U-statistics

Xiaohui Chen

Keywords: *random sample, sample variance, Cramér-von Mises, Kaplan-Meier, Nelson-Aalen, energy statistic, asymptotics, bootstrap.*

Abstract

Abstract: A U-statistic, calculated from a random sample of size n, is an average of a symmetric function calculated for all m-tuples in the sample. Examples include the sample variance, the Cramér-von Mises and energy statistics of goodness-of-fit, and the Kaplan-Meier and Nelson-Aalen estimators in survival analysis. Asymptotic properties are described.

1 Introduction and Examples

Given a **random sample** ⟨stat05945⟩ (a sequence of independent and identically distributed **random variables** ⟨stat04404⟩ $X_1, \ldots, X_n$ with common **distribution function** ⟨stat07524⟩ F), the study of the statistical properties of the **sample mean** ⟨stat00541⟩, $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$, is a well-established part of **probability theory** ⟨stat03979⟩. The notion of averaging over the observations has been generalized by **Hoeffding** ⟨stat01309⟩ [14] in the following way: given a **measurable** ⟨stat02290⟩ real-valued function h, symmetric in its m arguments, a *U-statistic* is obtained by averaging the outcomes $h(X_{i_1}, \ldots, X_{i_m})$ over all possible *ordered* m-tuples $I_{n,m} = \{(i_1, \ldots, i_m) : 1 \le i_1 < \ldots < i_m \le n\}$, i.e.,

$$U_n = \binom{n}{m}^{-1} \sum_{(i_1,\ldots,i_m)\in I_{n,m}} h(X_{i_1}, \ldots, X_{i_m}).$$

Then U_n is called a U-statistic with *kernel* h of *degree* m. We assume, of course, that $n \ge m$. Many statistics in estimation and testing theory can be represented as U-statistics. We give three examples.

University of Illinois at Urbana-Champaign, Illinois, USA. Email: xhchen@illinois.edu

The author would like to acknowledge a preliminary version of this article by Paul Janssen.

Example 1. *Assume* $0 < \sigma^2 = var(X_1) < \infty$*. The sample variance* $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$*, the* ***minimum variance unbiased estimator*** *⟨stat05910⟩ for* σ^2*, can be rewritten as*

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

Therefore, the sample variance is a U*-statistic with kernel* $h(x, y) = (x-y)^2/2$*. In general, we have that the minimum variance unbiased estimator of the* m*-th central* ***moment*** *⟨stat05913⟩ is a* U*-statistic with kernel of degree* m*. See, for example, Hoeffding [14, p. 295] and Serfling [21, p. 176] for details.*

Example 2. *The* ***Cramér-von Mises statistic*** *⟨stat01467⟩, a* ***goodness-of-fit*** *⟨stat05753⟩ statistic to test if the unknown distribution function* F *equals some specified distribution function* F_0*, is given by*

$$V_n = \int_{-\infty}^{+\infty} [F_n(x) - F_0(x)]^2 \mathrm{d}F_0(x),$$

where $F_n(x) = n^{-1}\sum_{i=1}^n \mathrm{I}\{X_i \le x\}$ *is the* ***empirical distribution function*** *⟨stat02712⟩ of the sample* $X_1, \ldots, X_n$*. Then we can write* $V_n = n^{-2}\sum_{i=1}^n \sum_{j=1}^n h(X_i, X_j)$ *as the* V*-statistic associated with the kernel*

$$h(x, y) = \int_{-\infty}^{+\infty} [\mathrm{I}\{x \le t\} - F_0(t)][\mathrm{I}\{y \le t\} - F_0(t)]\mathrm{d}F_0(t).$$

An asymptotically equivalent statistic is the U*-statistic*

$$U_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h(X_i, X_j).$$

See de Wet [7] for a detailed discussion.

