Week 9: Multidimensional Change of Variable

MATH697

Sahir Bhatnagar

October 31, 2017

McGill University

First Some Terminology

One-to-one and Onto

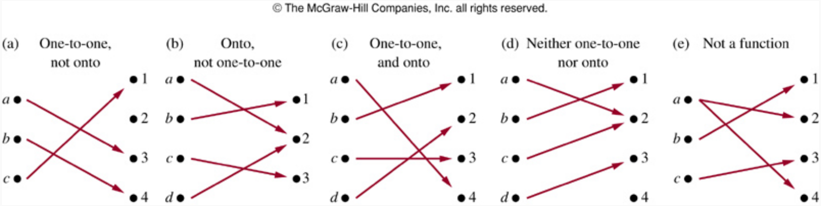

Figure 1

Multidimensional Change of Variable

- **Univariate Transformations**: We previously discussed the problem of starting with a single random variable X, forming some function of X, such as X^2 or e^X, to obtain a new random variable $Y = h(X)$

- **Univariate Transformations**: We previously discussed the problem of starting with a single random variable X, forming some function of X, such as X^2 or e^X, to obtain a new random variable $Y = h(X)$
- **Multivariate Transformations**: We now generalize this scenario by starting with more than a single random variable

Introduction

- Consider a system having a component that can be replaced just once before the system itself expires.

- Consider a system having a component that can be replaced just once before the system itself expires.
- X: lifetime of the original component

Introduction

- Consider a system having a component that can be replaced just once before the system itself expires.
- X: lifetime of the original component
- Y: lifetime of the replacement component

- Consider a system having a component that can be replaced just once before the system itself expires.
- X: lifetime of the original component
- Y: lifetime of the replacement component
- Any of the following functions of X and Y may be of interest to an investigator:

Introduction

- Consider a system having a component that can be replaced just once before the system itself expires.
- X: lifetime of the original component
- Y: lifetime of the replacement component
- Any of the following functions of X and Y may be of interest to an investigator:

    1. The total lifetime $X + Y$

Introduction

- Consider a system having a component that can be replaced just once before the system itself expires.
- X: lifetime of the original component
- Y: lifetime of the replacement component
- Any of the following functions of X and Y may be of interest to an investigator:

  1. The total lifetime $X + Y$
  2. The ratio of lifetimes X/Y; e.g. if the ratio is 2, the original component lasted twice as long as its replacement

- Consider a system having a component that can be replaced just once before the system itself expires.
- X: lifetime of the original component
- Y: lifetime of the replacement component
- Any of the following functions of X and Y may be of interest to an investigator:
    1. The total lifetime $X + Y$
    2. The ratio of lifetimes X/Y; e.g. if the ratio is 2, the original component lasted twice as long as its replacement
    3. The ratio $X/(X + Y)$; the proportion of system lifetime during which the original component operated

- Given two continuous random variables X and Y, consider forming two new random variables

$$U = g_1(X, Y) \qquad V = g_2(X, Y)$$

- Given two continuous random variables X and Y, consider forming two new random variables

$$U = g_1(X, Y) \qquad V = g_2(X, Y)$$

- We now focus on finding the joint distribution of these two new continuous variables (U, V)

Joint Distribution of Two New Random Variables

Recall, $U = g_1(X, Y)$ and $V = g_2(X, Y)$.

Joint Distribution of Two New Random Variables

Recall, $U = g_1(X, Y)$ and $V = g_2(X, Y)$. Let

$$f_{X,Y}(x, y) = \text{the joint pdf of the two original variables}$$
$$f_{U,V}(u, v) = \text{the joint pdf of the two new variables}$$

Joint Distribution of Two New Random Variables

Recall, $U = g_1(X, Y)$ and $V = g_2(X, Y)$. Let

$$f_{X,Y}(x, y) = \text{the joint pdf of the two original variables}$$
$$f_{U,V}(u, v) = \text{the joint pdf of the two new variables}$$

- $g_1(\cdot)$ and $g_2(\cdot)$ are functions that express the **new variables in terms of the original ones**

