ON THE ASYMPTOTIC DISTRIBUTION OF U-STATISTICS

by

Alan J. Lee*

University of Auckland
and
University of North Carolina at Chapel Hill

Abstract

The asymptotic distribution of a U-statistic is found in the case when the corresponding Von Mises functional is stationary of order 1. Practical methods for the tabulation of the limit distributions are discussed, and the results extended to certain incomplete U-statistics.

Key Words and Phrases: asymptotic distribution, stationary statistical functional, incomplete U-statistic

---

*The work of this author was partially supported by the Air Force Office of Scientific Research under Contract AFOSR-75-2796.

§1. Introduction.

Let $X_1,\ldots,X_n$ be independent random k-vectors with common distribution function F. Let $\phi(\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_m)$ be a function on $\mathbb{R}_{km}$ symmetric in the k-vectors $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_m$ and let

$$\theta(F) = \int_{\mathbb{R}_k}\cdots\int_{\mathbb{R}_k} \phi(x_1,\ldots,x_m) \prod_{i=1}^{m} F(dx_i) \tag{1}$$

be a functional defined on the space of d.f.s. on $\mathbb{R}_k$, assuming that the integral exists. An unbiased estimate of $\theta = \theta(F)$ is furnished by the U-statistic

$$U_n = \binom{n}{m}^{-1} \sum \phi(X_{i_1},\ldots,X_{i_m}) \tag{2}$$

where the sum extends over all $\binom{n}{m}$ subsets of the X_i. Following Hoeffding [10], define

$$\Phi_c(x_1,\ldots,x_c) = E[\phi(x_1,\ldots,x_c, X_{c+1},\ldots,X_m)] \qquad c = 1,2,\ldots,m$$

and $\Psi_c(x_1,\ldots,x_c) = \Phi_c(x_1,\ldots,x_c) - \theta$, assuming all expectations exist. Also define

$$\zeta_0 = 0\ , \qquad \zeta_c = \mathrm{Var}(\Psi_c(X_1,\ldots,X_c))\ , \qquad c = 1,2,\ldots,m\ .$$

If for some nonnegative integer $d < m$, $\zeta_{d+1} \neq 0$ but $\zeta_c = 0$ for $c = 0,\ldots,d$, the functional (1) is said to be stationary of order d at F. If $d = 0$, then (2) has an asymptotically normal distribution after suitable normalization [10]. If $d \neq 0$, the distribution is no longer normal and may be obtained by

applying the theory of differentiable statistical functionals due to Von Mises and Filippova ([13], [7]).

Let F_n be the empirical distribution function of $X_1,\ldots,X_n$. The distribution of

$$\theta(F_n) = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \phi(X_{i_1},\ldots,X_{i_m}) \tag{3}$$

is closely related to the distribution of U_n as described in §2, and the asymptotic distribution of $\theta(F_n)$ is given by the following theorem.

Theorem (Von Mises, Filippova). Let θ be a stationary functional of order d. Then the asymptotic distribution of $n^{\frac{d+1}{2}}(\theta(F_n)-\theta)$ is identical to that of

$$n^{\frac{d+1}{2}} \binom{m}{d} \int_{\mathbb{R}_k}\cdots\int_{\mathbb{R}_k} \psi_{d+1}(x_1,\ldots,x_d) \prod_{i=1}^{d} (F_n(dx_i)-F(dx_i)) \ . \tag{4}$$

In the case d = 1, the asymptotic distribution of (4) can be found using integral equation techniques. Filippova gives an expression for the characteristic function of the asymptotic distribution of (4) in terms of Fredholm determinants. Gregory [8] gives a more concrete representation of the asymptotic distribution of (4) as an infinite series of random variables, in the case m = 2. Below we combine these results to give explicit expressions for the c.f.s. of the limit distributions and discuss practical methods for the tabulation of the limit distributions. We also make some remarks about incomplete U-statistics.