Example 3. *The Cramér-von Mises statistic is not rotation invariant for multivariate distributions. Suppose that* X *and* Y *are independent random vectors in* $\mathbb{R}^p$ *such that* $E|X| < \infty$ *and* $E|Y| < \infty$*, where* $|\cdot|$ *is the Euclidean norm of* $\mathbb{R}^p$*. The energy distance between* X *and* Y *is defined as*

$$\mathcal{E}(X, Y) = 2E|X - Y| - E|X - X'| - E|Y - Y'|,$$

where X' *(or* Y'*) is an independent copy of* X *(or* Y*) [26]. It is known that* $\mathcal{E}(X, Y) \ge 0$*, where the equality is attained if and only if* X *and* Y *have the same distribution. Given a random sample* $X_1, \ldots, X_n$ *with an unknown distribution function* F*, a goodness-of-fit test for* $H_0 : F = F_0$ *based on the energy statistic is given by*

$$\mathcal{E}_n = \frac{2}{n}\sum_{i=1}^n E_Y|X_i - Y| - E|Y - Y'| - \frac{1}{n^2}\sum_{i,j=1}^n |X_i - X_j|,$$

where Y and Y' are iid with distribution F_0 (also independent of $X_1, \ldots, X_n$), and E_Y is the expectation taken with respect to Y. It is clear that $\mathcal{E}_n$ is a V-statistic, which is asymptotically equivalent to the unbiased U-statistic with the kernel

$$h(x, y) = E|x - Y| + E|y - Y'| - E|Y - Y'| - |x - y|.$$

For all examples above. we have that the **parameter** ⟨stat00676⟩ of interest is of the form

$$\theta(F) = Eh(X_1, X_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x, y) \mathrm{d}F(x) \mathrm{d}F(y).$$

With h as in Example 1 we have $\theta(F) = \sigma^2$. The goodness-of-fit parameter in Example 2 is $\theta(F) = \int_{-\infty}^{+\infty} [F(x) - F_0(x)]^2 \mathrm{d}F_0(x)$ and in Example 3 is

$$\theta(F) = 2 \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} |x-y| \mathrm{d}F(x) \mathrm{d}F_0(y) - \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} |y-y'| \mathrm{d}F_0(y) \mathrm{d}F_0(y') - \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} |x-x'| \mathrm{d}F(x) \mathrm{d}F(x').$$

Under the null hypothesis $H_0 : F = F_0$, we have $\theta(F_0) = 0$ in both cases. If, in general, a real-valued functional θ defined on a set $\mathcal{F}$ of distribution functions can be written as the expectation with respect to $F \in \mathcal{F}$ of a properly chosen kernel h of degree m, the functional θ is called a *regular functional*. Such functionals have U-statistics as minimum variance unbiased estimators. For more details, we refer to the book by Lee [19, Chapter 1], which includes a variety of further examples (Chapter 6).

Note that a naive estimator for $\theta(F)$ can be obtained by the plug-in method (replace F by F_n), i.e., use $\theta(F_n)$ as an estimator for $\theta(F)$. The resulting (biased) estimator is the von Mises statistic. The goodness-of-fit statistics, V_n in Example 2 and $\mathcal{E}_n$ in Example 3, are plug-in estimators. U-statistics and von Mises statistics are closely related.

A U-statistic with kernel of degree m can be written in terms of uncorrelated U-statistics of degree $1, \ldots, m$. Indeed, the *Hoeffding decomposition* (due to Hoeffding [15]) is given by

$$U_n - \theta(F) = \sum_{c=1}^{m} \binom{m}{c} U_n^{(c)},$$

where

$$U_n^{(c)} = \binom{n}{c}^{-1} S_n^{(c)} := \binom{n}{c}^{-1} \sum_{(i_1, \ldots, i_c) \in I_{n,c}} h_c(X_{i_1}, \ldots, X_{i_c})$$

and

$$\begin{aligned} h_c(x_1, \ldots, x_c) &= (\delta_{x_1} - F) \cdots (\delta_{x_c} - F) F^{m-c} h \\ &= \int \cdots \int h(u_1, \ldots, u_m) \prod_{i=1}^{c} (\mathrm{d}\delta_{x_i}(u_i) - \mathrm{d}F(u_i)) \prod_{i=c+1}^{m} \mathrm{d}F(u_i). \end{aligned}$$

