Lecture 21: Convergence of transformations and generating a random variable

If Z_n converges to Z in some sense, we often need to check whether $h(Z_n)$ converges to $h(Z)$ in the same sense.

Continuous mapping theorem.

If a sequence of random vectors Z_n converges in distribution to Z and h is a function continuous on A with $P_Z(A) = 1$, then $h(Z_n)$ converges in distribution to $h(Z)$.

- If random variables Z_n, $n = 1, 2, ...,$, converges in distribution to $Z \sim N(0,1)$, then Z_n^2 converges in distribution to $Z^2 \sim$ the chi-square distribution with degree of freedom 1.
- If (Z_n, Y_n) converges in distribution to (Z, Y) as random vectors, then $Y_n + Z_n$ converges in distribution to $Y + Z$, $Y_n Z_n$ converges in distribution to YZ, and Y_n / Z_n converges in distribution to Y/Z.
- In the previous situation if Z and Y are iid $\sim N(0,1)$, then Z_n/Y_n converges in distribution to the $Cauchy(0,1)$ distribution.

- In the previous situation if Z and Y are iid $\sim N(0,1)$, then $\max\{Z_n, Y_n\}$ converges in distribution to $\max\{Z, Y\}$, which has the cdf $[\Phi(x)]^2$ ($\Phi(x)$ is the cdf of $N(0,1)$).
- The condition that (Z_n, Y_n) converges in distribution to (Z, Y) cannot be relaxed to Z_n converges in distribution to Z and Y_n converges in distribution to Y, i.e., we need the convergence of the joint cdf of (Z_n, Y_n); e.g., $Z_n + Y_n$ may not converge in distribution to $Z + Y$ if we only have the marginal convergence.
- The next result, which plays an important role in statistics, establishes the convergence in distribution of $Z_n + Y_n$, $Z_n Y_n$, or Z_n / Y_n with no information regarding the joint cdf of (Z_n, Y_n).

Theorem 5.5.17 (Slutsky's theorem)

Let X_n converges in distribution to X and Y_n converges in distribution (probability) to c (a constant), then

(i) $X_n + Y_n$ converges in distribution to $X + c$;

(ii) $Y_n X_n$ converges in distribution to cX;

(iii) X_n / Y_n converges in distribution to X/c if $c \neq 0$.

Proof.

We prove (i) only. (The proofs of (ii) and (iii) are left as exercises.)
For $t \in \mathscr{R}$ and $\varepsilon > 0$ being fixed constants,

$$\begin{aligned} F_{X_n+Y_n}(t) &= P(X_n + Y_n \leq t) \\ &\leq P(\{X_n + Y_n \leq t\} \cap \{|Y_n - c| < \varepsilon\}) + P(|Y_n - c| \geq \varepsilon) \\ &\leq P(X_n \leq t - c + \varepsilon) + P(|Y_n - c| \geq \varepsilon) \end{aligned}$$

Similarly,

$$F_{X_n+Y_n}(t) \geq P(X_n \leq t - c - \varepsilon) - P(|Y_n - c| \geq \varepsilon).$$

If $t - c$, $t - c + \varepsilon$, and $t - c - \varepsilon$ are continuity points of F_X, then it follows from the previous two inequalities and the convergence of $X_n + Y_n$ that

$$F_X(t - c - \varepsilon) \leq \liminf_n F_{X_n+Y_n}(t) \leq \limsup_n F_{X_n+Y_n}(t) \leq F_X(t - c + \varepsilon).$$

Since ε can be arbitrary (why?),

$$\lim_{n\to\infty} F_{X_n+Y_n}(t) = F_X(t - c).$$

The result follows from $F_{X+c}(t) = F_X(t - c)$.

Example 5.5.18.

A typical application of Slutsky's theorem is in establishing the normal approximation with estimated variance.

