MATH/STAT 338: Probability

Multivariate Distributions

Anthony Scotina

Joint Distributions

Multiple Random Variables

So far, we have looked at a **single random variable**, Y.

- The distribution of Y provides *complete information* about **probabilities** associated with Y.

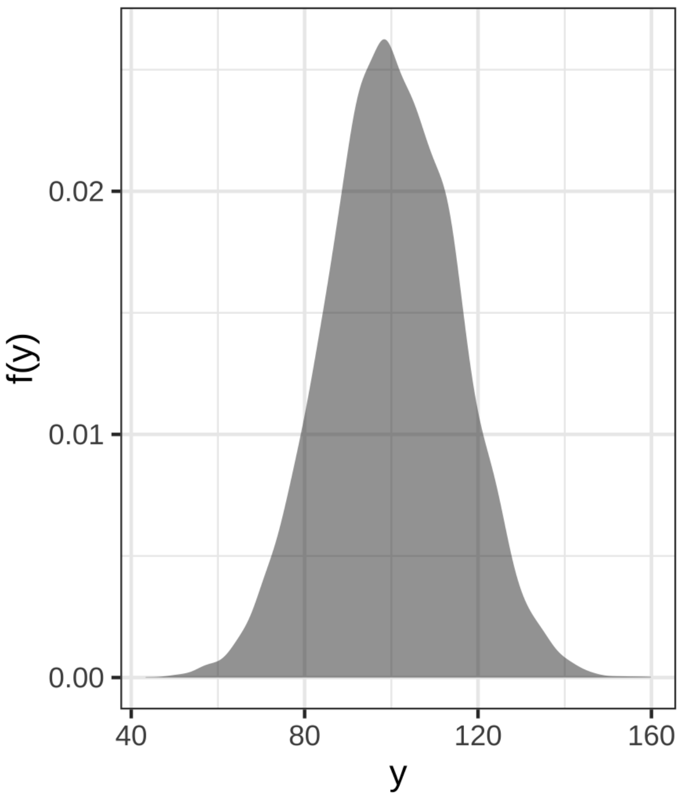
But in many applications, we care about the relationship between **multiple random variables** in the same experiment!

Joint/Multivariate distributions allow us to learn about how multiple random variables are related, by studying the probability distribution associated with the *intersection* of multiple random variables:

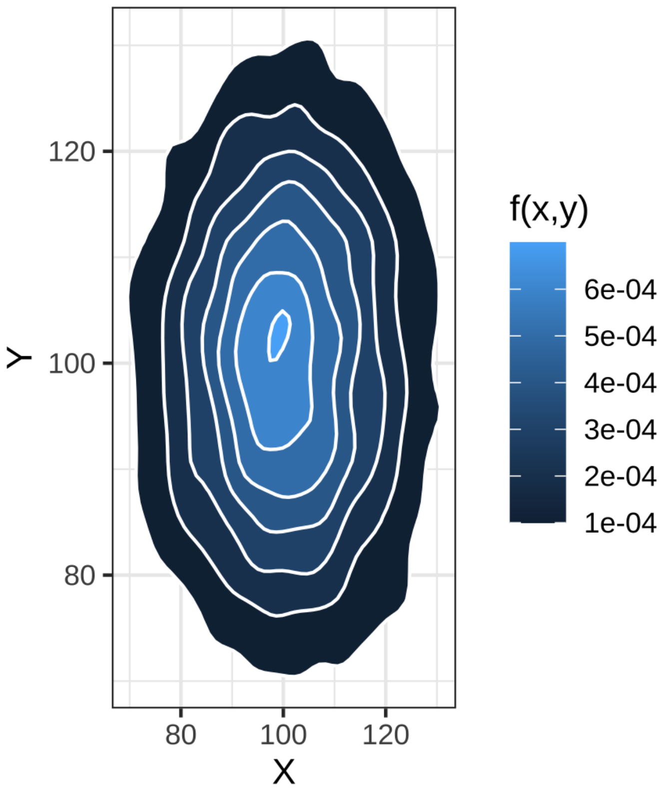
$$(Y_1 = y_1), (Y_2 = y_2), \ldots, (Y_n = y_n)$$

Extending to Multivariate Distributions

- $Y \sim \text{Normal}(100, 15)$



- $\binom{X}{Y} \sim \text{Multivariate Normal}$



Discrete Joint Distributions

Example

Suppose we make our way back to *Piéchart Casino* and play a game where we toss three *balanced* coins *independently* (no, this game isn't rigged):

- $X =$ number of heads
- $Y = \ldots$
  - \$1 if first head is on toss 1
  - \$2 if first head is on toss 2
  - \$3 if first head is on toss 3
  - -\$1 if no heads

Outcomes

Outcome	HHH	THH	HTH	HHT	TTH	THT	HTT	TTT
X	3	2	2	2	1	1	1	0
Y	1	2	1	1	3	2	1	-1

Joint PMF

Let X and Y be **discrete** random variables. The **joint (bivariate) probability function** for X and Y is given by

$$p(x, y) = P(X = x, Y = y).$$

Outcome	HHH	THH	HTH	HHT	TTH	THT	HTT	TTT
X	3	2	2	2	1	1	1	0
Y	1	2	1	1	3	2	1	-1

In the *Piéchart Casino* example, a visual of the joint PMF is given below:

Joint PMF

Similar to the single variable (**univariate**) case, a *valid* PMF must be nonnegative and sum to 1.

Theorem: If X and Y are discrete random variables with joint probability function $p(x, y)$, then...

1. $p(x, y) \geq 0$ for all x, y.
2. $\sum_{x,y} p(x, y) = 1$, where the sum is over all values (x, y) that are assigned non-zero probabilities.

🤔 Is our PMF from the *Piéchart Casino* example a *valid* PMF?