Here, δ_x is the **Dirac delta function** ⟨stat02228⟩. See [19, Section 1.6] or [6, Section 3.5] for further discussions. Other important structural properties are the *forward martingale* *structure* of $\{S_n^{(c)}, \mathcal{F}_n\}_{n \geq c}$ with $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, and the *reverse martingale structure* of $\{U_n, \tilde{\mathcal{F}}_n\}_{n \geq m}$ with $\tilde{\mathcal{F}}_n = \sigma(X_{(1):n}, \ldots, X_{(n):n}, X_{n+1}, X_{n+2}, \ldots)$ and $X_{(i):n}$ the i-th order statistic of $X_1, \ldots, X_n$ [19, Section 3.4] (see the discussion of **martingales** ⟨stat02941⟩ in the entry on **Counting Process Methods in Survival Analysis** ⟨stat06009⟩).

So far we have demonstrated that many statistics are U-statistics and we have discussed some structural properties. It is also highly relevant that U-statistics appear as terms in stochastic approximations of smooth statistics. U-statistics are, for example, extremely useful to approximate important estimators in nonparametric **density estimation** ⟨stat05843⟩ and **nonparametric regression** ⟨stat05768⟩ theory (see [13] and [20]) and **survival analysis** ⟨stat06060⟩ (see [5]). The basic idea is that the estimator of interest can be approximated by a sum of uncorrelated U-statistics. This idea is closely related to the Hoeffding decomposition of a U-statistic (see [19, Section 4.1] and [9]) and to von Mises expansions, a generalization of the projection method (a technique discussed in more detail in Section 2). For further reading we refer to [21, Chapter 6] and [10].

A more detailed discussion would require a number of technical concepts and definitions. We therefore restrict ourselves to one illustration.

Example 4. *Let $T_1, \ldots, T_n$ denote iid nonnegative survival times with a continuous distribution function F and let $C_1, \ldots, C_n$ denote iid nonnegative censoring times with a continuous distribution function G. For $i = 1, \ldots, n$, we denote $X_i = min(T_i, C_i)$ and $\delta_i = I\{T_i \leq C_i\}$. Let $\hat{F}_n(t)$ denote the product-limit or **Kaplan-Meier estimator** ⟨stat06033⟩ for $F(t)$. With $\hat{\Lambda}_n(t)$ the **Nelson-Aalen estimator** ⟨stat06045⟩ and $\Lambda(t)$ the cumulative **hazard function** ⟨stat04288⟩, a U-statistic representation has been established in [5] for $\hat{\Lambda}_n(t) - \Lambda(t)$. On the basis of the relation*

$$\hat{F}_n(t) - F(t) = \exp[-\Lambda(t)] \times \{1 - \exp[-(\Lambda_n(t) - \Lambda(t)]\}$$

*and using **Taylor expansion** ⟨stat00778⟩ ideas, a U-statistic representation for the Kaplan-Meier estimator can be obtained.*

2 Asymptotic Properties

A basic contribution to the study of the asymptotic behavior of U-statistics (*see* **Large-sample Theory** ⟨stat05876⟩) is the following result.

Theorem 1. *If $E|h(X_1, \ldots, X_m)| < \infty$, then $U_n \to \theta(F)$ almost surely (a.s.).*

This theorem states that the classical strong **law of large numbers** ⟨stat05877⟩ for the sample mean generalizes to U-statistics. Various proofs are available. They rely on the martingale structure of U-statistics mentioned above. For full proofs and references to the original papers, see Lee [19, Section 3.4].

Next, we briefly discuss the asymptotic distribution theory for U-statistics. The limit distribution of a (properly standardized) U-statistic will be Gaussian if we can obtain a stochastic approximation $\hat{U}_n$ of iid structure that is close to U_n (in the sense that U_n inherits the asymptotic distributional behavior of $\hat{U}_n$). The appropriate approximation is obtained from the projection technique, which is in fact the first term in the Hoeffding decomposition. We have

$$\hat{U}_n = \sum_{i=1}^{n} E(U_n|X_i) - (n-1)\theta(F).$$

With

$$h_1(x) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(x, x_2, \ldots, x_m) \mathrm{d}F(x_2) \cdots \mathrm{d}F(x_m) - \theta(F)$$

we can write

$$\hat{U}_n - \theta(F) = \frac{m}{n} \sum_{i=1}^{n} h_1(X_i).$$

If $h_1 \equiv 0$, then the U-statistic is said to be *degenerate* (of order 1); otherwise, the U-statistic is *nondegenerate*. Degenerate U-statistics do not admit an iid approximation, and as a consequence the limit distribution is not Gaussian. For nondegenerate U-statistics the following central limit result is valid.

