STA257: Probability and Statistics 1

Instructor: Katherine Daignault

Department of Statistical Sciences
University of Toronto

Week 10

Outline

Outline

Calculus Review - Partial Derivatives

- Because we are moving into the case where we are dealing with more than one random variable at a time, you will now need to be able to take partial derivatives.
- When dealing with a function of several variables, $f(x, y, \ldots)$, we can take the derivative with respect to each variable, as

$$\frac{\partial}{\partial x} f(x, y, \ldots) \text{ or } \frac{\partial}{\partial y} f(x, y, \ldots)$$

- We sometimes use a shorthand notation of f'_x or f'_y to denote the partial derivative with respect to x and y respectively.
- When taking partial derivatives, we just treat all other variables as constants, and all normal rules of differentiation apply.

Calculus Review - Multiple Integration

- Just as we need to take partial derivatives for multiple variables, we need to take multiple integrals over all the variables in our function.
- We define the double integral over $f(x, y)$ as

  $$\int_{supp(X)} \int_{supp(Y)} f(x, y) dy dx$$

  where we use the term $supp(X)$ to denote the support of X, or the values that X is defined over.
- In this case, you (1) take the integral with respect to Y, treating the x's as constants, then (2) take the integral of the result with respect to X.

Calculus Review - Multiple Integration

- Be careful about the order of the integrals!
  - If you want to integrate Y first, then the inner integral must correspond to the integral over y values
  - If you want to integrate X first, then you need to switch the order of the integrals to

$$\int_{supp(Y)} \int_{supp(X)} f(x, y) dx dy$$

- Since a single integral is the area under a curve, a double integral find the area **between** two curves.
- Often the support of one variable will be defined in terms of the other variable so you will need to be able to manipulate the bounds of integration.

Calculus Review - Multiple Integration

- If we are dealing with a function $f(x,y)$ defined on the region

$$F = \{(x,y) : a \leq x \leq b, p(x) \leq y \leq q(x)\}$$

  then we must take the integral in the following order

$$\int\int_F f(x,y)dA = \int_a^b \int_{p(x)}^{q(x)} f(x,y)dydx$$

- If I wanted to take the integral in the reverse order, then I would need to determine the functions $r(y)$ and $s(y)$ so that

$$\int\int_F f(x,y)dA = \int_c^d \int_{r(y)}^{s(y)} f(x,y)dxdy$$

  and $F = \{(x,y) : c \leq y \leq d, r(y) \leq x \leq s(y)\}$

Jacobian Matrices and Determinants

- We will be using Jacobian matrices this week.
- The Jacobian matrix represents all first order partial derivatives of a vector valued function, $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$.
- For example, in the 2-dimensional case, $\mathbf{g}(\mathbf{x}, \mathbf{y}) = (g_1(x, y), g_2(x, y))$, the Jacobian matrix is

$$\begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix}$$

- We will also need to take the determinant of the Jacobian, by

$$J(x, y) = \det \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix} = \left(\frac{\partial g_1}{\partial x} \times \frac{\partial g_2}{\partial y} \right) - \left(\frac{\partial g_2}{\partial x} \times \frac{\partial g_1}{\partial y} \right)$$

Outline

Lack of Independence

- So far, we have discussed how to determine whether two random variables X and Y are independent of each other.
- We saw that we could do this by checking either
  - if the joint PMF/PDF or joint CDF factors into their respective marginal distributions
  - if the conditional PMF/PDF of $X \mid Y$ (or $Y \mid X$) is equal to the marginal distribution of X (or Y)
- However, in many cases, we will have that X and Y are dependent variables.
- This lack of independence will play a roll in how we talk about the joint variation of the two variables.

Covariance of Two Random Variables

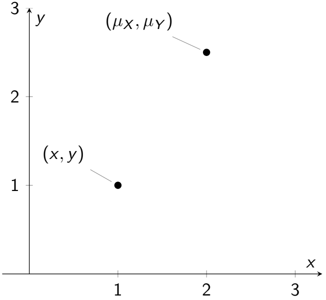
- Variance for a single random variable is a measure of the variability/spread of its distribution, i.e. variation in chances of taking on possible values.
- Because we have two variables that are jointly distributed, we need to measure how they vary jointly, as well as how strong is their association.
- First we will start with looking at how to express the variation in the joint distribution by using what we call the covariance.
- Recall first that $Var(X) = E[(X-\mu)^2]$, i.e. that variance measures how far values of X are from the centre of the distribution.

Variance in Univariate Case

- When we only have one variable and want to measure the variability from the mean, we compute the squared deviations $(X - \mu)^2$ and average them.
- The reason for the square is we are considering values of X that are both above and below the mean.
  - This results in negative distances for $x < \mu$ and positive distances for $x > \mu$.
  - Let's consider the standard Normal, with $\mu = 0$.
  - When we average over values that only differ in their sign (i.e. take the average of $(x - 0)$ and $(-x - 0)$), we get an average dispersion of 0.
  - This is nonsensical unless $x = 0$, and most values of x will not be equal to the mean.
  - In order for these distances to not "cancel out", they are squared.

