THE DELTA-METHOD AND ASYMPTOTICS OF SOME ESTIMATORS

The delta-method gives a way that asymptotic normality can be preserved under nonlinear, but differentiable, transformations. The method is well known; one version of it is given in J. Rice, *Mathematical Statistics and Data Analysis*, 2d. ed., 1995. A simple form of it using only a first derivative, for functions of one variable, will be given here. (A multidimensional version is used in Section 3.7 of Mathematical Statistics, 18.466 course notes by R. Dudley, on the MIT OCW website.)

Theorem. Let Y_n be a sequence of real-valued random variables such that for some μ and σ, $\sqrt{n}(Y_n - \mu)$ converges in distribution as $n \to \infty$ to $N(0, \sigma^2)$. Let f be a function from $\mathbb{R}$ into $\mathbb{R}$ having a derivative $f'(\mu)$ at μ. Then $\sqrt{n}[f(Y_n) - f(\mu)]$ converges in distribution as $n \to \infty$ to $N(0, f'(\mu)^2\sigma^2)$.

Remarks. In statistics, where μ is an unknown parameter, one will want f to be differentiable at all possible μ (and preferably, for f' to be continuous, although that is not needed in the proof).

Proof. We have $Y_n - \mu = O_p(1/\sqrt{n})$ as $n \to \infty$. Also, $f(y) = f(\mu) + f'(\mu)(y-\mu) + o(|y-\mu|)$ as $y \to \mu$ by definition of derivative. Thus

$$f(Y_n) = f(\mu) + f'(\mu)(Y_n - \mu) + o_p(|Y_n - \mu|),$$

so

$$\sqrt{n}[f(Y_n) - f(\mu)] = f'(\mu)\sqrt{n}(Y_n - \mu) + \sqrt{n}o_p(1/\sqrt{n}).$$

The last term is $o_p(1)$, so the conclusion follows. □

Let's say a distribution function F has a *good median* if F has a continuous density $F' = f$ with $f(m) > 0$ at m, the median of F. More precisely, $f(m) > 0$ and f continuous at m imply that F is strictly increasing in a neighborhood of m, so m is the unique x with $F(x) = 1/2$ and so the unique median. Let's find the asymptotic distribution of the sample median. First let $n = 2k + 1$ odd, so the nth sample median $m_n = X_{(k+1)}$. If F is the $U[0, 1]$ distribution, let its order statistics be $U_{(1)} < \cdots < U_{(n)}$. Recall that $U_{(j)}$ has a beta distribution $\beta_{j,n-j+1}$ for each j, so the sample median $U_{(k+1)}$ has a $\beta_{k+1,k+1}$ distribution. Its density is $x^k(1-x)^k/B(k+1, k+1)$ for $0 \le x \le 1$ and 0 elsewhere. The distribution has mean 1/2 and variance $1/[4(2k + 3)] = 1/[4(n + 2)]$.

This beta distribution is asymptotically normal with its mean and variance as $n \to \infty$ or equivalently $k \to \infty$. This fact is a special case of facts known since about 1920, but lacking a handy reference, I'll indicate a proof. Let $y = x - (1/2)$, so $|y| \le 1/2$ where the density is non-zero. On that interval,

$$x^k(1-x)^k = \left(\frac{1}{2} + y\right)^k \left(\frac{1}{2} - y\right)^k = \left(\frac{1}{4} - y^2\right)^k = 4^{-k}(1 - 4y^2)^k.$$

We have $(1 - 4y^2)^k \le \exp(-4ky^2)$ for all y with $|y| \le 1/2$, and for any constant c and $|y| \le c/\sqrt{k}$, $k\log(1-4y^2)+4ky^2 = O(k(4y^2)^2) = O(1/k) = O(1/n)$ as $n \to \infty$ and $k \to \infty$, so for such y (depending on k), $(1-4y^2)^k$ is asymptotic to $\exp(-4ky^2)$. It follows that $\beta_{k+1,k+1}$ is asymptotically normal with mean $1/2$ and variance $1/(8k)$ which is asymptotic to $1/(4n)$. In other words $\sqrt{n}[U_{(k+1)} - \frac{1}{2}]$ converges in distribution as $n \to \infty$ to $N(0, 1/4)$.

