BIOS 660/BIOS 672 (3 Credits): Probability and Statistical Inference I

Jianwen Cai

https://sakai.unc.edu/portal/site/bios660-bios672-3-credits

Notes 7

Functions of Random Variables

(C-B Chap 2.1 & Gut I.2)

If X is a rv with sample space $\mathcal{X} \subset \mathbb{R}$ and cdf $F_X(x)$ then any function of X, say $Y = g(X)$ is also a random variable. The new random variable Y has a new sample space $\mathcal{Y} = g(\mathcal{X}) \subset \mathbb{R}$. The objective is to find the cdf $F_Y(y)$ of Y.

Example: Suppose X is an exponential random variable with parameter 1, i.e. $F_X(x) = 1 - e^{-x}$, $f_X(x) = e^{-x}$. What is the distribution of $Y = X/\lambda$?

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(X/\lambda \le y) = P(X \le \lambda y) \\ &= F_X(\lambda y) = 1 - e^{-\lambda y} \end{aligned}$$

Y is distributed $\exp(\lambda)$ with density $f_Y(y) = \lambda e^{-\lambda y}$.

Simple example

Example: Change units – Fahrenheit to Celsius

$$C = \frac{5}{9}(F - 32)$$

Ranges:

$$20^\circ C < C < 30^\circ C \;\Leftrightarrow\; 68^\circ F < F < 86^\circ F$$

$$(\mathcal{Y}, \mathcal{B}, F_Y) \leftarrow g(\cdot) \leftarrow (\mathcal{X}, \mathcal{B}, F_X)$$

Probability mapping

For any Borel set A:

$$\begin{aligned} P(Y \in A) &= P(g(X) \in A) \\ &= P(\{x \in \mathcal{X} : g(x) \in A\}) \\ &= P(X \in g^{-1}(A)). \end{aligned}$$

where we have defined

$$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\}$$

Notice that $g^{-1}(A)$ is well defined even if $g(\cdot)$ is not bijective (one-to-one).

Example: Let $g(x) = x^2$.
Then

$$g^{-1}([-1,1]) = [-1,1]$$

But

$$g(g^{-1}([-1,1])) = [0,1]$$

Discrete RVs

Suppose that X is a discrete random variable with probability mass function $p(x) = P(X = x)$.
Then, the *pmf* of a 1-1 transformation $Y = g(X)$ is given by

$$P(Y = y) = P(g(X) = y) = P(\{x : g(x) = y\}) = \sum_{x:g(x)=y} p(x)$$

In practice, one never sees many general results about transformations of discrete random variables because the results are so simple!

Continuous RVs

Consider the transformation $Y = g(X)$ where $g(x)$ is strictly increasing (consequently a one-to-one transformation), and suppose g is differentiable. This means that we can also define the *inverse function*, $g^{-1}(y)$.

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

The *pdf* of Y is thus,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = F_X'[g^{-1}(y)] \frac{dg^{-1}(y)}{dy} = f_X(g^{-1}(y)) \frac{dx}{dy}$$

since

$$x = g^{-1}(y), \quad \text{so that } \frac{dx}{dy} = \frac{dg^{-1}(y)}{dy}$$

Continuous RVs

Suppose $Y = g(x)$ is still one-to-one, but decreasing instead of increasing.

$$F_Y(y) = P\{g(X) \leq y\} = P\{X > g^{-1}(y)\} = 1 - F_X(g^{-1}(y))$$

and

$$\begin{aligned} f_Y(y) &= -F_X'(g^{-1}(y)) \frac{dg^{-1}(y)}{dy} = -f_X(g^{-1}(y)) \frac{dx}{dy} \\ &= f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right| \end{aligned}$$

The last step follows because $\frac{dx}{dy}$ is negative.
Therefore, regardless of whether $Y = g(x)$ is increasing or decreasing, so long as it is monotonic, we have

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|$$

Linear Transformation

Given X with pdf $f_X(x)$, let

$$Y = a + bX, \qquad \frac{dy}{dx} = b$$

Then

$$f_Y(y) = f_X[g^{-1}(y)] \left|\frac{dx}{dy}\right| = f_X\left(\frac{y-a}{b}\right)\frac{1}{|b|}$$

This transformation is often used when X has mean 0 and standard deviation 1. The linear transformation above creates a rv Y with a distribution that has the same shape as that of X but has mean a and standard deviation b.
Conversely, if Y has mean a and standard deviation b, then $X = (Y-a)/b$ has mean 0 and standard deviation 1. This is called sometimes the *"Studentized"* transform.