Theorem 2. *[14]. If $Eh^2(X_1, \ldots, X_m) < \infty$ and $\zeta_1 = Var(h_1(X_1)) > 0$ (i.e., U_n is nondegenerate), then*

$$\frac{\sqrt{n}[U_n - \theta(F)]}{(m\zeta_1^{1/2})} \xrightarrow{d} Z,$$

where Z is a standard ***normal*** *⟨stat01090⟩ random variable.*

A simple calculation shows that

$$\zeta_1 = E\{[h(X_1, X_2, \ldots, X_m) - \theta(F)][h(X_1, X_{m+1}, \ldots, X_{2m-1}) - \theta(F)]\}.$$

For a degenerate U-statistic (i.e., the first term in the Hoeffding decomposition vanishes and $\zeta_1 = 0$) with $\zeta_2 = E\{[h(X_1, X_2, X_3, \ldots, X_m) - \theta(F)][h(X_1, X_2, X_{m+1}, \ldots, X_{2m-2}) - \theta(F)]\} > 0$, we have

$$U_n - \theta(F) = \tfrac{m(m-1)}{n(n-1)} \sum_{1 \le i < j \le n} h_2(X_i, X_j) + \sum_{c=3}^{m} \tbinom{m}{c} U_n^{(c)}.$$

For h_2, define the integral operator

$$Az(x) = \int_{-\infty}^{+\infty} h_2(x, y) z(y) \mathrm{d}F(y),$$

where z is square integrable with respect to F. Let λ_1, λ_2, … denote the real (not necessarily distinct) eigenvalues corresponding to the distinct solutions z_1, z_2, … of the equation $Az = \lambda z$.

Theorem 3. *[12]. If $E[h^2(X_1, \ldots, X_m)] < \infty$ and $\zeta_1 = 0 < \zeta_2$, then*

$$n[U_n - \theta(F)] \xrightarrow{d} \frac{m(m-1)}{2} Y,$$

where Y is a random variable of the form $Y = \sum_{j=1}^{\infty} \lambda_j [\chi_j^2(1) - 1]$, where $\chi_1^2(1), \chi_2^2(1), \ldots$ are independent $\chi^2(1)$ ⟨stat00936⟩ random variables (see ***Convergence in Distribution and in Probability*** *⟨stat02847⟩).*

Example 5. *For the sample variance an application of Theorem 5 yields (with μ_k the k-th central moment): if $\mu_4 < \infty$ and $\mu_4 - \mu_2^2 > 0$, then $\sqrt{n}(S_n^2 - \mu_2)$ has a limiting normal distribution with mean zero and variance $\mu_4 - \mu_2^2$.*

Example 6. *Under the null hypothesis $F = F_0$, the Cramér-von Mises statistic is a degenerate U-statistic. Then Theorem 6 holds with the eigenvalues $\lambda_j = (j\pi)^{-2}$. See [7] for details.*

3 Remarks and Extensions

1. For U-statistics with a kernel of degree $m > 2$, higher-order terms in the Hoeffding decomposition might vanish (i.e., higher-order degeneracy). Asymptotic distribution theory has been established in the literature. The resulting limit distributions are characterized in terms of multiple Wiener integrals [8].

2. We reviewed some basic results for one-sample U-statistics. Extensions to multi-sample or generalized U-statistics are available. See the books by Lee [19], Koroljuk & Borovskich [18] and Borovskikh [4] for details. These books also deal with other variations on the theme: incomplete U-statistics, random U-statistics, weighted U-statistics, generalized L-statistics, **Edgeworth expansions** ⟨stat05844⟩ for U-statistics, among many others.

3. **Bootstrap** ⟨stat02662⟩ theory for U-statistics is reviewed in Janssen [17]. Bickel & Freedman [3] is a basic reference.

