Application of Jackknife in the Analysis of Robust Designs

ASOKAN MULAYATH VARIYATH, BOVAS ABRAHAM and JIAHUA CHEN

University of Waterloo, Waterloo, ON N2L 3G1

Robust designs with performance measures as responses are common in industrial applications. The existing analysis methods often regard performance measures as sole response variables without replicates. Consequently, no degrees of freedom are left for error variance estimation in these methods. In reality, performance measures are obtained from replicated primary response variables. Precious information is hence lost. In this paper, we suggest a jackknife based approach on the replicated primary responses to provide an estimate of error variance of performance measures. The resulting tests for factor effects become easy to construct and more reliable. We compare the proposed method with some existing methods using two real examples and investigate the consistency of the jackknife variance estimate based on simulation studies.

Introduction

Experimental designs are widely used in industries to control and improve the quality of products. The basic purpose of these experiments is to arrive at combination of factor levels which optimize the response or to identify the important factors which control the characteristic of interest. Designed experiments are also used to reduce the variation of the response by identifying critical factors (Taguchi (1986)).

There are many situations where analysis is performed on summary statistics of the primary response variables. Here after we will refer to these summary statistics as performance measures. For example, in an experiment discussed in Wu and Hamada (2000, page 124), the primary response variable is the thickness of epitaxial layer on a silicon wafer.

The aim of the experiment is to find the level combinations of the 4 factors such that its variation is minimized. In this case, the performance measure is chosen as the log(sample variance) of the replicated observations at each level combination. If the performance measure is regarded as our primary response, for the purpose of data analysis, then no degrees of freedom are left for error variance estimation. In this situation, the general practice is to use the analysis methods for unreplicated factorial experiments. See Wu and Hamada (2000).

A detailed review of analysis of unreplicated factorial experiments is available in Hamada and Balakrishnan (1998). Some widely used methods are, (i) Normal/Half Normal probability plots (Daniel (1959)) to identify the active effects and then pool the non-active effects to arrive at an estimate of error variance. (ii) Pseudo Standard Error(PSE) estimation method by Lenth (1989). These simple methods will be discussed in more detail later with examples and/or in simulations.

The problem we are interested in is different from the analysis of unreplicated factorial experiments. We have replications for each treatment combination, but we are interested in a performance measure of these replications. When we use analysis methods for unreplicated factorial experiments for these performance measures, precious information in the replicated observations (for the primary response) is lost and hence the opportunity to obtain a proper estimate of the error variance is also lost. We suggest a method based on jackknife to recoup the information to estimate the error variance.

This paper is organized as follows. First, we discuss the most frequently used performance measures and introduce the jackknife method for analyzing the performance measures, which will be explained by two real examples. In the subsequent section, we compare the performance of jackknife method with Lenth's method and explore the consistency of jackknife method. Some concluding remarks are given in the last section.

Jackknife Method for Performance Measures

Consider a robust design experiment with n runs each replicated m times. Let y_{ij} be the j^{th} replicate of the i^{th} experimental run, where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Most commonly used performance measures include $\eta_1 = \bar{y}_i$ and $\eta_2 = \log_e(s_i^2)$ where $\bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij}$ and $s_i^2 = \frac{1}{m-1}\sum_{j=1}^{m}(y_{ij} - \bar{y}_i)^2$. Taguchi (1986) proposed a number of performance measures in the context of quality engineering based on the response of interest. These performance measures, referred to as Signal to Noise (SN) Ratios are also discussed in Phadke (1989). Some frequently used performance measures are given in Table 1.

