STAT 501 Fall 2025 Course Notes

Chapter 4: Multiple Random Variables

Motivation: Y is a RV that denotes a random numerical outcome of an "experiment". A random variable was defined to be a function from a sample space S into the real numbers. Often we are interested *simultaneously* in two or more outcomes of a random experiment like...

- $Y_1 =$ the number of eggs in a hen Mallard's nest, AND $Y_2 =$ the number of eggs that survive and hatch
- $Y_1 =$ GRE, AND $Y_2 =$ GPA of a prospective graduate student
- A 501 student's $Y_1 =$ number of hours spent on extra practice problems, AND $Y_2 =$ midterm score

In each of these situations, we are interested in the 2-dimensional *random vector* (Y_1, Y_2). A natural extension of this is to more than two dimensions. In fact, consider taking a random sample of n MSU students who live in the dorms and measure whether or not they are registered to vote (1 = yes, 0 = no; assume 18+ years old), define

$Y_i =$

Then, taken together, we have an *n-dimensional* random vector of voter registrations

$\mathbf{Y} =$

and the observed values as,

$\mathbf{y} =$

How could we represent the collection of these outcomes as an event?

Another way to write the intersection of n events is:

or even more simply as,

We can use *multivariate probability distributions* as models for random samples to make inference about the population from which the sample was drawn.

4.1 Joint and Marginal Distributions

DEF 4.1.1 (n-dimensional random vector) An *n-dimensional random vector* is a function from a sample space S into $\mathbb{R}^n$, n-dimensional Euclidean Space.

For simplicity, we'll first focus on *bivariate random variables*, denoted (X, Y).

DEF p. 147 (Bivariate cdf) Let X and Y be two random variables. The joint cumulative distribution function (**bivariate cdf**) of X and Y is

$$F_{XY}(x, y) =$$

This extends to n-dimensional vectors $\mathbf{X} = (X_1, \ldots, X_n)$:

$$F_{\mathbf{X}}(x_1, \ldots, x_n) =$$

Bivariate Die-roll Example Suppose we roll two 6-sided fair dice and let (X, Y) be the random vector of the two outcomes. The visualization of the cdf, $F_{X,Y}(x, y)$, below helps provide some intuition about properties of bivariate cdfs.
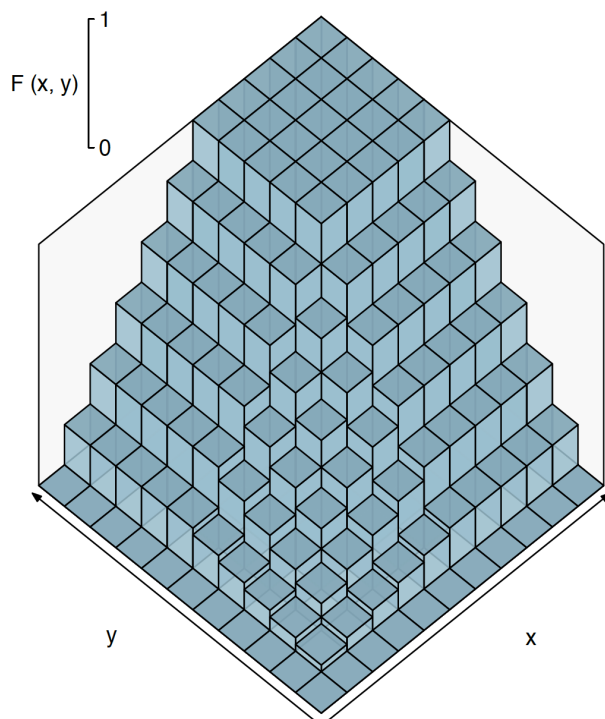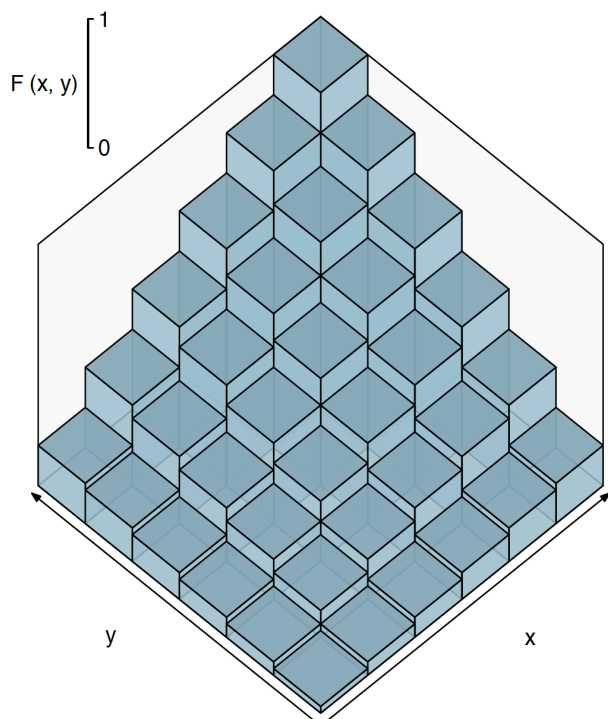
Bivariate cdf properties: The function $F_{XY}(x, y)$ is a bivariate cdf *iff* the following four conditions are met.

1. *If $a \leq b$ and $c \leq d$, then $F_{XY}(b, d) - F_{XY}(a, d) - F_{XY}(b, c) + F_{XY}(a, c) \geq 0$.*

   Write this expression as a probability:

2. $F_{XY}(x, y)$ *is right continuous in each variable, i.e.,* $\lim_{h \to 0^+} F_{XY}(x + h, y) = \lim_{h \to 0^+} F_{XY}(x, y + h) = F_{XY}(x, y)$ *for all x, y.*

3. $\lim_{x \to -\infty, y \to -\infty} F_{XY}(x, y) = \lim_{x \to -\infty} F_{XY}(x, y) = \lim_{y \to -\infty} F_{XY}(x, y) =$

4. $\lim_{x \to \infty, y \to \infty} F_{XY}(x, y) =$ \qquad ; $\lim_{x \to \infty} F_{XY}(x, y) =$ \qquad ; $\lim_{y \to \infty} F_{XY}(x, y) =$

3

Bivariate Discrete Random Variables

> **DEF 4.1.3 (Joint/Bivariate pmf)** Let (X, Y) be a discrete bivariate random vector. Then the function $f_{XY}(x, y)$ from $\mathbb{R}^2$ to $\mathbb{R}$ defined by $f_{XY}(x, y) = P(X = x, Y = y)$ is called the *joint probability mass function* — **joint pmf** — of (X, Y).

Note: A joint pmf can be used to compute the probability of any event in terms of (X, Y). Let A be any subset of $\mathbb{R}^2$. Then,

$$P((X, Y) \in A) = \sum_{\{(x,y):\ (x,y) \in A,\ f_{XY}(x,y) > 0\}} f_{XY}(x, y)$$

Properties of a Joint pmf:

1.