Joint Distribution Function

For any random variables X and Y, the **joint (bivariate) distribution function**, $F(x, y)$, is

$$F(x, y) = P(X \leq x, Y \leq y), \qquad -\infty < x < \infty, \ -\infty < y < \infty.$$

Practice

In the *Piéchart Casino* example, find $F(2, 1) = P(X \leq 2, Y \leq 1)$.

Outcome	HHH	THH	HTH	HHT	TTH	THT	HTT	TTT
X	3	2	2	2	1	1	1	0
Y	1	2	1	1	3	2	1	-1

Joint Distribution Function

For any random variables X and Y, the **joint (bivariate) distribution function**, $F(x, y)$, is

$$F(x, y) = P(X \leq x, Y \leq y), \qquad -\infty < x < \infty, \ -\infty < y < \infty.$$

Solution

In the *Piéchart Casino* example, find $F(2, 1) = P(X \leq 2, Y \leq 1)$.

Outcome	HHH	THH	HTH	HHT	TTH	THT	HTT	TTT
X	3	2	**2**	**2**	1	1	**1**	**0**
Y	1	2	**1**	**1**	3	2	**1**	**-1**

$$P(X \leq 2, Y \leq 1) = P(X = 0, Y = -1) + P(X = 1, Y = 1) + P(X = 2, Y = 1) = \frac{1}{8} + \frac{1}{8}$$

Continuous Joint Distributions

Let X and Y be **continuous** random variables with joint distribution function $F(x, y)$. If there exists a nonnegative function $f(x, y)$, such that

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t_1, t_2)\, dt_2 dt_1,$$

for all $-\infty < x < \infty$, $-\infty < y < \infty$, then X and Y are **jointly continuous random variables**.

- The function $f(x, y)$ is called the **joint probability density function (PDF)**.

Similar to joint discrete random variables, if X and Y are jointly continuous, then...

1. $f(x, y) \geq 0$ for all x, y.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dxdy = 1$

Continuous Joint Probabilities

Let X and Y be **continuous** random variables with joint distribution function $F(x, y)$. If there exists a nonnegative function $f(x, y)$, such that

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t_1, t_2)\, dt_2 dt_1,$$

for all $-\infty < x < \infty$, $-\infty < y < \infty$, then X and Y are **jointly continuous random variables**.

- The function $f(x, y)$ is called the **joint probability density function (PDF)**.

For two RVs X and Y with joint PDF $f(x, y)$, the *joint probability* $P(a_1 \leq X \leq a_2, b_1 \leq Y \leq b_2)$ is

$$P(a_1 \leq X \leq a_2, b_1 \leq Y \leq b_2) = \int_{b_1}^{b_2} \int_{a_1}^{a_2} f(x, y)\, dx dy$$

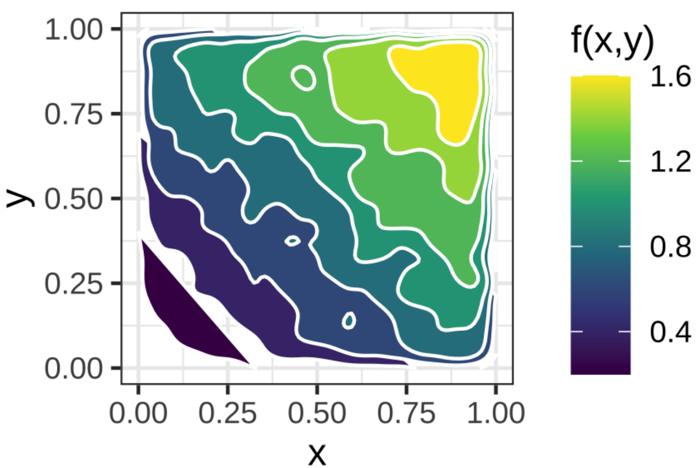
Continuous Joint Distributions

Example

Let X and Y be random variables with joint probability density function

$$f(x,y) = \begin{cases} x+y, & 0<x<1, 0<y<1 \\ 0, & \text{elsewhere.} \end{cases}$$

- Find $P(X > 0.5, Y < 0.5)$.



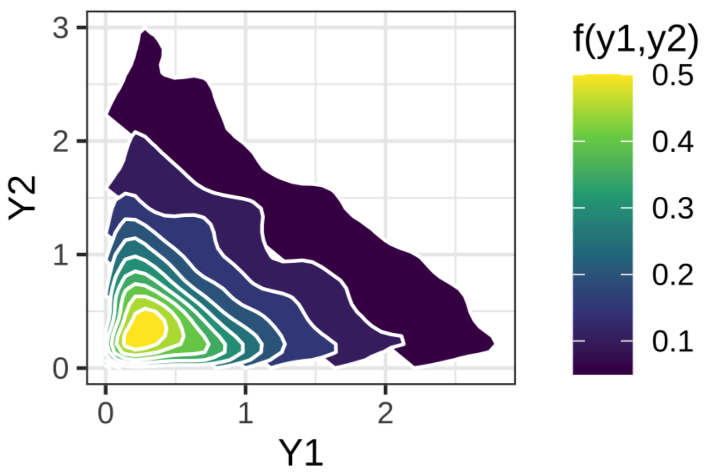
Continuous Joint Distributions

Example (WMS 5.7)

Let Y_1 and Y_2 have the joint probability density function

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & y_1 > 0, y_2 > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

1. Show that $P(Y_1 + Y_2 < 3) = 1 - 4e^{-3} = 0.8009.$
2. Find $P(Y_1 < 1, Y_2 > 5)$.



Marginal Distributions

Piéchart Casino

Recall the *Piéchart Casino* from earlier.

