Transformations

Feng-Chang Lin

Department of Biostatistics
University of North Carolina at Chapel Hill

flin@bios.unc.edu

(C&B §2.1, §4.3)

Introduction

In this unit we will learn how to answer the following questions:

- $X \sim \mathrm{N}(0,1)$. Find the distribution of X^2.
- $X \sim \mathrm{U}(0,1)$. Find the distribution of $-\log X$.
- $X \sim \mathrm{Exp}(\lambda)$. Find the distribution of $X_{(1)} \equiv \min\{X_1, \ldots, X_n\}$.
- $X \sim \mathrm{Exp}(\lambda)$, $Y \sim \mathrm{Exp}(\mu)$, and $X \perp Y$. Find the distribution of $X - Y$. Find the distribution of X/Y.
- $X \sim \mathrm{Exp}(\lambda)$, $Y \sim \mathrm{Exp}(\mu)$, and $X \perp Y$. Find the (joint) distribution of $(X - Y, X/Y)$.
- Table of common distributions in page 627 in C&B may help.

Why Such Questions?

- Summarize data to make statistical inferences.
- Examples include sample mean and sample variance.
- Need to know the distributions under a given *model* in order to use the summary statistics.
- Mathematically, we formulate these statistics as *transformations*.

Transformation of One Variable

- X is a random variable with pdf or pmf f_X and one wants to find the distribution of $Y = g(X)$ where g is a given function.
- We define $\mathcal{X}$ to be the sample space of X

$$\mathcal{X} := \{x : f(x) > 0\},$$

and $\mathcal{Y}$ to be the sample space of Y, where

$$\mathcal{Y} := \{y : g(x) = y \;\; \textit{for some } x \in \mathcal{X}\}.$$

- A set such as X or Y is called the support set of a distribution, or simply the support of the distribution.

PMF of Discrete Random Variables

- If X is discrete, the pmf of $g(X)$ is no more than simple enumeration.
- **Example**: X is Poisson(λ) and $Y = X^2$, i.e. $g(x) = x^2$. What is $P(Y = 25)$?

$$P(Y = 25) = P(X = 5) = e^{-\lambda}\lambda^5/5!$$

- How about $P(Y = 10)$?
- In this example $X = \{0, 1, 2, 3, 4, \ldots\}$ and $Y = \{0, 1, 4, 9, 16, \ldots\}$.
- For $y \geq 0$, we have $P(Y = y) = P(X = \sqrt{y})$. What if $\sqrt{y}$ is not an integer?

PMF of Discrete Random Variables (cont'd)

- **Example**: X is Poisson(λ) and $Y = X^2 - 7X + 12$. What is $P(Y = 0)$?

$$P(Y = 0) = P(X \in \{3, 4\}) = P(X = 3) + P(X = 4)$$

- For discrete X, to find $P(Y = y)$, find the set

$$A_y = \{x : g(x) = y, x \in \mathcal{X}\}.$$

- $P(Y = y) = P(X \in A_y)$, where A_y may be an empty set.

Transformations of Continuous Random Variables

- Enumeration does not work. Use either *cdf* or *Jacobian*.
- **Example**: Let $X \sim \text{Exp}(\lambda)$ and $Y = g(X) = X^{1/2}$. The cdf of X is $F_X(x) = 1 - e^{-x/\lambda}$, $x \geq 0$. The cdf of Y is

  $$F_Y(y) = P(Y \leq y) = P(X \leq y^2) = F_X(y^2) = 1 - e^{-y^2/\lambda},$$

  and the pdf is

  $$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{2y}{\lambda} e^{-y^2/\lambda}, \quad y \geq 0.$$

- $Y \sim$ Weibull $(2, \lambda)$ (C&B, page 627).

