A Simulation Experiment to Investigate the Distributional Behavior of Extreme Cook’s Distance for GEE to Models with Contaminated Binary Responses

Kunthel By*  John Preisser†
Bahjat Qaqish‡

*University of North Carolina at Chapel Hill, kby@bios.unc.edu
†University of North Carolina at Chapel Hill, jpreisse@bios.unc.edu
‡University of North Carolina at Chapel Hill, qaqish@bios.unc.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/uncbiostat/art9

Copyright ©2008 by the authors.
A Simulation Experiment to Investigate the Distributional Behavior of Extreme Cook’s Distance for GEE to Models with Contaminated Binary Responses

Kunthel By, John Preisser, and Bahjat Qaqish

Abstract

A simulation experiment was conducted to investigate the behavior of the largest (and second, third and fourth largest) cluster Cook’s distance statistics for marginal mean and marginal association parameters, separately, in the context of an application of the orthogonalized residuals estimating equations formulation of alternating logistic regressions. The behavior of extreme Cook’s distance statistics for the association parameters was responsive to random contamination, whereby binary responses were transposed with equal probability, and without regard to cluster membership. Specifically, for moderate (20) to large (50) cluster sizes, the distribution “shifted to the right” with increasing levels of contamination; however, shifting was not observed for small cluster sizes (5). For other situations - association parameters under a cluster-concentrated contaminated model, and mean parameters, generally - the behavior of the extreme Cook’s distance did not behave in the anticipated manner.
A Simulation Experiment to Investigate the Distributional Behavior of Extreme Cook’s Distance for GEE to Models with Contaminated Binary Responses

Kunthel By, John S. Preisser, and Bahjat F. Qaqish
Department of Biostatistics
University of North Carolina at Chapel Hill

Summary

Abstract

A simulation experiment was conducted to investigate the behavior of the largest (and second, third and fourth largest) cluster Cook’s distance statistics for marginal mean and marginal association parameters, separately, in the context of an application of the orthogonalized residuals estimating equations formulation of alternating logistic regressions. The behavior of extreme Cook’s distance statistics for the association parameters was responsive to random contamination, whereby binary responses were transposed with equal probability, and without regard to cluster membership. Specifically, for moderate (20) to large (50) cluster sizes, the distribution "shifted to the right" with increasing levels of contamination; however, shifting was not observed for small cluster sizes (5). For other situations - association parameters under a cluster-concentrated contaminated model, and mean parameters, generally - the behavior of the extreme Cook’s distance did not behave in the anticipated manner.
Contents

1 Background 4

2 Goals Of Simulation Experiment 5
   2.1 Contamination .................................................. 5
   2.2 The Model ...................................................... 5
   2.3 Constructing the design matrices ................................ 5
   2.4 Generating The Response ......................................... 7
   2.5 The Simulation .................................................. 7
   2.6 Analysis Outline ................................................. 9

3 Results 11

4 Conclusion 11

A QQ Of Random versus No Contamination: $D_{cls.a}(K)$ 13
B QQ Of Random versus No Contamination: $D_{cls.b}(K)$ 18
C QQ Of Cluster versus No Contamination: $D_{cls.a}(K)$ 23
D QQ Of Cluster versus No Contamination: $D_{cls.b}(K)$ 28
E QQ Of Random versus No Contamination: $D_{cls.a}(K-1)$ 33
F QQ Of Random versus No Contamination: $D_{cls.b}(K-1)$ 38
G QQ Of Cluster versus No Contamination: $D_{cls.a}(K-1)$ 43
H QQ Of Cluster versus No Contamination: $D_{cls.b}(K-1)$ 48
I QQ Of Random versus No Contamination: $D_{cls.a}(K-2)$ 53
J QQ Of Random versus No Contamination: $D_{cls.b}(K-2)$ 58
K QQ Of Cluster versus No Contamination: $D_{cls.a}(K-2)$ 63
L QQ Of Cluster versus No Contamination: $D_{cls.b}(K-2)$ 68
M QQ Of Random versus No Contamination: $D_{cls.a}(K-3)$ 73
N QQ Of Random versus No Contamination: $D_{cls.b}(K-3)$ 78
O QQ Of Cluster versus No Contamination: $D_{cls.a}(K-3)$ 83
P QQ Of Cluster versus No Contamination: $D_{cls.b}(K-3)$ 88
Q QQ Of Random versus No Contamination: $D_{cls.a}(K) - D_{cls.a}(K-1)$ 93
R QQ Of Random versus No Contamination: $D_{cls.b}(K) - D_{cls.b}(K-1)$ 98
S QQ Of Cluster versus No Contamination: $D_{\text{cls},\alpha}(K) - D_{\text{cls},\alpha}(K-1)$

T QQ Of Cluster versus No Contamination: $D_{\text{cls},\beta}(K) - D_{\text{cls},\beta}(K-1)$

U Tables: Summary Statistics
1 Background

Many interesting questions in medical research involve correlated binary data. From a biostatistical standpoint, these questions are often addressed by the use of marginal mean models that consider the effects of a set of explanatory variables on the response. These effects are often estimated by the use of generalized estimating equations (GEE).

While the marginal mean is often of interest, information on the within-cluster marginal association is also potentially useful to the investigator. For example, if we were studying some genetic disorder, we might be interested in knowing the concordance of child \( j \) and child \( k \) both from family \( i \) having the genetic disorder. The estimation of these marginal association parameters are often implemented through the use of second-order generalized estimating equations (GEE2) (Qaqish and Liang, 1992). Although GEE2 provides efficient estimates of the marginal association parameters, the computational barrier is too daunting when cluster sizes are large – a common occurrence in medical data. Compromises between computational feasibility and efficiency of the estimates are undesirable but often unavoidable.

Carey et al. (1993) proposed the method of alternating logistic regressions (ALR) as a means of maintaining efficiency of the estimates and at the same time overcoming some of the computational barriers. Alas, the efficiencies of ALR reported by Carey et al. (1993) were limited to studies with small cluster sizes and were not considered in more realistic settings. In addition, ALR suffers from several important drawbacks. First, the sandwich covariance matrix estimator of the association parameter estimates is not permutation-invariant. It is possible to get different standard error estimates of the association parameters just by rearranging the order in which the responses appear in the data matrix. Second, preliminary studies by Bahjat Qaqish and his students showed that the claimed efficiency of ALR relative to GEE2 is not realized for large and variable cluster sizes.

Alternatively, Zink (2003) and Zink and Qaqish (2008) developed a new approach to estimation called orthogonalized residuals (ORTH). It can be shown that ALR is a special case of ORTH. In this regard, ORTH generalizes ALR. The promising aspects of ORTH include the following:

1. computational feasibility,
2. it produces covariance estimates that are permutation-invariant,
3. it recasts ALR in a form that is more standard, and
4. in some settings, ORTH has been shown to have efficiency advantages over ALR.

(It cannot be said however that ORTH consistently offers efficic something that still needs to be addressed.) Preisser and Qaqish (1996) developed cluster Cook’s distance for the marginal mean using fast one-step approximation formulae. Out of the studies on ORTH, cluster Cook’s distance for the association parameters using similar one-step approximations were also developed. The details of these formulae and their discussions are presented fully in Preisser et al. (2008). The purpose of the simulation study described here is to address the distribution of the extreme cluster Cook’s distance based on these formulae. Our simulation study made heavy use of Zink and Qaqish’s ORTHRES SAS (SAS Institute Inc., 2004) macro with appropriate modifications as needed. The ORTHRES macro is an implementation of the estimation algorithm based on ORTH. An R (R Development Core Team, 2008) version is also available; details may be found in By et al. (2008). We also made heavy use of SAS routines for computing Cook’s distance developed by John Preisser and his student Jamie Perin (Preisser and Perin, 2007). Routines based on Qaqish’s conditional linear family (Qaqish, 2003) were used to generate the multivariate binary data. In the next section, we discuss our simulation experiment in detail.
2 Goals Of Simulation Experiment

The aim of the study is to consider the distribution of cluster Cook’s distance in the context of the ORTH formulation of alternating logistic regressions for the estimation of marginal mean parameters ($\beta$) and marginal association parameters ($\alpha$), respectively. The nature of multivariate binary data does not readily lend itself to analytical derivations. The simulation study was conducted as a way to circumvent this shortcoming. Specifically, the simulation study endeavors to answer the following questions. First, is the proposed Cook's distance measure responsive to contamination of the data? Second, if it is responsive to contamination, does it behave in some predictable way relative to the Cook’s distance for the uncontaminated data? We expect to see a shift in the distribution of Cook's distance under contamination relative to the uncontaminated data. Furthermore, we expect this shift to grow (to a point) as the level of contamination increases.

