 0 & otherwise \end{cases}$$

- Verify that this is a valid joint pmf.

- Find the marginals, $p_X(x)$ and $p_Y(y)$.

- Find $E(X)$ and $E(Y)$.

EXAMPLE Suppose a bag has 6 boxes. Three boxes have 3 darts, two of them have 4 darts, and one box has 5 darts. A player is told to pick a box at random, then shoot all the darts in the box at a target. Suppose the player is a 60% shooter, i.e., can hit the target 60% of the time. Let X be the number of darts in the box he picks, and Y the number of times the player hits the target. Find:

1. The joint distribution of X and Y, $p(x, y)$
2. How many times would you expect the player to hit the target, before he has even chosen any boxes?
3. Suppose you know the player has hit the target 3 times, but not how many darts were in the box he picked. Find the probability distribution of $p(x|y)$ and find $E(X|Y = 3)$. Are X and Y independent?

Multivariate Continuous Random Variables

Suppose we have a $k-$ dimensional random vector $\{Y_1, Y_2, ..., Y_k\}$ (*continuous OR discrete, or mixture*). Then, the CDF can be defined as follows:

$$F(y_1, y_2, ..., y_k) = P(Y_1 \leq y_1, Y_2 \leq y_2, ..., Y_k \leq y_k).$$

$\{Y_1, Y_2, ..., Y_k\}$ is said to be *jointly continuous* if there exists a nonnegative function $f(\cdot) \geq 0$ such that:

$$F(y_1, y_2, ..., y_k) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} ... \int_{-\infty}^{y_k} f(t_1, t_2, ..., t_k) dt_1 dt_2 ... dt_k.$$

The function $f(y_1, y_2, ..., y_k)$ is said to be the *joint probability density function*, or joint pdf.

Facts about the CDF (bivariate case):

If a function $F(y_1, y_2)$ is a bivariate CDF if:

1. $\lim_{y_1, y_2 \to -\infty} F(y_1, y_2) = 0$ $(F(-\infty, -\infty) = 0)$
2. $\lim_{y_1 \to -\infty} F(y_1, y_2) = 0 \; \forall y_2 \quad (F(-\infty, y_2) = 0)$
3. $\lim_{y_2 \to -\infty} F(y_1, y_2) = 0 \; \forall y_1 \quad (F(y_1, -\infty) = 0)$
4. $\lim_{y_1, y_2 \to \infty} F(y_1, y_2) = 1$ $(F(\infty, \infty) = 1)$
5. $F(\cdot, \cdot)$ is right-continuous: $\lim_{h \to 0^+} F(y_1 + h, y_2) = \lim_{h \to 0^+} F(y_1, y_2 + h) = F(y_1, y_2)$.

6. Marginal CDF: $F_1(y_1) = \lim_{y_2 \to \infty} = F(y_1, y_2)$; similarly $F_2(y_2) = \lim_{y_1 \to \infty} F(y_1, y_2)$.
7. Integral over a rectangular region: if $a > c$ and $b > d$, then $F(a, b) - F(a, d) - F(b, c) + F(c, d) \geq 0$

Picture:

Analogously to the discrete case, multivariate continuous random variables with joint pdfs have the following properties (reducing to bivariate case for simplicity):

1. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.
2. $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$; $f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$.
3. $f(x|y) = f(x, y)/f_X(x)$; $f(y|x) = f(x, y)/f_Y(y)$
4. X and Y are independent if and only if $f(x, y) = f_X(x) f_Y(y)$.

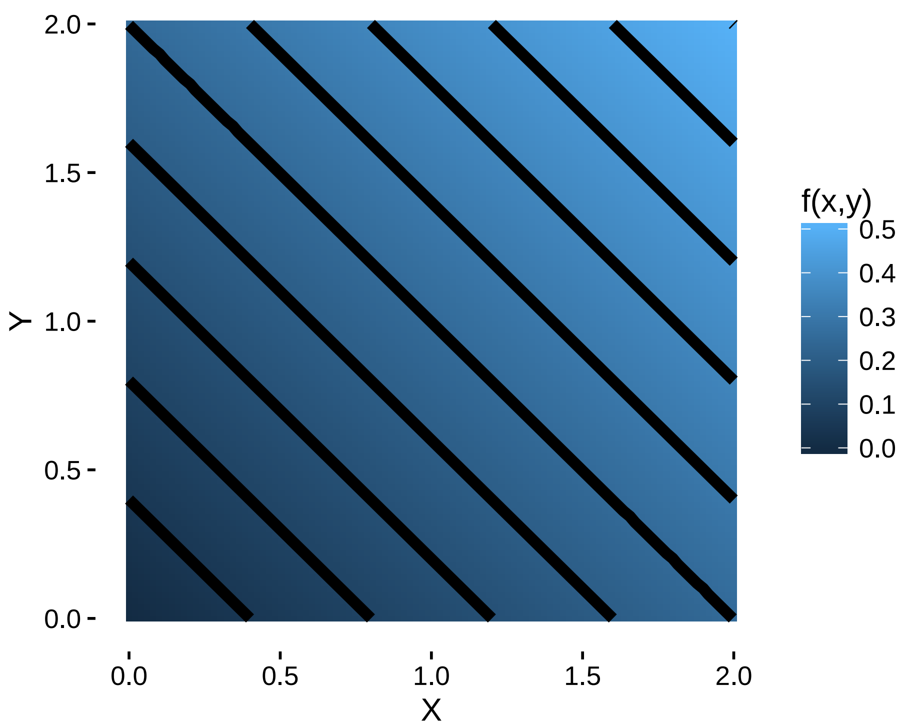
EXAMPLE

Let X and Y have the following joint pdf:

$$f(x,y) = \begin{cases} \frac{1}{8}(x+y) & 0 \leq x \leq 2; 0 \leq y \leq 2 \\ 0 & otherwise \end{cases}$$