- $X =$ number of heads
- $Y = \ldots$
  - \$1 if first head is on toss 1
  - \$2 if first head is on toss 2
  - \$3 if first head is on toss 3
  - -\$1 if no heads

Outcomes

Outcome	HHH	THH	HTH	HHT	TTH	THT	HTT	TTT
X	3	2	2	2	1	1	1	0
Y	1	2	1	1	3	2	1	-1

What if we only wanted to find a probability *associated with* X, such as $P(X = 2)$?

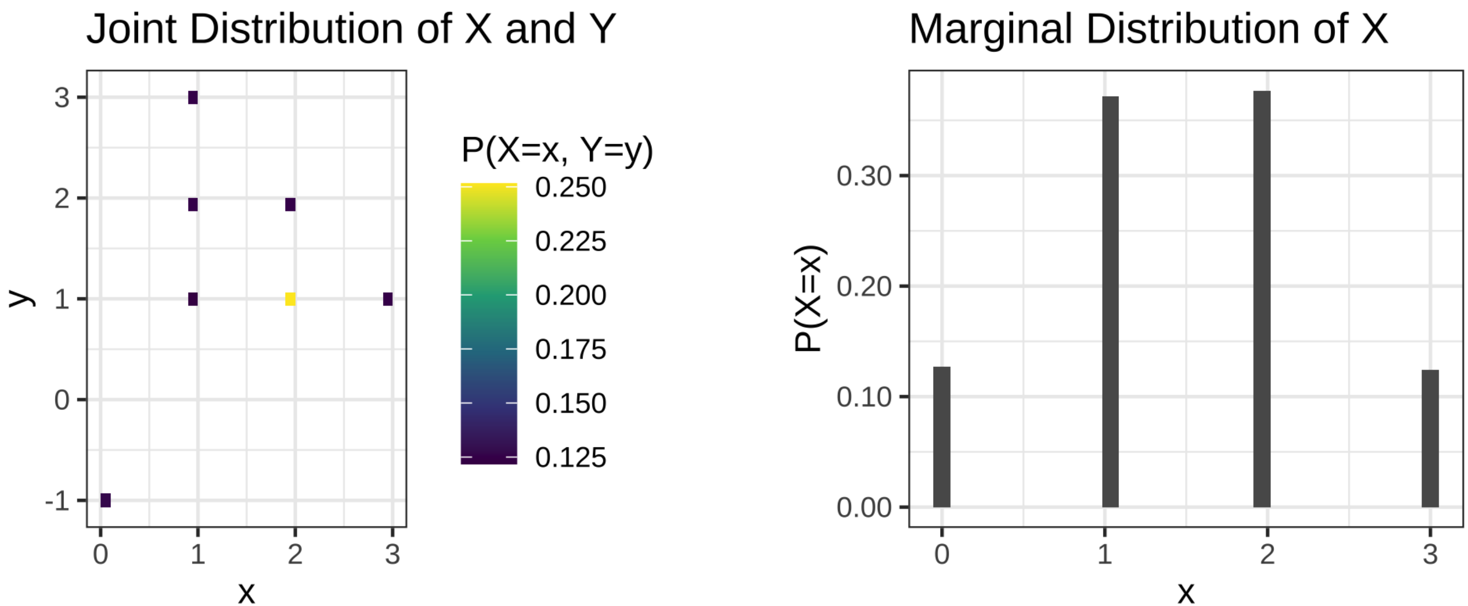
- To find the distribution of X **alone**, we need to sum over *all possible values of Y*: $P(X = 2) = p(2, -1) + p(2, 1) + p(2, 2) + p(2, 3)$

Marginal Probability Functions (Discrete)

We just found a *marginal probability*!

For **discrete** random variables X and Y, the **marginal probability function** of X is