Covariance

- When we transition to the multivariable case, we are now considering distances from the mean in two dimensions.
- So we can define a point in the space of (X, Y) values that corresponds to the mean of both variables (μ_X, μ_Y).
- For any point with an X coordinate x, we can determine how far away this from μ_x by calculating $x - \mu_x$
- We can similarly find the distance of any point with a Y coordinate y from μ_Y by $y - \mu_Y$



Covariance

- Putting these together, we can find the distance of any point (x, y) from (μ_X, μ_Y) by finding $(x - \mu_X)(y - \mu_Y)$
- Finally, as with variances, we want the average distance from the centre of the distribution, so we take the expected value.
- This gives the definition of the covariance of X and Y.

Definition of Covariance

If X and Y are jointly distributed random variables with expectations μ_X and μ_Y respectively, the covariance of X and Y is

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

provided the expectation exists.

Alternative Form of Covariance

- Like the variance in the univariate case, we have an alternative expression for the covariance that is simpler for computations.

Alternative Covariance Expression

The following is an equivalent expression for the covariance of jointly distributed X and Y:

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

Proof:

$\square$

Example: Gasoline example from week 8

Recall that X was the proportion of the tank with gasoline after delivery at the beginning of the week, and Y was the proportion of the tank that was sold by the end of the week. The joint PDF of X and Y was given by $f(x, y) = 3x$, if $0 \leq y \leq x \leq 1$. Find the covariance of X and Y.

- We will use the alternative expression for covariance because it will make the integration easier.
- So we need to find $E(XY)$, $E(X)$ and $E(Y)$.
- We previously found $E(X)$ in Week 9 as

$$E(X) = \int_0^1 \int_0^x x(3x)dydx = \int_0^1 3x^3 dx = \frac{3}{4}$$

and we found $E(Y)$ to be

$$E(Y) = \int_0^1 \int_0^x y(3x)dydx = \int_0^1 \frac{3x^3}{2} dx = \frac{3}{8}$$

Example: Gasoline (cont.)

- Now we are just left to find $E(XY)$ using the joint PDF.
- Recall from Week 9 that XY can be seen as a function $g(X, Y)$, and so we must use the result that

$$E(XY) = \int\int g(x,y)f_{XY}(x,y)dxdy$$

- So we can write out the integral as

$$E(XY) = \int_0^1 \int_0^x xy(3x)dydx = \int_0^1 \frac{3x^4}{2}dx = \left[\frac{3x^5}{10}\right]_0^1 = \frac{3}{10}$$

- Finally, we find the covariance to be

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{3}{10} - \frac{3}{4}\left(\frac{3}{8}\right) = 0.02$$

Exercise - Give it a try!

Recall we had 3 checkout counters, two customers that pick a counter at random. Let X denote the number of customers who choose counter 1, and Y the number of customers who choose counter 2. The joint and marginal PMFs are below. Find the covariance between X and Y.

	y			
x	0	1	2	$p_X(x)$
0	$1/9$	$2/9$	$1/9$	$4/9$
1	$2/9$	$2/9$	0	$4/9$
2	$1/9$	0	0	$1/9$
$p_Y(y)$	$4/9$	$4/9$	$1/9$	

Some Properties of Covariance

- Previously we saw that variances had the following property:

$$Var(a + bX) = b^2 Var(X)$$

- We can now develop analogous results for the covariance of linear combinations of X and Y.

Covariance Property 1

For jointly distributed random variables X and Y, and constant a,

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

Some Properties of Covariance

Covariance Property 2

For jointly distributed random variables X and Y, and constants a and b,

$$Cov(aX, bY) = abCov(X, Y)$$

Exercise - Give it a try!

Prove this property.

Some Properties of Covariance

Covariance Property 3

For jointly distributed random variables X, Y and Z,

$$Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$$

Some Properties of Covariance

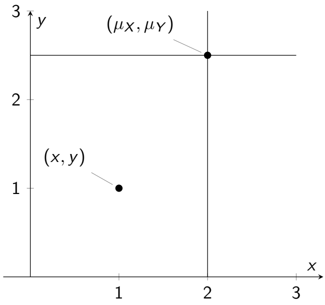
Covariance Property 4

For jointly distributed random variables X, Y, Z and W, and constants a, b, c and d,

$$
\begin{aligned}
Cov(aW + bX, cY + dZ) = &\, acCov(W, Y) + bcCov(X, Y) \\
&+ adCov(W, Z) + bdCov(X, Z)
\end{aligned}
$$

- This can be generalized to linear combinations of any number of random variables.
- From these, we get 3 important results about variance:
  1. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
  2. $Var\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j Cov(X_i, X_j)$
  3. $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$ if X_i are independent.
     - $\Rightarrow Cov(X_i, X_j) = 0$ if X_i and X_j are independent.