§2. Relationship between $\theta(F_n)$ and U_n.

Let I be the set of indices $\{(i_1,\ldots,i_m)\colon 1 \le i_\ell \le n,\ \ell = 1,\ldots,m\}$ and let I_j be the subset of I consisting of those m-tuples with exactly j distinct integers. Define for each $j = i,\ldots,m$ a symmetric kernel $\phi_{(j)}(x_1,\ldots,x_j)$ by $j!\phi_{(j)}(x_1,\ldots,x_j) = \sum_{(j)} \phi(x_{i_1},\ldots,x_{i_m})$ where the sum $\sum_{(j)}$ is taken over all m-tuples of indices $(i_1,\ldots,i_m)$ with $1 \le i_\ell \le j$ such that exactly j of the i_ℓ are distinct. Then

$$\theta(F_n) = \frac{1}{n^m} \sum_{I} \phi(X_{i_1},\ldots,X_{i_m})$$

$$= \frac{1}{n^m} \sum_{j=1}^{m} \sum_{I_j} \phi(X_{i_1},\ldots,X_{i_m})$$

$$= \frac{1}{n^m} \sum_{j=1}^{m} \sum_{1\le i_1\le\ldots\le i_j\le n} j!\phi_{(j)}(X_{i_1},\ldots,X_{i_j})$$

$$= \frac{1}{n^m} \sum_{j=1}^{m} j!\binom{n}{j} U_n^{(j)}$$

where $U_n^{(j)}$ is the U-statistic corresponding to the kernel $\phi_{(j)}$. Note that $U_n^{(m)} = U_n$. Thus

$$n(\theta(F_n)-\theta) = n(U_n-\theta) + Z_n \tag{5}$$

where

$$Z_n = nU_n\left(\frac{n!}{(n-m)!}\,n^{-m} - 1\right) + \sum_{j=1}^{m-1} \frac{n!}{(n-j)!n^{m-1}}\,U_n^{(j)} \,. \tag{6}$$

If θ is stationary of order 1, then $n(\theta(F_n)-\theta)$ converges to a nondegenerate distribution as seen in §3, and $\mathrm{Var}(nU_n)$ is bounded as $n \to \infty$ ([10]). Moreover, $E(Z_n)$ converges to the quantity $\frac{-m(m-1)}{2} + \theta^{(m-1)}$ where

$\theta^{(m-1)} = E[\phi_{(m-1)}(X_1,\ldots,X_{m-1})]$ and since Var $U_n = O\left(\frac{1}{n^2}\right)$ and Var $U_n^{(j)} = O\left(\frac{1}{n}\right)$ for $j < m$ it follows from (6) that Z_n converges in probability to $\frac{-m(m-1)}{2} + \theta^{(m-1)}$, and so the asymptotic distributions of $n(\theta(F_n)-\theta)$ and $n(U_n-\theta)$ differ only by a shift. A more explicit expression for $\theta^{(m-1)}$ is

$$\begin{aligned}\theta^{(m-1)} &= E[\phi_{(m-1)}(X_1,\ldots,X_m)] \\ &= \frac{1}{(m-1)!}\sum_{(m-1)} E(\phi(X_{i_1},\ldots,X_{i_m})) \\ &= \frac{1}{(m-1)!}\frac{m!(m-1)}{2} E(\phi(X_1,X_1,X_2,\ldots,X_{m-1})) \\ &= \frac{m(m-1)}{2} E(\Phi_2(X_1,X_1))\end{aligned}$$

and so we may write $\lim_n E(Z_n) = \lim_n E_n(\theta(F_n)-\theta) = \binom{m}{2} E(\Psi_2(X_1,X_1))$.

§3. Asymptotic distribution of $n(\theta(F_n)-\theta)$ and $n(U_n-\theta)$.