Table 1: Performance measures

Response Type	Notation	Formulae
Average	η_1	$\bar{y}_i$
Log Variance	η_2	$\log_e(s_i^2)$
Smaller the better	η_3	$-10\log_{10}(\frac{1}{m}\sum_{j=1}^{m} y_{ij}^2)$
Nominal the better	η_4	$10\log_{10}(\bar{y}_i^2/s_i^2)$
Larger the better	η_5	$-10\log_{10}(\frac{1}{m}\sum_{j=1}^{m} y_{ij}^{-2})$

Resampling methods are commonly used for constructing simple and efficient variance estimators. Our idea here is to obtain an appropriate estimator of the variance of a performance measure by resampling methods. Jackknife and Bootstrap are two widely used resampling methods. When the sample size is small, the bootstrap method is likely to produce unbalanced replicates while the jackknife is always balanced. Thus, we will only persue jackknife in this paper.

Assume that we have n experimental runs and each run is replicated m times. Let $y_i = (y_{i1}, y_{i2}, \ldots, y_{im})$ be the vector of replications from the i^{th} experimental run, with $c(y_i)$ be the corresponding performance measure. By deleting y_{ij} from y_i for $j = 1, \ldots, m$, we obtain m delete-one jackknife replicates of size $(m-1)$, $y_{i(j)}$. Hence, we obtain m jackknife replications of the performance measure $c(y_{i(j)}), j = 1, 2, \ldots, m$. The jackknife variance estimate of the estimator $c(y_i)$ is given by

$$\hat{V}_{ja}(c(y_i)) = \frac{m-1}{m} \sum_{j=1}^{m} (c(y_{i(j)}) - c(y_{i.}))^2 \tag{1}$$

where $c(y_{i.}) = 1/m \sum_{j=1}^{m} c(y_{i(j)})$. A pooled estimate of the error variance is,

$$\hat{V}_{pja}(c(y)) = \frac{1}{n} \sum_{i=1}^{n} \hat{V}_{ja}(c(y_i)). \tag{2}$$

Let us consider the "F-Statistics"

$$\mathrm{F} = \frac{\text{Mean Square for the Factor Effect}}{\hat{V}_{pja}(c(y))}.$$

We suggest to use this F to test the significance of factor effects in the ANOVA where the variance estimate has (m-1)n degrees of freedom. Some theoretical aspects of jackknife together with justifications for the F test is discussed in the Appendix.

Next we illustrate this method with two real examples from the literature together with existing analysis methods.

Example 1:

We refer to this example from Wu and Hamada (2000, page 124). The nominal value of the thickness of epitaxial layer on a silicon wafer is 14 μm with a specification of $\pm 0.5 \mu m$. The current process setting leads to excessive variation and a 2^4 factorial experiment is conducted with four process factors (A,B,C and D) and the experiment is replicated $m = 6$ times. In this experiment we like to identify the important factors that could be used to minimize epitaxial layer non-uniformity while maintaining average thickness close to the nominal value. Let us consider the performance measure $\eta_2 = \log_e(s^2)$. The design matrix together with values of the performance measure is given in Table 2.

Traditional half-normal plots are used to judge the significance of the factor effects. Factor A is judged as "significant" based on these plots. A formal test of effect significance

Table 2: Design matrix, performance measures and the jackknife variance estimates for Example 1

Run	Factors				η_2	$\hat{V}_{ja}(\eta_2)$
	A	B	C	D		
1	-	-	-	+	-5.77	0.6904
2	-	-	-	-	-5.31	0.1665
3	-	-	+	+	-5.70	0.6371
4	-	-	+	-	-6.98	0.8964
5	-	+	-	+	-5.92	0.4658
6	-	+	-	-	-5.49	0.9030
7	-	+	+	+	-4.11	0.1596
8	-	+	+	-	-6.24	0.5398
9	+	-	-	+	-1.54	0.2893
10	+	-	-	-	-2.12	0.1446
11	+	-	+	+	-1.58	0.1155
12	+	-	+	-	-1.49	0.2961
13	+	+	-	+	-1.92	0.2711
14	+	+	-	-	-2.43	0.2231
15	+	+	+	+	-1.12	0.1129
16	+	+	+	-	-2.65	0.1816

using Lenth's method (1989) was also provided in Wu and Hamada (2000). The method uses a robust estimator of the standard deviation of the factor effect θ_i. It is called pseudo standard error (PSE) and is defined as

$$PSE = 1.5\text{Median}_{(|\theta_i|<2.5s_0)}|\theta_i|$$

where median is computed among the $|\theta_i|$ with $|\theta_i| < 2.5s_0$ and $s_0 = 1.5\text{Median}|\theta_i|$.