2.

Recall: How do we use a univariate *pmf* to find a univariate *cdf*?

Joint *pmf* and *cdf* relationship: If X and Y are jointly distributed discrete random variables, then the support, $\{(x_1, y_1), (x_2, y_2), ...\}$, is countable, and

$$F_{XY}(x, y) =$$

where $f_{XY}(x, y)$ is a joint (bivariate) pmf.

Bivariate Continuous Random Variables

> **DEF 4.1.10 (Joint/Bivariate pdf)** A function $f_{XY}(x, y)$ from $\mathbb{R}^2$ to $\mathbb{R}$ is called a *joint probability density function* – **joint pdf** — of the continuous bivariate random vector (X, Y) if, for every $A \subset \mathbb{R}^2$,
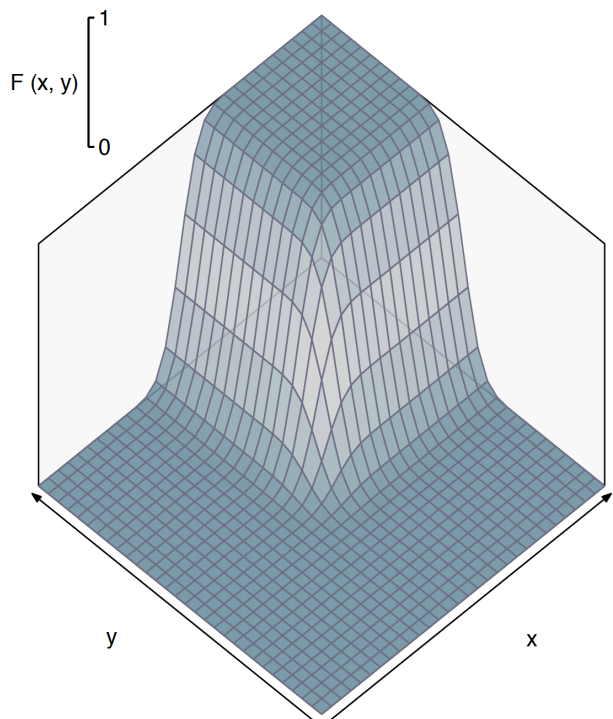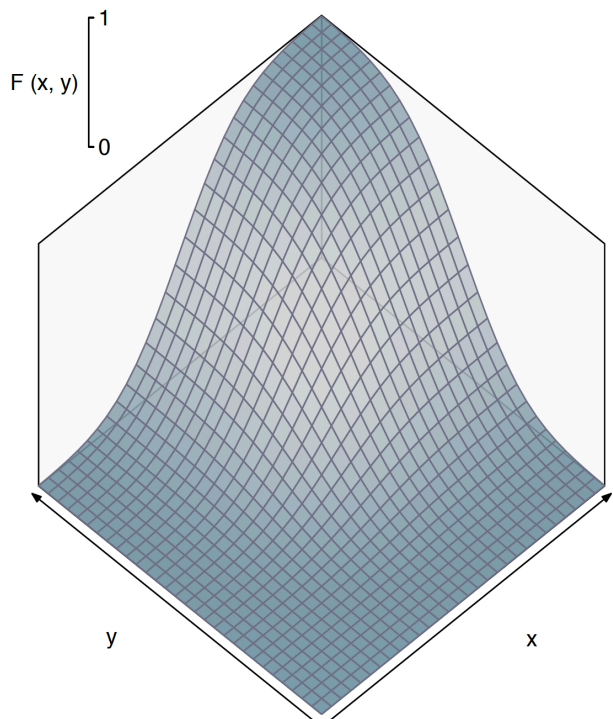> $$P((X, Y) \in A) = \int \int_A f_{XY}(x, y)$$
> .

Note The notation $\int \int_A$ means that the limits of integration are set so that the function is integrated over all $(x, y) \in A$.

Properties of a Joint pdf:

1.

2.

Example (Bivariate Normal Distribution) Let $(X,Y) \sim \mathcal{N}\left(\boldsymbol{\mu} = [0,0]', \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$. A (3-D) plot of the *pdf*, $f_{XY}(x,y)$, a (2-D) plot of a random sample of 500 (x,y) pairs, and a (3-D) plot of the *cdf*, $F_{XY}(x,y)$, are shown below.
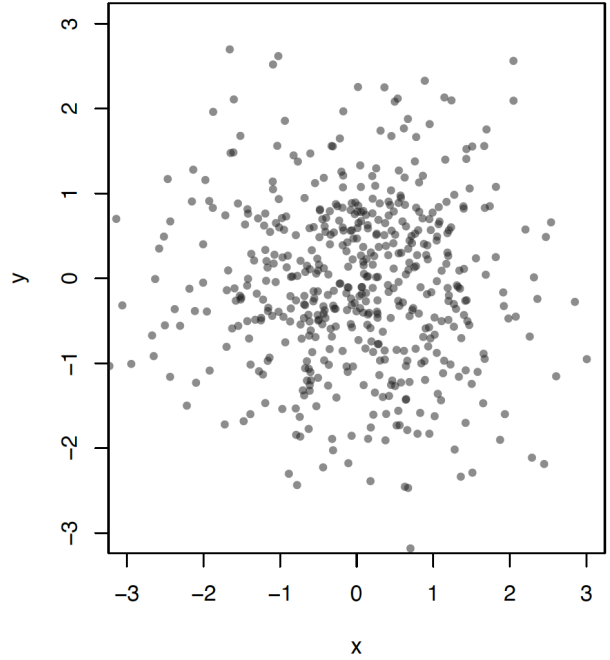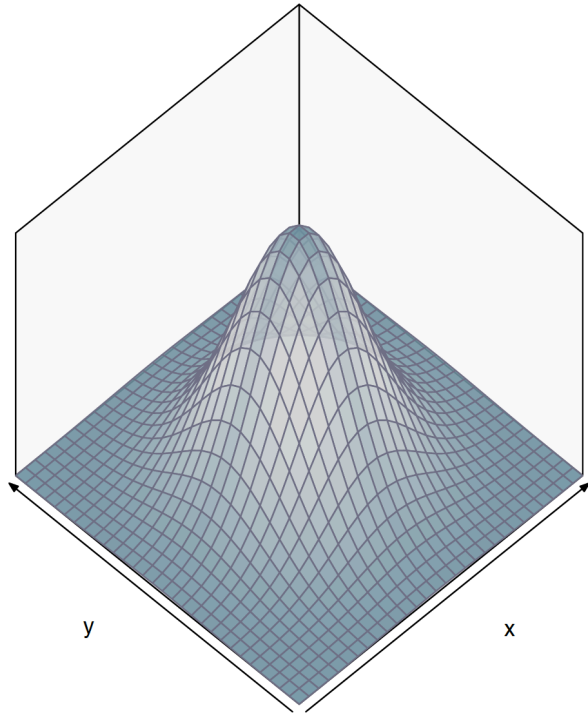
Recall: How do we use a univariate *pdf* to find a univariate *cdf*? How do we use a univariate *cdf* to find a univariate *pdf*?