Joint Distribution of Two New Random Variables

Recall, $U = g_1(X, Y)$ and $V = g_2(X, Y)$. Let

$$f_{X,Y}(x, y) = \text{the joint pdf of the two original variables}$$
$$f_{U,V}(u, v) = \text{the joint pdf of the two new variables}$$

- $g_1(\cdot)$ and $g_2(\cdot)$ are functions that express the **new variables in terms of the original ones**
- The general result presumes that these functions can be inverted to solve for the original variables in terms of the new ones:

$$X = h_1(U, V) \qquad Y = h_2(U, V)$$

- For example, if

$$u = g_1(x, y) = x + y \qquad v = g_2(x, y) = x - y$$

then

$$x = h_1(u, v) = \frac{1}{2}(u + v) \qquad y = h_2(u, v) = \frac{1}{2}(u - v)$$

Joint Distribution of Two New Random Variables

- For example, if

$$u = g_1(x, y) = x + y \qquad v = g_2(x, y) = x - y$$

  then

$$x = h_1(u, v) = \frac{1}{2}(u + v) \qquad y = h_2(u, v) = \frac{1}{2}(u - v)$$

- $\mathcal{A} = \{(x, y) : f(x, y) > 0\}$ is the region of positive density for the original variables

Joint Distribution of Two New Random Variables

- For example, if

$$u = g_1(x, y) = x + y \qquad v = g_2(x, y) = x - y$$

then

$$x = h_1(u, v) = \frac{1}{2}(u + v) \qquad y = h_2(u, v) = \frac{1}{2}(u - v)$$

- $\mathcal{A} = \{(x, y) : f(x, y) > 0\}$ is the region of positive density for the original variables
- $\mathcal{B} = \{(u, v) : u = g_1(x, y), v = g_2(x, y) \text{ for some } (x, y) \in \mathcal{A}\}$ is the region of positive density for the new variables

Joint Distribution of Two New Random Variables

- For example, if

$$u = g_1(x, y) = x + y \qquad v = g_2(x, y) = x - y$$

then

$$x = h_1(u, v) = \frac{1}{2}(u + v) \qquad y = h_2(u, v) = \frac{1}{2}(u - v)$$

- $\mathcal{A} = \{(x, y) : f(x, y) > 0\}$ is the region of positive density for the original variables
- $\mathcal{B} = \{(u, v) : u = g_1(x, y), v = g_2(x, y) \text{ for some } (x, y) \in \mathcal{A}\}$ is the region of positive density for the new variables
- $\mathcal{B}$ is the **image** of $\mathcal{A}$ under the transformation

- Suppose that the partial derivative of each $h_i(u, v)$ with respect to both u and v exists for every pair (u, v) in $\mathcal{B}$ and is continuous

Joint Distribution of Two New Random Variables

- Suppose that the partial derivative of each $h_i(u,v)$ with respect to both u and v exists for every pair (u,v) in $\mathcal{B}$ and is continuous
- Define the **Jacobian** as

$$\begin{aligned}
\mathcal{J} &= \det \begin{bmatrix} \frac{\partial h_1(u,v)}{\partial u} & \frac{\partial h_1(u,v)}{\partial v} \\ \frac{\partial h_2(u,v)}{\partial u} & \frac{\partial h_2(u,v)}{\partial v} \end{bmatrix} \\
&= \frac{\partial h_1(u,v)}{\partial u}\frac{\partial h_2(u,v)}{\partial v} - \frac{\partial h_1(u,v)}{\partial v}\frac{\partial h_2(u,v)}{\partial u}
\end{aligned}$$

Joint Distribution of Two New Random Variables

- Suppose that the partial derivative of each $h_i(u, v)$ with respect to both u and v exists for every pair (u, v) in $\mathcal{B}$ and is continuous
- Define the **Jacobian** as

$$\begin{aligned}\mathcal{J} &= \det \begin{bmatrix} \frac{\partial h_1(u,v)}{\partial u} & \frac{\partial h_1(u,v)}{\partial v} \\ \frac{\partial h_2(u,v)}{\partial u} & \frac{\partial h_2(u,v)}{\partial v} \end{bmatrix} \\ &= \frac{\partial h_1(u,v)}{\partial u}\frac{\partial h_2(u,v)}{\partial v} - \frac{\partial h_1(u,v)}{\partial v}\frac{\partial h_2(u,v)}{\partial u}\end{aligned}$$