An effect θ_i is declared significant if $|\frac{\theta_i}{PSE}|$ exceeds the critical values which can be found in Wu and Hamada (2000), reproduced from Ye and Hamada (2000). In this example, factor A is found to be significant.

Using jackknife replicates for each experimental run, we obtain a variance estimate $\hat{V}_{ja}(\eta_2)$ of the performance measure as per (1). For the first run, the primary responses are (14.812, 14.774, 14.772, 14.794,14.860, 14.914) (see Wu and Hamada(2000)). By deleting one observation at a time, we create 6 jackknife samples of size 5. The performance measure, $\log_e(s_i^2)$, for these jackknife samples are (-5.5537, -5.7338, -5.7518,-5.6052, -5.6719, -6.6432). Then jackknife variance estimate of $\log_e(s_i^2)$ for run 1 is computed as 0.6904. Similarly, we compute the jackknife variance estimate for each run and they are given in the last column of Table 2. The pooled jackknife variance estimate $\hat{V}_{pja} = 0.3808$ by (2), which is then used to test the factor effects. The ANOVA table for η_2 is given in Table 3.

We find that factors A and D are significant. The analysis based on Lenth's method finds only A being significant. Note that A is extremely large compared to other effects. Almost any method will declare it to be significant. Smaller effect, D, was not judged significant by Lenth's method.

Example 2:

This example is from Taguchi (1986, page 127). An experiment was planned and conducted to identify the factors that have strong effects on the wear on a slider pump. Five factors (A,B,C,D and E) and two interaction effects (AB and AC) were suspected to influence the

Table 3: Analysis of Variance table for η_2

Source	d.f	MS	F
A	1	58.8135	154.4
B	1	0.0245	0
AB	1	0.7321	1.7
C	1	0.0236	0
AC	1	1.829	0.5
BC	1	0.4394	1.2
ABC	1	0.4479	1.2
D	1	1.5961	4.2
AD	1	0	0
BD	1	0.3721	1.0
ABD	1	0.0295	0.1
CD	1	1.3535	3.6
ACD	1	0.9758	2.6
BCD	1	0.3947	1.0
ABCD	1	0.0472	0.1
$\hat{V}_{ja}$		0.3808	

wear and an Orthogonal Array Experiment $L_8(2^7)$ with $m = 8$ replicates was performed. The data consist of wear (in microns) at eight points on the slider of a pump and the goal is to reduce both the mean and variation of the wear. According to Taguchi, the SN Ratio of 'smaller the better' type, η_3 (see Table 1) is an appropriate choice for a performance measure. The design matrix and the values of η_3 for this experiment are given in Table 4. The corresponding Analysis of Variance (ANOVA) table is given in Table 5.

Table 4: Design matrix, performance measures and the jackknife variance estimates for Example 2

Run	Factors					η_3	$\hat{V}_{ja}(\eta_3)$
	A	B	C	D	E		
1	-	-	-	-	-	-21.8717	1.8395
2	-	-	+	+	+	-20.6023	5.6720
3	-	+	-	+	+	-14.7712	4.9053
4	-	+	+	-	-	-16.1278	1.3237
5	+	-	-	-	+	-24.1539	7.0389
6	+	-	+	+	-	-21.7136	9.9465
7	+	+	-	+	-	-22.9584	2.6745
8	+	+	+	-	+	-23.2710	7.1220

In Taguchi's (1986) analysis, the sum of squares due to the small effects C, AC and E are pooled to obtain the sum of squares of error. The contribution percentages ($\rho\%$) of the remaining effects are then calculated. The effects of A, B and AB are found "significant" by examining the magnitude of their contribution percentages.