Joint *pdf* and *cdf* relationship: If X and Y are jointly distributed (absoulutely) continuous random variables, then

$$F_{XY}(x, y) =$$

where $f_{XY}(x, y)$ is a joint (biviariate) pdf.

From the bivariate Fundamental Theorem of Calculus, this implies that

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

Example

Let X and Y have a joint density function given by $f_{XY}(x, y) = \begin{cases} kxy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & else \end{cases}$

- Find k.

- Find $P(X + Y \leq 1)$.

Example (cont)

- Find the joint cdf of X and Y. Use this cdf to obtain $f_{XY}(x, y)$ given above.

Marginal Distributions

THM 4.1.6 (Discrete Marginal Distributions) Let X and Y be jointly distributed discrete RVs with joint pmf $f_{XY}(x, y)$. Then, the **marginal pmfs** of X and Y are given by:

$f_X(x) =$

$f_Y(y) =$

THM 4.1.6 (Continuous Marginal Distributions) Let X and Y be jointly distributed continuous RVs with joint pdf $f_{XY}(x, y)$. Then, the **marginal pdfs** of X and Y are given by:

$f_X(x) =$

$f_Y(y) =$

7

Notes:

- If we sum/integrate one of the RVs out of the bivariate joint *pmf/pdf*, we are left with the *pmf/pdf* of the other variable. Visually, we could "smash down onto a single axis or margin," which is where the term **marginal probability density function** comes from.

- The marginal pmf/pdf of X or Y can be used to compute probabilities or expectations that involve only X or Y. However, to compute a probability or an expectation that simultaneously involves both X and Y, we must use the joint pmf/pdf of X and Y. In general, a joint distribution often tells us additional information about the distribution of X and Y that is not found in the marginal distributions. Therefore, we can find a marginal distribution from a joint distribution, but the converse may not be true!

DEF (Bivariate Expectations) Suppose $g(x, y)$ is a real-valued function. If X and Y are random variables with joint pmf/pdf $f_{XY}(x, y)$,

$E[g(X, Y)] =$

Example (cont)

Let X and Y have a joint density function given by $f_{XY}(x, y) = \begin{cases} 4xy & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & else \end{cases}$.

- Find the marginal density functions (pdfs) of X and Y.

Example (cont)

- Find $E(X^2)$ using both $f_{XY}(x,y)$ and $f_X(x)$ to verify you get the same result!

Example

Let X and Y have a joint density function given by $f_{XY}(x,y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{else} \end{cases}$

- Find $E(X^2Y)$.

Example (cont)

- Find $P(X + Y \geq 1)$.

Example

Let X and Y have a joint probability mass function given by

$$f_{XY}(x, y) = \begin{cases} \binom{y}{x} p^x (1-p)^{y-x} \dfrac{e^{-\lambda}\lambda^y}{y!} & y = 0, 1, 2, ...; x = 0, ..., y \\ 0 & else \end{cases}$$

- Use the joint *pmf* to show that $Y \sim Poisson(\lambda)$ and $X \sim Poisson(p\lambda)$.

Example (cont)

- Find $E(XY)$.

4.2 Conditional Distributions and Independence

Often, when two random variables are observed, the values of the two variables are related. For example, a randomly chosen person's height is typically related to that person's weight—we would think it more likely that a randomly selected person's weight is more than 200 pounds if we were told the person is 73 inches tall than if we were told the person is 41 inches tall. Knowledge about the value of a randomly chosen person's height, X, gives us some information about the value of that person's weight, Y, even if it doesn't tell us the exact value. In such cases, we are often interested in the conditional probability of Y given knowledge that $X = x$.

DEF 4.2.1 (Conditional pmf) Let (X, Y) be a discrete bivariate random vector with joint pmf $f_{XY}(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $P(X = x) = f_X(x) > 0$, the *conditional pmf* of Y given that $X = x$ is the function of y denoted by $f_{Y|X}(y|x)$ and defined by

$$f_{Y|X}(y|x) =$$

For any y such that $P(Y = y) > 0$, the *conditional pmf* of X given that $Y = y$ is the function of x denoted by $f_{X|Y}(x|y)$ and defined by

$$f_{X|Y}(x|y) =$$

DEF 4.2.2 (Conditional pdf) Let (X, Y) be a continuous bivariate random vector with joint pmf $f_{XY}(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$, the *conditional pdf* of Y given that $X = x$ is the function of y denoted by $f_{Y|X}(y|x)$ and defined by

$$f_{Y|X}(y|x) =$$

For any y such that $f_Y(y) > 0$, the *conditional pdf* of X given that $Y = y$ is the function of x denoted by $f_{X|Y}(x|y)$ and defined by

$$f_{X|Y}(x|y) =$$

DEF p. 150 (Conditional Expectation) Let X and Y be two random variables. If $g(Y)$ is a function of Y, then the conditional expected value of $g(Y)$ given $X = x$ is,

$$E(g(Y)|X = x) =$$

Example

Consider the joint pdf $f_{XY}(x, y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{else} \end{cases}$. We can show the marginal pdfs of X

and Y are $f_X(x) = \begin{cases} e^{-x} & 0 < x < \infty \\ 0 & \text{else} \end{cases}$ and $f_Y(y) = \begin{cases} ye^{-y} & 0 < y < \infty \\ 0 & \text{else} \end{cases}$, respectively. (*This is left as extra practice for you.*)

   a. For $x > 0$, find the conditional pdf of $Y|X = x$.

   b. For $y > 0$, find the conditional pdf of $X|Y = y$.

   c. Find $E(X|Y = y)$ and $Var(X|Y = y)$.

Sometimes the conditional distribution of Y given that $X = x$ is different for different values of x (i.e., $Y|X = x$ depends on x). However, in some situations, the knowledge that $X = x$ doesn't give us any more information about Y than what we already had. This important relationship between X and Y is called ***independence***.

DEF 4.2.5 (Independence) Let (X, Y) be a bivariate random vector with joint pdf or pmf $f_{XY}(x, y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are independent random variables if, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$

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

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

$$f_{Y|X}(y|x) =$$

regardless of the value of x. Similarly, $f_{X|Y}(x|y) =$

Example

Let X and Y be jointly distributed random variables with joint pdf

$$f_{XY}(x, y) = \begin{cases} 2x & 0 < x < 1; x \le y \le x + 1 \\ 0 & \text{else} \end{cases}$$

a. Are X and Y independent? Explain why or why not.

b. Find $E(e^Y | X = x)$.

c. For what values of X does the expectation in part b. hold?

Independence—Revisited

> **Lemma 4.2.7 (AKA "Factorization THM")** Let (X, Y) be a bivariate random vector with joint pdf or pmf $f_{XY}(x, y)$. Then X and Y are independent random variables if and only if there exist two functions $g(x)$ and $h(y)$ such that, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$,

<u>Note:</u> *g and h do not necessarily need to be pdfs or pmfs*

Proof: On your own .