Now for any distribution function F with a good median m, and $n = 2k+1$ odd, the sample median $m_n = X_{(k+1)}$ has the distribution of $F^{\leftarrow}(U_{(k+1)})$ because $F^{\leftarrow}$ is monotonic (non-decreasing, and strictly increasing in a neighborhood of $\frac{1}{2}$). We have $F^{\leftarrow}(1/2) = m$. So by the delta-method theorem above, $\sqrt{n}(m_n - m)$, being equal in distribution to $\sqrt{n}(F^{\leftarrow}(U_{(k+1)}) - F^{\leftarrow}(1/2))$, converges in distribution as $n \to \infty$ to $N(0, (F^{\leftarrow})'(1/2)^2/4) = N(0, 1/(4f(m)^2))$, as stated in Randles and Wolfe, p. 227, line 2, for symmetric distributions.

For $n = 2k$ even, $U_{(k)}$ and $U_{(k+1)}$ have $\beta_{k,k+1}$ and $\beta_{k+1,k}$ distributions respectively, and $|U_{(k+1)} - U_{(k)}| = O_p(1/n)$. For the sample median $m_{U,n} = [U_{(k)} + U_{(k+1)}]/2$, we then also have $|m_{U,n} - U_{(k)}| = O_p(1/n)$. By a small adaptation of the argument for the n odd case, we get that $\sqrt{n}(U_{(k)} - \frac{1}{2})$ converges in distribution to $N(0, 1/4)$ as $n = 2k \to \infty$, and so does $\sqrt{n}(m_{U,n} - \frac{1}{2})$. So, for a distribution F with a good median m and sample medians m_n, we get $\sqrt{n}(m_n - m)$ converging in distribution as $n \to \infty$ to $N(0, 1/(4f(m)^2))$, just as when n is odd and as stated by Randles and Wolfe.

Next, let's consider the Hodges-Lehmann estimator. In this case, beside assuming F has a good median m, we'll assume the distribution is symmetric around m. (If a distribution is symmetric around a point θ, then θ must be the median.) In other words, there is a density f_0 with $f_0(-x) = f_0(x)$ for all x, $f_0(0) > 0$, f_0 is continuous at 0, and the density f is $f_m(x) \equiv f_0(x-m)$, which is then symmetric around m. Given $X_1, ..., X_n$ i.i.d. with a distribution F satisfying the given conditions, but otherwise unknown, the Hodges-Lehmann estimator $\hat{\theta}_{HL}$ is the median of the numbers $(X_i + X_j)/2$ for $1 \le i \le j \le n$. There are $n(n+1)/2$ of these numbers (which are called *Walsh averages*). The sample median is an estimator of the unknown m, and $\hat{\theta}_{HL}$ is another which is often better. To look into it we'll consider some U-statistics. For any real x, x_1, and x_2 let $h_x(x_1, x_2) = \Psi(2x - x_1 - x_2)$. This kernel is symmetric under interchanging x_1 and x_2 for each x.

We want to find the asymptotic behavior of $\hat{\theta}_{HL} - m$, specifically, that it's asymptotically normal with mean 0 and variance C/n for some C depending on F. In doing this, we can assume $m = 0$, because subtracting m from all the observations makes $m = 0$ and doesn't change the distribution of the difference. So we can assume F is symmetric around 0.

