STAT:5100 (22S:193) Statistical Inference I
Week 11

Luke Tierney

University of Iowa

Fall 2015
Recap

- Change of variables examples
- General invertible linear transformations
- Marginal distributions of functions of two variables via change of variables
Example (continued)

- By symmetry the mean is zero (if it exists; it does for \( p > 1 \)).
- The variance is

\[
E[U^2] = E \left[ \left( \frac{X}{\sqrt{Y/p}} \right)^2 \right] = E \left[ p \frac{X^2}{Y} \right]
\]

\[
= pE[X^2] E \left[ \frac{1}{Y} \right] = pE \left[ \frac{1}{Y} \right]
\]

\[
= p \int_0^\infty \frac{1}{\Gamma(p/2)2^{p/2}} y^{p-1-1} e^{-y/2} dy
\]

\[
\begin{cases} 
\frac{\Gamma(p/2-1)}{\Gamma(p/2)^2} & p > 2 \\
\infty & p \leq 2
\end{cases} = \begin{cases} 
\frac{p}{p-2} & p > 2 \\
\infty & p \leq 2
\end{cases}
\]
Example

- Let $X \sim \text{Gamma}(\alpha, 1)$ and $Y \sim \text{Gamma}(\beta, 1)$ be independent.
- Find the marginal density of $X/Y$.
- Set $U = X/Y$ and $V = Y$.
- Then $X = UV$, $Y = V$, and

$$J(u, v) = \det \begin{pmatrix} v & u \\ 0 & 1 \end{pmatrix} = v.$$

- Therefore for $u > 0, v > 0$

$$f_{UV}(u, v) = f_{XY}(uv, v)v = \frac{(uv)^{\alpha-1}e^{-uv}v^{\beta-1}e^{-v}v}{\Gamma(\alpha)\Gamma(\beta)} = \frac{u^{\alpha-1}v^{\alpha+\beta-1}e^{-(1+u)v}}{\Gamma(\alpha)\Gamma(\beta)}. $$

- So for $u > 0$

$$f_U(u) = \int_0^\infty u^{\alpha-1}v^{\alpha+\beta-1}e^{-(1+u)v} \frac{dv}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{u^{\alpha-1}}{(1 + u)^{\alpha+\beta}}.$$

- This is related to the $F$ distribution, and to the Beta distribution.
Example

- The variables $U$ and $V$ are uniformly distributed over the set
  \[ A = \{(u, v) : 0 \leq u \leq 1, v^2 \leq -4u^2 \log u\} \]
- Find the density of $X = V/U$.
- The set $A$ is bounded.

- Let $|A|$ denote the area of $A$. 

![Diagram of the region $A$]

\[ u \quad 0.0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8 \quad 1.0 \]
\[ v \quad -0.5 \quad 0.0 \quad 0.5 \]
Example (continued)

• The joint density of $U, V$ is

\[
f_{U,V}(u, v) = \begin{cases} 
\frac{1}{|A|} & \text{if } 0 \leq u \leq 1, v^2 \leq -4u^2 \log u \\
0 & \text{otherwise.}
\end{cases}
\]

• Let $Y = U$.

• Then the set of possible $X, Y$ values is

\[
B = \{(x, y) : 0 \leq y \leq 1, x^2 \leq -4 \log y\}.
\]

• The range of possible $X$ values is $(-\infty, \infty)$.

• For a fixed value of $X = x$, the possible values of $Y$ are $[0, e^{-\frac{1}{4}x^2}]$. 
Example (continued)

- The inverse transformation is

\[ U = Y \]
\[ V = XY \]

- The Jacobian determinant of this inverse transformation is

\[ \det \begin{pmatrix} 0 & 1 \\ y & x \end{pmatrix} = -y \]

- The joint density of \( X, Y \) is thus

\[ f_{X,Y}(x, y) = \begin{cases} \frac{1}{|A|} y & 0 < y < e^{-\frac{1}{4}x^2} \\ 0 & \text{otherwise.} \end{cases} \]
Example (continued)

• The marginal density of $X$ is

$$f_X(x) = \frac{1}{|A|} \int_0^y e^{-\frac{1}{4}x^2} \, dy = \frac{1}{2|A|} \left[ e^{-\frac{1}{4}x^2} \right]_0^y = \frac{1}{2|A|} e^{-\frac{1}{2}x^2}$$

• Thus $X$ has a standard normal distribution.