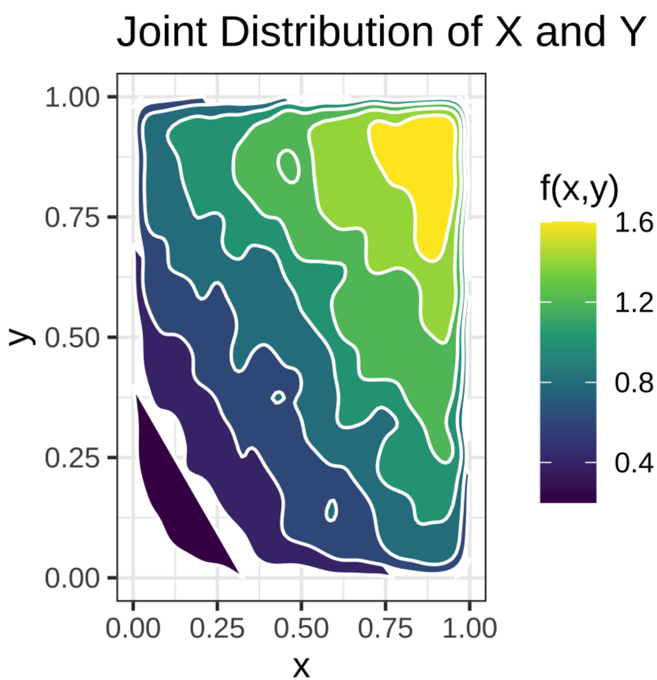
$$P(X=x)=\sum_y P(X=x,Y=y).$$

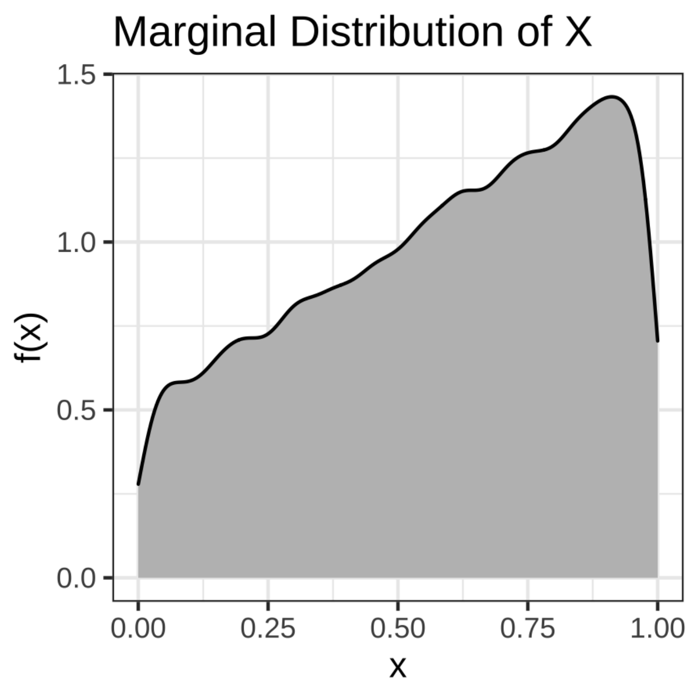


Marginal Probability Functions (Continuous)

For **continuous** random variables X and Y, the **marginal probability functions** of X and Y, respectively, are

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad \text{and} \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$





Practice

Let X and Y be random variables with joint probability density function

$$f(x,y) = \begin{cases} x+y, & 0<x<1, 0<y<1 \\ 0, & \text{elsewhere.} \end{cases}$$

- Find the marginal distribution for X, $f_X(x)$.

Marginal Probability Functions (Continuous)

Example

Let X and Y be random variables with joint probability density function

$$f(x, y) = \begin{cases} \frac{e^{-y}}{\sqrt{2\pi}} e^{-x^2/2}, & -\infty < x < \infty, y > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

- Find the marginal distributions $f_X(x)$ and $f_Y(y)$.

Solution

It turns out...

- $X \sim \text{Normal}(0, 1)$
- $Y \sim \text{Exponential}(1)$

Conditional Distributions

Condtional Probability

Suppose we have two random variables, X and Y. We *observe* some value of X, and we want to use this to *update* our distribution of Y to reflect this information.

- The **marginal distributions**, $P(Y = y)$ and $f_Y(y)$, do not take into account *any* information about X.

Recall some properties of **conditional probabilities** that we studied previously:

- $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
- $P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$
- $P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$ (**Bayes' Rule**)

Conditional Distributions

We can extend the properties of **conditional probability** to settings involving random variables!

Discrete case

If X and Y are jointly **discrete** random variables with *joint* probability function $p(x,y)$ and *marginal* probability functions $p_X(x)$ and $p_Y(y)$, respectively, then the **conditional probability function** of X *given* $Y=y$ is:

$$p(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x,y)}{p_Y(y)}$$

Continuous case

For **continuous** random variables X and Y with *joint* density function $f(x,y)$ and *marginal* density functions $f_X(x)$ and $f_Y(y)$, the **conditional density function** of X *given* $Y=y$ is:

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

Practice

(WMS 5.25)

Let Y_1 and Y_2 have the joint probability density function

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & y_1 > 0, y_2 > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

1. Find the marginal density function for Y_2, $f_{Y_2}(y_2)$.
2. For $y_2 > 0$, what is the conditional density function of Y_1 given that $Y_2 = y_2$?

Conditional Distributions (Continuous)

Example (WMS 5.31)

Let Y_1 and Y_2 have the joint probability density function

$$f(y_1, y_2) = \begin{cases} 30y_1y_2^2, & y_1 - 1 \le y_2 \le 1 - y_1, 0 \le y_1 \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

1. Show that the marginal density of Y_1 is a beta density with $\alpha = 2$ and $\beta = 4$.
2. Find the conditional density of Y_2 given $Y_1 = y_1$. Use this to show that $P(Y_2 > 0 \mid Y_1 = 0.75) = 0.5$.

Independent Random Variables

Independent Events

Two events are said to be **independent** if knowledge that one event occurs does not indicate whether the other event is more likely to occur.

Formally, two events are **independent** if *any* one of the following holds:

- $P(A \mid B) = P(A)$
- $P(B \mid A) = P(B)$
- $P(A \cap B) = P(A)P(B)$

Otherwise, the events are said to be *dependent*.

This was from the first month of class, but we can extend it to a setting involving *random variables*.

- In fact, we've already seen joint distributions where...
  - the probabilities associated with X depended on the observed value of Y
  - the probabilities associated with X are completely independent of Y

Independent Random Variables

The *technical* definition of **independence** involves *cumulative distribution functions*...

Let X have distribution function $F_X(x), Y$ have distribution function $F_Y(y)$, and X and Y have *joint* distribution function $F(x, y)$. Then X and Y are said to be **independent** if and only if

$$F(x, y) = F_X(x)F_Y(y) = P(X \leq x)P(Y \leq y).$$

If X and Y are not independent, they are said to be **dependent**.

This is equivalent to the following, involving *probability functions*:

- **Discrete RVs**: Independence $\iff p(x, y) = p_X(x)p_Y(y)$
- **Continuous RVs**: Independence $\iff f(x, y) = f_X(x)f_Y(y)$

Independence in 2x2 Tables

Are SSRIs associated with increased risk of bone fractures among the elderly?

- $n = 5008$ adults over the age of 50, followed for five years

Study Results

	Taking SSRI	No SSRI	Total
Experiences Fractures	14	244	258
No Fractures	123	4627	4750
Total	137	4871	5008

For each study participant, we can understand the events "*Taking SSRI*" and "*Experiences Fractures*" (and their counterparts) as results of **Bernoulli trials**.

- We can find *joint*, *marginal*, and *conditional* distributions using the 2x2 table!

Independence in 2x2 Tables

Are SSRIs associated with increased risk of bone fractures among the elderly?

- $n = 5008$ adults over the age of 50, followed for five years

Study Results

	Y = 1	Y = 0	Total
X = 1	14	244	258
X = 0	123	4627	4750
Total	137	4871	5008

> Are *X* and *Y* independent?