Examples

For each of the following joint distributions, determine whether or not X and Y are independent. Explain why or why not.

1. $f_{XY}(x, y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{else} \end{cases}$.

2. $f_{XY}(x, y) = \begin{cases} 4xy & 0 \leq x \leq 1; 0 \leq y \leq 1 \\ 0 & \text{else} \end{cases}$

> **THM 4.2.10 (Using Independence)** Let X and Y be independent random variables.
>
> - For any $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.
>
> - Let $g(x)$ be a function only of x and $h(y)$ a function only of y. Then
>
> $$E(g(X)h(Y)) =$$

<u>Proof:</u> On your own .

THM 4.2.12 (Using Independence in MGFs) Let X and Y be independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of the random variable $Z = X + Y$ is given by

$$M_Z(t) =$$

Theorems 4.2.10 and 4.2.12 extend to n independent random variables (see Theorems 4.6.6 and 4.6.7). That is, if $X_1, \ldots, X_n$ are mutually independent random variables with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$, and $g_1, \ldots, g_n$ are real-valued functions such that $g_i(x_i)$ is a function only of x_i, $i = 1, \ldots, n$, then

- $E\left[\prod_{i=1}^n g_i(X_i)\right] = \prod_{i=1}^n E(g_i(X_i))$, and
- $M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t)$.

Example

Let $X_1, \ldots, X_n \overset{iid}{\sim} Bernoulli(p)$ ("iid" = independent and identically distributed), and define $Y = \sum_{i=1}^n X_i$. What distribution does Y have? Find the mgf of Y.

4.4 Hierarchical Models and Mixture Distributions

Through the use of conditional distributions, we can often model complicated processes by a sequence of relatively simple models placed in a hierarchy.

> **DEF 4.4.4 (Mixture Distribution)** A random variable X is said to have a *mixture distribution* if the distribution of X depends on a quantity that also has a distribution.

Example (Random Sum)   An insect lays a large number of eggs, each surviving with probability p. Suppose the number of eggs laid by the insect, denoted by Y, follows a Poisson distribution with mean λ and assume that each egg's survival is independent of the others. Let X be a random variable denoting the number of survivors. Why is this a mixture distribution?

How could we find $E(X)$ and $Var(X)$?

At times, finding the mean and variance of a random variable in this manner may be difficult and/or tedious. Sometimes, such calculations can be greatly simplified using the following theorems:

> **THM 4.4.3 (Conditional Mean Identity)** For any two random variables X and Y,
>
> $$E(X) = \qquad\qquad\qquad \text{and} \quad E(Y) =$$
>
> provided the expectations exist.

Proof:

> **THM 4.4.7 (Conditional Variance Identity)** For any two random variables X and Y,
>
> $$Var(X) =$$
>
> provided the expectations exist.

Proof:

Example (Revisited) An insect lays a large number of eggs, each surviving with probability p. Suppose the number of eggs laid by the insect, denoted by Y follows a Poisson distribution with mean λ and assume the each egg's survival is independent of the others. Let X be a random variable denoting the number of survivors. Earlier, we stated $X|Y \sim Binomial(Y, p)$, where $Y \sim Poisson(\lambda)$. Use this information to find $E(X)$ and $Var(X)$.

Recall: Earlier in the notes (pg. 11), we used the following joint pmf

$$f_{XY}(x, y) = \begin{cases} \binom{y}{x} p^x (1-p)^{y-x} \frac{e^{-\lambda}\lambda^y}{y!} & y = 0, 1, 2, ...; x = 0, ..., y \\ 0 & else \end{cases}$$

to show that $Y \sim Poisson(\lambda)$ and $X \sim Poisson(\lambda p)$. We could use this to also show that $X|Y \sim Binomial(Y, p)$. This turns out to be the same mixture distribution that we used above! Use the marginal pmf of X to find $E(X)$ and $Var(X)$ and compare to the results above.

Example (Revisited)   Consider a generalization of the previous example, where instead of one mother insect, there are a large number of mothers and one mother is chosen at random. We are still interested in the number of survivors, but the number of eggs laid does not follow the same Poisson distribution for each mother, i.e., assume the rate at which eggs are laid for a randomly selected insect, Λ, follows an exponential distribution with mean $\beta > 0$. Let X be a random variable denoting the number of survivors, and let Y be a random variable denoting the number of eggs laid.

a. Use this information to find $E(X)$, $E(Y)$, and $Var(Y)$.

b. Show that $Y \sim Geometric^* \left(\frac{1}{1+\beta}\right)$[1] and use this result to find $E(Y)$ and $Var(Y)$.

---

[1]The asterisk denotes that we are modeling the number of "failures" until the first success rather than the number of trials.

4.5 Covariance and Correlation

How can we measure the *linear* association between two RVs? Consider the following figures, dashed lines represent the the true mean values (μ_X and μ_Y) for X and Y, respectively.
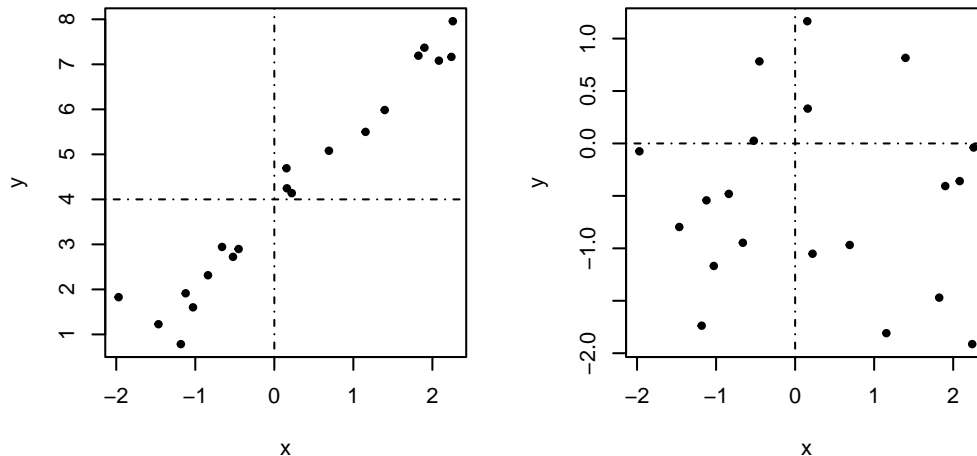
DEF 4.5.1 and THM 4.5.3 (Covariance) Let X and Y be bivariate random variables with $E(X) = \mu_X$, $E(Y) = \mu_Y$, $Var(X) = \sigma_X^2$ and $Var(Y) = \sigma_Y^2$. The *covariance* between X and Y is

$$Cov(X, Y) =$$

Note: $Cov(X, Y) = Cov(Y, X)$. And $Cov(X, X) =$

Proof: On own, (see pg 170 in text).