- Then

$$f_{U,V}(u, v) = \begin{cases} f_{X,Y}(h_1(u, v), h_2(u, v)) \times |\mathcal{J}| & u, v \in \mathcal{B} \\ 0 & else \end{cases}$$

Theorem 1 (Multivariate Transformations (2 random variables))

Let X and Y be jointly continuous, with PDF $f_{X,Y}(x,y)$ on the set $\mathcal{A} = \{(x,y) : f_{X,Y}(x,y) > 0\}$. Let $U = g_1(X,Y)$ and $V = g_2(X,Y)$ on the set $\mathcal{B} = \{(u,v) : u = g_1(x,y), v = g_2(x,y) \ for\ some\ (x,y) \in \mathcal{A}\}$, where $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$ are differentiable functions. Assume g_1, g_2 is one-to-one on $\mathcal{A}$, i.e., we can solve the equations $u = g_1(x,y), v = g_2(x,y)$ for x and y denoted by $x = h_1(u,v)$ and $y = h_2(u,v)$. Then U and V are also jointly continuous with PDF

$$f_{U,V}(u,v) = \begin{cases} f_{X,Y}(h_1(u,v), h_2(u,v)) \times |\mathcal{J}| & u,v \in \mathcal{B} \\ 0 & else \end{cases} \tag{1}$$

where

$$\mathcal{J} = det \begin{bmatrix} \frac{\partial h_1(u,v)}{\partial u} & \frac{\partial h_1(u,v)}{\partial v} \\ \frac{\partial h_2(u,v)}{\partial u} & \frac{\partial h_2(u,v)}{\partial v} \end{bmatrix}$$

Example 2 (Uniform Square Transformation)

Let X and Y be independent $Uniform(0, 1)$ random variables.

Consider the two random variables $U = X + Y$ and $V = X - Y$.

1. Determine the joint PDF of U and V
2. What is the marginal PDF of U and marginal PDF of V ?

Example 3 (Distribution of the product of beta variables)

Let $X \sim Beta(\alpha, \beta)$ and $Y \sim Beta(\alpha + \beta, \gamma)$ be independent random variables. The joint PDF of (X, Y) is

$$f_{X,Y}(x, y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta)\Gamma(\gamma)} y^{\alpha+\beta-1}(1-y)^{\gamma-1}$$

on the set $\mathcal{A} = \{(x, y) : 0 < x < 1, 0 < y < 1\}$ and 0 elsewhere. Consider the transformation $U = XY$ and $V = X$

1. Determine the joint PDF of U and V
2. What is the marginal PDF of U and marginal PDF of V ?

Multivariate Transformation Example

Example 4 (Sum and difference of normal variables)

Let $X \sim N(0,1)$ and $Y \sim N(0,1)$ be independent standard normal random variables. The joint PDF of (X, Y) is

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\{-x^2/2\} \exp\{-y^2/2\}$$

on the set $\mathcal{A} = \{(x,y) : -\infty < x < \infty, -\infty < y < \infty\}$.
Consider the transformation $U = X + Y$ and $V = X - Y$

1. Determine the joint PDF of U and V
2. What is the marginal PDF of U and marginal PDF of V ?

When the transformation is not one-to-one

- In many situations, the transformation of interest is not one-to-one

When the transformation is not one-to-one

- In many situations, the transformation of interest is not one-to-one
- Let $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$

When the transformation is not one-to-one

- In many situations, the transformation of interest is not one-to-one
- Let $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$
- Suppose $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ form a partition of $\mathcal{A}$ with the following properties:

When the transformation is not one-to-one

- In many situations, the transformation of interest is not one-to-one
- Let $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$
- Suppose $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ form a partition of $\mathcal{A}$ with the following properties:
    - The set A_0, which may be empty, satisfies $P\big((X, Y) \in \mathcal{A}_0\big) = 0$