Plotting this joint pdf using contours and rasters, we notice that it has the shape of a ramp: traveling in a straight diagonal from (0,0) to (2,2) is the steepest path of ascent, while we will stay flat if we walk, for example, from (2,0) to (0,2) in a straight diagonal.

```
library(ggplot2)
fxy <- function(x,y) {
 return((x+y)/8)
}
x <- seq(0,2,l=100)
y <- seq(0,2,l=100)
mydata <- expand.grid(x,y) #Get all x/y combinations
mydata$f <- fxy(mydata[,1],mydata[,2])
ggplot(aes(x=Var1,y=Var2), data=mydata) +  geom_raster(aes(fill=f)) +
  geom_contour(aes(z=f),color='black',size=2) + theme_classic() +
    scale_fill_continuous(name='f(x,y)') + ylab('Y') + xlab('X')
```



The region where $f(x,y) \geq 0$ is called the ***support***: in this case, the support is the Cartesian product $[0,2] \times [0,2]$

1. Show that this is a valid joint pdf.
2. Find the joint CDF.
3. Find the marginal pdfs and CDFs.
4. Are X and Y independent?
5. Find $f(x|y)$. Graph this pdf.

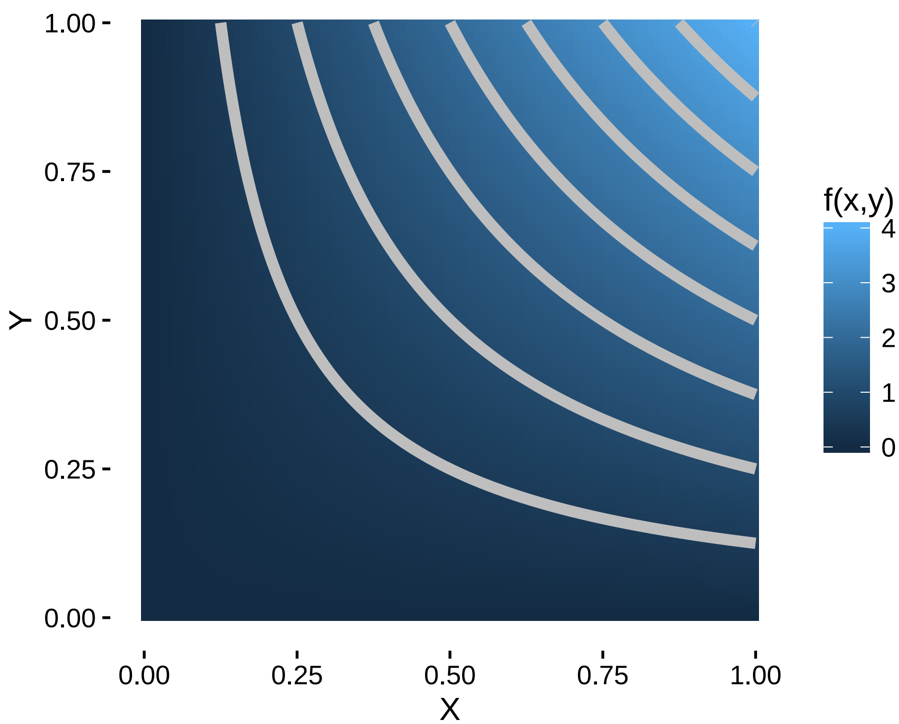
EXAMPLE

Let X and Y have the following joint pdf:

$$f(x,y) = \begin{cases} 4xy & 0 \leq x \leq 1; 0 \leq y \leq 1 \\ 0 & otherwise \end{cases}$$

Plot (it's like a convex skate ramp; note each contour line is a decrease of 0.5 in the height of $f(x,y)$):

```
library(ggplot2)
fxy <- function(x,y) {
 return(4*x*y)
}
x <- seq(0,1,l=100)
y <- seq(0,1,l=100)
mydata <- expand.grid(x,y) #Get all x/y combinations
mydata$f <- fxy(mydata[,1],mydata[,2])
ggplot(aes(x=Var1,y=Var2), data=mydata) +  geom_raster(aes(fill=f)) +
  geom_contour(aes(z=f),color='grey',size=2,binwidth=0.5) + theme_classic() +
    scale_fill_continuous(name='f(x,y)') + ylab('Y') + xlab('X')
```



1. Show that this is a valid joint pdf.
2. Find the joint CDF.
3. Find the marginal pdfs and CDFs.
4. Are X and Y independent?
5. Find $P(Y \leq 1/4|X = 1/2)$.