- What does a large covariance imply?

- What does a small covariance imply?

- How about a covariance of ≈ 0?

- Are covariances easy to compare? (i.e, $Cov(X, Y)$ vs. $Cov(U, V)$)

DEF 4.5.2 (Correlation) The correlation between X and Y is given by,

$$\rho_{XY} =$$

Where ρ_{XY} is also called the correlation coefficient.

- **Covariance** is a measure of direction of linear association between two quantitative RVs, whereas **correlation** measures the direction *and* strength of the linear association between two quantitative random variables.

Example   Consider the joint pdf $f_{XY} = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{else} \end{cases}$. The marginal pdfs of X and Y are $f_X(x) = \begin{cases} e^{-x} & 0 < x < \infty \\ 0 & \text{else} \end{cases}$ and $f_Y(y) = \begin{cases} ye^{-y} & 0 < y < \infty \\ 0 & \text{else} \end{cases}$, respectively. Find $Corr(X, Y)$.

Facts about Correlation, $\rho_{X,Y}$, (THM 4.5.7) For any random variables X and Y,

-

  - $|\rho_{XY} = 1|$ if and only if there exists numbers $a \neq 0$ and b such that $P(Y = aX + b) = 1$. If $\rho_{XY} = 1$, then $a > 0$ and if $\rho_{XY} = -1$, then $a < 0$.

The second property states that if there is a line $Y = aX + b$ with $a \neq 0$ such that the values of (X, Y) have a high probability of being near this line, then the correlation between X and Y will be near 1 or -1. But, if no such line exists, the correlation will be near 0.

Proof: See text 172–173 (on own).

> **THM 4.5.5 (Independence)** If X and Y are independent random variables, $Cov(X, Y) = 0 \implies Corr(X, Y) = 0$.

Proof:

- Is the converse true? That is does $Cov(X, Y) = \rho_{XY} = 0 \overset{?}{\implies} X$ independent of Y?

> **THM 4.5.6 (Variance of Linear Functions of Random Variables)** If X and Y are any two RVs and a and b are any two constants, then
>
> $Var(aX + bY) =$
>
> If X and Y are independent RVs, then
>
> $Var(aX + bY) =$

Note: In addition, if X and Y are any two random variables and a and b are any two constants, then

$Cov(aX, bY) =$

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

$Cov(X, aX + b) =$

Proofs: On own, see pg. 172 in text.

Example (Revisited)   Consider the joint pdf $f_{XY} = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{else} \end{cases}$. The marginal

pdfs of X and Y are $f_X(x) = \begin{cases} e^{-x} & 0 < x < \infty \\ 0 & \text{else} \end{cases}$ and $f_Y(y) = \begin{cases} ye^{-y} & 0 < y < \infty \\ 0 & \text{else} \end{cases}$, respectively.

Find $Var(3X - 4Y + 2)$.

4.3 Bivariate Transformations

Often, we are interested in finding the probability distribution of a function of two or more random variables and/or the joint distribution of functions of multiple random variables. For example, suppose two random variables X and Y have the joint distribution $f_{XY}(x, y)$. We are often interested in a new bivariate random vector (U, V) defined by $U = g_1(X, Y)$ and $V = g_2(X, Y)$, where we want to find either the cdf/pdf/pmf of U or V or the joint distribution of (U, V). In such instances, we can extend the methods used in the univariate case to the bivariate case.

Discrete Case

Suppose two discrete random variables X and Y have the joint pmf $f_{XY}(x, y)$. To find the joint distribution of $U = g_1(X, Y)$ and $V = g_2(X, Y)$:

1. Find the support for either

   - U and $V | U = u$ or
   - V and $U | V = v$.

2. Rewrite the joint pmf of (U, V) in terms of (X, Y).

   - Recall: $f_{XY}(x, y) = P(X = x, Y = y)$
   - No Jacobians!

   Let $B \subset \mathbb{R}^2$, and define A as the set (x, y) where $(g_1(x, y), g_2(x, y)) \in B$:

   $$A =$$

   Then

   $$P((U, V) \in B) =$$

Note: Generally (in this class), g_1 and g_2 will be one-to-one. That is, given $(u, v) \in \{(u, v) : u = g_1(x, y), v = g_2(x, y)$ for some (x, y) such that $f_{XY}(x, y) > 0\}$, the set $A_{UV} = \{(x, y) : u = g_1(x, y), v = g_2(x, y)\}$ will be a singleton set. In this case, define h_1 and h_2 as the single-valued inverse mappings of $u = g_1(x, y)$ and $g_2(x, y)$. Then

$$f_{UV}(u, v) =$$

Example Let $X \sim Pois(\theta)$ and $Y \sim Pois(\lambda)$ where X and Y are independent random variables. Define $U = X + Y$ and $V = Y$. Find the distribution (pmf) of $U = X + Y$.

Continuous Case

Now, suppose X and Y are (absolutely) continuous random variables. If there is a one-to-one transformation from the support of (X, Y) to the support of (U, V), the Jacobian method discussed in the univariate case can be extended to find the joint distribution, $f_{UV}(u, v)$, in the bivariate case.

THM p. 158 (Bivariate transformations) Suppose X and Y are (absolutely) continuous random variables with joint pdf $f_{XY}(x, y)$. Let $U = g_1(X, Y)$ and $V = g_2(X, Y)$. Also suppose that for all (x, y) such that $f_{XY}(x, y) > 0$, the transformation set $u = g_1(x, y)$ and $v = g_2(x, y)$ is one-to-one, and for each (u, v) in the support of (U, V), $x = h_1(u, v)$ and $y = h_2(u, v)$, where h_1 and h_2 are single-valued inverse mappings. If x and y have continuous partial derivatives with respect to u and v and **Jacobian**,

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} \neq 0,$$

then the joint pdf of U and V is

$$f_{UV}(u, v) =$$

Example Let X and Y be independent exponential random variables, both with mean β. Let $U = \dfrac{X}{X + Y}$ and $V = X + Y$. Find the marginal pdf of U.

Example (cont)

Example   Let the random variables X and Y, respectively, denote the premium and claims of a randomly selected insurance policy. Assume X and Y are independent random variables, where $X \sim Gamma(2, 5)$ and $Y \sim Exp(5)$.

    a. What is the probability a randomly chosen claim will be more than the premium?

b. Find the pdf of $Z = X/Y$. How would you interpret this variable?

What if the transformations are not one-to-one? If the transformations are not one-to-one, we can proceed similar to how we did with univariate transformations. That is, we can divide the support of (X, Y) into regions on which the transformations are one-to-one and sum those which have common support for (U, V). For an example of how to use this approach, please see Example 4.3.6 on p. 162 of the text.

In this situation, another option is to find the distribution of (U, V) by rewriting the cdf $F_{UV}(u, v) = P(U \leq u, V \leq v)$ in terms of X and Y.