By the Von Mises - Filippova theorem, the asymptotic distribution of $n(\theta(F_n)-\theta)$ coincides with that of

$$\binom{m}{2}\int_{\mathbb{R}_k}\int_{\mathbb{R}_k} \Psi_2(x_1,x_2)G_n(dx_1)G_n(dx_2) \tag{7}$$

where $G_n(x) = \sqrt{n}(F_n(x)-F(x))$. Consider the kernel $\tilde{\Psi}_2(x_1,x_2)$ where (all integrals are over $\mathbb{R}_k$)

$$\tilde{\Psi}_2(x_1,x_2) = \Psi_2(x_1,x_2) - \int \Psi_2(x_1,v)F(dv) - \int \Psi_2(u,x_2)F(du) + \int\int \Psi(u,v)F(du)F(dv)\,.$$

Then it is easily seen that

$$\int\int \Psi_2(x_1,x_2)G_n(dx_1)G_n(dx_2) = \int\int \tilde{\Psi}(x_1,x_2)G_n(dx_1)G_n(dx_2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{\Psi}_2(X_i,X_j) \ . \tag{8}$$

Now let λ_j, $j \geq 1$, be the (real) eigenvalues of the linear operator $L_2(\mathbb{R}_k,dF) \to L_2(\mathbb{R}_k,dF)$ with (symmetric) kernel $\tilde{\Psi}_2$. We note that by the assumption that ζ_2 exists, $\int\int |\Psi_2(u,v)|^2 F(du)F(dv) < \infty$. By the results of [8], we can see that

(i) the asymptotic distribution of (8) is the same as

$$\binom{m}{2}\left\{\sum_{j=1}^{\infty} \lambda_j(Y_j-1) + E(\tilde{\Psi}_2(X_1,X_1))\right\} \tag{9}$$

where the Y_j are independent x_1^2 random variables; and

(ii) if $\sum_{j=1}^{\infty} |\lambda_j| < \infty$ then the asymptotic distribution of (8) is that of

$$\binom{m}{2}\sum_{j=1}^{\infty} \lambda_j Y_j .$$

To prove these, set the functions h_n in Theorem 2.1 of [8] to be 0, then $\frac{1}{n}\sum_{i\neq j} \tilde{\Psi}_2(X_i,X_j)$ converges in distribution to $\sum_{j=1}^{\infty} \lambda_j(Y_j-1)$. Since

$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \Psi_2(X_i,X_j) = \frac{1}{n}\sum_{i\neq j} \tilde{\Psi}_2(X_i,X_j) + \frac{1}{n}\sum_{i=1}^{n} \tilde{\Psi}_2(X_i,X_i)$ and $\frac{1}{n}\sum_{i=1}^{n} \tilde{\Psi}_2(X_i,X_i)$ converges in probability to $E[\tilde{\Psi}_2(X_1,X_1)]$ by the WLLN, (9) follows upon noting that

$$E[\tilde{\Psi}_2(X_1,X_1)] = \int \Psi_2(x_1,x_1)F(dx_1) - E(\Psi_2(X_1,X_2))$$

$$= \int \Psi_2(x_1,x_1)F(dx_1)$$

$$= E[\Psi_2(X_1,X_1)] \ ,$$

since $E[\Psi_2(X_1,X_2)] = 0$ by [10]. (ii) follows as in 1.2.3 of [8].

Using (i) and (ii) above and (5) we see that the asymptotic distribution of $n(U_n-\theta)$ is that of $\binom{m}{2}\sum_{j=1}^{\infty}\lambda_j(Y_j-1)$, and if $\sum|\lambda_j| < \infty$, the distribution is that of $\binom{m}{2}\left[\sum_{j=1}^{\infty}\lambda_j Y_j - E[\Psi_2(X_1,X_1)]\right]$. The characteristic functions of the limit distributions are thus seen to be

$\phi(t) = \prod_{j=1}^{\infty} e^{-it\lambda'_j}(1-2i\lambda'_j t)^{-\frac{1}{2}}$ in the former case and

$\phi(t) = e^{-it\mu}\prod(1-2i\lambda'_j t)^{-\frac{1}{2}}$ in the latter where $\mu = \binom{m}{2}E[\Psi_2(X_1,X_1)]$ and $\lambda'_j = \binom{m}{2}\lambda_j$, $j \geq 1$.