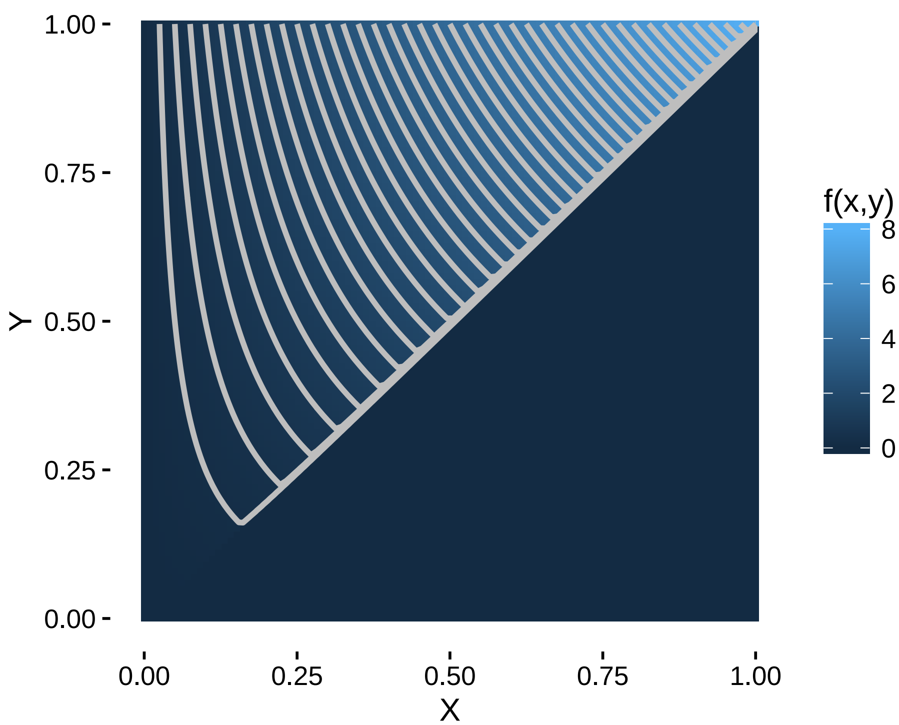
EXAMPLE

Let X and Y have the following joint pdf:

$$f(x,y) = \begin{cases} 8xy & 0 \le x \le y \le 1 \\ 0 & otherwise \end{cases}$$

Plot:

```
library(ggplot2)
fxy <- function(x,y) {
 z <- ifelse( 0 <=x & x <=y, 8*x*y,0)
 return(z)
}
x <- seq(0,1,l=100)
y <- seq(0,1,l=100)
mydata <- expand.grid(x,y) #Get all x/y combinations
mydata$f <- fxy(mydata[,1],mydata[,2])
ggplot(aes(x=Var1,y=Var2), data=mydata) +  geom_raster(aes(fill=f)) +
  geom_contour(aes(z=f),color='grey',size=1,binwidth=0.2) + theme_classic() +
    scale_fill_continuous(name='f(x,y)') + ylab('Y') + xlab('X')
```



1. Show that this is a valid joint pdf.
2. Find the joint CDF.
3. Find the marginal pdfs and CDFs.
4. Are X and Y independent?
5. Find $f(x|y)$ and $f(y|x)$. Use these to find $P(X > 1/4|Y = 1/2)$ and $P(Y < 3/4|X = 1/4)$.

EXAMPLE (5.13 and 5.31 in book)

Let X and Y have the following joint pdf:

$$f(x,y) = \begin{cases} 30xy^2 & x-1 \leq y \leq 1-x;\ 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

1. Plot the support.
2. Show that this is a valid joint pdf.
3. Define all possible regions for $F(x,y)$.
4. Find $F(1/2, 1/2)$; $F(1/2, 1)$; and $P(X > Y)$.
5. Find the marginal distribution of X; does it have a well-known form?
6. Find $P(Y > 0|X = 0.75)$

EXAMPLE (Le's GRE practice problem # 1)

Let X and Y be independent $U(0,1)$ random variables. Find the probability that the distance between X and Y is less than $1/2$.

EXAMPLE (Inspired by Le's GRE practice problem # 2)

Suppose

$$f(x,y) = \begin{cases} ke^{-y^2} & 0 \le x \le y < \infty \\ 0 & otherwise \end{cases}$$

1. Find the value of k that makes $f(x,y)$ a valid pdf.
2. What is the conditional distribution of $X|Y$? Verify with the following code:

```
fxy <- function(x,y) {
 z <- ifelse( 0 <=x & x <=y, 2*exp(-y^2),0)
 return(z)
}
x <- seq(0,1,l=100)
y <- seq(0,1,l=100)
mydata <- expand.grid(x,y) #Get all x/y combinations
mydata$f <- fxy(mydata[,1],mydata[,2])
ggplot(aes(x=Var1,y=Var2), data=mydata) +  geom_raster(aes(fill=f)) +
  geom_contour(aes(z=f),color='grey',size=1,binwidth=0.05) + theme_classic() +
    scale_fill_continuous(name='f(x,y)') + ylab('Y') + xlab('X')
```

Expectation of $g(Y_1, Y_2, ..., Y_k)$

We can define the expectation of a function of a multivariate random vector in a manner very similar to the univariate case.