For a random sample $X_1, ..., X_n$ with finite $\mu = E(X_1)$ and variance $\sigma^2 = \mathrm{Var}(X_1)$, the CLT shows that

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \text{ converges in distribution to } N(0, 1)$$

If σ^2 is "estimated" by the sample variance S^2, then by Example 5.5.3, S^2 converges in probability (distribution) to σ^2, and thus S converges in probability to σ.

Then Slutsky's theorem shows that

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \text{ converges in distribution to } N(0, 1)$$

Another typical application of Slutsky's theorem is given in the proof of the following important result.

Theorems 5.5.28 (the delta method)

Let $X_1, X_2, \ldots$ and Y be random k-vectors satisfying

$$a_n(X_n - c) \text{ converges in distribution to } Y,$$

where $c \in \mathscr{R}^k$ is a constant vector and a_n's are positive constants with $\lim_{n\to\infty} a_n = \infty$. (In many cases, $c = E(X_n)$ and $a_n = \sqrt{n}$.)
If g is a function from $\mathscr{R}^k$ to $\mathscr{R}$ and is differentiable at c, then

$$a_n[g(X_n) - g(c)] \text{ converges in distribution to } \nabla g(c)' Y,$$

where $\nabla g(x)'$ is the transpose of the k-dimensional vector of partial derivatives of g at x. In particular, if $Y \sim N(0, \Sigma)$, then

$$a_n[g(X_n) - g(c)] \text{ converges in distribution to } N\left(0, \nabla g(c)' \Sigma \nabla g(c)\right).$$

We only give a proof for the univariate case ($k = 1$), in which

$$a_n[g(X_n) - g(c)] \text{ converges in distribution to } g'(c) Y,$$

and if $Y \sim N(0, \sigma^2)$,

$$a_n[g(X_n) - g(c)] \text{ converges in distribution to } N(0, [g'(c)]^2 \sigma^2)$$

Proof of Theorem 5.5.28 with $k = 1$.

Let
$$Z_n = a_n[g(X_n) - g(c)] - a_n g'(c)(X_n - c).$$
If we can show that Z_n converges to in probability to 0, then the result follows from
$$a_n[g(X_n) - g(c)] = a_n g'(c)(X_n - c) + Z_n,$$
the condition that $a_n(X_n - c)$ converges in distribution to Y, and Slutsky's theorem.

The differentiability of g at c means that
$$\lim_{x \to c} \left| \frac{g(x) - g(c) - g'(c)(x - c)}{x - c} \right| = \lim_{x \to c} \left| \frac{g(x) - g(c)}{x - c} - g'(c) \right| = 0$$
i.e., for any $\varepsilon > 0$, there is a $\delta_\varepsilon > 0$ such that
$$|g(x) - g(c) - g'(c)(x - c)| \leq \varepsilon |x - c| \qquad \text{when } |x - c| < \delta_\varepsilon.$$
Then, for a fixed $\eta > 0$,
$$\begin{aligned} P(|Z_n| \geq \eta) &= P(|Z_n| \geq \eta, |X_n - c| \geq \delta_\varepsilon) + P(|Z_n| \geq \eta, |X_n - c| < \delta_\varepsilon) \\ &\leq P(|X_n - c| \geq \delta_\varepsilon) + P(a_n |X_n - c| \geq \eta/\varepsilon, |X_n - c| < \delta_\varepsilon) \end{aligned}$$

Since $a_n \to \infty$, $a_n(X_n - c)$ converges in distribution to Y and Slutsky's theorem imply that $X_n - c$ converges in probability to 0.

By the continuity mapping theorem, $a_n(X_n - c)$ converges in distribution to Y implies $a_n|X_n - c|$ converges in distribution to $|Y|$ (the function $|x|$ is continuous).

Without loss of generality, we assume that $F_{|Y|}$ is continuous at η/ε.

Then

$$\begin{aligned}\limsup_n P(|Z_n| \geq \eta) &\leq \lim_{n\to\infty} P(|X_n - c| \geq \delta_\varepsilon) + \lim_{n\to\infty} P(a_n|X_n - c| \geq \eta/\varepsilon) \\ &= P(|Y| \geq \eta/\varepsilon).\end{aligned}$$

and Z_n converges in probability to 0 since ε can be arbitrary.