§4. Tabulation of the limit distribution.

In this section we discuss practical means of computing the limit distribution of $n(U_n-\theta)$. We first consider the case when the eigenvalues λ_j are summable. The method employed to tabulate the limit distribution will depend on the amount of information we possess about the eigenvalues. There are three subcases.

(i) The eigenvalues must be determined numerically, and we have accurate determinations of λ_j for $1 \leq j \leq N$.

(ii) The eigenvalues can be explicitly determined for all j.

(iii) The characteristic function of $\sum_{j=1}^{\infty}\lambda_j Y_j$, namely $\phi(t) = \prod_{j=1}^{\infty}(1-2it\lambda_j)^{-\frac{1}{2}}$, can be expressed in closed form.

In subcase (iii), the limit distribution may most conveniently be found by numerical inversion of $\phi(t)$. Various methods are available (e.g. Davies [5], Bohman [3], Martynov [12]). In subcase (ii), computation of the inner product is necessary. A reasonable way to accomplish this is described in [2] and consists of expressing $\phi(t) = \prod_{j=1}^{N-1}(1-2it\lambda_j)^{-\frac{1}{2}}\phi_2(t)$ where

$\log \phi(t) = -\frac{1}{2} \sum_{j=N}^{\infty} \log(1-2it\lambda_j)$. Formally, we have

$\log \phi(t) = \frac{1}{2} \sum_{j=N}^{\infty} \sum_{k=1}^{\infty} (2it\lambda_j)^k/k = \frac{1}{2} \sum_{k=1}^{\infty} C_{N,k}(2it)^k/k$ where $C_{N,k} = \sum_{j=N}^{\infty} \lambda_j^k$.

The formal manipulations above are valid if the last power series converges. Since $|C_{Nk}| \leq \left(\sum_{j=N}^{\infty} |\lambda_j| \right)^k = C_N^k$, say, this convergence will take place if $|t| \leq \frac{1}{2C_N}$. The numerical inversion methods above will require the computation of ϕ_2 at a finite number M of ordinates t_m, so choose N so large that $C_N < (2 \max_m |t_m|)^{-1}$ and use the power series to compute ϕ_2.

In special cases other methods may be used. If the eigenvalues λ_j are all positive, the asymptotic results of Zolotarev [14] and Hoeffding [11] furnish good approximations for the tail probabilities. If the eigenvalues are all positive and of multiplicity unity, then Smirnov's formula may be used as an efficient numerical inversion method; see Martynov [12] for details.

In subcase (i) two possible approaches suggest themselves. The use of the distribution of $\sum_{j=1}^{N} \lambda_j Y_j$ to approximate that of $\sum_{j=1}^{\infty} \lambda_j Y_j$ will not be satisfactory in general for reasonable values of N. For a discussion of this point and bounds on the truncation error see [2]. Rather, some method of approximating the tail of the series is required. If all the λ_j for $j > N$ are positive, we may approximate (see [6]) the distribution of $\sum_{j=1}^{\infty} \lambda_j Y_j$ by that of $\sum_{j=1}^{N} \lambda_j Y_j + cY$ where Y is an x^2_ν variate and c and ν are chosen to make the mean and variance of the two distributions coincide. Thus c and ν are obtained from

$$\sum_{j=1}^{N} \lambda_j^2 + c^2\nu = \mathrm{Var}(\Psi_2(X_1,X_2)) = \zeta_2$$

$$\sum_{j=1}^{N} \lambda_j + c\nu = E[\Psi_2(X_1,X_1)]$$

yielding

$$c = \frac{\zeta_2 - \sum_{j=1}^{N} \lambda_j^2}{(E[\Psi_2(X_1,X_1)] - \sum_{j=1}^{N} \lambda_j)},$$

$$\nu = \frac{1}{c}\left\{E(\Psi_2(X_1,X_1)) - \sum_{j=1}^{N} \lambda_j\right\}.$$