Practice

(WMS 5.53)

Let Y_1 and Y_2 have the joint probability density function

$$f(y_1, y_2) = \begin{cases} 6(1 - y_2), & 0 \leq y_1 \leq y_2 \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

- By finding $f_1(y_1)$ and $f_2(y_2)$, show that Y_1 and Y_2 are **dependent**.

Solution:

- $f_1(y_1) = 3 - 6y_1 + 3y_1^2,\ 0 \leq y_1 \leq 1$
- $f_2(y_2) = 6y_2(1 - y_2),\ 0 \leq y_2 \leq 1$

Therefore, $f(y_1, y_2) \neq f_1(y_1)f_2(y_2)$.

One Simple Trick ™

Theorem: Let Y_1 and Y_2 have a joint density $f(y_1, y_2)$ that is positive if and only if $a \leq y_1 \leq b$ and $c \leq y_2 \leq d$, for constants a, b, c, and d. Then Y_1 and Y_2 are independent if and only if

$$f(y_1, y_2) = g(y_1)h(y_2),$$

where $g(y_1)$ is a nonnegative function of y_1 *alone* and $h(y_2)$ is a nonnegative function of y_2 *alone*.

- In some cases, we don't need to derive marginal densities to prove independence! We just need to break the joint density function into separate pieces involving y_1 and y_2 alone.

Example

Suppose Y_1 and Y_2 denote the lengths of life, in hundreds of hours, for components of types I and II, respectively, in an electronic system. The joint density of Y_1 and Y_2 is

$$f(y_1, y_2) = \begin{cases} (1/8)y_1 e^{-(y_1+y_2)/2}, & y_1 > 0, y_2 > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

- Using the previous theorem, prove that Y_1 and Y_2 are independent.

Solution: $f(y_1, y_2) = g(y_1)h(y_2)$, where...

- $g(y_1) = (1/8)y_1 e^{-y_1/2}$
- $h(y_2) = e^{-y_2/2}$

The Expected Value of a Function of Random Variables

Multivariate LOTUS

Multivariate LOTUS

Discrete

Let $g(Y_1, Y_2, \ldots, Y_k)$ be a function of discrete RVs with joint probability function $p(y_1, y_2, \ldots, y_k)$. Then the **expected value** of $g(Y_1, Y_2, \ldots, Y_k)$ is

$$E[g(Y_1, Y_2, \ldots, Y_k)] = \sum_{y_k} \cdots \sum_{y_2} \sum_{y_1} g(y_1, y_2, \ldots, y_k) p(y_1, y_2, \ldots, y_k).$$

Continuous

If $Y_1, Y_2, \ldots, Y_k$ are continuous RVs with joint density function $f(y_1, y_2, \ldots, y_k)$, then

$$E[g(Y_1, Y_2, \ldots, Y_k)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_k) f(y_1, y_2, \ldots, y_k)\, dy_1 dy_2 \ldots dy_k.$$

Multivariate LOTUS

Example

Certain electrical fuses operate *independently* with lifetimes $Y_i \sim \text{Exp}(\beta = 3)$, $i = 1, 2$. In other words,

$$f(y_1, y_2) = \begin{cases} \left(\frac{1}{3} e^{-y_1/3}\right)\left(\frac{1}{3} e^{-y_2/3}\right), & y_1 > 0, y_2 > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find $E(Y_1 + Y_2)$ and $E(Y_1 Y_2)$.

Two Simple Tricks ™

Theorems

1. Let Y_1 and Y_2 be random variables and $g_1(Y_1, Y_2), g_2(Y_1, Y_2), \ldots, g_k(Y_1, Y_2)$ be functions of Y_1 and Y_2. Then

$$\begin{aligned} &E[g_1(Y_1, Y_2) + g_2(Y_1, Y_2) + \cdots + g_k(Y_1, Y_2)] \\ &\quad = E[g_1(Y_1, Y_2)] + E[g_2(Y_1, Y_2)] + \cdots + E[g_k(Y_1, Y_2)] \end{aligned}$$

2. Let Y_1 and Y_2 be **independent** random variables and $g(Y_1)$ and $h(Y_2)$ be functions of *only* Y_1 and Y_2, respectively. Then

$$E[g(Y_1)h(Y_2)] = E[g(Y_1)]E[h(Y_2)],$$

provided that the expectations exist.

Practice

(WMS 5.81)

Suppose Y_1 and Y_2 denote the lengths of life, in hundreds of hours, for components of types I and II, respectively, in an electronic system. The joint density of Y_1 and Y_2 is

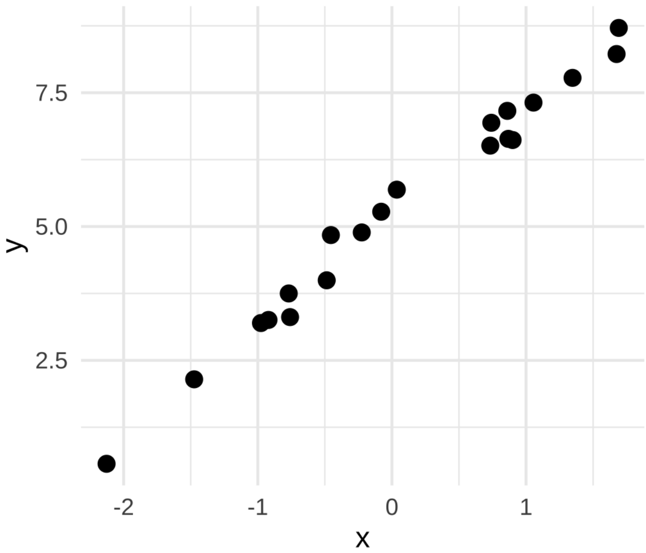
$$f(y_1, y_2) = \begin{cases} (1/8)y_1 e^{-(y_1+y_2)/2}, & y_1 > 0, y_2 > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

One way to measure the relative efficiency of the two components is to compute the ratio Y_2/Y_1. Find $E(Y_2/Y_1)$.