• The area of the set $A$ is $|A| = \frac{1}{2} \sqrt{2\pi} = \sqrt{\pi}/2$.

• This is the basis of the *ratio of uniforms* method for generating normal variates on a computer.

• Points can be uniformly sampled from $A$ using *rejection sampling* from an enclosing rectangle.
Example

- Let $X_1, \ldots, X_n$ be jointly continuous and let $a_1, \ldots, a_n$ be constants.
- Define the column vectors
  \[
  X = \begin{bmatrix}
  X_1 \\
  \vdots \\
  X_n
  \end{bmatrix}
  \quad \quad \quad \quad \quad a = \begin{bmatrix}
  a_1 \\
  \vdots \\
  a_n
  \end{bmatrix}
  \]

- Let $B$ be an $n \times n$ invertible constant matrix.
- Define
  \[
  U = \begin{bmatrix}
  U_1 \\
  \vdots \\
  U_n
  \end{bmatrix}
  \]

as $U = a + BX$. 
Example (continued)

• Then the inverse transformation is

\[ X = B^{-1}(U - a). \]

• The Jacobian determinant of the inverse transformation is

\[ J = \det B^{-1} = \frac{1}{\det B}, \]

• Therefore

\[ f_U(u) = \frac{1}{|\det B|} f_X(B^{-1}(u - a)). \]
**Example (continued)**

- Special case: Let $X_1, \ldots, X_n$ be independent standard normals.
- Then
  
  $$f_X(x) = \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} x^T x \right\}$$

- The joint density of $U$ is
  
  $$f_U(u) = \frac{1}{(2\pi)^{n/2} |\det B|} \exp \left\{ -\frac{1}{2} (u - a)^T (BB^T)^{-1} (u - a) \right\}$$

- This is the general multivariate normal density.
Example (continued)

• It is easy to see that $E[U_i] = a_i$, or

$$E[U] = a$$

• The covariances are

$$\text{Cov}(U_i, U_j) = \text{Cov} \left( \sum_k b_{ik} X_k, \sum_l b_{jl} X_l \right)$$

$$= \sum_k \sum_l b_{ik} b_{jl} \text{Cov}(X_k, X_l)$$

$$= \sum_k b_{ik} b_{jk}$$

$$= (BB^T)_{ij}$$
Example (continued)

- Let $C$ be the matrix of covariances, i.e.

$$C_{ij} = \text{Cov}(U_i, U_j).$$

- Then $C = BB^T$

- The determinant of $C$ is

$$\det C = (\det B)^2.$$

- Therefore

$$f_U(u) = \frac{1}{(2\pi)^{n/2}(\det C)^{1/2}} \exp \left\{ -\frac{1}{2}(u - a)^T C^{-1}(u - a) \right\}$$

- Given $C$, a suitable $B$ can be found by Cholesky factorization.
Multinomial Distribution

- Suppose every item can be classified into one of $n$ categories.
- The proportion of items in category $i$ is $p_i$, with $p_1 + \cdots + p_n = 1$.
- Suppose $m$ items are selected with replacement.
- For $i = 1, \ldots, n$ let
  \[ X_i = \text{number of selected items that belong to category } i. \]
- What is the joint PMF of $X_1, \ldots, X_n$?
• For $m = 10$ and $n = 3$ a particular draw might look like

\[ 2 \ 2 \ 3 \ 3 \ 1 \ 1 \ 1 \ 3 \ 2 \ 3 \]