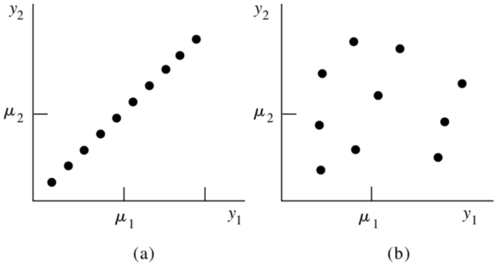
Values of Covariance



- The deviations $(x - \mu_X)$ and $(y - \mu_Y)$ will receive different signs depending on their position relative to (μ_X, μ_Y)
- If (x, y) is in the top right or bottom left quadrants (relative to (μ_X, μ_Y)), then the product of the deviations will be positive.
- If (x, y) is in the top left or bottom right quadrants (relative to (μ_X, μ_Y)), then the product of the deviations will be negative.
- So unlike variances, covariances can be negative or positive

Values of Covariance

- Covariances still give us some indication of how far (X, Y) are from (μ_X, μ_Y), i.e. distance.
- The fact that covariances can be negative or positive means that we have an idea of nature of the relationship between X and Y.
  - Positive association means that as the values of X increase, so do the values of Y
  - Negative association means that as the values of X increase, the values of Y decrease.



Strength of Association

- Specifically covariance can be used to measure how linearly dependent the variables are
- We could take the absolute value of the covariance to get an idea of the relative strength of the association between X and Y.
  - the larger the absolute value of the covariance, the stronger the linear dependence
  - the smaller the absolute value of the covariance, the weaker the linear dependence.
- However, the covariance on its own isn't enough to find an absolute measure for the strength of linear dependence
  - This occurs because each joint PDF will have a different spread of values (i.e. some will have large distances from the mean, some will have very small).

Correlation Coefficient

- If the spread of each distribution is preventing us from measuring the absolute strength of the dependence of X and Y, we can remove it be re-scaling the covariance.
  - Here, we re-scale by dividing by the variances of both X and Y (i.e. removing the units of spread)
  - By doing this, we get a measure that can be used for any two variables, regardless of the spread of each joint PMF/PDF

Correlation Coefficient

If X and Y are jointly distributed random variables and the variances and covariances of both X and Y exist, and the variances are non-zero, then the correlation of X and Y, denoted by ρ, is

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

Example: Gasoline

- Recall that the gasoline tank exercise had joint PDF

$$f(x, y) = 3x, \text{ if } 0 \leq y \leq x \leq 1$$

- We have already found that the covariance of proportion stocked and proportion sold is $Cov(X, Y) = 0.02$.
- To find the correlation between X and Y, we need to scale by the variances of both X and Y.
- We found in week 9 the expected values for X and Y: $E(X) = 3/4$ and $E(Y) = 3/8$
- So we just need to find $E(X^2)$ and $E(Y^2)$ to determine $Var(X)$ and $Var(Y)$.

Example: Gasoline (cont.)

- We can do this by simply integrating the joint PDF
  - $E(X^2) = \int_0^1 \int_0^x x^2(3x)dydx = \int_0^1 3x^4 dx = 3/5$
  - $E(Y^2) = \int_0^1 \int_0^x y^2(3x)dydx = \int_0^1 x^4 dx = 1/5$
- So we get the variances as
  - $Var(X) = E(X^2) - [E(X)]^2 = 3/5 - [3/4]^2 = 0.0375$
  - $Var(Y) = E(Y^2) - [E(Y)]^2 = 1/5 - [3/8]^2 = 0.059$
- Now we have all the information to find the correlation of X and Y:

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{0.02}{\sqrt{0.0375(0.059)}} = 0.425$$

Values of Correlation Coefficient

- So what did our answer in the previous example mean?
- Since $\rho > 0$, this means that there is a positive linear association (i.e. as X increases, so will Y)
  - But the covariance already told us this...
- The correlation tells us how strong this linear relationship is, and takes values as follows:

Values of ρ

$-1 \leq \rho \leq 1$. Further, $\rho = \pm 1$ if and only if $P(Y = a + bX) = 1$ for some constants a and b.

- so if $Y = a + bX$, we have a perfect linear relationship, and this results in $\rho = \pm 1$
- and $\rho = 0$ means no linear relationship.

