**U-statistics**

**Xiaohui Chen**

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

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

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

Then $U_n$ is called a $U$-statistic with kernel $h$ of degree $m$. We assume, of course, that $n \geq m$. 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.

This article was originally published online in 2005 in Encyclopedia of Biostatistics, © John Wiley & Sons, Ltd and republished in Wiley StatsRef: Statistics Reference Online, 2014.
**Example 1.** Assume $0 < \sigma^2 = \text{var}(X_1) < \infty$. The sample variance $S_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$, the **minimum variance unbiased estimator** \(\langle\text{stat05910}\rangle\) for $\sigma^2$, can be rewritten as

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

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 \(\langle\text{stat05913}\rangle\) 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** \(\langle\text{stat01467}\rangle\), a goodness-of-fit \(\langle\text{stat05753}\rangle\) statistic to test if the unknown distribution function $F$ equals some specified distribution function $F_0$, is given by

$$V_n = \int_{-\infty}^{+\infty} [F_n(x) - F_0(x)]^2 dF_0(x),$$ 

where $F_n(x) = n^{-1} \sum_{i=1}^{n} I\{X_i \leq x\}$ is the **empirical distribution function** \(\langle\text{stat02712}\rangle\) 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

$$h(x, y) = \int_{-\infty}^{+\infty} [I\{x \leq t\} - F_0(t)][I\{y \leq t\} - F_0(t)] dF_0(t).$$

An asymptotically equivalent statistic is the $U$-statistic

$$U_n = \left(\frac{n}{2}\right)^{-1} \sum_{1 \leq i < j \leq n} h(X_i, X_j).$$

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 $\mathcal{E}(X, Y) \geq 0$, 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

$$\mathcal{E}_n = 2 \frac{n}{n} \sum_{i=1}^{n} E_Y |X_i - Y| - E|Y - Y'| - \frac{1}{n^2} \sum_{i,j=1}^{n} |X_i - X_j|,$$
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 $\langle \text{stat00676} \rangle$ of interest is of the form

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

With $h$ as in Example 1 we have $\theta(F) = \sigma^2$. The goodness-of-fit parameter in Example 2 is $\theta(F) = \int_{-\infty}^{+\infty} [F(x) - F_0(x)]^2 dF_0(x)$ and in Example 3 is

$$\theta(F) = 2 \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} |x-y| dF_0(x) dF_0(y) - \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} |y-y'| dF_0(y) dF_0(y') - \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} |x-x'| dF(x) dF(x').$$

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 $\theta$ defined on a set $F$ of distribution functions can be written as the expectation with respect to $F \in F$ of a properly chosen kernel $h$ of degree $m$, the functional $\theta$ 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

$$U_n^{(c)} = \left( \begin{array}{c} n \\ c \end{array} \right)^{-1} S_n^{(c)} := \left( \begin{array}{c} n \\ c \end{array} \right)^{-1} \sum_{(i_1, \ldots, i_c) \in I_{n,c}} h_c(x_{i_1}, \ldots, x_{i_c})$$

and

$$h_c(x_1, \ldots, x_c) = (\delta_{x_1} - F) \cdots (\delta_{x_c} - F) F_0^{m-c} h$$

$$= \int \cdots \int h(u_1, \ldots, u_m) \prod_{i=1}^{c} (d\delta_{x_i}(u_i) - dF(u_i)) \prod_{i=c+1}^{m} dF(u_i).$$

Here, $\delta_x$ is the Dirac delta function $\langle \text{stat02228} \rangle$. 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 \( \{ U_n, \mathcal{F}_n \}_{n \geq m} \) with \( \mathcal{F}_n = \sigma(X_{(1):n}, \ldots, X_{(n):n}, X_{n+1}, X_{n+2}, \ldots) \) and \( X_{(i):n} \) the \( i \)-th order statistic of \( X_1, \ldots, X_n \) \cite[Section 3.4]{stat06009} (see the discussion of martingales \cite{stat02941} in the entry on Counting Process Methods in Survival Analysis \cite{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 \cite{stat05843} and nonparametric regression \cite{stat05768} theory \cite{13, 20} and survival analysis \cite{5} (see \cite{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 \cite[Section 4.1]{stat06009} and to von Mises expansions, a generalization of the projection method \cite{9} (a technique discussed in more detail in Section 2). For further reading we refer to \cite[Chapter 6]{21} and \cite{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 \cite{stat06033} for \( F(t) \). With \( \hat{\Lambda}_n(t) \) the Nelson-Aalen estimator \cite{stat06045} and \( \Lambda(t) \) the cumulative hazard function \cite{stat04288}, a \( U \)-statistic representation has been established in \cite{5} for \( \hat{\Lambda}_n(t) - \Lambda(t) \). On the basis of the relation

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

and using Taylor expansion \cite{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 \cite{Large-sample Theory} 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 \cite{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 \cite[Section 3.4]{stat06009}.

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

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

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 $E[h_2(X_1, \ldots, X_m)] < \infty$ and $\zeta_1 = Var(h_1(X_1)) > 0$ (i.e., $U_n$ is nondegenerate), then

$$\sqrt{n}[U_n - \theta(F)] \xrightarrow{d} Z,$$

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

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

For $h_2$, define the integral operator

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

where $z$ is square integrable with respect to $F$. Let $\lambda_1, \lambda_2, \ldots$ denote the real (not necessarily distinct) eigenvalues corresponding to the distinct solutions $z_1, z_2, \ldots$ 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

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

5
where $Y$ is a random variable of the form $Y = \sum_{j=1}^{\infty} \lambda_j [\chi_j^2(1) - 1]$, where $\chi_j^2(1), \chi^2(1), \ldots$ are independent $\chi^2(1)$ random variables (see Convergence in Distribution and in Probability).

Example 5. For the sample variance an application of Theorem 5 yields (with $\mu_k$ 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.


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