4.6 Multivariate Distributions

A (univariate) random variable is a function from a sample space S into the real numbers, $\mathbb{R}$. In the bivariate case, we are interested in a two-dimensional random vector (X, Y); that is, a function from a sample space S into $\mathbb{R}^2$. In this section, we will consider *n-dimensional random vectors*, or random vectors with more than two random variables, in more detail.

DEF 4.1.1 (n-dimensional random vector) An *n-dimensional random vector* is a function from a sample space S into $\mathbb{R}^n$, n-dimensional Euclidean Space.

In this case, many of the concepts discussed earlier—marginal and conditional distributions, independence, and so on—generalize from the bivariate to the multivariate setting.

Multivariate cdf

The joint **cumulative distribution function** of $\mathbf{X} = (X_1, X_2, ..., X_n)'$ on $\mathbb{R}^n$ is defined as,

$$F_{\mathbf{X}}(\mathbf{x}) =$$

Multivariate pmf (Eqn 4.6.1)

If $\mathbf{X} = (X_1, X_2, ..., X_n)'$ is a discrete random vector (the sample space is countable), then the **joint pmf** is the function

$$f_{\mathbf{X}}(\mathbf{x}) =$$

Then for any $A \in \mathbb{R}^n$, $P(\mathbf{X} \in A) = \sum_A f_{\mathbf{X}}(\mathbf{x})$.

Multivariate pdf (Eqn 4.6.2)

If $\mathbf{X} = (X_1, X_2, ..., X_n)'$ is an (absolutely) continuous random vector, then the **joint pdf** is the function $f_{\mathbf{X}}(\mathbf{x})$ that, for any $A \in \mathbb{R}^n$, satisfies

$$P(\mathbf{X} \in A) =$$

Note: These integrals are n-fold integrals with limits of integration set so that the integration is over all points $\mathbf{x} \in A$.

Marginal Distributions (Equations 4.64 & 4.65)

The marginal pdf or pmf of any subset of the coordinates of $\mathbf{X} = (X_1, \ldots, X_n)'$ can be computed by integrating or summing the joint pdf or pmf over all possible values of the other coordinates.

Continuous case: If $\mathbf{X} = (X_1, X_2, ..., X_k, X_{k+1}, ..., X_n)'$ is an (absolutely) continuous random vector, then the **marginal pdf** of $(X_1, X_2, ..., X_k)'$ is

$$f(x_1, ..., x_k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, ..., x_n) dx_{k+1} dx_{k+2} ... dx_n$$

for every $(x_1, ..., x_k)' \in \mathbb{R}^k$. That is, if we integrate $n - k$ ($k < n$) continuous random variables out of the multivariate *pdf*, we are left with the marginal *pdf* of the other k variable(s).

For example, the marginal pdf of the random variable X_j is

$$f_{X_j}(x_j) =$$

for every $x_j \in \mathbb{R}$.

Discrete case: Similarly, we can get the marginal pmf of a discrete random variable by summing the other random variables out of the joint pmf. If $\mathbf{X} = (X_1, X_2, ..., X_k, X_{k+1}, ..., X_n)'$ is discrete random vector, then the **marginal pmf** of $(X_1, X_2, ..., X_k)'$ is

$$f(x_1, ..., x_k) = \sum_{(x_{k+1}, x_{k+2}, ..., x_n) \in \mathbf{R}^{n-k}} f(x_1, ..., x_n)$$

for every $(x_1, ..., x_k) \in \mathbb{R}^k$.

For example, the marginal pmf of the random variable X_j is

$$f_{X_j}(x_j) =$$

for every $x_j \in \mathbb{R}$.

Multivariate Expectations (Eqn 4.6.3)

Suppose $g(\mathbf{x}) = g(x_1, ..., x_n)$ is a real-valued function defined on the sample space of $\mathbf{X} = (X_1, ..., X_n)'$. That is, $g : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \subset \mathbb{R}^n$ is the support of $\mathbf{X}$. Then the expected value of $g(\mathbf{X})$ is

Continuous case: If $f_{\mathbf{X}}(\mathbf{x})$ is the pdf of $\mathbf{X}$, then

$$E(g(\mathbf{X})) = E(g(X_1, .., X_n)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g(x_1, ..., x_n) f_{\mathbf{X}}(x_1, ..., x_n) dx_1, .., dx_n.$$

Discrete case: If $f_{\mathbf{X}}(\mathbf{x})$ is the pmf of $\mathbf{X}$, then

$$E(g(\mathbf{X})) = E(g(X_1, .., X_n)) = \sum_{\mathcal{X} \in \mathbb{R}^n} g(x_1, ..., x_n) f_{\mathbf{X}}(x_1, ..., x_n).$$

Conditional Distributions (Eqn 4.6.6)

The **conditional pdf or pmf** of a subset of the coordinates of $\mathbf{X} = (X_1, ..., X_n)'$ given the values of the remaining coordinates is obtained by dividing the joint pdf or pmf by the marginal pdf or pmf of the remaining coordinates. For example, if $f(x_1, x_2..., x_k) > 0$, the conditional pdf or pmf of $(X_{k+1}, ..., X_n)$ given $(X_1, ..., X_k)$ is defined by

$$f(x_{k+1}, ..., x_n | x_1, ..., x_k) = \frac{f(x_1, ..., x_n)}{f(x_1, ..., x_k)}.$$

Example   Let X_1, X_2, X_3 be continuous random variables with joint pdf

$$f(x_1, x_2, x_3) = \begin{cases} c & 0 < x_1 < x_2 < x_3 < 1 \\ 0 & else \end{cases}$$

a. Find c.
b. What is the marginal pdf of X_3?
c. Find $f_{X_1 X_2}(x_1, x_2)$.
d. Find $E(X_1 X_2 X_3)$.
e. Find the conditional pdf $f_{X_1 X_2 | X_3}(x_1, x_2 | x_3)$.
f. Find the conditional pdf $f_{X_3 | X_1 X_2}(x_3 | x_1, x_2)$.

Independence of Random Vectors

The concept of independence can also be extended from the bivariate case to the multivariate case.

DEF 4.6.5 (Mutual Independence) Let $\mathbf{X} = (X_1, ..., X_n)'$ be an n-dimensional random vector with joint pdf or pmf $f_\mathbf{X}(x_1, ..., x_n)$ and marginal pdfs or pmfs for each X_i denoted by $f_{X_i}(x_i)$. Then $X_1, ..., X_n$ are **mutually** **independent random variables** if, for every $(x_1, ..., x_n)' \in \mathbb{R}^n$,

$$f(x_1, x_2, ..., x_n) = f_{X_1}(x_1)...f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Note: Mutual independence implies that any pair of random variables is **pairwise independent**. However, the converse is not true—pairwise independence does not imply mutual independence!