• The probability of this particular draw, or any draw with 3 items from category 1, 3 items from category 2, and 4 items from category 3, is

\[ p_1^3 p_2^3 p_3^4. \]

• The number of different draws that consist of 3 items from category 1, 3 items from category 2, and 4 items from category 3, is

\[ \binom{10}{3} \binom{7}{3} \binom{4}{4} = \frac{10!}{3!7!} \times \frac{7!}{3!4!} \times 1 = \frac{10!}{3!3!4!} = \binom{10}{3 \ 3 \ 4}. \]

• The corresponding joint PMF value is therefore

\[ f(3, 3, 4) = P(X_1 = 3, X_2 = 3, X_3 = 4) = \frac{10!}{3!3!4!} p_1^3 p_2^3 p_3^4. \]
• The general form for \( m \) draws and \( n \) categories is

\[
f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n) = \frac{m!}{x_1! \ldots x_n!} p_1^{x_1} \ldots p_n^{x_n} = \binom{m}{x_1 \ldots x_n} p_1^{x_1} \ldots p_n^{x_n}.
\]

• This is called a *multinomial distribution*, because of the relation to

**Theorem (Multinomial Theorem)**

For any real numbers \( y_1, \ldots, y_n \)

\[(y_1 + \cdots + y_n)^m = \sum_{x_1 + \cdots + x_n = m} \binom{m}{x_1 \ldots x_n} y_1^{x_1} \cdots y_n^{x_n}\]

*with the sum ranging over nonnegative integers with \( x_1 + \cdots + x_n = m \).*
• Notation: \( X \sim \text{Multinomial}(m, p_1, \ldots, p_n) \).
• The marginal distribution of \( X_i \) is \( \text{Binomial}(m, p_i) \).
• For \( i \neq j \) the distribution of \((X_i, X_j, m - X_i - X_j)\) is
  \[
  (X_i, X_j, m - X_i - X_j) \sim \text{Multinomial}(m, p_i, p_j, 1 - p_i - p_j).
  \]
• Similar results hold for higher order marginals
• Conditional distributions are also Binomial/Multinomial.
Example

- Suppose $Y_1, Y_2, \ldots$ are independent Bernoulli random variables with success probability $p$ and let $X$ be the number of trials until the first success.

- One way to compute $E[X]$ is to derive an equation by conditioning on the result of the first trial.

- The expectation can be written as

$$E[X] = E[E[X|Y_1]] = pE[X|Y_1 = 1] + (1 - p)E[X|Y_1 = 0]$$

- Clearly $E[X|Y_1 = 1] = 1$.

- If $Y_1 = 0$ then the number of additional trials needed until a success is geometric and independent of $Y_1$.

- So

$$E[X|Y_1 = 0] = 1 + E[X].$$
Example (continued)

• This produces the equation

\[ E[X] = p + (1 - p)(1 + E[X]) \]

or

\[ E[X] = 1 + (1 - p)E[X] \]

• The unique \textit{finite} solution is \( E[X] = 1/p \).

• A separate argument is needed to show that

\[ E[X] = \sum_{k=1}^{\infty} kp(1 - p)^{k-1} < \infty \]

for \( p > 0 \).
Recap

- Marginal distributions of functions of two variables via change of variables
- General invertible linear transformations
- General multivariate normal as linear transformation of independent standard normals
- Multinomial distribution
Lemma

For positive numbers $a$, $b$, $p$, $q$ such that

$$\frac{1}{p} + \frac{1}{q} = 1$$

we have

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab.$$
\textit{Proof.}

- Fix $b$, $p$, and $q$, and let $g(a) = \frac{1}{p} a^p + \frac{1}{q} b^q - ab$.
- Then 
  \[
  \frac{d}{da} g(a) = a^{p-1} - b.
  \]
- This increases from $-b$ at $a = 0$ to $\infty$ at $a = \infty$.
- This means there is a unique root that is a minimum at 
  \[
  a = b^\frac{1}{p-1}
  \]
- Since $(p - 1)q = p$,
  \[
  g\left(b^\frac{1}{p-1}\right) = \frac{1}{p} b^{\frac{p}{p-1}} + \frac{1}{q} b^q - b^\frac{1}{p-1} + 1
  = \frac{1}{p} b^q + \frac{1}{q} b^q - b^q = 0.
  \]
**Theorem (Hölder’s Inequality)**