When the transformation is not one-to-one

- In many situations, the transformation of interest is not one-to-one
- Let $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$
- Suppose $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ form a partition of $\mathcal{A}$ with the following properties:
  - The set A_0, which may be empty, satisfies $P\big((X, Y) \in \mathcal{A}_0\big) = 0$
  - The transformation $U = g_1(X, Y)$ and $V = g_2(X, Y)$ is a one-to-one transformation from $\mathcal{A}_i$ onto $\mathcal{B}$ for each $i = 1, 2, \ldots, k$

When the transformation is not one-to-one

- In many situations, the transformation of interest is not one-to-one
- Let $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$
- Suppose $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ form a partition of $\mathcal{A}$ with the following properties:
  - The set A_0, which may be empty, satisfies $P\big((X, Y) \in \mathcal{A}_0\big) = 0$
  - The transformation $U = g_1(X, Y)$ and $V = g_2(X, Y)$ is a one-to-one transformation from $\mathcal{A}_i$ onto $\mathcal{B}$ for each $i = 1, 2, \ldots, k$
- Then for each i, the inverse functions from $\mathcal{B}$ to $\mathcal{A}$ can be found.

When the transformation is not one-to-one

- Denote the ith inverse by $x = h_{1i}(u, v)$ and $y = h_{2i}(u, v)$

When the transformation is not one-to-one

- Denote the ith inverse by $x = h_{1i}(u, v)$ and $y = h_{2i}(u, v)$
- This ith inverse gives, for $(u, v) \in \mathcal{B}$, the unique $(x, y) \in \mathcal{A}_i$ such that $(u, v) = (g_1(x, y), g_2(x, y))$.

When the transformation is not one-to-one

- Denote the ith inverse by $x = h_{1i}(u, v)$ and $y = h_{2i}(u, v)$
- This ith inverse gives, for $(u, v) \in \mathcal{B}$, the unique $(x, y) \in \mathcal{A}_i$ such that $(u, v) = (g_1(x, y), g_2(x, y))$.
- Let $\mathcal{J}_i$ denote the Jacobian computed from the i the inverse, then

$$f_{U,V}(u, v) = \sum_{i=1}^{k} f_{X,Y}(h_{1i}(u, v), h_{2i}(u, v))|\mathcal{J}_i| \tag{2}$$

Example 5 (Distribution of the ratio of normal variables)

Let $X \sim N(0,1)$ and $Y \sim N(0,1)$ be independent standard normal random variables. The joint PDF of (X, Y) is

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\{-x^2/2\} \exp\{-y^2/2\}$$

on the set $\mathcal{A} = \{(x,y) : -\infty < x < \infty, -\infty < y < \infty\}$.
Consider the transformation $U = X/Y$ and $V = |Y|$. Note that this transformation is not one-to-one since the points (x,y) and $(-x,-y)$ are both mapped into thesame (u,v) point. But if we restrict consideration to either positive or negative values of y, then the transformation is one-to-one.

1. Determine the joint PDF of U and V
2. What is the marginal PDF of U ?

Hierarchical Models and Mixture Distributions

- In the cases we have seen thus far, a random variable (RV) has a single distribution, possibly depending on parameters

- In the cases we have seen thus far, a random variable (RV) has a single distribution, possibly depending on parameters
- While in general, a RV can only have one distribution, it is often easier to model a situation by thinking of things in a hierarchy

- In the cases we have seen thus far, a random variable (RV) has a single distribution, possibly depending on parameters
- While in general, a RV can only have one distribution, it is often easier to model a situation by thinking of things in a hierarchy
- Perhaps the most classic hierarchical model is given in the following example

Example 6 (Binomial-Poisson hierarchy (revisited))

An chicken lays a large number of eggs, each surviving with probability p. On the average, how many eggs will survive? The large number of eggs laid is a random variable, often taken to be $Poisson(\lambda)$. Furthermore, if we assume that each egg's survival is independent, then we have Bernoulli trials. Therefore, if we let X=number of survivors and N=number of eggs laid, we have

$$X|N \sim Binomial(N, p) \quad \text{and} \quad N \sim Poisson(\lambda)$$

a hierarchical model.