- **Hint**: Use the fact that Y_1 and Y_2 are *independent*. The *Gamma* and *Exponential* PDFs will also help!

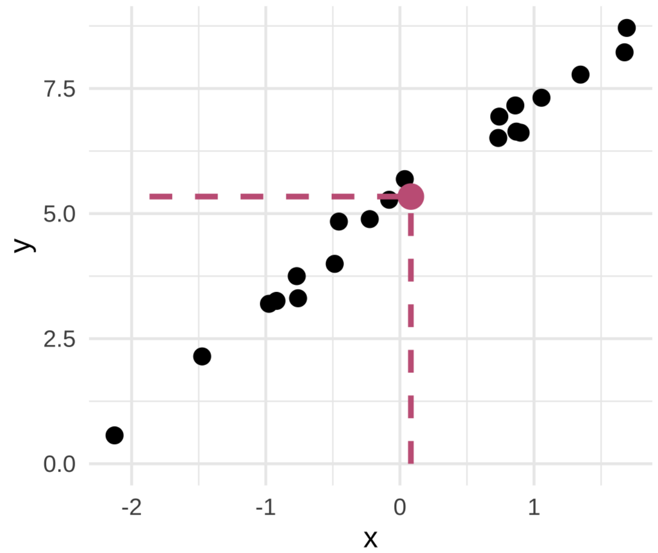
Covariance and Correlation

Dependent Random Variables



- $E(X) = \mu_X = 0.0818$
- $E(Y) = \mu_Y = 5.3403$

Dependent Random Variables

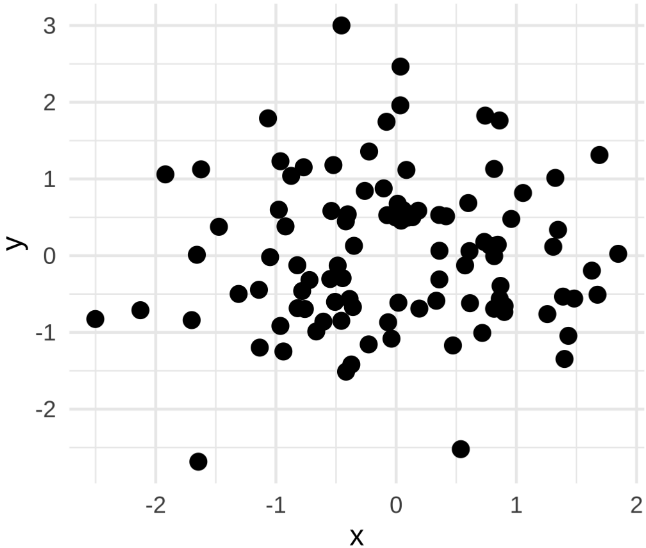


Because X and Y tend to move in the same direction...

- $X - E(X)$ and $Y - E(Y)$ with *both* be either **positive** or **negative**

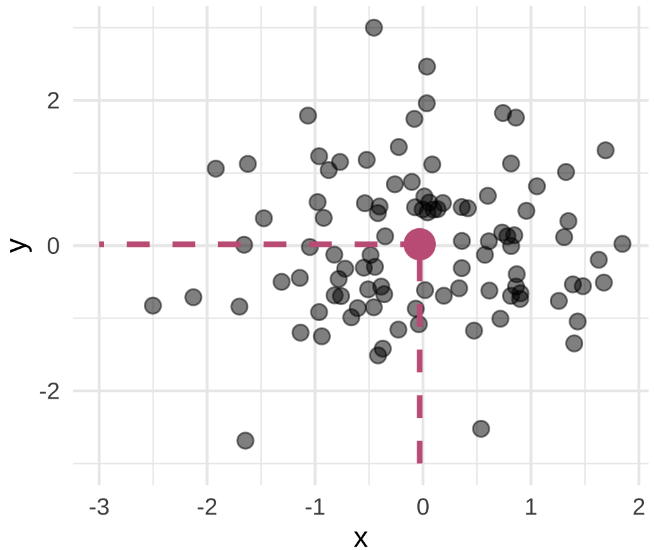
This implies that the **covariance** between X and Y will be positive!

Independent Random Variables



- $E(X) = \mu_X = -0.0309$
- $E(Y) = \mu_Y = 0.0188$

Independent Random Variables



Because X and Y don't really have any relationship in *either* direction...

- $X - E(X)$ and $Y - E(Y)$ with have the **same** sign for some points, and the **opposite** sign for others
  - This means that $(X - E(X))(Y - E(Y))$ will average to (approx.) zero.

This implies that the **covariance** between X and Y will be near zero!

Covariance

If X and Y are random variables with means μ_X and μ_Y, respectively, the **covariance** of X and Y is

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y).$$

We sometimes use the *standardized form* of the covariance, called the **correlation coefficient**:

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

where σ_X and σ_Y are the *standard deviations* of X and Y.

- ρ ranges between -1 and +1
  - $\rho = 1$ implies a *perfect positive (linear) correlation*
  - $\rho = -1$ implies a *perfect negative (linear) correlation*

Practice

Suppose that Y_1 and Y_2 are jointly continuous with PDF given by

$$f(y_1, y_2) = \begin{cases} 4y_1y_2, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $Cov(Y_1, Y_2)$.

> Is this result surprising? Why or why not?

Covariance and Independence

Theorem

If X and Y are **independent** random variables, then

$$Cov(X, Y) = 0.$$

Thus, the variables must be *uncorrelated*.