4. A further important topic, especially for applications in nonparametric density and regression estimation, is the study of U-statistics with the kernel depending on the sample size n. Key references are Jammalamadaka & Janson [16] and Mammen [20]. We also mention the work by Frees [11] on infinite order U-statistics.

5. In Serfling [22] the study of U-processes and U-quantiles is initiated. Important contributions on U-processes and U-quantiles include Arcones & Giné [2], Stute [23], and Arcones [1]. Keywords in the development of new results for U-processes are martingales and decoupling. For details we refer to the book by de la Peña & Giné [6].

6. Non-asymptotic rates of convergence of the Gaussian and bootstrap approximations for multivariate U-statistics (of degree 2) in high dimensions are derived in Chen [24]. Computational and statistical trade-off for distributional approximations of high-dimensional U-statistics can be found in Chen & Kato [25].

References

[1] Arcones, M. (1995). The asymptotic accuracy of the bootstrap U-quantiles, *Annals of Statistics* **23**, 1802 – 1822.

[2] Arcones, M. & Giné, E. (1993). Limit theorems for U-processes, *Annals of Probability* **21**, 1494 – 1542.

[3] Bickel, P. & Freedman, D. (1981). Some asymptotic theory for the bootstrap, *Annals of Statistics* **9**, 1196 – 1217.

[4] Borovskikh, Yu. V. (1996). *U-Statistics in Banach Spaces*. VSP, Utrecht.

[5] Chang, M. N. & Rao, P. V. (1989). Berry-Esseen bound for the Kaplan-Meier estimator, *Communications in Statistics – Theory and Methods* **18**, 4647 – 4664.

[6] de la Peña, V. & Giné, E. (1998). *An Introduction to Decoupling Inequalities with Applications*. Springer-Verlag, New York.

[7] de Wet, T. (1987). Degenerate U- and V-statistics, *South African Statistical Journal* **21**, 99 – 129.

[8] Dynkin, E. B. & Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals, *Annals of Statistics* **11**, 739 – 745.

[9] Efron, B. & Stein, C. (1981). The jackknife estimate of variance, *Annals of Statistics* **9**, 586 – 596.

[10] Fernholz, L. (1983). *Von Mises Calculus for Statistical Functionals*. Springer-Verlag, New York.

[11] Frees, E. W. (1989). Infinite order U-statistics, *Scandinavian Journal of Statistics* **16**, 29 – 45.

[12] Gregory, G. (1977). Large sample theory for U-statistics and tests of fit, *Annals of Statistics* **5**, 110 – 123.

[13] Härdle, W. & Stoker, T. (1989). Investigating smooth multiple regression by the method of average derivatives, *Journal of the American Statistical Association* **84**, 986 – 995.

[14] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics* **19**, 293 – 325.

[15] Hoeffding, W. (1961). The strong law of large numbers for U-statistics, *University of North Carolina Institute of Statistics Mimeo Series No. 302.*

[16] Jammalamadaka, S. R. & Janson, S. (1986). Limit theorems for a triangular scheme of U-statistics with applications to interpoint distances, *Annals of Probability* **14**, 1347 – 1358.

[17] Janssen, P. (1997). Bootstrapping U-statistics. *South African Statistical Journal*, to appear.

[18] Koroljuk, V. S. & Borovskich, Yu. V. (1994). *Theory of U-Statistics.* Kluwer Academic Publishers, Dordrecht.

[19] Lee, A. J. (1990). *U-Statistics.* Marcel Dekker, New York.

[20] Mammen, E. (1992). *When Does the Bootstrap Work?* Springer-Verlag, New York.

[21] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[22] Serfling, R. (1984). Generalized $L-$,M- and R-statistics, *Annals of Statistics* **12**, 76 – 86.

[23] Stute, W. (1994). U-statistic processes: a martingale approach, *Annals of Probability* **22**, 1725 – 1744.

[24] Chen, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Annals of Statistics* **46**, 642 – 678.

[25] Chen, X. & Kato, K. (2018+). Randomized incomplete U-statistics in high dimensions. *Annals of Statistics (accepted for publication).*

[26] Székely, Gábor J. and Rizzo, Maria L. (2017). The energy of data. *Annual Review of Statistics and Its Application* **4**, 447 – 479.