For $p, q > 0$ with $1/p + 1/q = 1$ and any random variables $X, Y$ with $E[|X|^p] < \infty$ and $E[|Y|^q] < \infty$,

\[ |E[XY]| \leq E[|XY|] \leq E[|X|^p]^{1/p} E[|Y|^q]^{1/q} \]
Proof.

- The first inequality in the theorem follows because
  \[-|XY| \leq XY \leq |XY|.

- If \(X\) or \(Y\) are identically zero the result holds trivially.

- So assume neither is identically zero.

- Define
  \[ a = \frac{|X|}{E[|X|^p]^{1/p}} \quad b = \frac{|Y|}{E[|Y|^q]^{1/q}}. \]

- Then from the Lemma,
  \[ \frac{|XY|}{E[|X|^p]^{1/p}E[|Y|^q]^{1/q}} \leq \frac{1}{p} \frac{|X|^p}{E[|X|^p]} + \frac{1}{q} \frac{|Y|^q}{E[|Y|^q]}. \]

- Therefore
  \[ \frac{E[|XY|]}{E[|X|^p]^{1/p}E[|Y|^q]^{1/q}} \leq \frac{1}{p} + \frac{1}{q} = 1. \]
• Equality holds if and only if one random variable is essentially a scalar multiple of the other.
• Special case: Cauchy-Schwartz inequality

\[ |E[XY]| \leq E[|XY|] \leq E[X^2]^{1/2} E[Y^2]^{1/2} \]

This is the result that implies that $|\rho| \leq 1$.
• Another corollary of Hölder’s inequality: If $|Y| \equiv 1$ then

\[ E[|X|] \leq E[|X|^p]^{1/p} \]

or

\[ E[|X|^p] \leq E[|X|^p] \]

for any $p \geq 1$. 

• Replace $|X|$ by $|X|^r$ for $r > 0$ to get

\[ E[|X|^r] \leq E[|X|^{rp}]^{1/p} \]

• Take $s = rp$, so $0 < r \leq s < \infty$. Then

\[ E[|X|^r]^{1/r} \leq E[|X|^s]^{1/s} \]

• This is Liapunov's inequality.
Theorem (Minkowski’s Inequality, Triangle Inequality)

Let $X, Y$ be two random variables and $p \geq 1$. If $E[|X|^p] < \infty$ and $E[|Y|^p] < \infty$, then

$$E[|X + Y|^p]^{1/p} \leq E[|X|^p]^{1/p} + E[|Y|^p]^{1/p}$$

- This inequality means that
  $$\|X\|_p = E[|X|^p]^{1/p}$$
  defines a length, or a *norm*, for random variables.
- This is called the $L_p$ norm.
- Norms are very useful in studying convergence of sequences of random variables.
Proof.

- For $p = 1$ this follows from $|a + b| \leq |a| + |b|$.
- For $p > 1$, by the triangle inequality

$$E[|X + Y|^p] = E[|X + Y||X + Y|^{p-1}]$$

$$\leq E[|X||X + Y|^{p-1}] + E[|Y||X + Y|^{p-1}]$$

- By Hölder’s inequality,

$$E[|X||X + Y|^{p-1}] \leq E[|X|^p]^{1/p} E[|X + Y|^{q(p-1)}]^{1/q}$$

$$= E[|X|^p]^{1/p} E[|X + Y|^p]^{1-1/p}$$

and similarly for the second term.
- Now divide through by $E[|X + Y|^{p-1}]^{1-1/p}$. 

\[\square\]
• The Hölder and Minkowski inequalities are *homogeneous*.
• Hence they also apply to integrals and sums.
A function $g(x)$ is convex if

$$g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y)$$

for all $x, y$ and all $\lambda \in [0, 1]$.