Normal Distribution

Let $X \sim N(0,1)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad -\infty < x < \infty$$

The transformation

$$Y = \mu + \sigma X, \qquad X = \frac{Y-\mu}{\sigma}$$

yields

$$f_Y(y) = f_X\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

More generally, a distribution is a member of the class of *location-scale distributions* if the distribution of a linear transformation of a random variable with that distribution has the same distribution, but with different parameters.

Square root of an exponential RV

We have already seen that a constant times an exponential random variable leads to another exponential random variable. Suppose $X \sim \exp(\lambda)$, so that

$$f_X(x) = \lambda e^{-\lambda x}\, 1(x \geq 0)$$

and consider the distribution of $Y = \sqrt{X}$.

The transformation

$$y = g(x) = \sqrt{x}, \qquad x \geq 0$$

is one-to-one and has an inverse $x = y^2$ with $dx/dy = 2y$. Thus

$$f_Y(y) = f_X(y^2) 2y = 2\lambda y e^{-\lambda y^2}, \qquad y \geq 0$$

This distribution is a particular form of the Rayleigh distribution and is a special case of the χ, Rice and Weibull distributions.

Probability Integral Transform

Let $X \sim F_X(x)$. Define the transformation

$$Y = F_X(X) \in [0, 1], \qquad X = F_X^{-1}(Y)$$

Here

$$\frac{dy}{dx} = F'_X(x) = f_X(x)$$
$$f_Y(y) = f_X[F_X^{-1}(y)]\, \frac{1}{f_X[F_X^{-1}(y)]} = 1$$

i.e. Y is uniform over $[0, 1]$.

Another way to see it is through the distribution function:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(F_X(X) \leq y) \\ &= P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y \end{aligned}$$

Probability Integral Transform (cont.)

The probability integral transform is useful in statistics for checking goodness of fit of a distribution to a set of data.

Example:

$$X \sim \exp(\lambda) \quad \Rightarrow \quad Y = 1 - \exp(-\lambda X) \sim U[0,1]$$

If one has data $X_1, \ldots, X_n$, one could compute the transformed data $Y_i = 1 - \exp(-\lambda X_i)$, $i = 1, \ldots, n$, and check whether the Y_i's appear uniformly distributed over the interval $[0,1]$.

Inverse Probability Integral Transform

We can also start from the uniform distribution and do the inverse procedure.
Suppose $X \sim U[0,1]$, so that $f_X(x) = 1$ and $F_X(x) = x$ for $x \in [0,1]$. Let

$$Y = F^{-1}(X), \qquad X = F(Y)$$

where $F(\cdot)$ is a non-decreasing absolutely continuous function $F : \mathbb{R} \to [0,1]$, $F(y) = \int_{-\infty}^{y} f(x)\,dx$.
Then

$$\frac{dx}{dy} = F'(y) = f(y) \quad \Rightarrow \quad f_Y(y) = f(y)$$

i.e. Y has the *pdf* corresponding to F.
Another way to see it is through the distribution function:

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

Inverse Probability Integral Transform (cont.)

The Inverse Probability Integral Transform is used extensively in simulation of random variables.

Example: Most random number generators generate random numbers uniformly in the interval $[0, 1]$. Suppose we want to generate random numbers from an exponential distribution with parameter λ.

