Lecture 14: Convergence of transformations, Slutsky's theorem and δ-method

Transformation is an important tool in statistics.
If X_n converges to X in some sense, is $g(X_n)$ converges to $g(X)$ in the same sense?
The following result (continuous mapping theorem) provides an answer to this question in many problems.

Theorem 1.10. Let $X, X_1, X_2, ...$ be random k-vectors defined on a probability space and g be a measurable function from $(\mathcal{R}^k, \mathcal{B}^k)$ to $(\mathcal{R}^l, \mathcal{B}^l)$. Suppose that g is continuous a.s. P_X. Then
(i) $X_n \to_{a.s.} X$ implies $g(X_n) \to_{a.s.} g(X)$;
(ii) $X_n \to_p X$ implies $g(X_n) \to_p g(X)$;
(iii) $X_n \to_d X$ implies $g(X_n) \to_d g(X)$.
Proof. (i) can be established using a result in calculus.
(iii) follows from Theorem 1.9(i): for any bounded and continuous h, $E[h(g(X_n))] \to E[h(g(X))]$, since $h \circ g$ is bounded and continuous.
To show (ii), we consider the special case of $X = c$ (a constant).
From the continuity of g, for any $\epsilon > 0$, there is a $\delta_\epsilon > 0$ such that $\|g(x) - g(c)\| < \epsilon$ whenever $\|x - c\| < \delta_\epsilon$. Hence,

$$\{\omega : \|g(X_n(\omega)) - g(c)\| < \epsilon\} \subset \{\omega : \|X_n(\omega) - c\| < \delta_\epsilon\}$$

and

$$P(\|g(X_n) - g(c)\| \geq \epsilon) \leq P(\|X_n - c\| \geq \delta_\epsilon).$$

Hence $g(X_n) \to_p g(c)$ follows from $X_n \to_p c$.
Is the previous arguement still valid when c is replaced by the random vector X in the general case? If not, how do we fix the proof?

Example 1.30. (i) Let $X_1, X_2, ...$ be random variables. If $X_n \to_d X$, where X has the $N(0,1)$ distribution, then $X_n^2 \to_d Y$, where Y has the chi-square distribution χ_1^2.
(ii) Let (X_n, Y_n) be random 2-vectors satisfying $(X_n, Y_n) \to_d (X, Y)$, where X and Y are independent random variables having the $N(0,1)$ distribution, then $X_n/Y_n \to_d X/Y$, which has the Cauchy distribution $C(0,1)$.
(iii) Under the conditions in part (ii), $\max\{X_n, Y_n\} \to_d \max\{X, Y\}$, which has the c.d.f. $[\Phi(x)]^2$ ($\Phi(x)$ is the c.d.f. of $N(0,1)$).

In Example 1.30(ii) and (iii), the condition that $(X_n, Y_n) \to_d (X, Y)$ cannot be relaxed to $X_n \to_d X$ and $Y_n \to_d Y$ (exercise); i.e., we need the convergence of the joint c.d.f. of (X_n, Y_n). This is different when $\to_d$ is replaced by $\to_p$ or $\to_{a.s.}$. The following result, which plays an important role in probability and statistics, establishes the convergence in distribution of $X_n + Y_n$ or $X_n Y_n$ when no information regarding the joint c.d.f. of (X_n, Y_n) is provided.

Theorem 1.11 (Slutsky's theorem). Let $X, X_1, X_2, ..., Y_1, Y_2, ...$ be random variables on a probability space. Suppose that $X_n \to_d X$ and $Y_n \to_p c$, where c is a constant. Then
(i) $X_n + Y_n \to_d X + c$;
(ii) $Y_n X_n \to_d cX$;
(iii) $X_n / Y_n \to_d X/c$ if $c \neq 0$.
Proof. We prove (i) only. The proofs of (ii) and (iii) are left as exercises.
Let $t \in \mathcal{R}$ and $\epsilon > 0$ be fixed constants. Then

$$\begin{aligned} F_{X_n+Y_n}(t) &= P(X_n + Y_n \le t) \\ &\le P(\{X_n + Y_n \le t\} \cap \{|Y_n - c| < \epsilon\}) + P(|Y_n - c| \ge \epsilon) \\ &\le P(X_n \le t - c + \epsilon) + P(|Y_n - c| \ge \epsilon) \end{aligned}$$

and, similarly,

$$F_{X_n+Y_n}(t) \ge P(X_n \le t - c - \epsilon) - P(|Y_n - c| \ge \epsilon).$$

If $t - c$, $t - c + \epsilon$, and $t - c - \epsilon$ are continuity points of F_X, then it follows from the previous two inequalities and the hypotheses of the theorem that

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

Since ϵ can be arbitrary (why?),

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

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

An application of Theorem 1.11 is given in the proof of the following important result.