Values of Correlation Coefficient

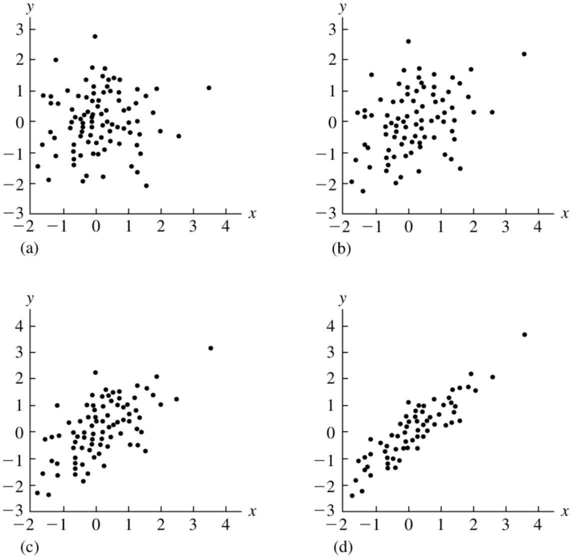


FIGURE 4.7 Scatterplots of 100 independent pairs of bivariate normal random variables, (a) $\rho = 0$, (b) $\rho = .3$, (c) $\rho = .6$, (d) $\rho = .9$.

Exercise - Give it a try!

Let Y_1 and Y_2 be discrete random variables with joint PMF below. Show that Y_1 and Y_2 are dependent, but have zero covariance.

y_2 \ y_1	-1	0	+1
-1	$^1/_{16}$	$^3/_{16}$	$^1/_{16}$
0	$^3/_{16}$	0	$^3/_{16}$
1	$^1/_{16}$	$^3/_{16}$	$^1/_{16}$

Outline

Transformations in the Multivariate Case

- Just like in the univariate case, we may be interested in finding the distribution for some function of joint random variables X and Y.
- Also like the univariate case, we have two methods of transforming random variables and determining their density functions:
  - Convolution method: similar to the brute force/direct method in the univariate case
  - Bivariate Transformation method: similar to the monotone transformation method.
- We will start by discussing the convolution method, which can be used when we are considering transformations that are sums or quotients of random variables.

Transformations Involving Sums

- Discrete case: suppose we have two discrete random variables X and Y and we want to find the probability mass function for $Z = X + Y$.
- We know that for any value $X = x$, we can express Y as a function of $X = x$ and the corresponding $Z = z$.
  - in particular, we have $X = x$ and $Y = z - x$, because $Z = z = x + y$.
- Therefore, we can write the joint PMF of X and Y as

$$p(x, y) = p(x, z - x)$$

- We know that we can find marginal distributions by summing over the values of the unwanted variable.
- So we now sum over x values to obtain the PMF of z:

$$p_Z(z) = \sum_x p(x, z - x)$$

Transformations Involving Sums

- In the case where X and Y are independent, we know that the joint PMF factors into the two marginals, so we will have

$$p_Z(z) = \sum_x p_X(x) p_Y(z - x)$$

- This sum is called the convolution of the sequences p_X and p_Y.
- The logic is quite similar to the direct method of transformation in the univariate case.
- Recall, if we want to make a transformation as $Y = X + c$, then we can use the CDF of X to get

$$P(Y \leq y) = P(X + c \leq y) = P(X \leq y - c) = F_X(y - c)$$

- The main difference is that x is not a constant here, so we need to consider the joint probability of x and $z - x$.

Transformations Involving Sums

- Continuous case: suppose we have two continuous random variables X and Y and we want the PDF of $Z = X + Y$.
- We can use the same procedure and write $X = x$ and $Y = z - x$ for some value of x.
- Now we can write the joint PDF of X and Y as

$$f(x, y) = f(x, z - x)$$

- Since we only want the PDF for Z, we can integrate over all values of X:

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x)dx$$

- If X and Y are independent, then the joint PDF will factor:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)dx$$

Transformations Involving Sums

- The previous result is derived using a similar argument to the direct method, where we start with the CDF.
- Since we have the relation $Z = X + Y$ and we want CDF for Z, we would be interested in integrating over the region where $X + Y \leq Z$.
- So we can write $F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) dydx$
- If we make a change of variables, $y = v - x$, then

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z} f(x, v-x) dvdx = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f(x, v-x) dxdv$$

- But we want the PDF of Z, so we can take derivative w.r.t. v:

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x) dx$$

Example: Lifetime of Component

Suppose that the lifetime of a component is exponentially distributed and an identical and independent backup is available. The system operates as long as one of the components functions. Thus the total lifetime of the system is a sum of exponentials. Let T_1 and T_2 be both $Exp(\lambda)$ and let $S = T_1 + T_2$. Find the density of S.