Another approach is to use the traditional analysis where the error sum of squares is constructed by pooling the small effects after a visual inspection of the normal probability

Table 5: Analysis of Variance table

Source	d.f	SS	MS	F_{pooled}	ρ %	F_{jack}
A	1	43.82	43.82	115.3	52.9	8.65
B	1	15.72	15.72	41.4	18.7	3.10
AB	1	17.81	17.81	46.9	21.2	3.51
C	1	0.52	0.52	1.37	0	0.10
AC	1	0.61	0.61	1.60	0	0.12
D	1	3.61	3.61	9.5	3.9	0.71
E	1	0	0	0	0	0
(e)	(3)	1.13	0.38	-	3.3	
Total	7	82.09			100.0	

plots. Here it has 3 degrees of freedom and the F tests confirm that A, B and AB are significant. Analysis based on Lenth's method indicate that only factor A is significant.

Using the jackknife replicates for each experiment, we obatin jackknife variance $\hat{V}_{ja}(\eta_2)$ (see Table 4). Pooled jackknife variance estimate $\hat{V}_{pja} = 5.06$. This is used to test the significance of the factor effects in the ANOVA table (see Table 5). We find that only the main effect A is significant at the 5% level. It is seen that our jackknife error variance estimate for η_3 is larger than the estimate obtained by pooling the 3 smaller effects.

Performance of Jackknife

Adapting methods from unreplicated factorial experiments to analyze robust designs requires the pooling of sum of squares for smaller effects as indicated in Examples 1 and 2. Tests based on jackknife error variance estimate make use of the information from the primary response variables. Thus, it is likely a better approach. We will use simulations to compare the jackknife method to Lenth's method under two performance measures and explore the consistency of the jackknife variance estimator.

Comparison under performance measure η_1

We compare the jackknife method to Lenth's method for the simple performance measure average (η_1), where the true variance is known. Let us consider a 2^3 factorial experiment with factors A, B and C where all factors and interactions have some effects. Let y_{ij} be the response for the j^{th} replicate of the i^{th} run where i=1,...,n and j=1,...,m, which is modelled as,

$$y_{ij} = 10 + 0.2A + 0.05B + 0.1C + 0.1AB + 0.075AC + 0.03BC + 0.001ABC + e_{ij} \quad (3)$$

where A, B and C takes values of ± 1 depending on the levels, e_{ij} are normally distributed with mean 0 and σ=0.5 and AB, AC, BC and ABC represent the values of the variables associated with the interactions. Var($\bar{y}_i$) for the i^{th} run is σ^2/m and the variance of an estimated effect = $\sigma^2/2m$. The jackknife error variance estimate for each run $\hat{V}_{ja}(\bar{y}_i)$ is unbiased for Var($\bar{y}_i$) = σ^2/m. Thus, $\hat{V}_{pja}/2$ is an unbiased estimate of $\sigma^2/2m$. Since the expected value of PSE is complicated, we carry out 1000 simulations as per model (3) to approximate its expected value. For each simulation, we estimate PSE and $\hat{V}_{ja}$, and perform the significance tests based on both jackknife and Lenth's methods. Based on the 1000 simulations, we compute the percentage of times each effect is declared significant and the results are summarized in Table 6.

It is very clear that Lenth's method over estimate the standard error of the factor effects for this performance measure. Jackknife variance estimate from the simulations is almost the same as the true variance. Lenth's method does not pick up the smaller effects as often as the jackknife method. For larger sample sizes, jackknife performs far superior to Lenth's method. i.e. power of the test increases when sample size increases.