Example Suppose you flip a fair coin twice. Define the random vector $\mathbf{X} = (X_1, X_2, X_3)'$ as the trio of indicator variables,

- $X_1 = 1$ if head on the first toss; 0 otherwise
- $X_2 = 1$ if head on the second toss; 0 otherwise
- $X_3 = 1$ if same outcome on both tosses; 0 otherwise

Are X_1, X_2, and X_3 pairwise independent? mutually independent?

> **THM 4.6.11 (Generalization of Lemma 4.2.7—Factorization THM; Random Variable Version)** Let $\mathbf{X} = (X_1, ..., X_n)'$ be an n-dimensional random vector with joint pdf or pmf $f_{\mathbf{X}}(x_1, ..., x_n)$. Then, $X_1, ..., X_n$ are mutually independent random variables *if and only if* there exist functions $g_i(x_i)$ such that, for every $\mathbf{x} = (x_1, x_2, ..., x_n)' \in \mathbb{R}^n$, the joint pdf or pmf of $\mathbf{X} = (X_1, ..., X_n)'$ can be written as
>
> $$f_{\mathbf{X}}(x_1, x_2, ..., x_n) = g_1(x_1)g_2(x_2)...g_n(x_n).$$

Note: g_i are not necessarily pdfs or pmfs, and the supports must also factor!

Properties of Mutually Independent Random Variables

Mutually independent random variables have many nice properties! These properties are particularly helpful when working with random samples from a large population, where $X_1, \ldots, X_n$ are independent and identically distributed according to some probability model.

> **THM 4.6.12 (Generalization of THM 4.3.5—Random Variable Version)** Let $X_1, ..., X_n$ be independent random variables, and let $g_i(x_i)$ be a function of x_i only. Then the random variables $U_i = g_i(X_i)$ are mutually independent.

Variance of Linear Functions of Random Variables

Let $(X_1, ..., X_n)$ be an n-dimensional random vector, and let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be real-valued constants. In general,

$$Var\left(\sum_{i=1}^{n} a_i X_i + b_i\right) =$$

However, if $X_1, ..., X_n$ are *mutually independent* random variables, then

$$Var\left(\sum_{i=1}^{n} a_i X_i + b_i\right) =$$

Multivariate Expectations—Mutually Independent Random Variables

> **THM 4.6.6 (Generalization of THM 4.2.12)** Let $X_1, ..., X_n$ be mutually independent random variables, and let $g_i(x_i)$ be real-valued functions such that $g_i(x_i)$ is a function of x_i only. Then
>
> $$E(g_1(X_1)g_2(X_2)...g_n(X_n)) =$$

Multivariate Transformations for Continuous Random Variables (One-to-One Version)

Suppose $X_1, ..., X_n$ are (absolutely) continuous random variables with joint pdf $f_{X_1...X_n}(x_1, ..., x_n)$. Let $A = \{\mathbf{x} : f_{\mathbf{X}}(\mathbf{x}) > 0\}$ be the support of $\mathbf{X}$. Consider the random vector $(U_1, ..., U_n)$, defined by $U_i = g_i(X_1, ..., X_n)$. Suppose that the transformation $(U_1, ..., U_n) = (g_1(\mathbf{X}), ..., g_n(\mathbf{X}))$ is a one-to-one transformation from A to $B = \{\mathbf{u} : (u_1, ..., u_n) = (g_1(\mathbf{x}), ..., g_n(\mathbf{x}))$ for some $\mathbf{x}\}$. Then, for each i, if the ith inverse exists, $x_i = h_i(u_1, ..., u_n)$, $i = 1, ..., n$, and if $x_1, x_2, ..., x_n$ have continuous partial derivatives with respect to $u_1, u_2, ..., u_n$ and Jacobian,

$$
J = \begin{vmatrix}
\frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \cdots & \frac{\partial x_1}{\partial u_n} \\
\frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \cdots & \frac{\partial x_2}{\partial u_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \cdots & \frac{\partial x_n}{\partial u_n}
\end{vmatrix} \neq 0,
$$

then the joint pdf of $\mathbf{U} = (U_1, U_2, \ldots, U_n)$ is,

$$
f_{U_1...U_n}(\mathbf{u}) = \begin{cases} f_{X_1...X_n}(h_1(\mathbf{u}), \ldots h_n(\mathbf{u}))|J| & \mathbf{u} \in B \\ 0 & else \end{cases}.
$$

Although we can use this approach, we are often interested in obtaining the distribution of $Y = \sum_{i=1}^{n} X_i$. In some of these instances, moment generating functions can be very helpful.

DEF: Multivariate Moment Generating Function Suppose $X_1, ..., X_n$ are random variables with joint pdf $f_{X_1,...,X_n}(x_1, ..., x_n)$. The **multivariate moment generating function** of $X_1, ..., X_n$ is,

$$M_{X_1,...,X_n}(t_1, \ldots, t_n) = $$

for all t_i near zero ($|t_i| < h_i, i = 1, 2, ...n$) where the expectation exists.

Multivariate MGFs can be used to. . .

Find moments:

$$E(X_i^k) = $$

- Example:

$$\frac{\partial}{\partial t_1} M_{X_1,\dots,X_n}(\mathbf{t}) \bigg|_{\mathbf{t}=\mathbf{0}} =$$

- Example:

$$\frac{\partial^2}{\partial t_1^2} M_{X_1,\dots,X_n}(\mathbf{t}) \bigg|_{\mathbf{t}=\mathbf{0}} =$$

- Example:

$$\frac{\partial^2}{\partial t_1 t_3} M_{X_1,\dots,X_n}(\mathbf{t}) \bigg|_{\mathbf{t}=\mathbf{0}} =$$

Find univariate MGFs:

$$M_{X_1,\dots,X_n}(t_1 = t, t_2 = 0, \dots, t_n = 0) =$$

Conduct transformations:

Suppose $\mathbf{X} = (X_1, \dots, X_n)'$ is an n-dimensional random vector. Often, we are interested in the distribution of $Y = \mathbf{X}' \cdot \mathbf{1} = \sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 + \dots + X_n$. In this case, the MGF of the transformed RV, Y is

$$M_Y(t) =$$

THM 4.6.7: In the case where $X_1, \dots, X_n$ are **mutually independent**,

$$M_Y(t) =$$

THM 4.6.7: In the very special case where $X_1, \dots, X_n$ are **independent and identically distributed (iid)**,

$$M_Y(t) =$$

Example Suppose $X_1, ..., X_n$ are independent random variables where $X_i \sim Poisson(\lambda_i)$. Use mgfs to find the distribution of $Y = \sum_{i=1}^{n} X_i$.

Example Suppose $X_1, ..., X_n$ are independent and identically distributed random variables where $X_i \sim Normal(\mu, \sigma^2)$ Use mgfs to find the distribution of $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

38

Special Multivariate Distributions

In the next part, we explore two different multivariate distributions: the bivariate normal distribution (Section 4.5) and the multinomial distribution (Section 4.6). The multivariate normal distribution is the foundation for most modern statistical theory. In general, the multivariate normal distribution is defined for n continuous random variables, $X_1, ..., X_n$. Because of its complexity, we look only at the bivariate normal distribution.