Transformations Involving Quotients

- Now let's consider a slightly different transformation, where we are finding the quotient of two continuous random variables.
- Suppose X and Y are continuous with joint PDF $f(x, y)$ and let $Z = {}^{Y}\!/_{X}$
- We can start by finding the CDF $F_Z(z) = P(Z \leq z)$ by noting that this gives the probability of the set (x, y) resulting in ${}^{y}\!/_{x} \leq z$.
  - We must consider two cases: $x > 0$ and $x < 0$.
  - When $x > 0$, then $y \leq xz$, whereas when $x < 0$, then $y \geq xz$
- My CDF can be written using composition as

$$F_Z(z) = \int_{-\infty}^{0} \int_{xz}^{\infty} f(x, y) dy dx + \int_{0}^{\infty} \int_{-\infty}^{xz} f(x, y) dy dx$$

Transformations Involving Quotients

- We can now make the change of variables $y = xv$ to remove x from the inner integral bound

$$F_Z(z) = \int_{-\infty}^{0} \int_{z}^{\infty} xf(x, xv)dvdx + \int_{0}^{\infty} \int_{-\infty}^{z} xf(x, xv)dvdx$$

- By flipping the inner integral of the first term to make it the negative version of the second term, we get

$$F_Z(z) = \int_{-\infty}^{0} \int_{-\infty}^{z} (-x)f(x, xv)dvdx + \int_{0}^{\infty} \int_{-\infty}^{z} xf(x, xv)dvdx$$

- Now we group the two terms by using $|x|$ for $-\infty < x < \infty$

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z} |x| \, f(x, xv)dvdx = \int_{-\infty}^{z} \int_{-\infty}^{\infty} |x| \, f(x, xv)dxdv$$

Transformations Involving Quotients

- Lastly, assuming continuity at z, we differentiate the last line to get the PDF

$$f_Z(z) = \frac{d}{dv} \int_{-\infty}^{z} \int_{-\infty}^{\infty} |x| \, f(x, xv) dx dv = \int_{-\infty}^{\infty} |x| \, f(x, xz) dx$$

- As before, if X and Y are independent, the joint PDF factors to

$$f_Z(z) = \int_{-\infty}^{\infty} |x| \, f_X(x) f_Y(xz) dx$$

- So just like the univariate transformations, we start from the CDF of X and Y, replace one variable with a function of Z, manipulate, and finally differentiate to get PDF

Example: Ratio of Standard Normals

Suppose X and Y are both independent $N(0,1)$ and we want the PDF of $Z = {}^{Y}\!/\!_{X}$.

- Recall that PDF of a standard Normal is $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
- Express y as a function of x and z as $y = xz$
- Use convolution formula for quotients:

$$f_Z(z) = \int_{-\infty}^{\infty} |x| \, f_X(x) f_Y(xz) dx = \int_{-\infty}^{\infty} |x| \, \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-(xz)^2/2}$$

- Simplify and use fact that $N(0,1)$ is symmetric at 0:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{|x|}{2\pi} e^{-x^2(z^2+1)/2} dx = 2 \int_{0}^{\infty} \frac{x}{2\pi} e^{-x^2(z^2+1)/2} dx$$

Example: Ratio of Standard Normals (cont.)

- Now we need to employ a change of variables $u = x^2$:

$$f_Z(z) = 2\int_0^\infty \frac{x}{2\pi} e^{-x^2(z^2+1)/2} dx = \frac{1}{2\pi}\int_0^\infty e^{-u(z^2+1)/2} du$$

- This integral can be solved in two ways:
  1. Brute force: the $(z^2+1)/2$ term is just a constant.
  2. Recognize that if $\lambda = (z^2+1)/2$, we could use that

$$\int_0^\infty \lambda e^{-\lambda u} du = 1$$

  because it's the PDF of an Exponential

- Choosing option 2, we have

$$f_Z(z) = \frac{1}{2\pi}\frac{1}{\lambda}\int_0^\infty \lambda e^{-\lambda u} du = \frac{1}{\pi(z^2+1)}, \quad -\infty < z < \infty$$

- This is the PDF of a Cauchy random variable.

Outline

Bivariate Transformation Method

- In the univariate case, the monotone transformation method was a more concise way to find the density of a transformed variable, compared to the direct method.
- We have an analogous result for the multivariate case.
- For start, we will consider the bivariate situation, with jointly distributed continuous X and Y variables
- Recall that the density for variable $U = h(Y)$ can be found by

$$f_U(u) = f_Y(h^{-1}(u)) \left| \frac{dh^{-1}(u)}{du} \right|$$

- We can extend this idea to deal with finding densities from functions of two variables

Bivariate Transformation Method

- What kinds of transformations can this be used with?
  - Can be used to find the joint PDF of $U = g_1(X, Y)$ and $V = g_2(X, Y)$, i.e. $f(u, v)$ by using $f(x, y)$
  - Can also be used to find marginal PDFs of $U = g(X, Y)$ by $f(u) = \int f(u, v)dv$
- Both follow the same principle as the monotone transformation method:
  - Determine functions g_1 and g_2 and find their inverse functions g_1^{-1} and g_2^{-1}
  - Find the derivatives of both functions with respect to both X and Y
  - Combine with the original joint PDF of X and Y