2.1 Contamination

Contamination will be considered under two scenarios: (1) random contamination and (2) cluster concentrated contamination. Under random contamination, we fixed the level of contamination, say $p_c$. If $K$ is the number of clusters and $m$ is the size of each cluster, then each of the $Km$ observations is contaminated with probability $p_c$. If we let $Y_{ij,c}$ be the contaminated observation and $Y_{ij}$ be the original observation, the contamination is done as follows:

$$Y_{ij,c} = 1 - Y_{ij}$$

Under cluster contamination, each cluster is first chosen with probability $2p_c$. Once the cluster is chosen, the $m$ observations within the chosen cluster are contaminated with probability $0.50$.

2.2 The Model

The basic model that we chose for our simulation is motivated by a genetic epidemiology data analyzed by Qaqish and Liang (1992) and may be described by

$$\logit(\mu_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}, \quad (1)$$

$$\log(\psi_{ijk}) = \alpha_0 + \alpha_1 z_{1ijk} + \alpha_2 z_{2ijk} \quad (2)$$

where $\psi_{ijk}$ is the pairwise odds ratio between the $j$-th and $k$-th observations in cluster $i$. The variables $x_{1ij}$ and $x_{2ij}$ in the mean model are continuous. For ease of understanding $x_{1ij}$ can be motivated as a cluster level baseline covariate and $x_{2ij}$ can be thought of as observation time. In the association model, the association variables are defined by

$$z_{1ijk} = x_{1ij} \quad \text{and} \quad z_{2ijk} = |x_{2ij} - x_{2ik}| .$$

Under this setting, it is natural to think of $\alpha_2$ as the association between two observations within a cluster that are $|x_{2ij} - x_{2ik}|$ units apart in time.

2.3 Constructing the design matrices

The covariates $x_{1ij}$ and $x_{2ij}$ were constructed by partitioning the interval $[-1, 1]$ into equal parts based on the number of clusters and the cluster size. Let $K$ be the number of clusters and $m$ denote the cluster size (we assumed equal cluster size). Then the design matrix for the mean is generated by algorithm 1. Algorithm 2 constructs $z_{2ijk}$. To construct the design matrix for the association parameters, we proceed based on algorithm 3.
**Routine**: generateX($K, m$)
**Input**: Number of clusters $K$ and cluster size $m$
**Output**: Design matrix $X$

for $i=1$ to $K$
    for $j=1$ to $m$
        $x_{1ij} = \frac{2(i-1)}{K-1} - 1$
        $x_{2ij} = \frac{2(j-1)}{m-1} - 1$
    end
end

**Algorithm 1**: Constructing the design matrix for the mean.

**Routine**: generateZ2i($x_{2i}, m$)
**Input**: $x_{2i}, m$
**Output**: $z_{2i}$

for $j=1$ to $(m-1)$
    for $k=(j+1)$ to $m$
        $z_{2ijk} = |x_{2ij} - x_{2ik}|$
    end
end

**Algorithm 2**: Constructing $z_{2ijk}$.

**Routine**: generateZi($x_i, m$)
**Input**: $x_i, m$
**Output**: $Z_i$

for $i=1$ to $K$
    for $j=1$ to $m$
        $n = \binom{m}{2}$
        Note $x_i = [1_m \ x_{1i} \ x_{2i}]$. Let $x_i[1,2]$ denote the first element of $x_{1i}$.
        Then
        $z_{1i} = 1_n x_i[1, 2]$
        $z_{2i} = $ generate using algorithm 2 with $x_{2i}$ as your argument.
        $z_i = [1_n \ z_{1i} \ z_{2i}]_{n \times 3}$
    end
end

**Algorithm 3**: Constructing $z_i$
2.4 Generating The Response

Once the design matrices for both the mean and the association models are generated, the \( m \)-variate binary response vector \( Y_i, i = 1, \ldots, K \) is generated using the conditional linear family (Qaqish, 2003) with

\[
\begin{bmatrix}
-0.80 \\
0.27 \\
0.20
\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}
1.05 \\
0.35 \\
-0.35
\end{bmatrix}
\]  

(3)

These parameters were chosen so as to induce response vectors whose correlation has values that had practical meaning. For example, we wanted positive correlation and at the same time, we wanted them to decrease over time–something akin to an autoregressive correlation in a longitudinal data setting.

2.5 The Simulation

Our simulation experiment involves the series of steps outlined in algorithm 4. We make the following definitions: \( KK = \{50, 100, 200\} \), \( mm = \{5, 20, 50\} \), \( p_c = \{0, 0.02, 0.05, 0.10\} \) and \( cc = \{RC, CC\} \) where RC and CC denotes random and cluster contamination respectively. Let \( r \) denote \( r \)-th simulation. Algorithm 4 fully describes the crux of our simulation experiment. Note that we have a total of 63 scenarios. Table 1 enumerates all 63 scenarios.

\begin{verbatim}
for k in KK do
  for m in mm do
    for p in pc do
      for c in cc do
        for r=1 to 500 do
          1. Generate \( X^{(r)} \) and \( Z^{(r)} \) based on algorithms 1 to 3.
          2. Generate \( Y^{(r)} \) based on the method of Qaqish (2003) using the above values of \( \beta \) and \( \alpha \) given in (3).
          3. Contaminate \( Y^{(r)} \) with probability \( p \). (Note that if \( p = 0 \), the phrases random contamination and cluster contamination does not really apply).
          4. Estimate the parameters for models (1) and (2) using \( X^{(r)} \) and \( Z^{(r)} \) based on ORTH which we will denote by \( \hat{\beta}^{(r)} \) and \( \hat{\alpha}^{(r)} \).
          5. Compute cluster Cook’s distance for both \( \alpha \) and \( \beta \) which we will denote by \( D^{(r)}_{\text{cls},\alpha,i} \) and \( D^{(r)}_{\text{cls},\beta,i} \), \( i = 1, \ldots, k \).
        end
      end
    end
  end
end
\end{verbatim}

\textbf{Algorithm 4}: Simulating the experiment
Table 1: All scenarios of the simulation with each scenario having 500 replications

<table>
<thead>
<tr>
<th>$K$</th>
<th>$m$</th>
<th>Type</th>
<th>Level</th>
<th>$K$</th>
<th>$m$</th>
<th>Type</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>200</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>random</td>
<td>0.02</td>
<td>200</td>
<td>50</td>
<td>random</td>
<td>0.02</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>random</td>
<td>0.05</td>
<td>200</td>
<td>50</td>
<td>random</td>
<td>0.05</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>random</td>
<td>0.10</td>
<td>200</td>
<td>50</td>
<td>random</td>
<td>0.10</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>50</td>
<td>5</td>
<td>cluster</td>
<td>0.02</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>random</td>
<td>0.02</td>
<td>50</td>
<td>5</td>
<td>cluster</td>
<td>0.05</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>random</td>
<td>0.05</td>
<td>50</td>
<td>5</td>
<td>cluster</td>
<td>0.10</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>random</td>
<td>0.10</td>
<td>100</td>
<td>5</td>
<td>cluster</td>
<td>0.02</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>100</td>
<td>5</td>
<td>cluster</td>
<td>0.05</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>random</td>
<td>0.02</td>
<td>100</td>
<td>5</td>
<td>cluster</td>
<td>0.10</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>random</td>
<td>0.05</td>
<td>200</td>
<td>5</td>
<td>cluster</td>
<td>0.02</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>random</td>
<td>0.10</td>
<td>200</td>
<td>5</td>
<td>cluster</td>
<td>0.05</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>200</td>
<td>5</td>
<td>cluster</td>
<td>0.10</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>random</td>
<td>0.02</td>
<td>50</td>
<td>20</td>
<td>cluster</td>
<td>0.02</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>random</td>
<td>0.05</td>
<td>50</td>
<td>20</td>
<td>cluster</td>
<td>0.05</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>random</td>
<td>0.10</td>
<td>50</td>
<td>20</td>
<td>cluster</td>
<td>0.10</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>100</td>
<td>20</td>
<td>cluster</td>
<td>0.02</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>random</td>
<td>0.02</td>
<td>100</td>
<td>20</td>
<td>cluster</td>
<td>0.05</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>random</td>
<td>0.05</td>
<td>100</td>
<td>20</td>
<td>cluster</td>
<td>0.10</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>random</td>
<td>0.10</td>
<td>200</td>
<td>20</td>
<td>cluster</td>
<td>0.02</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>200</td>
<td>20</td>
<td>cluster</td>
<td>0.05</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>random</td>
<td>0.02</td>
<td>200</td>
<td>20</td>
<td>cluster</td>
<td>0.10</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>random</td>
<td>0.05</td>
<td>50</td>
<td>50</td>
<td>cluster</td>
<td>0.02</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>random</td>
<td>0.10</td>
<td>50</td>
<td>50</td>
<td>cluster</td>
<td>0.05</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>50</td>
<td>50</td>
<td>cluster</td>
<td>0.10</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>random</td>
<td>0.02</td>
<td>100</td>
<td>50</td>
<td>cluster</td>
<td>0.02</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>random</td>
<td>0.05</td>
<td>100</td>
<td>50</td>
<td>cluster</td>
<td>0.05</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>random</td>
<td>0.10</td>
<td>100</td>
<td>50</td>
<td>cluster</td>
<td>0.10</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>200</td>
<td>50</td>
<td>cluster</td>
<td>0.02</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>random</td>
<td>0.02</td>
<td>200</td>
<td>50</td>
<td>cluster</td>
<td>0.05</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>random</td>
<td>0.05</td>
<td>200</td>
<td>50</td>
<td>cluster</td>
<td>0.10</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>random</td>
<td>0.10</td>
<td>200</td>
<td>50</td>
<td>cluster</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Figure 1: Distribution of cluster Cook’s distance for $\alpha$ under random contamination for $K = 50$ and $m = 5$. $C2$ denotes 2% contamination, $C5$ denotes 5% contamination, and so on. If the QQ plot lies on the 45° degree line, then the distribution under contamination is no different than under no contamination.