If $\{Y_1, ..., Y_k\}$ is a vector of discrete random variables, then:

$$E(g(Y_1, ..., Y_k)) = \sum_{all\ y_1} \sum_{all\ y_2} ... \sum_{all\ y_k} g(y_1, y_2, ..., y_k) p(y_1, y_2, ..., y_k)$$

Similarly, if $\{Y_1, ..., Y_k\}$ is a vector of continuous random variables, then:

$$E(g(Y_1, ..., Y_k)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g(y_1, y_2, ..., y_k) f(y_1, y_2, ..., y_k)$$

Some common examples of $g(\cdot)$ in the bivariate case would be:

- $X + Y$, $X - Y$, X/Y, Y/X, XY: Some of these are best obtained via transformations; i.e. letting $W = X/Y$ and finding $E(W)$; this is the topic of Chapter 6.
- X, X^2, Y, Y^2 (for deriving variances)
- $(X - E(X))(Y - E(Y))$: **Covariance**

We also have the following (reducing to the bivariate case for simplicity, though these hold more generally as well):

- $E(cg(X, Y)) = cE(g(X, Y))$
- $E(g_1(X, Y) + g_2(X, Y) + ... + g_k(X, Y)) = E(g_1(X, Y)) + E(g_2(X, Y)) + ... + E(g_k(X, Y))$
- If X and Y are independent, and $g(X)$ and $h(Y)$ are functions of X and Y only, then:

$$E(g(X)h(Y)) = E(g(X))E(h(Y)),$$

provided these expectations exist.

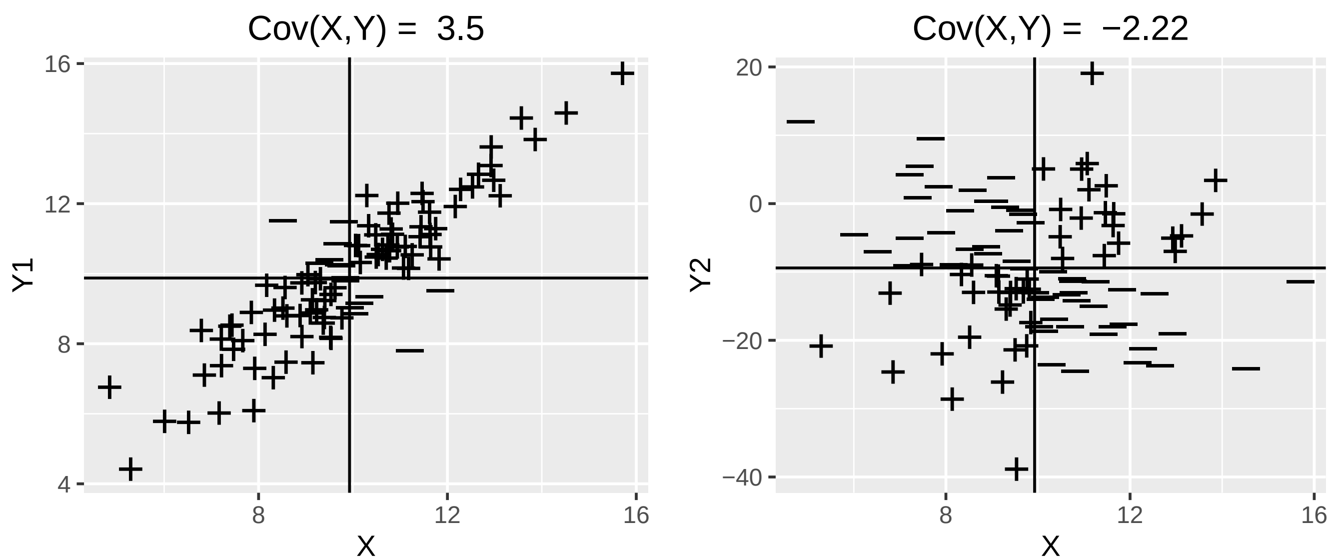
Proof for continuous case:

Studying the covariance We now turn to studying the covariance of two random variables.

Definition:

$$Cov(X,Y) = E\left[(X - E(X))(Y - E(Y))\right] = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

The covariance is very important in linear regression, and is often used to measure the strength of *linear* association between X and Y. Note the covariances of X and Y in the following graph, how they depend on the size and sign of $(X - \mu_X)(Y - \mu_Y)$:



The covariance is often written in the following simpler form:

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Proof:

Theorem: If X and Y are independent, then $Cov(X,Y) = 0$.

Proof:

Note that the converse is *NOT* true; $Cov(X,Y) = 0$ does *NOT* necessarily imply that X and Y are independent.

Variance of sums and differences of X and Y

The covariance plays an important role in finding means and variances of linear combinations of random variables.