Bivariate Transformation Method

Bivariate Transformation Method

Suppose X and Y are continuous random variables with joint PDF $f_{XY}(x, y)$. Let U and V be transformations of X and Y by the mapping

$$u = g_1(x, y), \quad v = g_2(x, y)$$

and assume that the transformations can be inverted to obtain

$$x = h_1(u, v), \quad y = h_2(u, v).$$

Also assume that g_1 and g_2 have continuous partial derivatives and the Jacobian

$$J(x, y) = \det \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix} = \frac{\partial g_1}{\partial x} \times \frac{\partial g_2}{\partial y} - \frac{\partial g_2}{\partial x} \times \frac{\partial g_1}{\partial y} \neq 0$$

for all x and y.

Bivariate Transformation Method (cont.)

Under these assumptions, the joint PDF of U and V is

$$f_{UV}(u,v) = f_{XY}\left(h_1(u,v), h_2(u,v)\right)\left|J^{-1}(h_1(u,v), h_2(u,v))\right|$$

for u and v such that $u = g_1(x,y)$ and $v = g_2(x,y)$ for some (x,y) and 0 elsewhere.

- The Jacobian represents the partial derivatives with respect to x and y of g_1 and g_2, then evaluated at $x = h_1(u,v)$ and $y = h_2(u,v)$.
- This can of course be extended to the case of more than two variables, but for our purposes, we will only work with the bivariate situation.

Example: Independent Standard Normals

Let X and Y be independent $N(0,1)$ random variables. We define the variables $U = X + Y$ and $V = X - Y$. What is the joint PDF of U and V?

- We know the PDF for both X and Y:

$$f_Y(y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}} \quad \text{and} \quad f_X(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

- Independence of X and Y implies $f(x,y) = \frac{e^{-y^2/2 - x^2/2}}{2\pi}$
- The transformation we are interested in taking is

$$u = x + y = g_1(x,y) \quad \text{and} \quad v = x - y = g_2(x,y)$$

- The inverse of these transformations is

$$x = \frac{u+v}{2} = h_1(u,v) \quad \text{and} \quad y = \frac{u-v}{2} = h_2(u,v)$$

Example: Independent Standard Normals (cont.)

- Now we need to find all the partial derivatives:
  - $\frac{\partial g_1(x,y)}{\partial x} = \frac{\partial}{\partial x}(x+y) = 1, \quad \frac{\partial g_1(x,y)}{\partial y} = \frac{\partial}{\partial y}(x+y) = 1$
  - $\frac{\partial g_2(x,y)}{\partial x} = \frac{\partial}{\partial x}(x-y) = 1, \quad \frac{\partial g_2(x,y)}{\partial y} = \frac{\partial}{\partial y}(x-y) = -1$
- So my Jacobian will be $J = \det \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = -1 - 1 = -2$
- Finally plug relevant terms into the formula:

  $$f_{UV}(u,v) = f_{XY}\left(\tfrac{u+v}{2}, \tfrac{u-v}{2}\right)\left|\frac{1}{-2}\right| = \frac{e^{-\frac{1}{2}\left(\frac{u+v}{2}\right)^2 - \frac{1}{2}\left(\frac{u-v}{2}\right)^2}}{2\pi}\left(\frac{1}{2}\right)$$

- A little algebraic manipulation gives

  $$f_{UV}(u,v) = \frac{e^{-u^2/4}}{\sqrt{2}\sqrt{2\pi}}\frac{e^{-v^2/4}}{\sqrt{2}\sqrt{2\pi}}, -\infty < u, v < \infty$$

  which is the product of two independent $N(0,2)$

Example: Using Bivariate Transformation to find Marginal

Find the PDF for $V = XY$ when random variables X and Y have joint PDF

$$f_{XY}(x,y) = \begin{cases} 2(1-x), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

- To use the bivariate transformation method, we need to create a transformation $U = g_1(x,y)$ so that we can get a one-to-one function of (x,y) to (u,v).
- The simplest option is to choose an identity function, as

$$u = x = g_1(x,y) \quad \text{and} \quad v = xy = g_2(x,y)$$

- Then the inverse functions are

$$x = u = h_1(u,v) \quad \text{and} \quad y = {}^{v}\!/_{u} = h_2(u,v)$$

Example: Marginal (cont.)