Table 6: Comparison of Jackknife method to Lenth's method for η_1 based on % times factors significant

Factor Effects	m=6		m=10		m=20		m=50	
	Jack	Lenth	Jack	Lenth	Jack	Lenth	Jack	Lenth
A	80.9	26.9	94.6	24.3	100	25.7	100	22.8
B	11.7	2.6	15.3	1	24.7	0.2	51.4	0.1
AB	27.1	5.1	44.3	4.8	73	3.7	97.2	0.9
C	28.3	5.2	41.8	4.6	72.9	3.8	97.9	1.1
AC	17	3.2	29	2	48.3	1.2	84.4	0
BC	7.9	1.3	8.2	0.6	11.4	0	20.3	0
ABC	6.9	0.6	6.1	0.2	5.3	0.1	4.8	0
Standard Error of Factor Effect	0.1445	0.2417	0.1119	0.2311	0.0791	0.2193	0.05	0.2136
True Standard Error of Factor Effect	0.1443	0.1443	0.1118	0.1118	0.0791	0.0791	0.05	0.05

Comparison under performance measure η_2

Our next simulation, compares jackknife method to Lenth's method under performance measure η_2. We construct a linear model for a 2^3 factorial experiment with performance measure η_2 as response so that factors A and C have significant effects on η_2 and other factor effects including interactions are negligible. That is, we use a linear model

$$y_{ij} = \mu_i + e_{ij} \tag{4}$$

for the primary response of the i^{th} run and the j^{th} replicate. Suppose that we are interested in the effect of factors on the performance measure $\log(s_i^2)$ which estimates $\log(\sigma_i^2)$. Let μ represents the overall mean, α is the effect of factor A, β is the effect of factor C. A suitable model for the performance measure is to let $\sigma_i^2 = var(e_{ij})$ such that

$$\log(\sigma_i^2) = \mu + \alpha A + \beta C + \epsilon_i \tag{5}$$

where A and C take values of ± 1 depending on their levels in the i^{th} run and ϵ_i is some white noise such that $\epsilon_i \sim N(0, \sigma^2)$.

In our simulation, we set $\sigma^2 = 0.05^2$, $\alpha = 1.05, \beta = 0.95$ and $\mu = 1$. We first generate ϵ_i and determine a value of σ_i^2 according to (5). We then obtain $m = 6$ primary responses according to (4) by generating e_{ij} from $N(0, \sigma_i^2)$ for the i^{th} run. The performance measures $\log_e(s_i^2)$ and its jackknife variance estimate are computed thereafter. We then test the significance of factor effects at 5% level based on jackknife variance estimate and Lenth's method. The whole process is repeated 1,000 times. The percentage of times each factor is significant at 5% level is shown in Table 7.

According to Table 7, Lenth's method has lower power to capture the significance of the factors A and C than the jackknife method. For instance, factor A is found significant 95.1% of times by using jackknife method while it is only 70.3% by using Lenth's method. In addition, the main effect of B is found significant only 3% of the times which is in line

Table 7: Percentage of time factors significant

Factor	Lenth's Method	Jackknife Method
A	70.3	95.1
B	2.4	3.0
AB	2.4	4.7
C	61.5	87.4
AC	3.3	3.0
BC	2.9	2.7
ABC	2.6	2.6

with the 5% significant level. We note that Type I errors of the jackknife method are comparable to those of Lenth's method, and are near the 5% nominal value.

Consistency

We now explore the consistency of the jackknife variance estimate of the performance measures. Jackknife variance estimator for the performance measure η is strongly consistent if

$$\frac{\hat{V}_{ja}}{\sigma_n^2(\eta)} \rightarrow_{a.s.} 1$$

as $n \rightarrow \infty$ where $\sigma_n^2(\eta)$ is the variance of the performance measure and a.s denotes almost sure convergence (Shao and Tu (1995 page 25)).

We postulate a simple linear model for a 2^3 full factorial design with all three factors having significant effects:

$$y_{ij} = \mu + \alpha A + \beta B + \gamma C + e_{ij} \tag{6}$$

with $\mu = 10, \alpha = \beta = \gamma = 1, e_{ij} \sim N(0, 0.3^2)$, and A, B, C equal ± 1 depending on the levels of the three factors.