Let G be the distribution function of $X_1 + X_2$. Then G has a density g given by the convolution of f with itself, $g(x) = \int_{-\infty}^{\infty} f(x-y)f(y)dy$. We have for all x

$$Eh_x(X_1, X_2) = P(X_1 + X_2 < 2x) = G(2x).$$

The quantity called ζ_1, entering into the asymptotic variance of the U-statistic formed from the kernel h_x, is given by

$$\zeta_1 = P(X_1 + X_2 < 2x, X_1 + X_3 < 2x) - G(2x)^2.$$

We are interested especially in $x = 0$ since that is now the median and center of symmetry of F and of G. For $x = 0$ we get

$$P(X_1 + X_2 < 0, X_1 + X_3 < 0) = \int_{-\infty}^{\infty} F(-u)^2 dF(u) =$$

$$\int_{-\infty}^{\infty}[1-F(u)]^2 dF(u) = \int_0^1 (1-t)^2 dt = 1/3,$$

and $Eh_0 = 1/2$, so $\zeta_1 = 1/12$. We have a kernel of order $r = 2$, and the asymptotic variance of a U-statistic is $r^2\zeta_1$. Defining a U-statistic depending on x we have

$$U_{(x)}^{(n)} = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \Psi(x - X_i - X_j).$$

For $x = 0$, bearing in mind that under symmetry around 0, $-X_i - X_j$ is equal in distribution to $X_i + X_j$, this becomes the U-statistic that Randles and Wolfe call U_4 and is closely related to the Wilcoxon signed-rank statistic. We get that $\sqrt{n}(U_{(x=0)}^{(n)} - \frac{1}{2})$ converges in distribution as $n \to \infty$ to $N(0, 1/3)$.

If we included all the terms with $i = j$ in the sum defining the U-statistic, giving another statistic $V^{(n)}$, it would make a difference of $O(n)$ in the sum, thus $O(1/n)$ in $U^{(n)}$, thus $O(1/\sqrt{n})$ in $\sqrt{n}U^{(n)}$, so $\sqrt{n}(V^{(n)} - \frac{1}{2})$ also has a distribution converging to $N(0, 1/3)$. In other words, $V^{(n)} = \frac{1}{2} + Z_n/\sqrt{3n} + o_p(1/\sqrt{n})$ where Z_n converges in distribution to $N(0, 1)$ as $n \to \infty$.

The Hodges-Lehmann estimate $\hat{\theta}_{HL}$ is an x for which $V_{(x)}^{(n)} = \frac{1}{2} + O(1/n^2)$. For x near 0, specifically $|x| = O(1/\sqrt{n})$, $Eh_x = G(2x)$ which will be within $O(1/\sqrt{n})$ of 1/2. The asymptotic variance of $V_{(x)}^{(n)}$ will still be $1/(3n)$ plus smaller terms that don't affect the asymptotic distribution. So we will have, where again Z_n is asymptotically $N(0, 1)$,

$$V_{(x)}^{(n)} = G(2x) + Z_n/\sqrt{3n} + o_p(1/\sqrt{n}).$$

If this equals 1/2 (within $O(1/n^2)$), then

$$\hat{\theta}_{HL} = x = \frac{1}{2}G^{\leftarrow}\left(\frac{1}{2} - (Z_n/\sqrt{3n})\right) + o_p(1/\sqrt{n}).$$

It follows by the delta-method that the distribution of $\sqrt{n}(\hat{\theta}_{HL} - m) = \sqrt{n}\hat{\theta}_{HL}$ converges to $N(0, \sigma^2)$ where

$$\sigma^2 = (G^{\leftarrow})'(1/2)^2/12 = 1/(12G'(0)^2) = 1/(12g(0)^2),$$

and by convolution $g(0) = \int_{-\infty}^{\infty} f(0-x)f(x)dx = \int_{-\infty}^{\infty} f(x)^2 dx$ by symmetry. So the asymptotic variance of the Hodges-Lehmann statistic is $1/[12n\{\int_{-\infty}^{\infty} f(x)^2\,dx\}^2]$, as indicated by Randles and Wolfe on p. 228, (7.3.12) and (7.3.14).

Note. We considered a family of U-statistics indexed by a parameter x. There is a theory of such families, called U-processes, begun in some papers by Deborah Nolan and David Pollard in Annals of Statistics. In the present case, since $U_{(x)}^{(n)}$ is non-decreasing in x, we have a relatively simple U-process, but still, the argument was incomplete.