🚨 🚨 The *converse* is not true!!! 🚨 🚨

- If $Cov(X, Y) = 0$, then X and Y are *not necessarily* independent!

Example

Let $X \sim \text{Normal}(0, 1)$ and $Y = X^2$. Prove that $Cov(X, Y) = 0$ but X and Y are **not** independent.

- We'll use the fact that $\text{Skew}(X) = 0$ for a $\text{Normal}(0, 1)$ distribution.

Covariance in R

Let's simulate X and Y from the previous example!

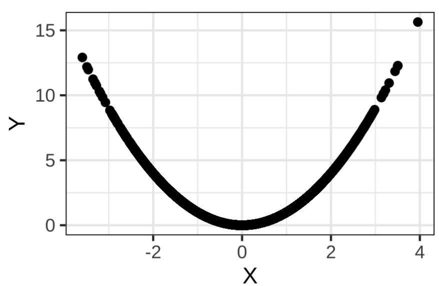
```
X = rnorm(10000)
Y = X^2
data = data.frame(X, Y)
```

Using the `cov()` function on a *data frame* will produce a **covariance matrix**:

```
cov(data)
```

```
##              X            Y
## X  0.991641551 -0.003753336
## Y -0.003753336  1.969326686
```

Clearly not independent!



The Expected Value and Variance of Linear Functions of Random Variables

Coupon Collector's Problem

Moose's **favorite** toys are pictured below.

- Suppose there are **12** types of toys, which you collect one by one, with the goal of getting a complete set.
- When collecting toys, the toy types are **random** and *equally likely*.

  > What is the expected number of toys I need to collect until Moose has a complete set?

Coupon Collector's Problem

In *this* problem that we worked through earlier in the semester, we let

$$N = N_1 + \cdots + N_{12}$$

- $N =$ number of toys needed until we get all 12
- $N_i =$ number of toys needed, starting from N_{i-1}, until we get i toys

Then

$$E(N) = E(N_1) + \cdots + E(N_{12})$$

- **Note**: The N_i are **independent** from one another!

Linear Functions of Random Variables

For random variables $Y_1, Y_2, \ldots, Y_n$ and constants $a_1, a_2, \ldots, a_n$, a **linear function** of the Y_i random variables is:

$$U_1 = a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n = \sum_{i=1}^{n} a_i Y_i$$

- Many problems in *statistical inference* (see, 😎**STAT 339**😎) involve estimators that are linear functions of measurements in a sample.

We might be interested in:

- The **expected value** and **variance** of U_1
- The **covariance** between U_1 and a second linear combination of RVs, U_2.

Linear RVs Theorem

Let $Y_1, Y_2, \ldots, Y_n$ and $X_1, X_2, \ldots, X_m$ be random varibles with $E(Y_i) = \mu_i$ and $E(X_i) = \xi_i$. Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{and} \quad U_2 = \sum_{j=1}^{m} b_j X_j$$

for constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Then:

1. $E(U_1) = \sum_{i=1}^{n} a_i E(Y_i)$
2. $Var(U_1) = \sum_{i=1}^{n} a_i^2 Var(Y_i) + 2 \sum\sum_{1 \leq i < j \leq n} a_i a_j Cov(Y_i, Y_j)$, where the double sum is over all pairs (i, j) with $i < j$.
3. $Cov(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(Y_i, X_j)$

Example

Let $Y_1, Y_2, \ldots, Y_n$ be *independent* random variables with $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2$. Define the **sample mean** as

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Show that $E(\bar{Y}) = \mu$ and $Var(\bar{Y}) = \sigma^2/n$.

- $\bar{Y} = \left(\frac{1}{n}\right) Y_1 + \cdots + \left(\frac{1}{n}\right) Y_n$

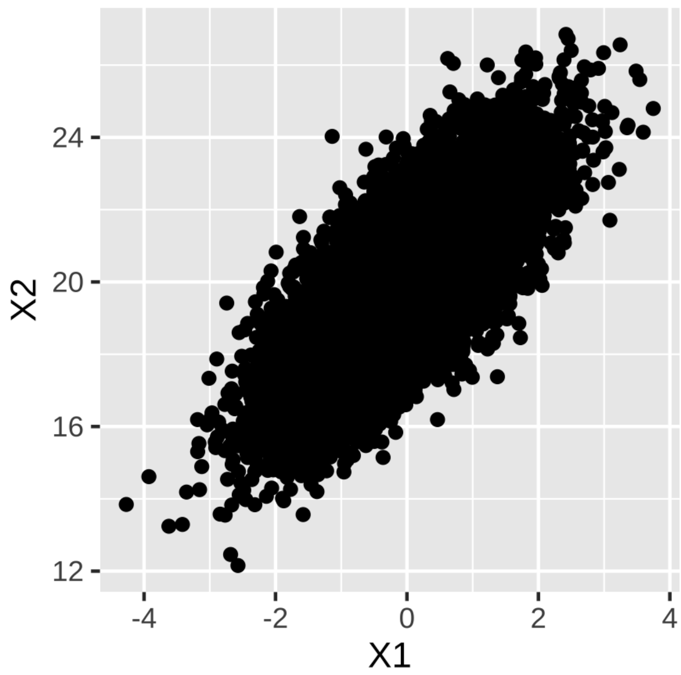
Correlated Normal RVs

Let's *simulate* correlated Normal random variables!

```
library(faux)
fake_data = rnorm_multi(
  n = 10000,
  mu = c(0, 20),
  sd = c(1, 2),
  r = 0.75
)
```

- $X_1 \sim Normal, X_2 \sim Normal$
- $E(X_1) = 0, Var(X_1) = 1$
- $E(X_2) = 20, Var(X_2) = 4$
- $Corr(X_1, X_2) = 0.75 \implies$
    - $Cov(X_1, X_2) = 0.75\sqrt{1}\sqrt{4} = 1.5$

```
gf_point(X2 ~ X1, data = fake_data)
```



Correlated Normal RVs

First, find $Var(X_1 - X_2)$ and $Var(X_1 + X_2)$ by hand.