In statistics, we often need a nondegenerated limiting distribution of $a_n[g(X_n) - g(c)]$ so that probabilities involving $a_n[g(X_n) - g(c)]$ can be approximated by the cdf of $\nabla g(c)'Y$, using the delta method.

In some situations, however, $\nabla g(c)$ so that the limiting distribution of $a_n[g(X_n) - g(c)]$ becomes degenerated, in which cases we need to consider a high order delta method.

Theorem 5.5.26 (mth order delta method)

Let X_n and Y be random k-vectors and c and a_n be constants satisfying the conditions in Theorem 5.5.28.
Suppose that g has continuous partial derivatives of order $m > 1$ in a neighborhood of c, with all the partial derivatives of order j, $1 \le j \le m-1$, vanishing at c, but with the mth-order partial derivatives not all vanishing at c. Then

$$a_n^m[g(X_n) - g(c)] \text{ converges in distribution to}$$

$$\frac{1}{m!} \sum_{i_1=1}^{k} \cdots \sum_{i_m=1}^{k} \left. \frac{\partial^m g}{\partial x_{i_1} \cdots \partial x_{i_m}} \right|_{x=c} Y_{i_1} \cdots Y_{i_m}$$

where Y_j is the jth component of Y. When $k = 1$,

$$a_n^m[g(X_n) - g(c)] \text{ converges in distribution to } \frac{g^{(m)}(c)}{m!} Y^m$$

Example.

Let $\{X_n\}$ be a sequence of random variables satisfying

$$\sqrt{n}(X_n - c) \text{ converges in distribution to } Z \sim N(0,1).$$

Consider the function $g(x) = x^2$, $g'(x) = 2x$.
If $c \neq 0$, then an application of the delta method gives that

$$\sqrt{n}(X_n^2 - c^2) \text{ converges in distribution to } \sim N(0, 4c^2).$$

If $c = 0$, $g'(c) = 0$ but $g''(c) = 2$.
Hence, an application of the 2nd order delta method gives that

$$nX_n^2 \text{ converges in distribution to } Z^2,$$

which has the chi-square distribution with degree of freedom 1.
This result can also be obtained by applying the continuity mapping theorem.

The next example involves 2-dimensional random vectors.

Example 5.5.27 (ratio estimator)

Let $Z_n = (X_n, Y_n)$, $n = 1, 2, \ldots$, be iid bivariate random vectors with finite 2nd order moments.

Let $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, $\bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i$, $\mu_x = E(X_1)$, $\mu_y = E(Y_1) \neq 0$, $\mu = (\mu_x, \mu_y)$, $\sigma_x^2 = \mathrm{Var}(X_1)$, $\sigma_y^2 = \mathrm{Var}(Y_1)$, $\sigma_{xy} = \mathrm{Cov}(X_1, Y_1)$.

By the CLT,

$$\sqrt{n}(Z_n - \mu) \text{ converges in distribution to } N\left(0, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right)$$

By the delta method, $g(x,y) = x/y$, $\partial g/\partial x = y^{-1}$, $\partial g/\partial y = -xy^{-2}$

$$\sqrt{n}\left(\frac{\bar{X}_n}{\bar{Y}_n} - \frac{\mu_x}{\mu_y}\right) \text{ converges in distribution to } N(0, \sigma^2)$$

$$\sigma^2 = (\mu_y^{-1} \quad -\mu_x\mu_y^{-2}) \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \begin{pmatrix} \mu_y^{-1} \\ -\mu_x\mu_y^{-2} \end{pmatrix} = \frac{\sigma_x^2}{\mu_y^2} - \frac{2\mu_x\sigma_{xy}}{\mu_y^3} + \frac{\mu_x^2\sigma_y^2}{\mu_y^4}$$

Example 5.5.27 (more discussions)

What if $\mu_y = 0$ in the previous discussion?