$g$ is strictly convex if the inequality is strict when $x \neq y$ and $0 < \lambda < 1$.

A function $g$ is concave if $-g$ is convex.
• If $g$ is convex then for any point $y$ in its domain there exists a linear function $\ell_y(x)$ such that

$$\ell_y(y) = g(y)$$

and for all $x$ in the domain

$$\ell_y(x) \leq g(x).$$

• If $g$ is strictly convex then $\ell_y(x) < g(x)$ for all $x \neq y$.

• This is a property of convex sets called the supporting hyperplane theorem.
Theorem (Jensen’s Inequality)

For any random variables $X$ if $g(x)$ is convex and $E[|X|]$ is finite, then

$$E[g(X)] \geq g(E[X])$$

If $g$ is strictly convex and $X$ is not constant with probability one, then the inequality is strict.
Proof.

- The proof follows from the supporting hyperplane property of convex functions.

- There exists a linear function \( \ell_{E[X]}(x) \) such that

\[
\ell_{E[X]}(E[X]) = g(E[X])
\]

and \( \ell_{E[X]}(x) \leq g(x) \) for all \( x \).

- If \( g \) is strictly convex, then

\[
\ell_{E[X]}(x) < g(x)
\]

for all \( x \neq E[X] \).

- Therefore

\[
E[g(X)] \geq E[\ell_{E[X]}(X)] = \ell_{E[X]}(E[X]) = g(E[X]).
\]

- The inequality is strict if \( g \) is strictly convex and \( X \) is not constant.
Example

- Jensen’s inequality gives an alternative proof of Liapunov’s inequality.
- For $0 < r \leq s < \infty$,

\[
E[|X|^s] = E[(|X|^r)^{s/r}] \geq E[|X|^r]^{s/r}
\]

since $x^{s/r}$ is convex on $(0, \infty)$ for $s/r \geq 1$. 
**Example**

- Suppose $X$ has PDF or PMF $f$.
- Let $g$ be another PDF or PMF with $g(x) = 0$ whenever $f(x) = 0$.
- Then, since the logarithm is concave, Jensen's inequality implies

  $$E \left[ \log \frac{g(X)}{f(X)} \right] \leq \log E \left[ \frac{g(X)}{f(X)} \right]$$

- Now for the continuous case

  $$E \left[ \frac{g(X)}{f(X)} \right] = \int \frac{g(x)}{f(x)} f(x) dx = \int g(x) dx = 1.$$  

- An analogous result holds in the discrete case.
- Therefore

  $$E \left[ \log \frac{g(X)}{f(X)} \right] \leq \log(1) = 0.$$  

- The inequality is strict unless $g(X) = f(X)$ almost surely.
- This inequality arises quite often when studying statistical procedures.
A useful identity:

**Lemma (Stein’s Lemma)**

Let $X \sim N(\mu, \sigma^2)$, and let $g$ be a differentiable function satisfying $E[|g'(X)|] < \infty$. Then

$$E[g(X)(X - \mu)] = \sigma^2 E[g'(X)]$$

**Proof.**

This can be shown by two approaches:

- integration by parts
- writing the normal density as the integral of its derivative and changing order of integration
Example

• Suppose $X \sim N(0, 1)$.

• Apply Stein’s lemma to $E[X^n]$ to get

$$E[X^n] = E[X^{n-1}X] = E[g(X)X]$$
$$= E[g'(X)] = (n - 1)E[X^{n-2}].$$

• Applied repeatedly this gives

$$E[X^n] = (n - 1) \times (n - 3) \times \cdots \times 3 \times 1$$

for even $n$. 
Second Midterm Exam

- The exam will cover the material covered in class and in assignments through Friday, October 30.
- The exam is closed book.
- The exam will include some information on distributions along the lines of the *Table of Common Distributions* in the text.