2.6 Analysis Outline

To address whether the one-step cluster Cook’s distance approximation for $\beta$ and $\alpha$ is sensitive to contamination, we took each scenario as described in Table 1 and obtained the four largest order statistics for each replication:

$$D^{(r)}_{\text{cls}, \alpha, (K)} > D^{(r)}_{\text{cls}, \alpha, (K-1)} > D^{(r)}_{\text{cls}, \alpha, (K-2)} > D^{(r)}_{\text{cls}, \alpha, (K-3)}$$

and

$$D^{(r)}_{\text{cls}, \beta, (K)} > D^{(r)}_{\text{cls}, \beta, (K-1)} > D^{(r)}_{\text{cls}, \beta, (K-2)} > D^{(r)}_{\text{cls}, \beta, (K-3)}$$

with $r = 1$ to 500. For each scenario, we approximated the distribution of $t$-th ordered statistics, for $\alpha$ say, by

$$D^{(1)}_{\text{cls}, \alpha, (t)}, D^{(2)}_{\text{cls}, \alpha, (t)}, \ldots, D^{(500)}_{\text{cls}, \alpha, (t)}.$$ 

For each $K$ and for each $m$, shifts in the distribution due to contamination were evaluated by constructing QQ plots under no contamination against 2%, 5% and 10% contamination for a chosen type of contamination. Figure 1, for example, compares the QQ plots of the largest order statistics for cluster Cook’s distance for $\alpha$ under $K = 50$ and $m = 5$. The rest of the comparisons are given in the appendices. In addition, shifts in the distribution of the differences of the order statistics due to contamination were also considered. We specifically looked at the distributions of

$$D_{\text{cls}, \alpha, (K)} - D_{\text{cls}, \alpha, (K-1)} \text{ and } D_{\text{cls}, \beta, (K)} - D_{\text{cls}, \beta, (K-1)}.$$
Figure 2: Shifts in the distribution of the difference of cluster Cook’s distance for $\alpha$ with $K = 50$ and $m = 5$. $\text{diffA} 2$ means $D_{\text{cls}, \alpha}(K) - D_{\text{cls}, \alpha}(K-1)$ under 2% contamination with the rest defined similarly.

Figure 2 is such an example. The rationale for looking at the difference $D(K) - D(K-1)$ under the various levels of contamination is based on our expectation that the distance between the largest and the second largest ordered Cook’s distance would increase, possibly monotonically, as the level of contamination increases. The rest of the comparisons are provided in the appendices. In addition, summary statistics for the largest order statistics were computed and are also provided in the appendix.
3 Results

Figures 3 through 11 in appendix A show QQ plots of the largest ordered Cook’s distance for $\alpha$ under contaminated data relative to uncontaminated data. This is done under random contamination for all combinations of $K$ and $m$. Some salient features are the following. First, for cluster size 5, there doesn’t appear to be any difference between the ordered cluster Cook’s distance under contamination relative to no contamination. Second, as the cluster size increases to 20 and 50, we do see evidence that the contaminated Cook’s distance increases almost monotonically (Figures 6 to 11). This behavior can be seen by noting a couple of things. One, as the level of contamination increases, the higher the distribution of Cook’s distance lies above the 45 degree line. Two, the lines exhibit clear separation between the different levels of contamination. Table 2 in appendix U corroborates these observations. At least for $\alpha$, the cluster Cook’s distance under random contamination seem to behave in the manner that we expected it to behave—by shifting to the right as the level of contamination increases. The second largest, third largest, and fourth largest cluster Cook’s distance for $\alpha$ also exhibit this behavior (see appendices E, I, and M).

This behavior is not consistent in general. In fact, for every other situations other than that described in the previous paragraph, the QQ plots either show that the distribution of the contaminated Cook’s distance is no different than the uncontaminated or that if a shift is present, it is shifting to the left as the contamination increases. In figures 12 to 20, we considered cluster Cook’s distance for $\beta$ under random contamination. For small cluster sizes ($m = 5$), the distribution seems to shift further and further to the left as as the level of contamination increases. For the other cluster sizes ($m = 20$ and $m = 50$), there is no obvious difference between the contaminated and uncontaminated distribution. The summary statistics given in Table 3 also reflect these observations. The reader may observe the rest of the QQ plots for the other scenarios in the appendices.

4 Conclusion

In sum, our simulation study does show some evidence of shifting (to the right) of the distribution of the extreme cluster Cook’s distance. However, this behavior is limited specifically to the cluster Cook’s distance for $\alpha$ under random contamination and is most notable for cluster sizes of 20 and 50. For the other situations, the behavior of the distribution of the extreme Cook’s distance does not behave in the manner that we had anticipated prior to running the simulations. We saw that the distributions can either remain unchanged under contamination or shift to the left. The latter is most notable under cluster concentrated contamination. The reason for this behavior is not clear.
References


A QQ Of Random versus No Contamination: $D_{\text{cls}, \alpha}(K)$

Figure 3: $K = 50$ and $m = 5$

Figure 4: $K = 100$ and $m = 5$
Figure 5: $K = 200$ and $m = 5$

QQ: No Contamination vs 2%, 5%, and 10% random contamination under $K=200$ and $m=5$

Largest Order Statistics

Cluster Cook's D: ALPHAs

0.05 0.10 0.15 0.20
C0 C5 C10
Legend Under Statistics

Figure 6: $K = 50$ and $m = 20$

QQ: No Contamination vs 2%, 5%, and 10% random contamination under $K=50$ and $m=20$

Largest Order Statistics

Cluster Cook's D: ALPHAs

0.05 0.10 0.15 0.20 0.25 0.30
C0 C5 C10
Legend Under Statistics
Figure 7: $K = 100$ and $m = 20$

Figure 8: $K = 200$ and $m = 20$
Figure 9: $K = 50$ and $m = 50$

![Graph showing QQ plots for no contamination and 2%, 5%, and 10% random contamination under $K=50$ and $m=50$.]

Figure 10: $K = 100$ and $m = 50$

![Graph showing QQ plots for no contamination and 2%, 5%, and 10% random contamination under $K=100$ and $m=50$.]
Figure 11: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5%, and 10% random contamination under $K=200$ and $m=50$
B QQ Of Random versus No Contamination: $D_{cls, \beta}(K)$

Figure 12: $K = 50$ and $m = 5$

![Figure 12: QQ plot for $K = 50$ and $m = 5$]

Figure 13: $K = 100$ and $m = 5$

![Figure 13: QQ plot for $K = 100$ and $m = 5$]
Figure 14: $K = 200$ and $m = 5$

![Graph 1](image1)

QQ: No Contamination vs 2%, 5%, and 10% random contamination under $K=200$ and $m=5$

Largest Order Statistics

- Cluster Cook's D: BETAs
- $C_0$
- $C_2$
- $C_5$
- $C_{10}$

Figure 15: $K = 50$ and $m = 20$

![Graph 2](image2)

QQ: No Contamination vs 2%, 5%, and 10% random contamination under $K=50$ and $m=20$

Largest Order Statistics

- Cluster Cook's D: BETAs
- $C_0$
- $C_2$
- $C_5$
- $C_{10}$
Figure 16: $K = 100$ and $m = 20$

![Figure 16: $K = 100$ and $m = 20$](image)

Figure 17: $K = 200$ and $m = 20$

![Figure 17: $K = 200$ and $m = 20$](image)
Figure 18: $K = 50$ and $m = 50$

Figure 19: $K = 100$ and $m = 50$
Figure 20: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5%, and 10% random contamination under $K=200$ and $m=50$
C  QQ Of Cluster versus No Contamination: $D_{\text{cls},\alpha}(K)$

Figure 21: $K = 50$ and $m = 5$

QQ: No Contamination vs 2% and 5% cluster concentrated contamination under $K=50$ and $m=5$

Largest Order Statistics

Cluster Cook's $D$: ALPHAs

0.05 0.10 0.15 0.20 0.25 0.30

0.05 0.10 0.15 0.20 0.25

C2  C5  C10

Figure 22: $K = 100$ and $m = 5$

QQ: No Contamination vs 2% and 5% cluster concentrated contamination under $K=100$ and $m=5$