Transformations of Continuous RV (cont'd)

- **Example**: Let $X \sim N(0,1)$ and $Y = g(X) = X^2$. For $y > 0$, the cdf of Y is

$$F_Y(y) = P(Y \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}),$$

and the pdf is

$$f_Y(y) = \{\phi(\sqrt{y}) + \phi(-\sqrt{y})\}\frac{d}{dy}\sqrt{y} = \frac{1}{\sqrt{2\pi y}}e^{-y/2}, \quad y > 0.$$

- $Y \sim \chi^2(1)$ (C&B, page 626).

Inverse Probability Integral Transform

- **Example**: Suppose that F is a *continuous* and *strictly increasing* cdf, and suppose that U is uniform on (0,1). The distribution of $Y = F^{-1}(U)$ is

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(F^{-1}(U) \leq y) = P(F(F^{-1}(U)) \leq F(y)) \\ &= P(U \leq F(y)) = F(y). \end{aligned}$$

- Useful in some computer simulation. For example, $X \sim \text{Exp}(1)$ with $F(x) = 1 - e^{-x}$ and $F^{-1}(u) = -log(1 - u)$.
- One may generate U from $U(0, 1)$ and get $-log(1 - U)$ following Exp(1).
- May not work for a normal distribution since F^{-1} is not easy to compute.

Transformation Using Jacobian

- Suppose g is *monotone increasing*. That implies *one-to-one and onto* from $\mathcal{X}$ to $\mathcal{Y}$.
- Then g^{-1} is well-defined *monotone increasing* function. If $X = g^{-1}(Y)$, then

$$P(Y \leq y) = P(g^{-1}(Y) \leq g^{-1}(y)) = P(X \leq g^{-1}(y)).$$

- Hence $F_Y(y) = F_X(g^{-1}(y))$. The pdf of Y is

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$

- If g is *monotone decreasing*, then

$$f_Y(y) = -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$

Transformation Using Jacobian (cont'd)

- For monotone g (either increasing or decreasing),

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|, & y \in \mathcal{Y}, \\ 0, & \textit{otherwise}. \end{cases}$$

- The factor $\frac{d}{dy}g^{-1}(y)$ is called *Jacobian* of g^{-1}.
- Only works for monotone g.
- Require $\frac{d}{dy}g^{-1}(y)$ be continuous on $\mathcal{Y}$.
- See Theorems 2.1.3 and 2.1.5 in C&B.

Transformation Using Jacobian (cont'd)

- **Example** Let $X \sim \text{Exp}(\lambda)$ and $Y = g(X) = X^{1/2}$. Note that $f_X(x) = \lambda^{-1} e^{-x/\lambda}$, $\mathcal{X} = [0, \infty)$.
- The function g is monotone on $\mathcal{X}$, $\mathcal{Y} = [0, \infty)$, and $g^{-1}(y) = y^2$ for $y \in \mathcal{Y}$.
- The derivative of $g^{-1}(y)$ is $2y$.
- The density of Y is

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = \frac{2y}{\lambda} e^{-y^2/\lambda},$$

for $y \geq 0$ and $f_Y(y) = 0$ for $y < 0$.

Transformation Using Jacobian (cont'd)

- What if g^{-1} is not monotone, such as $g(x) = x^2$ on $\mathcal{X} = (-\infty, \infty)$?
- Take advantage of partition: monotone over $(-\infty, 0)$ and monotone over $(0, \infty)$.
- Apply the method of Jacobian to each piece and add up the contributions from all the pieces.

Transformation Using Jacobian (cont'd)

- **Theorem 2.1.8 in C&B**: Suppose that there exists a partition, $A_0, A_1, \ldots, A_k$ of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i, $i = 1, \ldots, k$.
- Further suppose that the function g is monotone over each A_i.
- Let g_i denote the restriction of g to $x \in A_i$, $i > 0$, and suppose that

$$\mathcal{Y} = \{y : g_i(x) = y \textit{ for some } x \in A_i\}, \ 1 \leq i \leq k.$$

- Knowing that $g_i^{-1}(y)$ must be in A_i for $y \in \mathcal{Y}$, one can have

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y)) |\frac{d}{dy} g_i^{-1}(y)|, & y \in \mathcal{Y}, \\ 0, & \textit{otherwise}. \end{cases}$$

- A_0 is a set for interval endpoints which have zero probability.