- Now we need to find all the partial derivatives:
  - $\frac{\partial g_1(x,y)}{\partial x} = \frac{\partial}{\partial x}(x) = 1$
  - $\frac{\partial g_1(x,y)}{\partial y} = \frac{\partial}{\partial y}(x) = 0$
  - $\frac{\partial g_2(x,y)}{\partial x} = \frac{\partial}{\partial x}(xy) = y = {}^{v}/_{u}$
  - $\frac{\partial g_2(x,y)}{\partial y} = \frac{\partial}{\partial y}(xy) = x = u$
- So my Jacobian will be

  $J(h_1(u,v), h_2(u,v)) = \det \begin{bmatrix} 1 & 0 \\ {}^{v}/_{u} & u \end{bmatrix} = u - 0 = u$
- Plug relevant terms into the formula to get joint PDF of U and V:

$$f_{UV}(u,v) = f_{XY}\left(u, {}^{v}/_{u}\right)\left|\frac{1}{u}\right| = \frac{2(1-u)}{u}$$

Example: Marginal (cont.)

- Of course, we still need to specify the bounds for u and v.
  - The bounds for u are easy: $0 \le x \le 1 \Rightarrow 0 \le u \le 1$
  - We find the bounds for v as:
    $0 \le y \le 1 \Rightarrow 0 \le {}^{v}/_{u} \le 1 \Rightarrow 0 \le v \le u$
  - Combined we get $0 \le v \le u \le 1$
- The question is asking though for the marginal PDF of V.
- To find this, just integrate out U from the joint PDF:

$$f_V(v) = \int_v^1 \frac{2(1-u)}{u} du = \int_v^1 \left(\frac{2}{u} - 2\right) du = 2(-\ln(v)+v-1)$$

defined on $0 \le v \le 1$.

Exercise - Give it a try!

Suppose that X and Y are independent standard Normal random variables. Find the joint PDF of $U = X$ and $V = X + Y$.

Outline

Order Statistics

- Many functions of random variables of interest in practice depend on the relative magnitudes of the observed values.
  - For example, we may jointly consider the lap times of each race car on a track.
  - However our interest may only be in the fastest lap time, instead of the joint lap times of all cars.
- To do this, we need to order the random variables according to their magnitudes.
- The ordered variables are called order statistics.
- However because we don't have observed values for each random variable, we must manipulate the joint distribution to find each order statistic.

The Maximum of $X_1, \ldots, X_n$

- Suppose like in the race car example, we are interested in finding the density function of the maximum lap time (i.e. slowest car).
  - We shall call this U
- So we have n cars, each with X_i denoting the i^{th} car's lap time, with common $F_X(x)$
- To find the PDF representing the distribution of lap times for the slowest car, we need to work with the joint CDF of all cars' lap times, $F(x_1, \ldots, x_n)$
- Let's start with finding the CDF of U, which is the probability that the slowest car has a lap time smaller than u:

$$F_U(u) = P(U \leq u)$$

The Maximum of $X_1, \ldots, X_n$

- The important piece to realize is, if U is the maximum lap time, then that means that the lap times for all other cars must be smaller.
- We can now rewrite the CDF of U using the joint CDF of the X's:
$$F_U(u) = P(U \leq u) = P(X_1 \leq u, X_2 \leq u, \ldots, X_n \leq u)$$
- When each car's lap times are independent, we can factor this to
$$F_U(u) = P(U \leq u) = \prod_{i=1}^{n} P(X_i \leq u)$$
- Because each lap time has common CDF $F_X(x)$, we can conclude
$$F_U(u) = \prod_{i=1}^{n} P(X_i \leq u) = \prod_{i=1}^{n} P(X \leq u) = \left[F_X(u)\right]^n$$

The Maximum of $X_1, \ldots, X_n$

- We can now find the PDF of the maximum of the n random variables by differentiating:

$$f_U(u) = \frac{d}{dx}\left[F_X(x)\right]^n = nf_X(u)\left[F_X(u)\right]^{n-1}$$

- Generally, we use the notation $X_{(n)}$ to refer to the maximum of $X_1, \ldots, X_n$
- The sequence of ordered random variables can be written as $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, where
  - $X_{(1)}$ is the minimum of $X_1, \ldots, X_n$,
  - $X_{(2)}$ is the second smallest of $X_1, \ldots, X_n$, etc.
  - This implies that $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$
- We can follow the same logic as deriving the PDF of $X_{(n)}$ to derive the PDF of any $X_{(i)}$

Exercise - Give it a try!

Find the PDF of $V = X_{(1)}$, the minimum of $X_1, \ldots, X_n$, when the X_i are independent with common PDF and CDF.

Example: Lifetime of Components in Series

Suppose the n components are connected in series, so that the whole system fails if any one of them fails. The lifetimes of each component $T_1, \ldots, T_n$ are independent exponential random variables with parameter λ. What is the PDF for the lifetime of the system, which is the minimum lifetime of the n components?