The c.f. of the approximating r.v. is

$$\phi_0(t) = \prod_{j=1}^{N} (1-2i\lambda_j t)^{-\frac{1}{2}} (1-2ict)^{-\nu/2} \tag{10}$$

and a numerical inversion of (10) furnishes the desired approximation.

If all but a finite number of the λ_j are not of the same sign, one may compute the density of $\sum_{j=1}^{N-1} \lambda_j Y_j$ by e.g. the integral equation method of Grenander et al. [9] and approximate the distribution function of the remainder $\sum_{j=N}^{\infty} \lambda_j Y_j$ by means of a Cornish-Fisher expansion. The cumulants of the remainder are easily seen to be

$$\kappa_r = 2^{r-1}(r-1)!C_{N,r}$$

so $\kappa_1 = E[\Psi_2(X_1,X_1)] - \sum_{j=1}^{N-1} \lambda_j, \kappa_2 = \zeta_2 - \sum_{j=1}^{N-1} \lambda_j^2$ and for $r \geq 3$ the κ_r may be approximated reasonably well by computing a few more eigenvalues. The limit df may then be obtained by a numerical convolution.

If all but a finite number of eigenvalues are positive (or negative) then the kernel $\tilde{\Psi}_2(x_1,x_2)$ can be expressed as the sum of a degenerate and a positive definite (negative definite) kernel and hence its eigenvalues will be summable. Thus in the nonsummable case, an infinite number of both positive and negative eigenvalues will be encountered. If only a finite number are known, we may employ the Cornish-Fisher method suggested above. If all are known, we may compute $\phi_2(t) = \prod_{j=N}^{\infty} e^{-it\lambda_j}(1-2i\lambda_j t)^{-\frac{1}{2}}$ by $\log \phi_2(t) = -\frac{1}{2}\sum_{k=2}^{\infty} C_{N,k}(2it)^k/k$ and proceed as before.

§5. Some remarks on incomplete U-statistics.

The calculation of the U-statistic (2) requires the averaging of $\binom{n}{m}$ terms, which may not be practical if m and n are not small. To reduce the volume of computation Blom [1] and Brown and Kildea [4] have proposed the use of incomplete U-statistics of the form

$$U = \frac{1}{N}\sum \phi(X_{i_1},\ldots,X_{i_m}) \tag{11}$$

when the sum in (11) is taken over N specified or randomly selected m-subsets of the indices. The asymptotic distributions of nondegenerate incomplete U-statistics ($\zeta_1 > 0$) are studied in the references above. For the degenerate case $\zeta_1=0,\zeta_2>0$ some of their results remain true while others need modification.

Denote by ϕ_i the r.v. $\phi(X_{i_1},\ldots,X_{i_m})$ where $(i_1,\ldots,i_m)$ is the i-th subset of indices in the summation (11). As in [1], let p_c denote the proportion of the N^2 pairs (ϕ_i,ϕ_j) having c indices in common. p_c will be a fixed constant or an r.v. depending on the method of subset selection. From [1] we have

$$\text{Var } U = \sum_{c=2}^{m} E(p_c)\zeta_c$$

and the assumption is made that n Var U $\to \beta$ where β is some nonnegative constant. Note that the $E(p_c)$ do not depend on ϕ but only on the method of subset selection.