For each experimental run, we generate 6 samples based on the above model and all

Table 8: Comparison of $\hat{R}(\eta)$ for different performance measures

Run	η_1	η_2	η_3	η_4	η_5
1	1.03	1.48	1.01	1.42	1.02
2	0.89	1.53	0.97	1.70	1.07
3	0.96	1.51	0.97	1.65	0.98
4	0.99	1.73	0.96	1.30	1.04
5	1.03	1.53	0.90	1.51	1.01
6	0.97	1.48	1.06	1.49	1.04
7	1.02	1.60	0.99	1.46	1.02
8	0.99	1.53	0.92	1.58	0.93
Pooled	0.98	1.55	0.97	1.51	1.01

the five performance measures are considered. We construct 6 jackknife samples from each run and compute jackknife variance estimate for each of the performance measures. This procedure is repeated 1000 times. The average of the jackknife variance estimates over 1000 simulations is computed and denoted as $\hat{V}_{ja}(\eta)$. The variances of performance measures based on 1000 repeatitions are denoted as $\hat{V}(\eta)$.

Let

$$\hat{R}(\eta) = \frac{\hat{V}_{ja}(\eta)}{\hat{V}(\eta)}. \tag{7}$$

If the jackknife variance estimator provides a sensible error variance estimate, we should have $\hat{R}(\eta)$ close to 1. The jackknife variance estimator overestimates when $\hat{R}(\eta) > 1$ while it underestimates when $\hat{R}(\eta) <1$. The simulation results on $\hat{R}(\eta)$ run-wise and design-wise are given in Table 8.

Table 8 reveals that the value of $\hat{R}(\eta)$ varies for different performance measures. It is close to 1 for η_3 and η_5, and around 1.5 for η_2 and η_4 indicating overestimation.

Table 9: Comparison of $\hat{R}(\eta)$ for different replication (m) for η_2 and η_4

Run	η_2				η_4			
	m=6	m=10	m=20	m=50	m=6	m=10	m=20	m=50
1	1.48	1.32	1.08	1.03	1.42	1.25	1.19	1.04
2	1.53	1.28	1.12	1.04	1.70	1.30	1.05	0.98
3	1.51	1.44	1.17	0.96	1.65	1.15	1.07	1.09
4	1.73	1.26	1.12	1.02	1.30	1.18	1.17	0.98
5	1.53	1.29	1.21	1.07	1.51	1.42	1.13	1.08
6	1.48	1.30	1.16	1.00	1.49	1.37	1.14	0.95
7	1.60	1.26	1.10	0.99	1.46	1.24	1.07	1.16
8	1.53	1.34	1.12	0.95	1.58	1.35	1.05	0.98
Pooled	1.55	1.31	1.13	1.01	1.51	1.28	1.11	1.03

It is well known that the jackknife variance estimator is consistent under very general conditions (Shao and Tu (1995)). The overestimation problem is likely a small sample problem. To make the point, we repeat the simulations with $m = 10$, 20 and 50 for performance measures η_2 and η_4. We expect $\hat{R}(\eta)$ to get close to 1 as sample size increases. These results are given in Table 9.

Table 9 confirms that the jackknife variance estimate is very close to the true variance when m=50. Thus, when m is small, tailor made adjustments need to be considered. In general, this adjustment factor can simply be determined with simulation. For example, for performance measure η_4 with $m = 6$, a factor of 1.51 be appropriate according to Table 10.