Let Y_1, Y_2,...,Y_n be random variables with $E(Y_i) = \mu_i$. Let:

$$U = \sum_{i=1}^{n} a_i Y_i.$$

Then:

1. $E(U) = \sum_{i=1}^{n} a_i \mu_i$
2. $Var(U) = \sum_{i=1}^{n} a_i^2 Var(Y_i) + 2 \sum \sum_{1 \leq i < j \leq n} a_i a_j Cov(Y_i, Y_j)$

Proof:

Important corollaries:

- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$

Correlation

The covariance is scale-dependent, meaning one can blow it up just by multiplying X and/or Y by a large scalar. This is easy to see mathematically:

$$Cov(aX, bY) =$$

For example, let $\{X, Y\}$ be children's age and height measured in years and feet, and $\{X^*, Y^*\}$ be the ages and heights of the same children, but measured in minutes and inches instead. Then $Cov(X, Y) << Cov(X^*, Y^*)$.

Accordingly, we are often interested in the **correlation** of X and Y:

$$\rho = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Given data pairs $(X_1, Y_1), ..., (X_n, Y_n)$, the correlation is estimated in the following intuitive manner:

$$\hat{\rho} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Facts about ρ:

1. $|\rho| \leq 1$; i.e. $-1 \leq \rho \leq 1$
   Proof by construction:

2. $\rho = \pm 1$ if and only if $Y = aX + b$ (for example; Y = temperature in ^{o}C and X = temperature in ^{o}F)

Example

We previously considered the following joint pdf:

$$f(x,y) = \begin{cases} 8xy & 0 \le x \le y \le 1 \\ 0 & otherwise \end{cases}$$

1. Find μ_X and μ_Y.
2. Find $Cov(X,Y)$.
3. Find $Var(X)$ and $Var(Y)$.
4. Find ρ.

5. Find $Var(X+Y)$.

Conditional expectation and variance

Conditional expectation and variance is a fundamental concept in regression, where the intent is to model the mean of Y conditional on X. E.g., modeling mean height given age. These expectations are defined as follows:

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

if (X, Y) are jointly continuous, and

$$E(g(Y)|X = x) = \sum_{all\ \ y} g(y)p(y|x)$$

if (X, Y) are jointly discrete.

It is obvious that with $g(Y) = Y$, the conditional expectation $E(Y|X) = \int_{-\infty}^{\infty} yf(y|x)dy$ in the continuous case (the discrete case is analogous).

Note from here that $E(Y|X = x)$ is a function of x; similarly $E(X|Y = y)$ is a function of y.

We can similarly define the conditional variance. Letting $\mu_{Y|X} \equiv E(Y|X)$, and $g(Y) = (Y - \mu_{Y|X})^2$, we have (for the continuous case):

$$Var(Y|X = x) = E[(Y - \mu_{Y|X})^2|X = x] = \int_{-\infty}^{\infty} (y - \mu_{Y|X})^2 f(y|x)dy$$

The following theorems are hugely important for deriving marginal means and variances from conditional means and variances (using subscripts to emphasize the pdf over which to integrate):

1. $E_Y(Y) = E_X[E_{Y|X}(Y|X)]$
2. $Var_Y(Y) = E_X(Var_{Y|X}(Y|X)) + Var_X(E_{Y|X}(Y|X))$

Proof of 1

Proof of 2

Importance of (2) in regression

A very important result follows from the fact that $Var_Y(Y) = E_X(Var_{Y|X}(Y|X)) + Var_X(E_{Y|X}(Y|X))$. From this, we can show that $Var_Y(Y) \geq Var_X(E_{Y|X}(Y|X))$; i.e., that *conditioning on X reduces the variability in Y. This is essentially the entire point of regression; to reduce as much as possible the unexplained variability of Y.

Proof:

Example: 3.202 revisited

Let X be the number of cars driving past a parking area in a one-minute interval. Assume $X \sim POI(\lambda)$. Conditional on X, $Y \equiv$ the number of cars that decide to park, follow a binomial distribution: $Y|X = x \sim BIN(x,p)$. What is the *unconditional* expected number of cars that decide to park in any one-minute interval? What is the unconditional variance?

Example

Let N denote the number of insurance claims per month. Assume $N \sim POI(\lambda)$. Let Y_i denote the dollar amount of each claim, and assume that the dollar amounts $Y_1, Y_2, ..., Y_N$ are independent. Suppose $Y_i \sim EXP(\mu)$ where $E(Y_i) = \mu$. Let $T = \sum_{i=1}^{N} Y_i$ denote the total dollar amount of all claims in a given month. Find $E(T)$ and $Var(T)$.