Let $X \sim U[0, 1]$. We want Y with distribution function $F(y) = 1 - \exp(-\lambda y)$. Then we need the transformation

$$Y = F^{-1}(X) = \frac{-1}{\lambda} \log(1 - X)$$

The recipe is

1. Generate random numbers X_i uniformly over $[0, 1]$.
2. Compute $Y_i = -\log(1 - X_i)/\lambda$.

Example: Cauchy Distribution

Let θ be distributed uniformly between $(-\pi/2, \pi/2)$:

$$f(\theta) = \frac{1}{\pi}, \qquad -\pi/2 < \theta < \pi/2$$

Consider $Y = \tan\theta$.

$$\begin{aligned} \frac{dy}{d\theta} = \sec^2\theta &= 1 + \tan^2\theta = 1 + y^2 \\ f_Y(y) = \frac{1}{\pi}\left|\frac{d\theta}{dy}\right| &= \frac{1}{\pi}\frac{1}{(1+y^2)} \qquad -\infty < y < \infty \end{aligned}$$

The distribution with this density is known as the Cauchy distribution.

One-to-many

What if the transformation is not 1-1? The trick is to start with the cdf of the transformed random variable.

Example: Let $Y = |X|$, and assume X is continuous.

$$\begin{aligned} F_Y(y) &= P\{Y \le y\} = P\{-y \le X \le y\} = F_X(y) - F_X(-y) \\ f_Y(y) &= F'_X(y) - F'_X(-y)(-1) = f_X(y) + f_X(-y) \end{aligned}$$

Suppose

$$X \sim N(0,1), \qquad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Then

$$f_Y(y) = \frac{2}{\sqrt{2\pi}} e^{-y^2/2}, \qquad 0 < y < \infty$$

Quadratic transformation

Let

$$Y = X^2, \qquad \frac{dy}{dx} = 2x, \qquad \left|\frac{dy}{dx}\right| = 2\sqrt{y}$$

Then

$$\begin{aligned} F_Y(y) &= P\{Y \le y\} = P\{X^2 \le y\} = P\{-\sqrt{y} < X \le \sqrt{y}\} \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ f_Y(y) &= F'_X(\sqrt{y}) \left(\frac{1}{2} y^{-\frac{1}{2}}\right) - F'_X(-\sqrt{y}) \left(-\frac{1}{2} y^{-\frac{1}{2}}\right) \\ &= \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})], \qquad y > 0 \end{aligned}$$

Example

Suppose $X \sim N(0,1)$, $Y = X^2$:

$$
\begin{aligned}
f_X(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty \\
f_Y(y) &= \frac{1}{2\sqrt{y}} \left[\frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \right] \\
&= \frac{(y/2)^{-\frac{1}{2}} e^{-\frac{y}{2}}}{2\sqrt{\pi}}, \quad y > 0
\end{aligned}
$$

This is the density of a χ^2 distribution with 1 degree of freedom.

Result: If $X \sim N(0,1)$, then $X^2 \sim \chi^2(1)$.

General Result–Theorem C-B 2.1.8

Suppose $Y = g(X)$ is not 1-1, but there are disjoint sets $A_1, ...A_k$ that span the domain (sample space) of X such that $g(.) = g_j(.)$ is continuous and 1-1 on each A_j. This means that the inverse, $x = g_j^{-1}(y)$ exists on each A_j. Then

$$
f_Y(y) = \sum_{j=1}^{k} f(g_j^{-1}(y)) \left| \frac{dg_j^{-1}(y)}{dy} \right|
$$

Example: A wrapped distribution

Suppose $X \in \mathbb{R}$ with density $f_X(x)$ represents a random angle of rotation (in radians) from the x-axis on the unit circumference. The observed angle is

$$\Theta = X \mod 2\pi, \qquad \Theta \in [0, 2\pi)$$

because it is impossible to tell if the rotation involved more than one full turn. In this case

$$f_\Theta(\theta) = \sum_{j=-\infty}^{\infty} f(\theta + 2\pi j), \qquad 0 \le \theta < 2\pi$$

If $X \sim N(0, \sigma^2)$, then

$$f_\Theta(\theta) = \sum_{j=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\theta + 2\pi j)^2}{2\sigma^2}\right)$$