Since the function $g(x,y) = x/y$ is not differentiable at $y = 0$, the delta method does not apply.

However, the early discussion says that $\bar{X}_n/\bar{Y}_n$ converges in distribution to $Cauchy(0,1)$ if $\mu_x = \mu_y = 0$.

By the delta method, we can similarly show that, if at least one of μ_x and μ_y is not 0, then

$$\sqrt{n}(\bar{X}_n\bar{Y}_n - \mu_x\mu_y) \text{ converges in distribution to } N(0,\sigma^2)$$

with

$$\sigma^2 = (\mu_y \;\; \mu_x)\begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix} = \mu_y^2\sigma_x^2 + 2\mu_x\mu_y\sigma_{xy} + \mu_x^2\sigma_y^2$$

If $\mu_x = \mu_y = 0$, then an application of the 2nd order delta method or continuous map theorem shows that

$$n\bar{X}_n\bar{Y}_n \text{ converges in distribution to } Z_1Z_2$$

where Z_1 and Z_2 are independent, $Z_1 \sim N(0,1)$, and $Z_2 \sim N(0,1)$.

Generating a random variable

We study how to generate random variables from a given distribution.

This may be useful in applications, or in statistical research when we carry out simulation studies, or to approximate integrals that do not have explicit forms.

Theorem 2.1.10

Let X have continuous cdf F_X and $Y = F_X(X)$. Then Y is uniformly distributed on $(0,1)$, that is, $P(Y \leq y) = y$, $0 < y < 1$.

Direct method

If we want to generate a random value X from a cdf F_X whose inverse F_X^{-1} exists, then we need only to generate a Y from the uniform distribution on $(0,1)$ and let $X = F_X^{-1}(Y)$.
Some cdf's may not have an inverse.

The inverse of a cdf

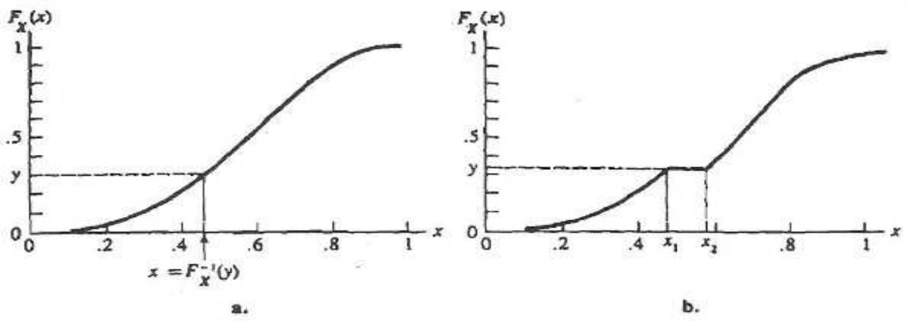
If F_X is strictly increasing, then its inverse F_X^{-1} is well defined by

$$F_X^{-1}(y) = x \qquad \text{iff} \qquad F_X(x) = y$$

If F_X is flat on an interval $(x_1, x_2]$ (see the figure), then we define

$$F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}, \qquad 0 \leq y \leq 1$$

- F_X^{-1} is increasing in y.
- $F_X(F_X^{-1}(y)) = y$ when F_X is continuous at $F_X^{-1}(y)$.
- $F_X^{-1}(y) \leq x$ iff $y \leq F_X(x)$.



In part b of the figure, $F_X^{-1}(y) = x_1$.

Proof of Theorem 2.1.10

For $0 < y < 1$,

$$\begin{aligned} P(Y \le y) &= P(F_X(X) \le y) \\ &= P(F_X^{-1}(F_X(X)) \le F_X^{-1}(y)) && (F_X^{-1} \text{ is increasing}) \\ &= P(X \le F_X^{-1}(y)) && (P(F_X^{-1}(F_X(X)) < X) = 0) \\ &= F_X(F_X^{-1}(y)) \\ &= y && (\text{continuity of } F_X) \end{aligned}$$

Probability integral transform

If U is uniformly distributed on $(0,1)$ and F is a cdf on $\mathscr{R}$, then the random variable $X = F^{-1}(U)$ has cdf F.