Largest Order Statistics

Cluster Cook's $D$: ALPHAs

0.05 0.10 0.15 0.20 0.25 0.30

0.05 0.10 0.15 0.20 0.25

C2  C5  C10

Hosted by The Berkeley Electronic Press
Figure 23: $K = 200$ and $m = 5$

QQ: No Contamination vs 2% and 5% cluster concentrated contamination under $K=200$ and $m=5$

Largest Order Statistics

Cluster Cook's D: ALPHAs

0.05 0.10 0.15

0.05 0.10 0.15 0.20

C0 C2 C5 C10

Legend: Order Statistics

Figure 24: $K = 50$ and $m = 20$

QQ: No Contamination vs 2% and 5% cluster concentrated contamination under $K=50$ and $m=20$

Largest Order Statistics

Cluster Cook's D: ALPHAs

0.05 0.10 0.15 0.20 0.25 0.30

0.10 0.15 0.20 0.25 0.30

C0 C5 C10

Legend: Order Statistics
Figure 25: $K = 100$ and $m = 20$

QQ: No Contamination vs 2% and 5% cluster concentrated contamination under $K=100$ and $m=20$

Figure 26: $K = 200$ and $m = 20$

QQ: No Contamination vs 2% and 5% cluster contamination under $K=200$ and $m=20$
Figure 27: $K = 50$ and $m = 50$

QQ: No Contamination vs 2% and 5% cluster concentrated contamination under $K=50$ and $m=50$

Largest Order Statistics

Cluster Cook's D: ALPHAs

0.05 0.10 0.15 0.20 0.25 0.30

C2 C5 C10

Legend: A = Black
B = Red
C = Blue
D = Green

Figure 28: $K = 100$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster concentrated contamination under $K=100$ and $m=50$

Largest Order Statistics

26
Figure 29: \( K = 200 \) and \( m = 50 \)

QQ: No Contamination vs 2% and 5% cluster concentrated contamination under \( K=200 \) and \( m=50 \)

Largest Order Statistics

Cluster Cook's D: ALPHAs

0.05 0.10 0.15 0.20

C0  C5  C10

Hosted by The Berkeley Electronic Press
D  QQ Of Cluster versus No Contamination: $D_{\text{cls}, \beta_i(K)}$

Figure 30: $K = 50$ and $m = 5$

![Graph showing QQ of Cluster versus No Contamination for $K=50$ and $m=5$]

Figure 31: $K = 100$ and $m = 5$

![Graph showing QQ of Cluster versus No Contamination for $K=100$ and $m=5$]

http://biostats.bepress.com/uncbiostat/art9
Figure 32: $K = 200$ and $m = 5$

Figure 33: $K = 50$ and $m = 20$
Figure 34: $K = 100$ and $m = 20$

![Figure 34](image1)

Figure 35: $K = 200$ and $m = 20$

![Figure 35](image2)
Figure 36: \( K = 50 \) and \( m = 50 \)

QQ: No Contamination vs 2% and 5% cluster concentrated contamination under \( K=50 \) and \( m=50 \)

Largest Order Statistics

Cluster Cook's D: BETAs

0.1 0.2 0.3 0.4 0.5 0.6 0.7

C0 C2 C5 C10

Figure 37: \( K = 100 \) and \( m = 50 \)

QQ: No Contamination vs 2%, 5% and 10% cluster concentrated contamination under \( K=100 \) and \( m=50 \)

Largest Order Statistics

Cluster Cook's D: BETAs

0.05 0.10 0.15 0.20

C0 C2 C5 C10
Figure 38: $K = 200$ and $m = 50$

QQ: No Contamination vs 2% and 5% cluster concentrated contamination under $K=200$ and $m=50$

Largest Order Statistics

Cluster Cook's D: BETAs

0.02 0.04 0.06 0.08 0.10
C0  C5  C10

http://biostats.bepress.com/uncbiostat/art9
QQ Of Random versus No Contamination: $D_{\text{cls,\alpha}}(K-1)$

Figure 39: $K = 50$ and $m = 5$

Figure 40: $K = 100$ and $m = 5$
Figure 41: $K = 200$ and $m = 5$

![Graph showing QQ plots for no contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=5$.]

Figure 42: $K = 50$ and $m = 20$

![Graph showing QQ plots for no contamination vs 2%, 5% and 10% random contamination under $K=50$ and $m=20$.]
Figure 43: $K = 100$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=100$ and $m=20$

Figure 44: $K = 200$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=20$
Figure 45: $K = 50$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=50$ and $m=50$

2nd Largest Order Statistics

Cluster Cook's D: ALPHAs

0.05 0.10 0.15 0.20 0.25

C0 C3 C5 C10

Figure 46: $K = 100$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=100$ and $m=50$

2nd Largest Order Statistics
Figure 47: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K = 200$ and $m = 50$
QQ Of Random versus No Contamination: $D_{cls, \beta,(K-1)}$

Figure 48: $K = 50$ and $m = 5$

Figure 49: $K = 100$ and $m = 5$
Figure 50: $K = 200$ and $m = 5$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=5$

![Graph showing QQ plot](image)

Figure 51: $K = 50$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=50$ and $m=20$

![Graph showing QQ plot](image)
Figure 52: $K = 100$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=100$ and $m=20$

2nd Largest Order Statistics
Cluster Cook's D: BETAs

0.04 0.06 0.08 0.10

Figure 53: $K = 200$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=20$

2nd Largest Order Statistics
Cluster Cook's D: BETAs

0.02 0.03 0.04 0.05 0.06
Figure 54: $K = 50$ and $m = 50$

![Graph showing QQ plot for 2%, 5%, and 10% random contamination under $K=50$ and $m=50$.]

Figure 55: $K = 100$ and $m = 50$

![Graph showing QQ plot for 2%, 5%, and 10% random contamination under $K=100$ and $m=50$.]
Figure 56: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=50$
QQ Of Cluster versus No Contamination: $D_{\text{cls},\alpha,(K-1)}$

Figure 57: $K = 50$ and $m = 5$

Figure 58: $K = 100$ and $m = 5$
Figure 59: \( K = 200 \) and \( m = 5 \)

![Graph showing QQ: No Contamination vs 2%, 5% and 10% cluster contamination under \( K=200 \) and \( m=5 \).]

Figure 60: \( K = 50 \) and \( m = 20 \)

![Graph showing QQ: No Contamination vs 2%, 5% and 10% cluster contamination under \( K=50 \) and \( m=20 \).]
Figure 61: $K = 100$ and $m = 20$

![Graph showing QQ plots for no contamination vs 2%, 5%, and 10% cluster contamination under $K=100$ and $m=20$.]

Figure 62: $K = 200$ and $m = 20$

![Graph showing QQ plots for no contamination vs 2% and 5% cluster contamination under $K=200$ and $m=20$.]
Figure 63: $K = 50$ and $m = 50$

Figure 64: $K = 100$ and $m = 50$
Figure 65: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=200$ and $m=50$
QQ Of Cluster versus No Contamination: $D_{\text{cls}, \beta, (K-1)}$
Figure 68: $K = 200$ and $m = 5$

Figure 69: $K = 50$ and $m = 20$
Figure 70: $K = 100$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=100$ and $m=20$

2nd Largest Order Statistics

Cluster Cook’s D: BETAs

Figure 71: $K = 200$ and $m = 20$

QQ: No Contamination vs 2% and 5% cluster contamination under $K=200$ and $m=20$

2nd Largest Order Statistics
Figure 72: $K = 50$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=50$ and $m=50$

Figure 73: $K = 100$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=100$ and $m=50$
Figure 74: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=200$ and $m=50$
I QQ Of Random versus No Contamination: $D_{\text{cls, } \alpha, (K-2)}$

Figure 75: $K = 50$ and $m = 5$

Figure 76: $K = 100$ and $m = 5$
Figure 77: $K = 200$ and $m = 5$

Figure 78: $K = 50$ and $m = 20$
Figure 79: $K = 100$ and $m = 20$

Figure 80: $K = 200$ and $m = 20$
Figure 81: $K = 50$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=50$ and $m=50$

Figure 82: $K = 100$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=100$ and $m=50$
Figure 83: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=50$
Figure 84: $K = 50$ and $m = 5$

Figure 85: $K = 100$ and $m = 5$
Figure 86: $K = 200$ and $m = 5$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=5$

3rd Largest Order Statistics

Cluster Cook's D: BETAs

Figure 87: $K = 50$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=50$ and $m=20$
Figure 88: $K = 100$ and $m = 20$

Figure 89: $K = 200$ and $m = 20$
Figure 90: $K = 50$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=50$ and $m=50$

Figure 91: $K = 100$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=100$ and $m=50$
Figure 92: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=50$

Cluster Cook’s D: BETAs

$C_0$  $C_2$  $C_5$  $C_{10}$
Figure 93: $K = 50$ and $m = 5$

Figure 94: $K = 100$ and $m = 5$
Figure 95: $K = 200$ and $m = 5$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=200$ and $m=5$