Example

Once again, let's consider the following example:

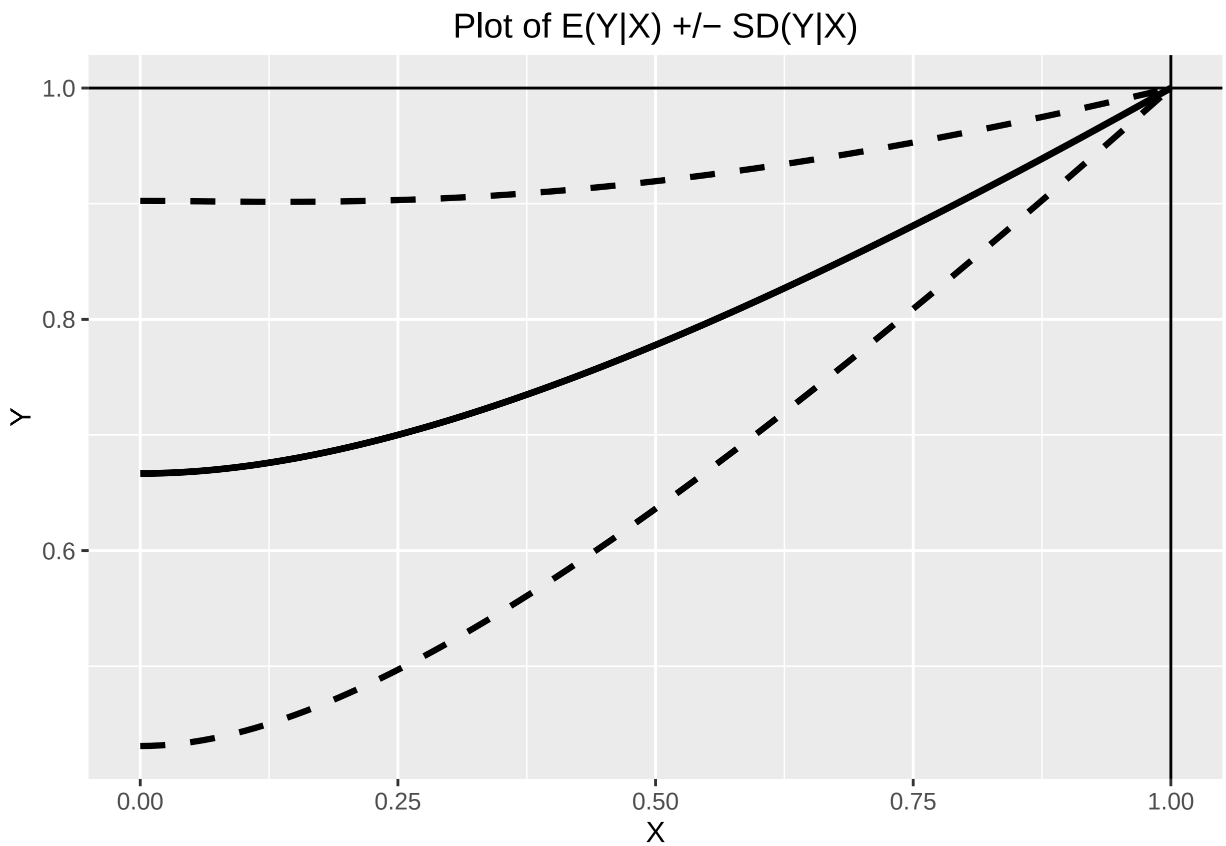
$$f(x,y) = \begin{cases} 8xy & 0 \leq x \leq y \leq 1 \\ 0 & otherwise \end{cases}$$

Find:

1. $E(Y|X)$
2. $Var(Y|X)$

Visualizing the regression line:

```
x <- seq(0,1,l=100)
cond.mean <- 2*(x^2+x+1)/(3*(x+1))
cond.var <- (1+x^2)/2 - cond.mean^2
mydata = data.frame(x = x,
                    E.YgivenX = cond.mean,
                    V.YgivenX = cond.var)
ggplot(data=mydata) + geom_line(aes(x=x,y=cond.mean),size=1.2) +
  geom_line(aes(x=x,y=(cond.mean-sqrt(cond.var))),linetype=2,size=1.1) +
  geom_line(aes(x=x,y=(cond.mean+sqrt(cond.var))),linetype=2,size=1.1) +
  geom_hline(yintercept=1) + geom_vline(xintercept=1) + ylab('Y') + xlab('X') +
  ggtitle('Plot of E(Y|X) +/- SD(Y|X)')
```



Note the non-constant variance of Y conditioned on X, and the slightly non-linear mean of Y as a function of X.

The multinominal distribution

Given our thorough treatment of multivariate distributions, we turn now to a specific multivariate joint distribution: the *multinomial distribution*.

Definition 5.11 describes a multinomial experiment:

1. The experiment consists of n identical, independent trials.
2. The outcome of each trial falls into one of k classes or cells.
3. The probability that the outcomes falls into cell i is p_i, for $i = 1, 2, ..., k$, and remains the same from trial to trial. *Wemusthave*p_1 + p_2 + . . . + p_k = 1$.

The random variables here are the vector $\{Y_1, Y_2, ..., Y_K\}$, where Y_i is the number of trials that fall into cell i. Notice that $Y_1 + Y_2 + ... + Y_k = n$.