We use the property $F^{-1}(y) \le x$ iff $y \le F(x)$.

- If $y \le F(x)$, then $x \ge \inf\{t : F(t) \ge y\} = F^{-1}(y)$.
- If $F^{-1}(y) = \inf\{t : F(t) \ge y\} \le x$, then $F(F^{-1}(y)) \le F(x)$. But

$$F(F^{-1}(y)) \begin{array}{ll} = y & \text{if } F \text{ is continuous at } y \\ > y & \text{if } F \text{ has a jump at } y \end{array}$$

Thus, we must have $y \le F(F^{-1}(y)) \le F(x)$.

Then, for any $x \in \mathscr{R}$,

$$P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$$

Thus, if we can generate $U \sim uniform(0,1)$ (there are many algorithms for this), then $Y = F^{-1}(U)$ is a random variable from the desired cdf.

This method requires $F^{-1}(u) = \inf\{x : F(x) \ge u\}$ has an explicit form.

Applying the direct method

Suppose that we want to generate random variables from *exponential*$(0,2)$, which has cdf $F(x) = 1 - e^{-x/2}$, $x \ge 0$.
Then $F^{-1}(u) = -2\log(1-u)$, $0 < u < 1$.
Using $Y = -2\log(1-U)$ for the 10,000 generated U values, the authors obtained that the average of Y values is 2.0004 and the sample variance of Y values is 4.0908, close to $E(Y) = 2$ and $\text{Var}(Y) = 4$.

Indirect methods

When no simple transformation is available to apply a direct method, the following indirect method can often provide a solution.

Theorem 5.6.8 (Accept/Reject Algorithm).

Let f and g be pdf's with $M = \sup_{t\in\mathscr{R}} \frac{f(t)}{g(t)} < \infty$.
The following algorithm generates a random variable $Y \sim f$:

(a) generate independent $U \sim uniform(0,1)$ and $W \sim g$;

(b) if $U \le M^{-1} f(W)/g(W)$, set $Y = W$; otherwise, return to step (a).

Proof.

The generated Y satisfies

$$
\begin{aligned}
P(Y \le y) &= P\left(W \le y \middle| U \le M^{-1} f(W)/g(W)\right) \\
&= \frac{P\left(W \le y, U \le M^{-1} f(W)/g(W)\right)}{P\left(U \le M^{-1} f(W)/g(W)\right)} \\
&= \frac{\int_{-\infty}^{y} \int_{0}^{M^{-1} f(w)/g(w)} du g(w) dw}{\int_{-\infty}^{\infty} \int_{0}^{M^{-1} f(w)/g(w)} du g(w) dw} \\
&= \frac{\int_{-\infty}^{y} f(w) dw}{\int_{-\infty}^{\infty} f(w) dw} = \int_{-\infty}^{y} f(w) dw
\end{aligned}
$$

There are cases, however, where the target pdf f is unbounded and it is difficult to find a pdf g such that $M = \sup_{t \in \mathscr{R}} \frac{f(t)}{g(t)} < \infty$.
In such cases we can consider the following algorithm, which is a special case of the so-called Markov Chain Monte Carlo methods.

Metropolis Algorithm

Let f and g be pdf's. Consider the following algorithm:

(0) Generate $V \sim g$ and set $Z_0 = V$.
For $i = 1, 2, ...,$

(1) Generate independent $U_i \sim uniform(0,1)$ and $V_i \sim g$ and compute

$$\rho_i = \min\left\{ \frac{f(V_i)g(Z_{i-1})}{g(V_i)f(Z_{i-1})}, 1 \right\}$$

(2) Set

$$Z_i = \begin{cases} V_i & U_i \le \rho_i \\ Z_{i-1} & U_i > \rho_i \end{cases}$$

Then, as $i \to \infty$, Z_i converges in distribution to Y.