Theorem 1.12. Let $X_1, X_2, ...$ and Y be random k-vectors satisfying

$$a_n(X_n - c) \to_d Y, \tag{1}$$

where $c \in \mathcal{R}^k$ and $\{a_n\}$ is a sequence of positive numbers with $\lim_{n\to\infty} a_n = \infty$. Let g be a function from $\mathcal{R}^k$ to $\mathcal{R}$.
(i) If g is differentiable at c, then

$$a_n[g(X_n) - g(c)] \to_d [\nabla g(c)]^\tau Y, \tag{2}$$

where $\nabla g(x)$ denotes the k-vector of partial derivatives of g at x.
(ii) 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

$$a_n^m[g(X_n) - g(c)] \to_d \frac{1}{m!} \sum_{i_1=1}^{k} \cdots \sum_{i_m=1}^{k} \left. \frac{\partial^m g}{\partial x_{i_1} \cdots \partial x_{i_m}} \right|_{x=c} Y_{i_1} \cdots Y_{i_m}, \tag{3}$$

where Y_j is the jth component of Y.

Proof. We prove (i) only. The proof of (ii) is similar. Let

$$Z_n = a_n[g(X_n) - g(c)] - a_n[\nabla g(c)]^\tau (X_n - c).$$

If we can show that $Z_n = o_p(1)$, then by (1), Theorem 1.9(iii), and Theorem 1.11(i), result (2) holds.

The differentiability of g at c implies that for any $\epsilon > 0$, there is a $\delta_\epsilon > 0$ such that

$$|g(x) - g(c) - [\nabla g(c)]^\tau (x - c)| \leq \epsilon \|x - c\| \tag{4}$$

whenever $\|x - c\| < \delta_\epsilon$. Let $\eta > 0$ be fixed. By (4),

$$P(|Z_n| \geq \eta) \leq P(\|X_n - c\| \geq \delta_\epsilon) + P(a_n \|X_n - c\| \geq \eta/\epsilon).$$

Since $a_n \to \infty$, (1) and Theorem 1.11(ii) imply $X_n \to_p c$. By Theorem 1.10(iii), (1) implies $a_n \|X_n - c\| \to_d \|Y\|$. Without loss of generality, we can assume that η/ϵ is a continuity point of $F_{\|Y\|}$. Then

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

The proof is complete 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 c.d.f. of $[\nabla g(c)]^\tau Y$, if (2) holds. Hence, result (2) is not useful for this purpose if $\nabla g(c) = 0$, and in such cases result (3) may be applied.

A useful method in statistics, called the *delta-method*, is based on the following corollary of Theorem 1.12.

Corollary 1.1. Assume the conditions of Theorem 1.12. If Y has the $N_k(0, \Sigma)$ distribution, then

$$a_n[g(X_n) - g(c)] \to_d N\left(0, [\nabla g(c)]^\tau \Sigma \nabla g(c)\right).$$

Example 1.31. Let $\{X_n\}$ be a sequence of random variables satisfying $\sqrt{n}(X_n - c) \to_d N(0, 1)$. Consider the function $g(x) = x^2$. If $c \neq 0$, then an application of Corollary 1.1 gives that $\sqrt{n}(X_n^2 - c^2) \to_d N(0, 4c^2)$. If $c = 0$, the first-order derivative of g at 0 is 0 but the second-order derivative of $g \equiv 2$. Hence, an application of result (3) gives that $nX_n^2 \to_d [N(0, 1)]^2$, which has the chi-square distribution χ_1^2 (Example 1.14). The last result can also be obtained by applying Theorem 1.10(iii).