The asymptotic distribution of U depends on the ratio $^n/N$. If $^n/N \to \infty$ as n,N both $\to \infty$; the quantities ϕ_i are asymptotically independent; and the degeneracy of the complete U-statistic is irrelevant to the limit distribution of $\sqrt{n}(U-\theta)$, which will be normal with mean zero and variance ζ_m. On the other hand, if $^n/N \to 0$, the limit distribution of U will coincide with that of the complete U-statistic under certain conditions. For example, from [1] we have (denoting the complete U-statistic by U_0)

$$\text{Var } U - \text{Var } U_0 = E(U-U_0)^2 ,$$

so $n(U_0-\theta)$ and $n(U-\theta)$ will have the same asymptotic distribution if $n^2(\text{Var } U - \text{Var } U_0)$ converges to zero. Now n^2 Var U_0 converges to $2\binom{m}{2}^2 \zeta_2$ so a sufficient condition for the coincidence of the asymptotic distributions is

$$\lim_{n\to\infty} n^2 E(p_c) = \begin{cases} 2\binom{m}{2}^2 & c = 2 , \\ 0 & c > 2 . \end{cases}$$

Alternatively, if the N subsets are chosen at random, with replacement from the $\binom{n}{m}$ possible subsets, then ([1])

$$\text{Var } U - \text{Var } U_0 = \frac{1}{N}(\zeta_m - \text{Var } U_0)$$

so the asymptotic distributions coincide if n^2/N converges to zero. Thus the incomplete U-statistic based on $n^{2+\varepsilon}$ randomly chosen subsets will have the same asymptotic distribution as that of the complete U-statistic based on $\binom{n}{m}$ subsets.

References

[1] Blom, G. (1976). Some properties of incomplete U-statistics. *Biometrika, 63,* pp. 573-580.

[2] Blum, J.R., Kiefer, J., and Rosenblatt, M. (1961). Distribution-free tests of independence based on the sample distribution function. *Ann. Math. Statist., 32,* pp. 485-498.

[3] Bohman, H. (1972). From the characteristic function to the distribution function via Fourier Analysis. *BIT, 12,* pp. 279-283.

[4] Brown, B.M. and Kildea, D.G. (1978). Reduced U-statistics and the Hodges-Lehmann estimator. *Ann. Statist., 6,* pp. 828-835.

[5] Davies, R.B. (1973). Numerical inversion of the characteristic function. *Biometrika, 60,* pp. 415-417.

[6] Durbin, J. and Knott, M. (1972). Components of Cramér - Von Mises statistics I. *J.R. Statist. Soc., 34,* pp. 290-307.

[7] Filippova, A.A. (1962). Von Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theor. Prob. Appl., 7,* pp. 24-57.

[8] Gregory, G.G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist., 5,* pp. 110-115.

[9] Grenander, U., Pollak, H.O., and Slepian, D. (1959). The distribution of quadratic forms in normal variates. *J.S.I.A.M., 7,* pp. 374-401.

[10] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist., 19,* pp. 293-325.

[11] Hoeffding, W. (1964). On a theorem of V.M. Zolotarev. *Theor. Prob. Appl., 9,* pp. 89-91.

[12] Martynov, G.V. (1975). Computation of distribution functions of quadratic forms in normally distributed random variables. *Theor. Prob. Appl., 20,* pp. 782-793.

[13] Von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functionals. *Ann. Math. Statist.*, *18*, pp. 309-348.

[14] Zolotarev, V.M. (1961). Concerning a certain probability problem. *Theor. Prob. Appl.*, *6*, pp. 201-204.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* On the Asymptotic Distribution of U-Statistics		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL 6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1255
7. AUTHOR(s) Alan J. Lee		8. CONTRACT OR GRANT NUMBER(s) AFOSR-75-2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling AFB, DC 20332		12. REPORT DATE September 1979 13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release - distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Asymptotic distribution, stationary statistical functional, incomplete U-statistic

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The asymptotic distribution of a U-statistic is found in the case when the corresponding Von Mises functional is stationary of order 1. Practical methods for the tabulation of the limit distributions are discussed, and the results extended to certain incomplete U-statistics.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*