Cluster Cook's D: ALPHAs

Figure 96: $K = 50$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=50$ and $m=20$
Figure 97: $K = 100$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=100$ and $m=20$

Cluster Cook's D: ALPHAs

Figure 98: $K = 200$ and $m = 20$

QQ: No Contamination vs 2% and 5% cluster contamination under $K=200$ and $m=20$
Figure 99: $K = 50$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=50$ and $m=50$

Figure 100: $K = 100$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=100$ and $m=50$
Figure 101: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=200$ and $m=50$
QQ Of Cluster versus No Contamination: $D_{cls\beta,(K-2)}$

Figure 102: $K = 50$ and $m = 5$

Figure 103: $K = 100$ and $m = 5$
Figure 104: $K = 200$ and $m = 5$

Figure 105: $K = 50$ and $m = 20$
Figure 106: $K = 100$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=100$ and $m=20$

Cluster Cook's D: BETAs

Figure 107: $K = 200$ and $m = 20$

QQ: No Contamination vs 2% and 5% cluster contamination under $K=200$ and $m=20$
Figure 108: $K = 50$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=50$ and $m=50$

Figure 109: $K = 100$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=100$ and $m=50$
Figure 110: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=200$ and $m=50$
QQ Of Random versus No Contamination: $D_{\text{cls,}\alpha,(K-3)}$

Figure 111: $K = 50$ and $m = 5$

Figure 112: $K = 100$ and $m = 5$
Figure 113: $K = 200$ and $m = 5$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=5$

Figure 114: $K = 50$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=50$ and $m=20$
Figure 115: $K = 100$ and $m = 20$

Figure 116: $K = 200$ and $m = 20$
Figure 117: $K = 50$ and $m = 50$

Figure 118: $K = 100$ and $m = 50$
Figure 119: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=50$
Figure 120: $K = 50$ and $m = 5$

Figure 121: $K = 100$ and $m = 5$
Figure 122: $K = 200$ and $m = 5$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=5$

Figure 123: $K = 50$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=50$ and $m=20$
Figure 124: $K = 100$ and $m = 20$

QQ: No Contamination vs 2%, 5%, and 10% random contamination under $K=100$ and $m=20$

Figure 125: $K = 200$ and $m = 20$

QQ: No Contamination vs 2%, 5%, and 10% random contamination under $K=200$ and $m=20$
Figure 126: $K = 50$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=50$ and $m=50$

4th Largest Order Statistics

Cluster Cook's D: BETAs

0.04
0.06
0.08
0.10

0.04 0.05 0.06 0.07 0.08 0.09 0.10

C0 C2 C5 C10

Figure 127: $K = 100$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=100$ and $m=50$

4th Largest Order Statistics

Cluster Cook's D: BETAs

0.03
0.04
0.05
0.025 0.030 0.035 0.040 0.045 0.050

81

Hosted by The Berkeley Electronic Press
Figure 128: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% random contamination under $K=200$ and $m=50$
QQ Of Cluster versus No Contamination: $D_{cls, \alpha, (K-3)}$

Figure 129: $K = 50$ and $m = 5$

Figure 130: $K = 100$ and $m = 5$
Figure 131: $K = 200$ and $m = 5$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=200$ and $m=5$

Figure 132: $K = 50$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=50$ and $m=20$
Figure 133: $K = 100$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under K=100 and m=20

4th Largest Order Statistics

Cluster Cook's D: ALPHAs

Figure 134: $K = 200$ and $m = 20$

QQ: No Contamination vs 2% and 5% cluster contamination under K=200 and m=20

4th Largest Order Statistics
Figure 135: $K = 50$ and $m = 50$

Figure 136: $K = 100$ and $m = 50$
Figure 137: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=200$ and $m=50$
P QQ Of Cluster versus No Contamination: \( D_{cls,\beta,(K-3)} \)

Figure 138: \( K = 50 \) and \( m = 5 \)

Figure 139: \( K = 100 \) and \( m = 5 \)

http://biostats.bepress.com/uncbiostat/art9
Figure 140: $K = 200$ and $m = 5$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=200$ and $m=5$

4th Largest Order Statistics

Cluster Cook’s D: BETAs

Figure 141: $K = 50$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=50$ and $m=20$
Figure 142: $K = 100$ and $m = 20$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=100$ and $m=20$

Figure 143: $K = 200$ and $m = 20$

QQ: No Contamination vs 2% and 5% cluster contamination under $K=200$ and $m=20$
Figure 144: $K = 50$ and $m = 50$

Figure 145: $K = 100$ and $m = 50$
Figure 146: $K = 200$ and $m = 50$

QQ: No Contamination vs 2%, 5% and 10% cluster contamination under $K=200$ and $m=50$
QQ Of Random versus No Contamination: $D_{\text{cls.}\alpha.(K)} - D_{\text{cls.}\alpha.(K-1)}$

Figure 147: $K = 50$ and $m = 5$

Figure 148: $K = 100$ and $m = 5$
Figure 149: $K = 200$ and $m = 5$

Figure 150: $K = 50$ and $m = 20$
Figure 151: $K = 100$ and $m = 20$

Cluster Concentrated Difference of Cook's $D$: ALPHAs

Figure 152: $K = 200$ and $m = 20$

Cluster Concentrated Difference of Cook's $D$: ALPHAs
Figure 153: $K = 50$ and $m = 50$

Figure 154: $K = 100$ and $m = 50$
Figure 155: $K = 200$ and $m = 50$

QQ of $D(K) - D(K-1)$: No Contamination vs 2%, 5% and 10% RANDOM contamination under $K=200$ and $m=50$
QQ Of Random versus No Contamination: $D_{\text{cls},\beta, (K)} - D_{\text{cls},\beta, (K-1)}$

Figure 156: $K = 50$ and $m = 5$

Figure 157: $K = 100$ and $m = 5$
Figure 158: $K = 200$ and $m = 5$

![Graph showing QQ of $D_K - D_{K-1}$: No Contamination vs 2%, 5% and 10% RANDOM contamination under $K=200$ and $m=5$.]

Figure 159: $K = 50$ and $m = 20$

![Graph showing QQ of $D_K - D_{K-1}$: No Contamination vs 2%, 5% and 10% RANDOM contamination under $K=50$ and $m=20$.]
Figure 160: $K = 100$ and $m = 20$

Figure 161: $K = 200$ and $m = 20$
Figure 162: $K = 50$ and $m = 50$

Figure 163: $K = 100$ and $m = 50$
Figure 164: $K = 200$ and $m = 50$

QQ of $D_{(K)} - D_{(K-1)}$: No Contamination vs 2%, 5% and 10% RANDOM contamination under $K=200$ and $m=50$
QQ Of Cluster versus No Contamination: $D_{\text{cls},\alpha,(K)} - D_{\text{cls},\alpha,(K-1)}$

Figure 165: $K = 50$ and $m = 5$

Figure 166: $K = 100$ and $m = 5$
Figure 167: $K = 200$ and $m = 5$

QQ of $D_K - D_{K-1}$: No Contamination vs 2%, 5% and 10% CLUSTER contamination under $K=200$ and $m=5$

Figure 168: $K = 50$ and $m = 20$

QQ of $D_K - D_{K-1}$: No Contamination vs 2%, 5% and 10% CLUSTER contamination under $K=50$ and $m=20$
Figure 169: $K = 100$ and $m = 20$

Figure 170: $K = 200$ and $m = 20$
Figure 171: $K = 50$ and $m = 50$

DiffA0

Cluster Concentrated

Figure 172: $K = 100$ and $m = 50$

DiffA0

Cluster Concentrated

http://biostats.bepress.com/uncbiostat/art9
Figure 173: $K = 200$ and $m = 50$
QQ Of Cluster versus No Contamination: \( D_{\text{cls}, \beta (K)} - D_{\text{cls}, \beta (K-1)} \)

Figure 174: \( K = 50 \) and \( m = 5 \)

Figure 175: \( K = 100 \) and \( m = 5 \)
Figure 176: $K = 200$ and $m = 5$

![Graph showing QQ of $D_K - D_{K-1}$: No Contamination vs 2%, 5% and 10% CLUSTER contamination under $K=200$ and $m=5$.]