To verify that the adjustment factor is not sensitive to σ^2 chosen in the model (6), we repeat the simulation study with $\sigma = 0.3$, 0.6, 1, 1.5 and 2 for m=6, 10, 20 and 50. We

Table 10: Sensitivity of $\hat{R}(\eta)$ for different white noise (σ) for η_2 and η_4

	η_2					η_4				
σ	0.3	0.6	1	1.5	2	0.3	0.6	1	1.5	2
m=6	1.55	1.58	1.54	1.59	1.60	1.51	1.60	1.56	1.54	1.53
m=10	1.25	1.28	1.22	1.25	1.26	1.23	1.23	1.30	1.24	1.26
m=20	1.16	1.12	1.13	1.11	1.10	1.13	1.10	1.07	1.10	1.12
m=50	1.06	1.04	1.03	1.03	1.08	1.04	1.08	1.07	1.05	1.03

found that the ratio $\hat{R}(\eta)$ does not depend on the value of σ and takes very similar values among the experimental runs. For brevity, we only present the ratio for pooled variance in Table 10 for these two performance measures.

Incorporating the correction factor for η_2 in Example 2, the jackknife variance estimate is reduced to 0.2539. This results in the significance of interaction effects CD and ACD at 5% level. In the study of comparison under performance measure η_2, the percentage of times when factor A is significant is improved to 97.4; the corresponding percentage for factor C is improved to 95%. At the same time the percentages of times when other factors are found not significant are still as high as 93.5%.

Discussions and Conclusions

When replications are available for experimental runs in robust design experiments, it is better to use this information to estimate the variance of the performance measures of interest. The proposed jackknife method is simple to use and efficient in estimating the error variance. It gives an opportunity to obtain an estimate of the within run variance which leads to a more reliable test for the factorial effects. Other usual methods (including Lenth's) use variance estimators based on between effects variation. In fact, Lenth's method is devised for the analysis of unreplicated factorials when there is effect sparsity. If the ratio of number of active factors to the number of runs is large, then the commonly used methods for analyzing unreplicated factorials are not very effective. Increasing the number of replications in each run improves the power of detecting significant effects in the proposed method while it has little impact in the methods commonly used. We feel that the proposed method can be a great tool whenever one is analyzing a robust design with a performance measure such as the average or $\log(s^2)$.

Apppendix

Let $y = (y_1, y_2, \ldots, y_m)$ be a sample of size m. Suppose that the parameter ξ is estimated by $\hat{\xi} = c(y)$ with c being a continuous function of y. We wish to estimate the bias and variance of $\hat{\xi}$. The jackknife focuses on the construction of pseudo samples by leaving out one observation at a time. The i^{th} jackknife sample consists of data with the i^{th} observation removed:

$$y_{(i)} = (y_1, y_2, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m), i = 1, 2, \ldots, m. \tag{8}$$

Let $\hat{\xi}_{(i)} = c(y_{(i)})$ be the i^{th} jackknife replication of the estimate $\hat{\xi}$. The estimates of the bias and variance of this estimator are,

$$\begin{aligned} \hat{\text{bias}}_{ja} &= (m-1)(\hat{\xi}_{(\cdot)} - \hat{\xi}), \\ \hat{\text{V}}_{ja} &= \frac{m-1}{m} \sum_{i=1}^{m} (\hat{\xi}_{(i)} - \hat{\xi}_{(\cdot)})^2 \end{aligned} \tag{9}$$

where $\hat{\xi}_{(\cdot)} = \sum_{i=1}^{m} \hat{\xi}_{(i)}/m$ (Efron and Tibshirani (1993)).

As it is clear from (8) that two jackknife samples differ only by two data points hence are very similar. Thus simple standard deviation of the jackknife replications does not represent the standard error of the original estimator. Because of this, the total variation in jackknife replicates provides an approximate estimate of the variance of $\hat{\xi}$ as given in (9). For a class of estimators, it can be shown that $\hat{V}_{ja}$ is asymptotically unbiased as m $\to \infty$

(Theorem 2.1, Shao and Tu (1995, page 25)). One may refer to Efron and Tibshirani(1993) and Shao and Tu (1995) for more detailed discussion.