- We showed in Example 25 (Week 7), that the marginal distribution of X is $Poisson(\lambda p)$ with Y playing no part at all

- Introducing Y into the hierarchy was mainly to aid our understanding of the model

Introduction

- Introducing Y into the hierarchy was mainly to aid our understanding of the model
- The answer to the original question is easy to compute $E(X) = \lambda p$. So on average λp eggs will survive

- Introducing Y into the hierarchy was mainly to aid our understanding of the model
- The answer to the original question is easy to compute $E(X) = \lambda p$. So on average λp eggs will survive
- If we were interested only in this mean and did not need the distribution, we could have used properties of **conditioal expectation**

- Introducing Y into the hierarchy was mainly to aid our understanding of the model
- The answer to the original question is easy to compute $E(X) = \lambda p$. So on average λp eggs will survive
- If we were interested only in this mean and did not need the distribution, we could have used properties of **conditioal expectation**
- Sometimes, **calculations can be greatly simplified** by using the following theorem. Recal that $E(X|y)$ is a function of y and $E(X|Y)$ is a random variable whose value depends on the value of Y

Theorem 7 (Conditional Expectation)

If X and Y are any two random variables, then

$$E(X) = E(E(X|Y)) \tag{3}$$

provided that the expectations exist

Proof: on board

Mixture Distribution

- The term *mixture distribution* in the title of this section refers to a distribution arising from a hierarchical structure.

- The term *mixture distribution* in the title of this section refers to a distribution arising from a hierarchical structure.

Mixture Distribution

- The term *mixture distribution* in the title of this section refers to a distribution arising from a hierarchical structure.

Definition 8 (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

Mixture Distribution

- The term *mixture distribution* in the title of this section refers to a distribution arising from a hierarchical structure.

Definition 8 (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

- Thus in Example 6, the $Poisson(\lambda p)$ distribution is a mixture distribution since it is the result of combining a $Binomial(N, p)$ with a $Poisson(\lambda)$.

- The term *mixture distribution* in the title of this section refers to a distribution arising from a hierarchical structure.

Definition 8 (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

- Thus in Example 6, the $Poisson(\lambda p)$ distribution is a mixture distribution since it is the result of combining a $Binomial(N, p)$ with a $Poisson(\lambda)$.
- In general we can say that hierarchical models lead to mixture distributions

Generalization

Example 9 (Binomial-Poisson-Exponential hierarchy)

Consider a generalization of Example 6 where instead of one mother chicken, there are a large number of mothers, and one mother is chosen at random. We are still interested in knowing the average number of survivors, but it is no longer clear that the number of eggs laid follows the same Poisson distribution for each mother. The following three stage hierarchy may be more appropriate. Let X = number of survivors. Then

$$\begin{aligned} X|N &\sim Binomial(N, p) \\ N|\Lambda &\sim Poisson(\Lambda) \\ \Lambda &\sim Exponential(\beta) \end{aligned}$$

where the last stage of hierarchy accounts for the variability across different mothers.

Theorem 10 (Conditional Variance)

For any two random variables X and Y

$$Var(X) = E\big(Var(E|Y)\big) + Var\big(E(X|Y)\big) \tag{4}$$

Proof: on board

Example 11 (Beta-binomial hierarchy)

One generalization of the binomial is to allow the success probability to vary according to a distribution. A standard model for this situation is

$$X|P \sim Binomial(n, P)$$
$$P \sim beta(\alpha, \beta)$$

code

```r
rnorm(10)
```

```
##  [1] -0.96345134 -0.90577660 -0.72820359 -0.67367372 -0.07154115
##  [6] -0.51256126  0.03757083  0.77479372 -1.28611637  1.17296755
```

Correlation Example 2

Example 12 (Correlation)

Let $X \sim Uniform(-1, 1)$ and $Z \sim Uniform(0, 1/10)$. Let X and Z be independent. Let $Y = X^2 + Z$ and consider the random vector (X, Y).

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

Find the correlation of X and Y