Figure 177: $K = 50$ and $m = 20$

![Graph showing QQ of $D_K - D_{K-1}$: No Contamination vs 2%, 5% and 10% CLUSTER contamination under $K=50$ and $m=20$.]
Figure 178: $K = 100$ and $m = 20$

QQ of $D_K - D_{K-1}$: No Contamination vs 2%, 5% and 10% CLUSTER contamination under $K=100$ and $m=20$

Figure 179: $K = 200$ and $m = 20$

QQ of $D_K - D_{K-1}$: No Contamination vs 2%, 5% and 10% CLUSTER contamination under $K=200$ and $m=20$
Figure 180: $K = 50$ and $m = 50$

QQ of $D_{(K)} - D_{(K-1)}$: No Contamination vs 2%, 5% and 10% CLUSTER contamination under $K=50$ and $m=50$

Figure 181: $K = 100$ and $m = 50$

QQ of $D_{(K)} - D_{(K-1)}$: No Contamination vs 2%, 5% and 10% CLUSTER contamination under $K=100$ and $m=50$
Figure 182: $K = 200$ and $m = 50$

QQ of $D_{(K)} - D_{(K-1)}$: No Contamination vs 2%, 5% and 10% CLUSTER contamination under $K=200$ and $m=50$
Table 2: Summary description of the largest order statistics for the cluster Cooks distance for $\alpha$ under random contamination

<table>
<thead>
<tr>
<th>$K$</th>
<th>$m$</th>
<th>Contam</th>
<th>Level</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>$F^{-1}(.50)$</th>
<th>$F^{-1}(.75)$</th>
<th>$F^{-1}(.90)$</th>
<th>$F^{-1}(.95)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.138</td>
<td>0.050</td>
<td>0.32858</td>
<td>0.13325</td>
<td>0.16933</td>
<td>0.20836</td>
<td>0.227930</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.138</td>
<td>0.050</td>
<td>0.31449</td>
<td>0.13156</td>
<td>0.16796</td>
<td>0.20732</td>
<td>0.230000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.141</td>
<td>0.050</td>
<td>0.28937</td>
<td>0.13524</td>
<td>0.17552</td>
<td>0.21481</td>
<td>0.233124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.146</td>
<td>0.061</td>
<td>0.31151</td>
<td>0.13577</td>
<td>0.17624</td>
<td>0.21631</td>
<td>0.240407</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.101</td>
<td>0.036</td>
<td>0.26346</td>
<td>0.09117</td>
<td>0.12119</td>
<td>0.15629</td>
<td>0.185911</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.100</td>
<td>0.034</td>
<td>0.25841</td>
<td>0.09209</td>
<td>0.12021</td>
<td>0.15753</td>
<td>0.181967</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.104</td>
<td>0.033</td>
<td>0.25479</td>
<td>0.09725</td>
<td>0.12608</td>
<td>0.16233</td>
<td>0.192327</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.104</td>
<td>0.031</td>
<td>0.26237</td>
<td>0.09924</td>
<td>0.12909</td>
<td>0.15695</td>
<td>0.175917</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.152</td>
<td>0.066</td>
<td>0.30659</td>
<td>0.14135</td>
<td>0.18595</td>
<td>0.22165</td>
<td>0.247883</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.156</td>
<td>0.071</td>
<td>0.30751</td>
<td>0.14730</td>
<td>0.19088</td>
<td>0.22540</td>
<td>0.250864</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.161</td>
<td>0.070</td>
<td>0.32014</td>
<td>0.15441</td>
<td>0.19104</td>
<td>0.23070</td>
<td>0.259640</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.169</td>
<td>0.061</td>
<td>0.31873</td>
<td>0.16493</td>
<td>0.19766</td>
<td>0.24148</td>
<td>0.268005</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>0.114</td>
<td>0.044</td>
<td>0.26234</td>
<td>0.10613</td>
<td>0.13556</td>
<td>0.17113</td>
<td>0.207480</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.117</td>
<td>0.049</td>
<td>0.26226</td>
<td>0.10794</td>
<td>0.13935</td>
<td>0.17909</td>
<td>0.204771</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.121</td>
<td>0.047</td>
<td>0.27982</td>
<td>0.11453</td>
<td>0.14481</td>
<td>0.17566</td>
<td>0.207378</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.124</td>
<td>0.055</td>
<td>0.29814</td>
<td>0.11483</td>
<td>0.14786</td>
<td>0.18778</td>
<td>0.210817</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>0.077</td>
<td>0.033</td>
<td>0.20833</td>
<td>0.07150</td>
<td>0.09167</td>
<td>0.11239</td>
<td>0.131313</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.079</td>
<td>0.029</td>
<td>0.21625</td>
<td>0.07375</td>
<td>0.09169</td>
<td>0.11627</td>
<td>0.136236</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.082</td>
<td>0.033</td>
<td>0.19985</td>
<td>0.07758</td>
<td>0.09700</td>
<td>0.12119</td>
<td>0.137655</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.087</td>
<td>0.028</td>
<td>0.24714</td>
<td>0.07914</td>
<td>0.10700</td>
<td>0.12516</td>
<td>0.136374</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.156</td>
<td>0.064</td>
<td>0.32892</td>
<td>0.14705</td>
<td>0.18596</td>
<td>0.23192</td>
<td>0.253310</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.159</td>
<td>0.064</td>
<td>0.32872</td>
<td>0.14919</td>
<td>0.19048</td>
<td>0.23573</td>
<td>0.259069</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.167</td>
<td>0.063</td>
<td>0.32346</td>
<td>0.15862</td>
<td>0.19919</td>
<td>0.24530</td>
<td>0.266341</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.175</td>
<td>0.071</td>
<td>0.33144</td>
<td>0.16606</td>
<td>0.21264</td>
<td>0.25424</td>
<td>0.277163</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.114</td>
<td>0.047</td>
<td>0.28971</td>
<td>0.10582</td>
<td>0.13755</td>
<td>0.17125</td>
<td>0.193330</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.118</td>
<td>0.046</td>
<td>0.28519</td>
<td>0.11093</td>
<td>0.13892</td>
<td>0.17210</td>
<td>0.198636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.123</td>
<td>0.044</td>
<td>0.28696</td>
<td>0.11678</td>
<td>0.14851</td>
<td>0.18105</td>
<td>0.207132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.128</td>
<td>0.058</td>
<td>0.29177</td>
<td>0.12100</td>
<td>0.15431</td>
<td>0.18447</td>
<td>0.197179</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.081</td>
<td>0.031</td>
<td>0.25708</td>
<td>0.07731</td>
<td>0.09399</td>
<td>0.12361</td>
<td>0.139310</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.083</td>
<td>0.033</td>
<td>0.25738</td>
<td>0.07732</td>
<td>0.09956</td>
<td>0.12460</td>
<td>0.143962</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.086</td>
<td>0.034</td>
<td>0.23842</td>
<td>0.08048</td>
<td>0.10078</td>
<td>0.12710</td>
<td>0.148484</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.091</td>
<td>0.040</td>
<td>0.23671</td>
<td>0.08489</td>
<td>0.11055</td>
<td>0.13298</td>
<td>0.155214</td>
</tr>
</tbody>
</table>
Table 3: Summary description of the largest order statistics for the cluster Cooks distance for $\beta$ under random contamination