In many applications, the unknown parameter, say, ξ, is a smooth function of other easily estimable parameters such as the mean and the 2nd moment. The corresponding estimator is a function of the sample moments. This includes all performance measures discussed. Let $y_{i1}, y_{i2}, ..., y_{im}$ be a set of independent and identically distributed observations. For simplicity, let $\xi_1 = g(\mu_1)$, $\hat{\xi}_1 = g(\bar{y}_1)$ with $\mu_1 = E(y_{11})$ and $\bar{y}_1 = \frac{1}{m}\sum_{j=1}^{m} y_{1j}$. Let $s_1^2 = \frac{1}{m-1}\sum_{j=1}^{m}(y_{1j} - \bar{y}_1)^2$. Approximately, we have,

$$\hat{\xi}_1 - \xi_1 = g'(\mu_1)[\bar{y}_1 - \mu_1] + o_p(\bar{y}_1 - \mu_1) \tag{10}$$

where $o_p(.)$ is a negligible quantity when m is large and the jackknife variance estimate of $\hat{\theta}_1$ is approximately

$$\hat{V}_{ja} = [g'(\mu_1)]^2 \sum_{i=1}^{m}(y_{1(i)} - \bar{y}_1)^2 = \frac{1}{m}[g'(\mu_1)]^2 s_1^2. \tag{11}$$

Under normality, $\bar{y}_1 - \mu_1$ and s_1^2 are independent. Without the normality assumption, they are approximately independent when m becomes large. Hence,

$$\frac{(\hat{\xi}_1 - \xi_1)^2}{\hat{V}_{ja}} \approx \frac{m(\bar{y}_1 - \mu_1)^2}{s_1^2}$$

which has an F-distribution with degrees of freedom 1 and $m - 1$. Applied to our performance measure, the F-Statistic is approximately the ratio of two independent quadratic forms. Thus the F-distribution with 1 and n(m-1) degrees of freedom is a reasonable choice.

When the distribution of individual observations are nearly symmetric, the normal approximation is very good even when the sample size is not large. However, the Taylor expansion in (10) may not be very precise for the typical size of m in experimental designs. This may result in some bias in $\hat{V}_{ja}$ and some adjustment may be required as seen in Table 9.

References

1. Daniel, C. (1959). "Use of Half-Normal Plots in Interpreting Factorial Two Level Experiments", *Technometrics* 1. pp.311-340.

2. Efron, B and Tibshirani, R. J. (1993). *An introduction to the Bootstrap*, New York: Chapman & Hall.

3. Hamada, M and Balakrishnan, N. (1998). "Analyzing unreplicated factorial experiments: A review with some new proposals", *Statistica Sinica*, 8, pp.1-41.

4. Lenth, R. V. (1989). "Quick and easy analysis of unreplicated factorials", *Technometrics*, 31,pp.469-473.

5. Phadke, S. M. (1989). *Quality Engineering using Robust Design*, Englewood Cliffs, New Jersey: Prentice Hall.

6. Shao, J and Tu, D. (1995). *The Jackknife and Bootstrap*, New York: Springer.

7. Taguchi, G. (1986). *Introduction to Quality Engineering: Designing Quality into Products and Processes*, Tokyo: Asian Productivity Organization.

8. Wu, J. C. F and Hamada, M. (2000). *Experiments: planning, analysis and parameter design optimization*, New York: Wiley.

9. Ye, K and Hamada, M. (2000). "Critical values of the Lenth Method for Un-replicated Factorial Designs", *Journal of Quality Technology*, 32, pp.57-66.

Key Words: Lenth's Method, Pseudo Standard Error, Performance Measure, Unreplicated Factorial Experiments

Biographical footnote:

Mr. Variyath is a PhD Student in the Department of Statistics and Actuarial science. He is a Member of ASQ. His email address is mvasokan@uwaterloo.ca

Dr. Abraham is a Professor in the Department of Statistics and Actuarial Science. He is a Senior Member of ASQ. His email address is babraham@uwaterloo.ca.

Dr. Chen is a Professor in the Department of Statistics and Actuarial Science. His email address is jhchen@uwaterloo.ca.