Transformation Using Jacobian (cont'd)

- **Example**: Let $X \sim N(0,1)$ and $Y = g(X) = X^2$. Note that $\mathcal{X} = (-\infty, \infty)$ and $\mathcal{Y} = [0, \infty)$.
- We can take $A_1 = (-\infty, 0)$ and $g_1(x) = x^2$ on A_1 and $g_1^{-1}(y) = -\sqrt{y}$. We take $A_2 = (0, \infty)$ and $g_2(x) = x^2$ on A_2 and $g_2^{-1}(y) = \sqrt{y}$.
- The pdf of Y is

$$
\begin{aligned}
f_Y(y) &= \phi(-\sqrt{y})|-\frac{1}{2\sqrt{y}}| + \phi(\sqrt{y})\}|\frac{1}{2\sqrt{y}}| \\
&= \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad y > 0.
\end{aligned}
$$

- The method of Jacobian does NOT apply to discrete random variables.

Bivariate and Multivariate Transformations

- X is Poisson(λ_1), Y is Poisson(λ_2), and $X \perp Y$. Let $U = X + Y$.
- Find $P(U = 3)$: The event $\{U = 3\}$ arises when $\{X = 0, Y = 3\}$, $\{X = 1, Y = 2\}$, $\{X = 2, Y = 1\}$, and $\{X = 3, Y = 0\}$.
- Since these four events are mutually exclusive,

$$\begin{aligned} P(U = 3) = {} & P(X = 0, Y = 3) + P(X = 1, Y = 2) \\ & + P(X = 2, Y = 1) + P(X = 3, Y = 0). \end{aligned}$$

- By independence, $P(X = x, Y = y) = P(X = x)P(Y = y)$. Then,

$$P(U = 3) = \sum_{x=0}^{3} P(X = x, Y = 3 - x) = \sum_{x=0}^{3} P(X = x)P(Y = 3 - x).$$

Method of Jacobian

- The method of Jacobian applies with a small adjustment.
- Suppose that the random vector (X, Y) has pdf $f_{X,Y}(x, y)$ and sample space $\mathcal{S}$.
- Consider the transformation of (X, Y) into (U, V) through

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

- We write $(U, V) = g(X, Y)$. It requires
  - (i) g is one-to-one on $\mathcal{S}$, so its inverse exists and is well-defined.
  - (ii) g has continuous partial derivatives on $\mathcal{S}$.
  - (iii) The Jacobian of g is not zero on $\mathcal{S}$.
- Let h denote the inverse function of g and $x = h_1(u, v)$ and $y = h_2(u, v)$. The density of (U, V) is given by

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|J|.$$

Method of Jacobian (cont'd)

- J is the Jacobian of h; $|\cdot|$ is the *determinant* of the matrix of partial derivatives as

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}.$$

- What is the determinant of a 2×2 matrix?
- **Example** Suppose $X \sim \text{Gamma}(\alpha_1, 1)$, $Y \sim \text{Gamma}(\alpha_2, 1)$, and $X \perp Y$. Let $U = X + Y$ and $V = X/(X + Y)$. The joint pdf of X and Y is

$$f_{X,Y}(x, y) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-x-y} x^{\alpha_1 - 1} y^{\alpha_2 - 1}, \quad x > 0, \quad y > 0.$$

- Let $(u, v) = g(x, y) = (x + y, x/(x + y))$ and its range is $\{(u, v); u > 0, 0 < v < 1\}$.