The pmf is a generalization of the binomial, and is given by:

$$p(y_1, y_2, ..., y_k) = \frac{n!}{y_1!y_2!...y_k!} p_1^{y_1} p_2^{y_2} ... p_k^{y_k},$$

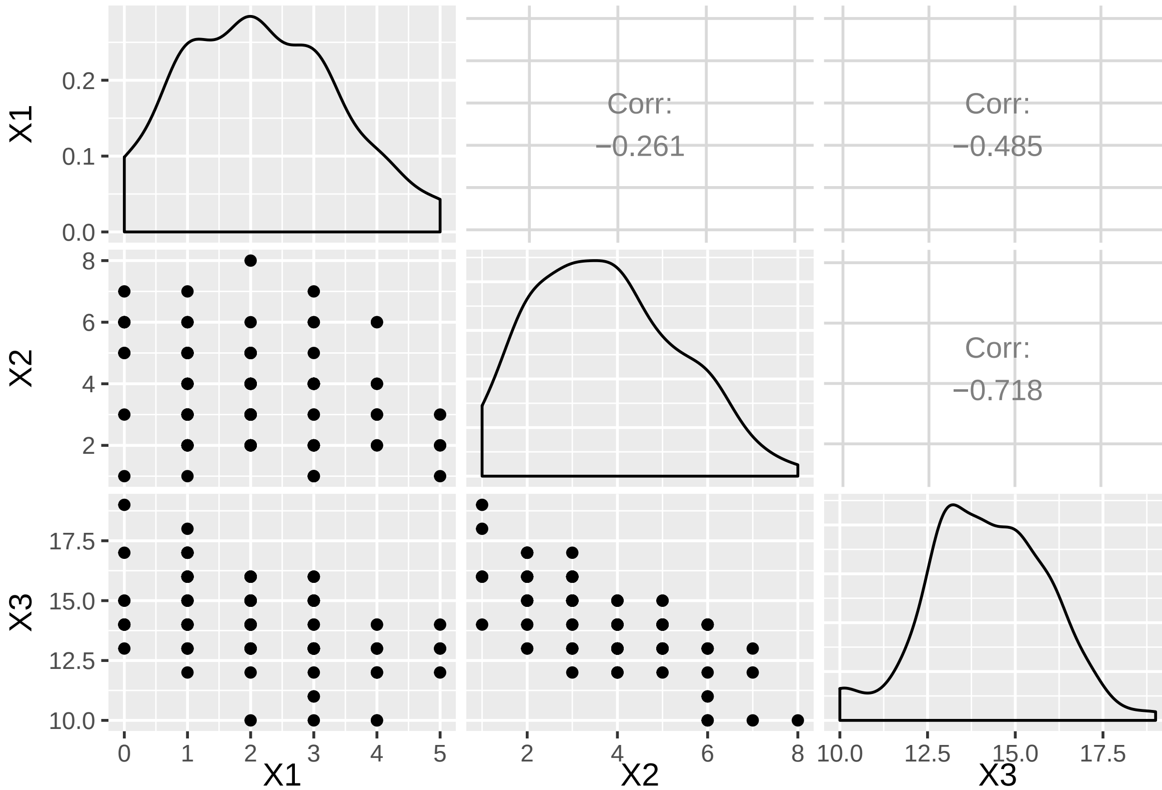
again with the restrictions that $\sum_{i=1}^{k} p_i = 1$ and $\sum_{i=1}^{k} y_i = n$.

Marginal pmf of Y_i

Joint pmf of (Y_i, Y_j)

It can also be shown (Theorem 5.13) that $Cov(Y_i, Y_j) = -np_ip_j$. I.e., the covariance between any pair is negative, and is most negative when p_i and p_j are both large. Simulating some multinomial data and plotting:

```
library(GGally)
set.seed(171)
random.data <- rmultinom(100,size = 20, prob = c(.1,.2,.7))
mydata <- data.frame(t(random.data))
ggpairs(mydata)
```


X1
X2
X3
Corr:
−0.261
Corr:
−0.485
Corr:
−0.718
0.0
0.1
0.2
2
4
6
8
10.0
12.5
15.0
17.5
0
1
2
3
4
5
2
4
6
8
10.0
12.5
15.0
17.5
X1
X2
X3

Bivariate normal distribution

A very important joint continuous distribution is the **multivariate normal** distribution. In the $k-$dimensional case, we let $\vec{y}$ denote a $k-$dimensional vector with covariance matrix Σ and mean vector $\vec{\mu}$. Then the pdf is:

$$f(\vec{y}) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{y}-\vec{\mu})^T\Sigma^{-1}(\vec{y}-\vec{\mu})}.$$

Note the similarities of this distribution to the normal distribution discussed in Chapter 4, the case when $k = 1$:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

The **bivariate normal** distribution is the case where $k = 2$. If $(X, Y) \sim BVN(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, then:

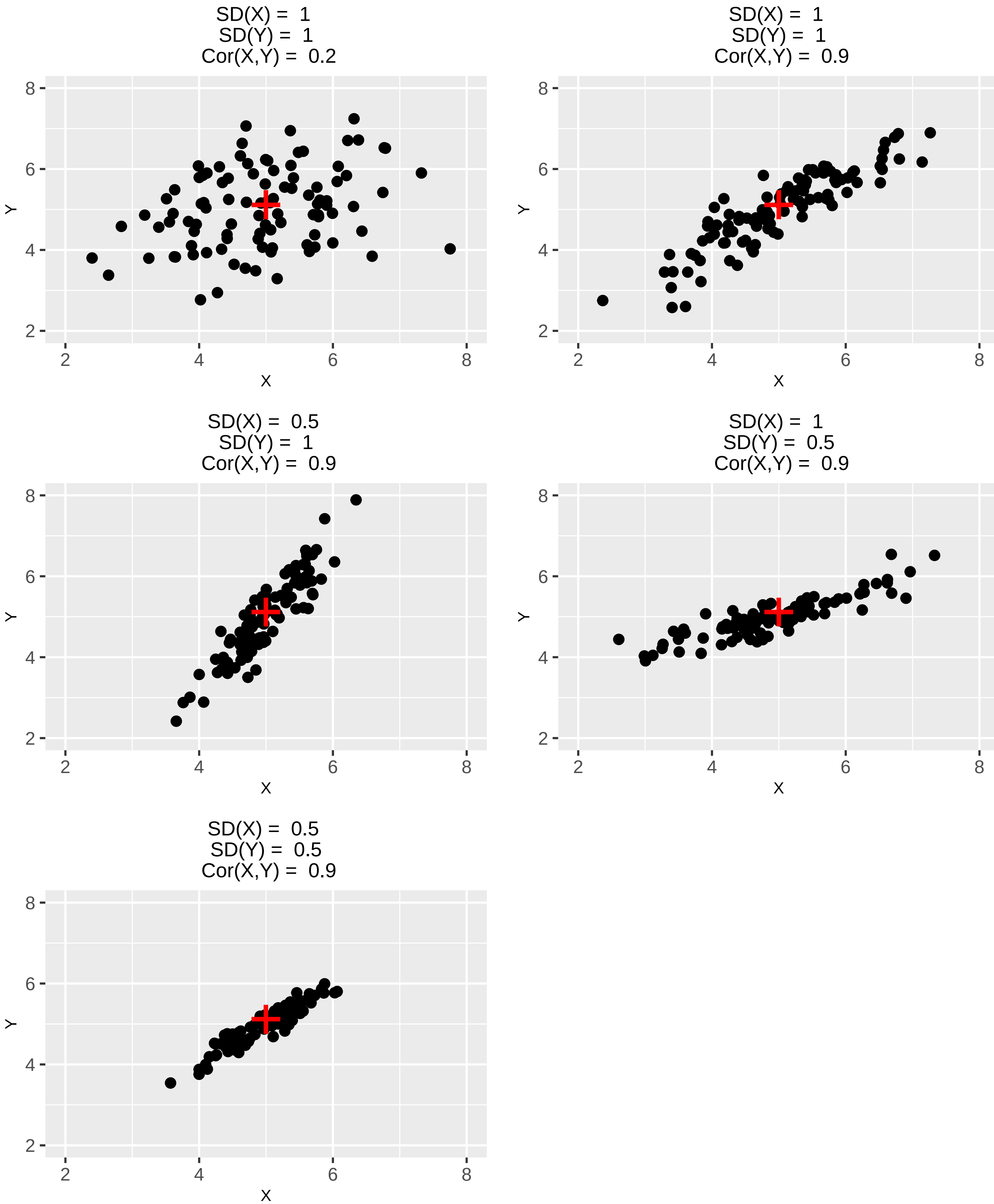
$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]}, \; -\infty < x < \infty; \; -\infty < y < \infty$$

Important feature of the bivariate normal: $\rho = Cor(X, Y) = 0$ if *and only if* X and Y are independent!

Plotting some random realizations of bivariate normal data:

```
library(MASS)
plot.bvnorm <- function(mux = 5, muy = 5, sigx,sigy,rho) {
Sig1 <- matrix(c(sigx^2,rho*sigx*sigy,rho*sigx*sigy,sigy^2),2,2)
random.data <- mvrnorm(100,mu=c(5,5),Sigma = Sig1)
mydata <- data.frame(random.data)
ggplot(data=mydata) + geom_point(aes(x = X1, y = X2), size = 2) + xlab('X') + ylab('Y')+
 ggtitle(paste('SD(X) = ',sigx,'\n SD(Y) = ',sigy,'\n Cor(X,Y) = ',rho)) +
  geom_point(aes(x = mux, y = muy),shape = 43,size=10,color='red') +
  xlim(c(2,8)) + ylim(c(2,8))
}

set.seed(724)
plot.bvnorm(sigx = 1, sigy = 1, rho = .2)
plot.bvnorm(sigx = 1, sigy = 1, rho = .9)
plot.bvnorm(sigx = .5, sigy = 1, rho = .9)
plot.bvnorm(sigx = 1, sigy = .5, rho = .9)
plot.bvnorm(sigx = .5, sigy = .5, rho = .9)
```


SD(X) = 1
SD(Y) = 1
Cor(X,Y) = 0.2
SD(X) = 1
SD(Y) = 1
Cor(X,Y) = 0.9
SD(X) = 0.5
SD(Y) = 1
Cor(X,Y) = 0.9
SD(X) = 1
SD(Y) = 0.5
Cor(X,Y) = 0.9
SD(X) = 0.5
SD(Y) = 0.5
Cor(X,Y) = 0.9
X
Y
2
4
6
8

1. Show that the marginal of X is $\sim N(\mu_X, \sigma_X^2)$

(continued)

2. Find and study the conditional distribution of $Y|X$.

(continued)