<table>
<thead>
<tr>
<th>$K$</th>
<th>$m$</th>
<th>Contam</th>
<th>Level</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>$F^{-1}(0.50)$</th>
<th>$F^{-1}(0.75)$</th>
<th>$F^{-1}(0.90)$</th>
<th>$F^{-1}(0.95)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.107</td>
<td>0.052</td>
<td>0.27892</td>
<td>0.10117</td>
<td>0.12470</td>
<td>0.15599</td>
<td>0.174572</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.106</td>
<td>0.049</td>
<td>0.35723</td>
<td>0.10075</td>
<td>0.12301</td>
<td>0.15010</td>
<td>0.167269</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.105</td>
<td>0.049</td>
<td>0.24713</td>
<td>0.09977</td>
<td>0.12504</td>
<td>0.14652</td>
<td>0.162036</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.105</td>
<td>0.052</td>
<td>0.23073</td>
<td>0.09960</td>
<td>0.11881</td>
<td>0.14662</td>
<td>0.165335</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.062</td>
<td>0.026</td>
<td>0.14489</td>
<td>0.05787</td>
<td>0.07076</td>
<td>0.08880</td>
<td>0.103129</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.061</td>
<td>0.031</td>
<td>0.14027</td>
<td>0.05722</td>
<td>0.06977</td>
<td>0.08377</td>
<td>0.095301</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.062</td>
<td>0.028</td>
<td>0.18850</td>
<td>0.05839</td>
<td>0.07183</td>
<td>0.08629</td>
<td>0.100849</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.060</td>
<td>0.030</td>
<td>0.15494</td>
<td>0.05761</td>
<td>0.06912</td>
<td>0.08172</td>
<td>0.094232</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.112</td>
<td>0.051</td>
<td>0.40909</td>
<td>0.10481</td>
<td>0.12791</td>
<td>0.15978</td>
<td>0.180466</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.113</td>
<td>0.056</td>
<td>0.77904</td>
<td>0.10392</td>
<td>0.12718</td>
<td>0.16123</td>
<td>0.182519</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.131</td>
<td>0.050</td>
<td>7.94838</td>
<td>0.10608</td>
<td>0.13505</td>
<td>0.16782</td>
<td>0.188912</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.114</td>
<td>0.056</td>
<td>0.31563</td>
<td>0.10562</td>
<td>0.13279</td>
<td>0.15898</td>
<td>0.185884</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>0.068</td>
<td>0.032</td>
<td>0.24039</td>
<td>0.06286</td>
<td>0.07914</td>
<td>0.10113</td>
<td>0.111527</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.068</td>
<td>0.031</td>
<td>0.23269</td>
<td>0.06165</td>
<td>0.07781</td>
<td>0.09736</td>
<td>0.111308</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.068</td>
<td>0.031</td>
<td>0.21949</td>
<td>0.06168</td>
<td>0.07769</td>
<td>0.09576</td>
<td>0.112529</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.067</td>
<td>0.034</td>
<td>0.15621</td>
<td>0.06100</td>
<td>0.07672</td>
<td>0.09670</td>
<td>0.107481</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>0.040</td>
<td>0.018</td>
<td>0.12444</td>
<td>0.03748</td>
<td>0.04623</td>
<td>0.05752</td>
<td>0.064038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.040</td>
<td>0.017</td>
<td>0.11469</td>
<td>0.03711</td>
<td>0.04567</td>
<td>0.05631</td>
<td>0.064376</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.040</td>
<td>0.020</td>
<td>0.09113</td>
<td>0.03745</td>
<td>0.04591</td>
<td>0.05686</td>
<td>0.063982</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.039</td>
<td>0.021</td>
<td>0.08591</td>
<td>0.03588</td>
<td>0.04470</td>
<td>0.05616</td>
<td>0.062812</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.123</td>
<td>0.055</td>
<td>0.37273</td>
<td>0.10962</td>
<td>0.14366</td>
<td>0.17696</td>
<td>0.232888</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.122</td>
<td>0.055</td>
<td>0.30798</td>
<td>0.11016</td>
<td>0.14170</td>
<td>0.17787</td>
<td>0.210755</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.122</td>
<td>0.055</td>
<td>0.40527</td>
<td>0.11008</td>
<td>0.13765</td>
<td>0.18274</td>
<td>0.231667</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.122</td>
<td>0.050</td>
<td>0.95400</td>
<td>0.11161</td>
<td>0.13764</td>
<td>0.17457</td>
<td>0.201059</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.070</td>
<td>0.036</td>
<td>0.20198</td>
<td>0.06306</td>
<td>0.07997</td>
<td>0.10062</td>
<td>0.120855</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.070</td>
<td>0.033</td>
<td>0.20428</td>
<td>0.06248</td>
<td>0.08021</td>
<td>0.10299</td>
<td>0.121418</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.070</td>
<td>0.033</td>
<td>0.22152</td>
<td>0.06229</td>
<td>0.08248</td>
<td>0.10435</td>
<td>0.120647</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.071</td>
<td>0.035</td>
<td>0.24011</td>
<td>0.06381</td>
<td>0.08152</td>
<td>0.10552</td>
<td>0.130785</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.041</td>
<td>0.019</td>
<td>0.13520</td>
<td>0.03847</td>
<td>0.04762</td>
<td>0.05803</td>
<td>0.061122</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.02</td>
<td>0.041</td>
<td>0.020</td>
<td>0.13417</td>
<td>0.03840</td>
<td>0.04826</td>
<td>0.05706</td>
<td>0.063042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.05</td>
<td>0.041</td>
<td>0.021</td>
<td>0.12308</td>
<td>0.03817</td>
<td>0.04764</td>
<td>0.05839</td>
<td>0.066215</td>
</tr>
<tr>
<td></td>
<td></td>
<td>obsv</td>
<td>0.10</td>
<td>0.042</td>
<td>0.019</td>
<td>0.13011</td>
<td>0.03865</td>
<td>0.04790</td>
<td>0.05833</td>
<td>0.066181</td>
</tr>
</tbody>
</table>
Table 4: Summary description of the largest order statistics for the cluster Cooks distance for \( \alpha \) under cluster level contamination

<table>
<thead>
<tr>
<th>( K )</th>
<th>( m )</th>
<th>Contam</th>
<th>Level</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>( F^{-1}(0.50) )</th>
<th>( F^{-1}(0.75) )</th>
<th>( F^{-1}(0.90) )</th>
<th>( F^{-1}(0.95) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.138</td>
<td>0.050</td>
<td>0.32858</td>
<td>0.13325</td>
<td>0.16933</td>
<td>0.20836</td>
<td>0.227930</td>
</tr>
<tr>
<td>clus</td>
<td>0.02</td>
<td>0.136</td>
<td>0.051</td>
<td>0.33491</td>
<td>0.13005</td>
<td>0.16797</td>
<td>0.20357</td>
<td>0.226914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.05</td>
<td>0.132</td>
<td>0.051</td>
<td>0.29300</td>
<td>0.12513</td>
<td>0.16191</td>
<td>0.19845</td>
<td>0.216316</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.10</td>
<td>0.128</td>
<td>0.054</td>
<td>0.30606</td>
<td>0.11998</td>
<td>0.15629</td>
<td>0.19166</td>
<td>0.215929</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.101</td>
<td>0.036</td>
<td>0.26346</td>
<td>0.09117</td>
<td>0.12119</td>
<td>0.15629</td>
<td>0.185911</td>
</tr>
<tr>
<td>clus</td>
<td>0.02</td>
<td>0.099</td>
<td>0.035</td>
<td>0.26010</td>
<td>0.08931</td>
<td>0.12014</td>
<td>0.15559</td>
<td>0.180696</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.05</td>
<td>0.097</td>
<td>0.035</td>
<td>0.26852</td>
<td>0.09017</td>
<td>0.11542</td>
<td>0.15143</td>
<td>0.177766</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.10</td>
<td>0.092</td>
<td>0.027</td>
<td>0.21236</td>
<td>0.08464</td>
<td>0.11169</td>
<td>0.14497</td>
<td>0.161570</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.152</td>
<td>0.066</td>
<td>0.30659</td>
<td>0.14135</td>
<td>0.18595</td>
<td>0.22165</td>
<td>0.247883</td>
</tr>
<tr>
<td>clus</td>
<td>0.02</td>
<td>0.150</td>
<td>0.065</td>
<td>0.30786</td>
<td>0.13976</td>
<td>0.17824</td>
<td>0.22157</td>
<td>0.244428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.05</td>
<td>0.145</td>
<td>0.063</td>
<td>0.30144</td>
<td>0.13790</td>
<td>0.17148</td>
<td>0.20861</td>
<td>0.229517</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.10</td>
<td>0.142</td>
<td>0.061</td>
<td>0.30417</td>
<td>0.13588</td>
<td>0.16535</td>
<td>0.20262</td>
<td>0.224237</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>0.114</td>
<td>0.044</td>
<td>0.26234</td>
<td>0.10613</td>
<td>0.13556</td>
<td>0.17113</td>
<td>0.207480</td>
</tr>
<tr>
<td>clus</td>
<td>0.02</td>
<td>0.111</td>
<td>0.042</td>
<td>0.25415</td>
<td>0.10255</td>
<td>0.13171</td>
<td>0.16281</td>
<td>0.190164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.05</td>
<td>0.103</td>
<td>0.038</td>
<td>0.25848</td>
<td>0.09329</td>
<td>0.12573</td>
<td>0.15529</td>
<td>0.178171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.10</td>
<td>0.100</td>
<td>0.043</td>
<td>0.24790</td>
<td>0.08840</td>
<td>0.11575</td>
<td>0.15345</td>
<td>0.178626</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>0.077</td>
<td>0.033</td>
<td>0.20833</td>
<td>0.07150</td>
<td>0.09167</td>
<td>0.11239</td>
<td>0.131313</td>
</tr>
<tr>
<td>clus</td>
<td>0.02</td>
<td>0.074</td>
<td>0.030</td>
<td>0.18365</td>
<td>0.06873</td>
<td>0.08771</td>
<td>0.10900</td>
<td>0.131639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.05</td>
<td>0.070</td>
<td>0.029</td>
<td>0.19956</td>
<td>0.06419</td>
<td>0.08265</td>
<td>0.10249</td>
<td>0.117528</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.10</td>
<td>0.064</td>
<td>0.028</td>
<td>0.17199</td>
<td>0.05947</td>
<td>0.07590</td>
<td>0.09629</td>
<td>0.109375</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.156</td>
<td>0.064</td>
<td>0.32892</td>
<td>0.14705</td>
<td>0.18596</td>
<td>0.23192</td>
<td>0.253310</td>
</tr>
<tr>
<td>clus</td>
<td>0.02</td>
<td>0.154</td>
<td>0.068</td>
<td>0.31712</td>
<td>0.14201</td>
<td>0.18094</td>
<td>0.23109</td>
<td>0.251295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.05</td>
<td>0.149</td>
<td>0.072</td>
<td>0.32475</td>
<td>0.13887</td>
<td>0.17560</td>
<td>0.21932</td>
<td>0.247127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.10</td>
<td>0.150</td>
<td>0.067</td>
<td>0.32961</td>
<td>0.13984</td>
<td>0.18013</td>
<td>0.22058</td>
<td>0.243865</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.114</td>
<td>0.047</td>
<td>0.28971</td>
<td>0.10582</td>
<td>0.13755</td>
<td>0.17125</td>
<td>0.193330</td>
</tr>
<tr>
<td>clus</td>
<td>0.02</td>
<td>0.111</td>
<td>0.038</td>
<td>0.28542</td>
<td>0.10295</td>
<td>0.13395</td>
<td>0.16327</td>
<td>0.185109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.05</td>
<td>0.109</td>
<td>0.042</td>
<td>0.29169</td>
<td>0.09962</td>
<td>0.12791</td>
<td>0.16498</td>
<td>0.185685</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.10</td>
<td>0.104</td>
<td>0.035</td>
<td>0.24597</td>
<td>0.09652</td>
<td>0.12053</td>
<td>0.15569</td>
<td>0.180185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.081</td>
<td>0.031</td>
<td>0.25708</td>
<td>0.07731</td>
<td>0.09399</td>
<td>0.12361</td>
<td>0.139310</td>
</tr>
<tr>
<td>clus</td>
<td>0.02</td>
<td>0.078</td>
<td>0.031</td>
<td>0.17248</td>
<td>0.07283</td>
<td>0.09109</td>
<td>0.11968</td>
<td>0.137218</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.05</td>
<td>0.074</td>
<td>0.034</td>
<td>0.16555</td>
<td>0.07000</td>
<td>0.08854</td>
<td>0.11129</td>
<td>0.131026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clus</td>
<td>0.10</td>
<td>0.069</td>
<td>0.032</td>
<td>0.19639</td>
<td>0.06448</td>
<td>0.08029</td>
<td>0.09928</td>
<td>0.113796</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5: Summary description of the largest order statistics for the cluster Cook’s distance for $\beta$ under cluster level contamination