- Here we just need to use our previous result to find the PDF of $X_{(1)}$
- Since each T_i is $Exp(\lambda)$, we know that
  - $f_T(t) = \lambda e^{-\lambda t}$ is the PDF
  - $F_T(t) = 1 - e^{-\lambda t}$ is the CDF
  - So using the previous result:

$$f_{(X_{(1)})}(v) = n\left(\lambda e^{-\lambda v}\right)\left[1 - (1 - e^{-\lambda v})\right]^{n-1} = n\lambda e^{-\lambda vn}$$

A Differential Argument

- Another way to think about order statistics and their respective distribution functions is with the use of a differential argument.
- Suppose we consider $U = X_{(n)}$, and want to find $f_U(u)$.
  - Note that $u \leq U \leq u + du$ if one of the X_i falls in the interval $(u, u + du)$, where du is a small increment above u.
    - This happens with probability $f_X(u)$
  - If exactly one X_i is in this interval, then necessarily the remaining $(n-1)$ X_i's must fall below/to the left of u
    - This happens with probability $\left[F_X(u)\right]^{(n-1)}$
  - So there is probability $f_X(u)\left[F_X(u)\right]^{(n-1)}$ of arranging the X_i such that one is in $(u, u + du)$ and $n-1$ are to the left of u.
  - Finally, using counting arguments, there are n such arrangements of X_i's possible, which yields $nf_X(u)\left[F_X(u)\right]^{n-1}$

A General Form of PDF of Order Statistics

- We may be interested in finding the PDF of any $X_{(k)}$.
  - e.g. We may want to find the distribution of the median.
  - If n is odd, say $n = 2m + 1$, then $X_{(m+1)}$ is the median of the X_i.
- The same heuristic argument can be used to understand the following result.

PDF of the kth Order Statistic

The density of $X_{(k)}$, the kth order statistic, is

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} f_X(x) \left[F_X(x)\right]^{k-1} \left[1 - F_X(x)\right]^{n-k}$$

Example: Working with Uniform(0,1)

Suppose each of n X_i's are independent Uniform(0, 1) random variables. Find the density for the kth order statistic, $X_{(k)}$.

- Recall that the PDF and CDF of Uniform(0,1) are

$$f_X(x) = 1, 0 \leq x \leq 1 \quad \text{and} \quad F_X(x) = x, 0 \leq x \leq 1$$

- Using the previous result and these distribution functions, we get

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}$$

- This can be recognized as a $Beta(k, n-k+1)$.

Joint PDF of $X_{(1)}$ and $X_{(n)}$

- Using the same heuristic argument as before, we can find the joint distribution of the minimum and maximum of $X_1, \ldots, X_n$.
- We note that $x \leq X_{(1)} \leq x + dx$ and $y \leq X_{(n)} \leq y + dy$ if we have
  - one $X_i \in (x, x + dx)$ with probability $f_X(v)$
  - one $X_i \in (y, y + dy)$ with probability $f_X(u)$
  - and the remaining $n - 2$ X_i's in $[x, y]$ with probability $F_X(u) - F_X(v)$
- Finally there are $n(n-1)$ ways in which to pick the minimum and maximum X_i's
- So the joint PDF will be

$$f_{X_{(1)}X_{(n)}}(u, v) = n(n-1)f_X(u)f_X(v)\left[F_X(u) - F_X(v)\right]^{n-2}, \quad u \geq v$$

Example: Order Statistics of Uniform(0, 1)

What is the joint PDF of the minimum and maximum of $X_1, \dots, X_n$, when each X_i are independent Uniform(0, 1) random variables?

- Again we have the PDF and CDF of Uniform(0,1) are

$$f_X(x) = 1, 0 \leq x \leq 1 \quad \text{and} \quad F_X(x) = x, 0 \leq x \leq 1$$

- Using the result from previous slide,

$$f_{X_{(1)} X_{(n)}}(u, v) = n(n-1)[u - v]^{n-2}, \quad 1 \geq u \geq v \geq 0$$

- Because this is still just a joint PDF, we can still use it as we would any other joint PDF (same goes for any $f_{x_{(k)}}(x)$).
  - e.g. find a marginal PDF from the joint, find the distribution of a bivariate transformation of order statistics, find means and covariances, etc.

Example: Order Statistics of Uniform(0, 1) (cont.)

Let $R = X_{(n)} - X_{(1)}$ be the range of $X_{(1)}, \dots, X_{(n)}$. Find the PDF of R.

- Just realize this is a sum transformation, so use convolution method.
- We write $u = r + v$ and substitute into the joint PDF:

$$f_{UV}(u, v) = f_{UV}(r + v, v) = n(n-1)(r)^{n-2}, \quad 0 \le v \le 1 - r$$

- To get the marginal distribution of R, just integrate out v:

$$f_R(r) = \int_0^{1-r} n(n-1)r^{n-2}dv = n(n-1)r^{n-2}(1-r), \quad 0 \le r \le 1$$