Bivariate Normal Distribution

DEF 4.5.10 (Bivariate Normal Distribution) Let $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, $0 < \sigma_X$, $0 < \sigma_Y$, and $-1 < \rho_{XY} < 1$ be five real numbers. The **bivariate normal pdf** with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2 and correlation ρ_{XY} is the bivariate pdf given by,

$$
f_{XY}(x,y) = \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}\right)^{-1} \times
$$
$$
\exp\left\{\frac{-1}{2(1-\rho_{XY}^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}
$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$. We denote this as $(X,Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$, or, in matrix notation,

$$
\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right).
$$

Properties of the Bivariate Normal Distribution

Suppose the random vector (X,Y) has a bivariate normal distribution with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2 and correlation ρ_{XY}. Then the following properties hold.

1. $E(X) = \mu_X$, $E(Y) = \mu_Y$, $Var(X) = \sigma_X^2$, $Var(Y) = \sigma_Y^2$, and $Cor(X,Y) = \rho_{XY}$.

2. X and Y both have marginal normal distributions: $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$.

3. For any constants a, b, and c, the distribution of $aX + bY + c$ is normal:

$$
aX + bY + c \sim N\left(a\mu_X + b\mu_Y + c, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho_{XY}\sigma_X\sigma_Y\right).
$$

4. The conditional density of $Y|X = x$ is normal:

$$
Y|X = x \sim N\left(\mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho_{XY}^2)\right).
$$

Note: All of the normal marginal and conditional distributions are derived from the starting point of a bivariate normal distribution. The derivation does not go in the opposite direction—marginal normality does not imply joint normality!

Proof: Show the conditional distribution of $Y|X = x$ is:

$$Y|X = x \sim N\left(\mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho_{XY}^2)\right).$$

Proof: Show $Cor(X, Y) = \rho_{XY} = \rho$.

Bivariate Normal Distribution—Independence

For any random variables X and Y, we have shown that if X and Y are independent, then $\rho_{XY} = Cor(X, Y) = 0$. However, this result does not go in the other direction—$\rho_{XY} = 0$ does *not* imply that X and Y are independent... unless (X, Y) has a bivariate normal distribution.

Let (X, Y) be a bivariate normal random vector and assume $\rho_{XY} = 0$. Show that X and Y are independent.

Summary of result: Assume (X, Y) has a bivariate normal distribution where $\rho_{XY} = Cor(X, Y)$. Then X and Y are independent *if and only if* $\rho_{XY} = 0$!

Multinomial Distribution

DEF 4.6.2 (Multinomial Distribution) Let m and n be positive integers, and let $p_1, \ldots, p_n$ be numbers satisfying $0 \leq p_i \leq 1$ for each $i = 1, \ldots, n$ and $\sum_{i=1}^{n} p_i = 1$. Then the random vector $\mathbf{X} = (X_1, \ldots, X_n)'$ has a **multinomial distribution** with m trials and cell probabilities $p_1, \ldots, p_n$ if the joint pmf of $\mathbf{X}$ is

$$f_{\mathbf{X}}(x_1, \ldots, x_n | p_1, \ldots, p_n) = \frac{m!}{x_1! x_2! \cdots x_n!} p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n} I_A(\mathbf{x}) = \binom{m}{x_1, x_2, \ldots, x_n} \prod_{i=1}^{n} p_i^{x_i} I_A(\mathbf{x}),$$

where $A = \{\mathbf{x} \in \mathbb{R}^n | x_i \in \{0, 1, \ldots, m\}, \sum_{i=1}^{n} x_i = m\}$. The quantity $\binom{m}{x_1, x_2, \ldots, x_n} = \frac{m!}{x_1! x_2! \cdots x_n!}$ is called a **multinomial coefficient**.

The multinomial distribution models the following scenario:

1. m independent, identical trials.

2. Each trial results in one of n possible outcomes.

3. An outcome of type i has probability p_i, where $\sum_{i=1}^{n} p_i = 1$ and p_i remains the same from trial to trial.

4. The random variables of interest are $X_1, \ldots, X_n$, where X_i equals the number of trials that result in outcome i, so the sum $\sum_{i=1}^{n} X_i = m$.

Hmmm... this should sound vaguely familiar... Why?

Notes:

- This distribution is extremely useful for understanding the distribution of an order statistic.

- Often, the number of possible outcomes per trial is more than two. For example, experiments that involve blood typing have at least four possible outcomes per trial. In addition, surveys might have multiple choices for a single question.

Example When commercial aircraft are inspected, wing cracks are reported as nonexistent, detectable, or critical. The history of a particular fleet indicates that 70% of the planes inspected have no wing cracks, 25% have detectable wing cracks, and 5% have critical wing cracks. Five planes are randomly selected.

- Find the probability that one has a critical crack, two have detectable cracks, and two have no cracks.

- Find the probability that at least one plane has critical cracks.

4.7 Inequalities

In Section 3.6, we encountered probabilistic inequalities derived using probabilistic arguments (e.g., Chebychev's Inequality). Now, we're going to discuss inequalities that apply to probabilities and expectations, but are based on arguments that use properties of functions and numbers.

THM 4.7.3 (Cauchy-Schwarz Inequality) For any two random variables X and Y,

$$|E(XY)| \leq E[|XY|] \leq (E|X|^2)^{1/2}(E|Y|^2)^{1/2} = \sqrt{E(X^2)E(Y^2)}.$$

Note: The Cauchy-Schwarz Inequality is a special case of Hölder's Inequality when $p = q = 2$. See Theorem 4.7.2 on p. 187 for the general version of this inequality.

Example   Suppose X and Y have means μ_X and μ_Y and variances σ_X^2 and σ_Y^2, respectively. Use the Cauchy-Schwarz Inequality to prove $0 \leq \rho_{XY}^2 \leq 1$.

THM 4.7.7 (Jensen's Inequality) For any random variable X, if $g(x)$ is a convex function, then

$$E(g(X)) \geq g(E(X)).$$

On the other hand, if $g(x)$ is a concave function, then

$$E(g(X)) \leq g(E(X)).$$

Recall: A function $g(x)$ is **convex** if

$$g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda)g(x_2)$$

for all x_1 and x_2, and $0 < \lambda < 1$. The function $g(x)$ is **concave** if $-g(x)$ is convex. Basically, this means a convex function lies above all of its tangent lines. To check whether $g(x)$ is convex, we can find the second derivative of $g(x)$, $g''(x)$. If $g''(x) \geq 0$ for all x, then $g(x)$ is convex; if $g''(x) \leq 0$ for all x, then $g(x)$ is concave.

Examples

- Consider the function $g(x) = x^2$. Is this function convex?

- For $x > 0$, consider the function $g(x) = \dfrac{1}{x}$. Is this function convex over $\{x : x > 0\}$?