- $Var(X_1 - X_2) = Var(X_1) + Var(X_2) - 2Cov(X_1, X_2) = 2$
- $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2) = 8$

Check:

```
var(fake_data$X1 - fake_data$X2)
```

```
## [1] 1.994268
```

```
var(fake_data$X1 + fake_data$X2)
```

```
## [1] 8.124726
```

Hypergeometric Variance

Another one of Moose's **favorite** toys is pictured below. Her *favorite* color toy is red.

If this box of size $N = r + b$ is filled with r red toys and b blue toys, and we draw n toys **without replacement**, then the number of **red toys**, Y follows a...

- $\text{HGeom}(N, r, n)$ distribution!

🤔 Let's derive $Var(Y)$.

Hypergeometric Variance

First, represent Y as a sum of **indicator variables**,

$$Y = I_1 + \cdots + I_n,$$

where I_j is the indicator of the jth toy in the sample being red ($I_j = 1$ if red, $I_j = 0$ if blue).

- For each I_j, $E(I_j) = r/N$, which we'll define as p.
  - Therefore, $Var(I_j) = p(1-p)$ if we think of the I_j as *Bernoulli* trials.

Therefore,

$$\begin{aligned} Var(Y) &= Var\left(\sum_{j=1}^{n} I_j\right) \\ &= Var(I_1) + \cdots + Var(I_n) + 2\sum_{i<j} Cov(I_i, I_j) \\ &= np(1-p) + 2\binom{n}{2} Cov(I_1, I_2), \end{aligned}$$

since there are $\binom{n}{2}$ combinations of I_i and I_j, and the covariance between each of the pairs is the same.

Hypergeometric Variance

For $Cov(I_1, I_2)$, use the following:

$$\begin{aligned} Cov(I_1, I_2) &= E(I_1 I_2) - E(I_1)E(I_2) \\ &= P(\text{1st and 2nd toys are both red}) - P(\text{1st toy red})P(\text{2nd toy red}) \\ &= \left(\frac{r}{N}\right)\left(\frac{r-1}{N-1}\right) - p^2 \end{aligned}$$

Plugging this in for $Cov(I_1, I_2)$, the variance simplifies to:

$$Var(Y) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

- The $\frac{N-n}{N-1}$ is known as a **finite population correction**, and adjusts the *binomial variance* for the fact that we are now sampling *without replacement.*

Conditional Expectation

and Hierarchical Models

Conditional Expectation

If Y_1 and Y_2 are any two random variables, the **conditional expectation** of $g(Y_1)$, given that $Y_2 = y_2$, is

- $E(g(Y_1) \mid Y_2 = y_2) = \int_{-\infty}^{\infty} g(y_1) f(y_1 \mid y_2)\, dy_1$ if Y_1 and Y_2 are **jointly continuous**.
- $E(g(Y_1) \mid Y_2 = y_2) = \sum_{y_1} g(y_1) p(y_1 \mid y_2)$ if Y_1 and Y_2 are **jointly discrete**.

Laws of Total Expectation and Variance

Let Y_1 and Y_2 denote random variables. Then:

1. $E(Y_1) = E[E(Y_1 \mid Y_2)]$
2. $Var(Y_1) = E[Var(Y_1 \mid Y_2)] + Var[E(Y_1 \mid Y_2)]$

Beta-Binomial Hierarchy

Sometimes when working with the **binomial** distribution, we allow the *success probability*, p, to vary according to a **beta** distribution:

- $Y \mid P \sim Binomial(n, P)$
- $P \sim Beta(\alpha, \beta)$.

Using the **Laws of Total Expectation and Variance**,

- $E(Y) = E[E(Y \mid P)] = E(nP) = nE(P) = n\left(\frac{\alpha}{\alpha+\beta}\right)$
- $Var(Y) = n\left(\frac{\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}\right)$
    - Let's show this...

Binomial-Poisson Hierarchy

An insect lays a large number of eggs, each surviving with probability p. Let $X =$ the number of eggs that survived, and $Y =$ the number of eggs laid.

Assuming that each egg's survival is independent from the next, we can treat them as *Bernoulli* trials with survival probability p, so the total number of survivors out of Y eggs laid is **binomial**.

- We model Y with a $Poisson(\lambda)$ distribution to reflect the variability in eggs laid across litters from the same mother.

Assume the following hierarchy:

- $X \mid Y \sim Binomial(Y, p)$
- $Y \sim Poisson(\lambda)$

🤔 Find $E(X)$ and $Var(X)$.

Binomial-Poisson Hierarchy

Assume the following hierarchy:

- $X \mid Y \sim Binomial(Y, p)$
- $Y \sim Poisson(\lambda)$

```
p = 0.75
lambda = 100

Y = rpois(10000, lambda = 100)
X = rbinom(10000, size = Y, prob = 0.75)

mean(X)
```

```
## [1] 74.9719
```

```
var(X)
```

```
## [1] 74.38135
```

Binomial-Poisson-Exponential Hierarchy

What if instead we selected one mother insect from a large sample of mothers?

Assume the following hierarchy, where we account for variability across different mothers with a *third stage*:

- $X \mid Y \sim Binomial(Y, p)$
- $Y \mid \Lambda \sim Poisson(\Lambda)$
- $\Lambda \sim Exponential(\beta)$

🤔 Find $E(X)$.

```
p = 0.75
beta = 2

L = rexp(10000, rate = 1/2)
Y = rpois(10000, lambda = L)
X = rbinom(10000, size = Y, prob = 0.75)

mean(X)
```

```
## [1] 1.4835
```