<table>
<thead>
<tr>
<th>$K$</th>
<th>$m$</th>
<th>Contam</th>
<th>Level</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>$F^{-1}(0.50)$</th>
<th>$F^{-1}(0.75)$</th>
<th>$F^{-1}(0.90)$</th>
<th>$F^{-1}(0.95)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.107</td>
<td>0.052</td>
<td>0.27892</td>
<td>0.10117</td>
<td>0.12470</td>
<td>0.15599</td>
<td>0.174572</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.02</td>
<td>0.105</td>
<td>0.050</td>
<td>0.29564</td>
<td>0.09937</td>
<td>0.12162</td>
<td>0.14524</td>
<td>0.164175</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.05</td>
<td>0.103</td>
<td>0.050</td>
<td>0.25730</td>
<td>0.09603</td>
<td>0.12031</td>
<td>0.14517</td>
<td>0.160875</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.10</td>
<td>0.101</td>
<td>0.056</td>
<td>0.31944</td>
<td>0.09469</td>
<td>0.11496</td>
<td>0.13947</td>
<td>0.158575</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.062</td>
<td>0.026</td>
<td>0.14489</td>
<td>0.05787</td>
<td>0.07076</td>
<td>0.08880</td>
<td>0.103129</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.02</td>
<td>0.061</td>
<td>0.029</td>
<td>0.13809</td>
<td>0.05707</td>
<td>0.06955</td>
<td>0.08446</td>
<td>0.096307</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.05</td>
<td>0.060</td>
<td>0.032</td>
<td>0.13261</td>
<td>0.05668</td>
<td>0.06898</td>
<td>0.08356</td>
<td>0.099469</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.10</td>
<td>0.057</td>
<td>0.028</td>
<td>0.12704</td>
<td>0.05438</td>
<td>0.06537</td>
<td>0.07748</td>
<td>0.086974</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>none</td>
<td>0.00</td>
<td>0.037</td>
<td>0.020</td>
<td>0.08359</td>
<td>0.03427</td>
<td>0.04294</td>
<td>0.05189</td>
<td>0.057868</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.02</td>
<td>0.037</td>
<td>0.018</td>
<td>0.08393</td>
<td>0.03379</td>
<td>0.04231</td>
<td>0.05121</td>
<td>0.057471</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.05</td>
<td>0.035</td>
<td>0.017</td>
<td>0.07017</td>
<td>0.03273</td>
<td>0.04029</td>
<td>0.04827</td>
<td>0.053835</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.10</td>
<td>0.033</td>
<td>0.016</td>
<td>0.06717</td>
<td>0.03125</td>
<td>0.03691</td>
<td>0.04381</td>
<td>0.048076</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>0.112</td>
<td>0.051</td>
<td>0.40909</td>
<td>0.10481</td>
<td>0.12791</td>
<td>0.15978</td>
<td>0.180466</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.02</td>
<td>0.111</td>
<td>0.052</td>
<td>0.81076</td>
<td>0.10222</td>
<td>0.12454</td>
<td>0.15504</td>
<td>0.176717</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.05</td>
<td>0.103</td>
<td>0.054</td>
<td>0.24977</td>
<td>0.09764</td>
<td>0.11725</td>
<td>0.13981</td>
<td>0.154573</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.10</td>
<td>0.101</td>
<td>0.053</td>
<td>0.27006</td>
<td>0.09343</td>
<td>0.11664</td>
<td>0.13936</td>
<td>0.159773</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>0.068</td>
<td>0.032</td>
<td>0.24039</td>
<td>0.06286</td>
<td>0.07914</td>
<td>0.10113</td>
<td>0.111527</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.02</td>
<td>0.066</td>
<td>0.030</td>
<td>0.20085</td>
<td>0.06185</td>
<td>0.07493</td>
<td>0.09351</td>
<td>0.10719</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.05</td>
<td>0.061</td>
<td>0.032</td>
<td>0.15397</td>
<td>0.05795</td>
<td>0.06991</td>
<td>0.08404</td>
<td>0.096392</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.10</td>
<td>0.058</td>
<td>0.029</td>
<td>0.13151</td>
<td>0.05462</td>
<td>0.06561</td>
<td>0.08035</td>
<td>0.090445</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>none</td>
<td>0.00</td>
<td>0.040</td>
<td>0.018</td>
<td>0.12444</td>
<td>0.03748</td>
<td>0.04623</td>
<td>0.05752</td>
<td>0.064038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.02</td>
<td>0.038</td>
<td>0.019</td>
<td>0.09988</td>
<td>0.03568</td>
<td>0.04369</td>
<td>0.05295</td>
<td>0.059590</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.05</td>
<td>0.036</td>
<td>0.020</td>
<td>0.08387</td>
<td>0.03325</td>
<td>0.04067</td>
<td>0.04794</td>
<td>0.053614</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.10</td>
<td>0.032</td>
<td>0.017</td>
<td>0.07037</td>
<td>0.03076</td>
<td>0.03664</td>
<td>0.04336</td>
<td>0.047663</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.123</td>
<td>0.055</td>
<td>0.37273</td>
<td>0.10962</td>
<td>0.14366</td>
<td>0.17696</td>
<td>0.232888</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.02</td>
<td>0.117</td>
<td>0.049</td>
<td>0.47444</td>
<td>0.10549</td>
<td>0.13448</td>
<td>0.16530</td>
<td>0.210813</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.05</td>
<td>0.112</td>
<td>0.055</td>
<td>0.79929</td>
<td>0.10148</td>
<td>0.12431</td>
<td>0.15647</td>
<td>0.185496</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.10</td>
<td>0.105</td>
<td>0.048</td>
<td>0.24908</td>
<td>0.09846</td>
<td>0.11955</td>
<td>0.14784</td>
<td>0.172983</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.070</td>
<td>0.036</td>
<td>0.20198</td>
<td>0.06306</td>
<td>0.07997</td>
<td>0.10062</td>
<td>0.120855</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.02</td>
<td>0.066</td>
<td>0.033</td>
<td>0.18552</td>
<td>0.05966</td>
<td>0.07359</td>
<td>0.09637</td>
<td>0.110233</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.05</td>
<td>0.062</td>
<td>0.030</td>
<td>0.15849</td>
<td>0.05664</td>
<td>0.06983</td>
<td>0.08828</td>
<td>0.099209</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.10</td>
<td>0.056</td>
<td>0.033</td>
<td>0.13917</td>
<td>0.05349</td>
<td>0.06262</td>
<td>0.07426</td>
<td>0.088979</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>none</td>
<td>0.00</td>
<td>0.041</td>
<td>0.019</td>
<td>0.13520</td>
<td>0.03847</td>
<td>0.04762</td>
<td>0.05803</td>
<td>0.061122</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.02</td>
<td>0.039</td>
<td>0.019</td>
<td>0.11042</td>
<td>0.03626</td>
<td>0.04383</td>
<td>0.05295</td>
<td>0.057426</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.05</td>
<td>0.035</td>
<td>0.020</td>
<td>0.08525</td>
<td>0.03304</td>
<td>0.03922</td>
<td>0.04580</td>
<td>0.053695</td>
</tr>
<tr>
<td></td>
<td></td>
<td>clus</td>
<td>0.10</td>
<td>0.031</td>
<td>0.019</td>
<td>0.08658</td>
<td>0.02960</td>
<td>0.03458</td>
<td>0.03971</td>
<td>0.045054</td>
</tr>
</tbody>
</table>