Method of Jacobian (cont'd)

- The inverse function is $h(u, v) = (uv, u - uv)$ and the Jacobian is

$$J = \begin{vmatrix} v & u \\ 1 - v & -u \end{vmatrix} = -uv - u(1 - v) = -u.$$

- The joint pdf of (U, V) is

$$f_{U,V}(u, v) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-u}(uv)^{\alpha_1 - 1}(u - uv)^{\alpha_2 - 1}u, \;\; u > 0, \;\; 0 < v < 1$$

- We can write

$$f_{U,V}(u, v) = \frac{e^{-u}u^{\alpha_1 + \alpha_2 - 1}}{\Gamma(\alpha_1 + \alpha_2)} \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1 - 1}(1 - v)^{\alpha_2 - 1}.$$

- We may claim $U \sim$ Gamma$(\alpha_1 + \alpha_2, 1)$, $V \sim$ Beta(α_1, α_2), and $U \perp V$.

Method of CDF

- **Example** A random point in the unit disc has coordinates X and Y where (X, Y) has density

$$f_{X,Y}(x,y) = 1/\pi, \ \textit{for} \ (x,y) \in \mathcal{S},$$

where $\mathcal{S} = \{(x,y) : x^2 + y^2 < 1\}$. The length of the line from the origin to (X, Y) is

$$U = \sqrt{X^2 + Y^2} = g(X, Y).$$

- The cdf of U is

$$F_U(u) = P(U \leq u) = P(\sqrt{X^2 + Y^2} \leq u) = P(X^2 + Y^2 \leq u^2) = u^2.$$

- The pdf U is $f_U(u) = \frac{d}{du} F_U(u) = 2u$.
- How about the method of Jacobian?

Convolution Formula

- Suppose X and Y are independent continuous random variables with pdf f_X and f_Y. One way to find the density of $Z = X + Y$ is to introduce another variable W so that the transformation from (X, Y) to (Z, W) is one-to-one.
- Choose $W = X$. The inverse transformation, from (Z, W) to (X, Y), is $X = W$ and $Y = Z - W$. The Jacobian is -1.
- Then the density of (Z, W) is $f_{Z,W}(z, w) = f_X(w) f_Y(z - w)$.
- The density of Z is obtained by integrating out w,

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,W}(z, w) dw = \int_{-\infty}^{\infty} f_X(w) f_Y(z - w) dw,$$

  which is called *convolution formula*.
- Be careful about the range of W.

Location-Scale Family

- Derivation can be simplified by shifting and scaling.
- Suppose that random variable Z has pdf f, and let $X = \mu + \sigma Z$ where $-\infty < \mu < \infty$ and $0 < \sigma < \infty$.
- Say, $X = g(Z)$ and $Z = g^{-1}(X) = (X - \mu)/\sigma$ with Jacobian $1/\sigma$. The density of X is

$$f_X(x) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma}).$$

- Starting with a given density f, the set of distributions generated by all possible (μ, σ) is known as a location-scale family.
- **Example** If $Z \sim N(0,1)$ with density $f(z) = (1/\sqrt{2\pi})e^{-z^2/2}$, then $X = \mu + \sigma Z$ has density

$$f_X(x) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

Sums of Independent Random Variables

- Use moment generating function (mgf) method.
- Recall that $M_X(t) = Ee^{tX}$.
- **Example** $X \perp Y$ and $X, Y \sim N(0,1)$; $M_X(t) = M_Y(t) = e^{t^2/2}$. Let $U = aX + bY + c$.
- The mgf of U is

$$M_U(t) = Ee^{taX+tbY+tc} = e^{t^2a^2/2}e^{t^2b^2/2}e^{tc} = e^{ct+(a^2+b^2)t^2/2},$$

  which is mgf of $N(c, a^2+b^2)$. Therefore, $U \sim N(c, a^2+b^2)$.
- **Example** $X_1, \ldots X_n$ are mutually independent Bernoulli(θ) random variables. The mgf of each X_i is $M_{X_i}(t) = 1 - \theta + \theta e^t$. The mgf of $U = X_1 + \cdots + X_n$ is $M_U(t) = (1 - \theta + \theta e^t)^n$, which is the mgf of what distribution?