Alaska
Fisheries Science
Center

National Marine
Fisheries Service

U.S. DEPARTMENT OF COMMERCE

AFSC PROCESSED REPORT 92-02

Statistical Analyses of Vessel Incentive Program Data

March 1992

ERRATA NOTICE

STATISTICAL ANALYSES OF VESSEL INCENTIVE PROGRAM DATA

by

Russell F. Kappenman

Resource Ecology and Fisheries Management Division
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way NE, BIN C15700
Seattle, WA 98115-0070

March 1992

1. Introduction

A prohibited species monthly bycatch rate for a vessel is the ratio of either the total weight or total number of the prohibited species caught by the vessel during a month to the total weight of fish taken by the vessel during the month. To make statistical inferences about a vessel's bycatch rate, some of the hauls of fish, that the vessel makes during the month, are randomly selected for sampling by an observer aboard the vessel. For each selected haul, the observer either weighs or counts all of the prohibited species present in the haul or weighs or counts all of the prohibited species present in several baskets of fish taken from randomly selected parts of the haul.

We will restrict our attention, from this point on, to the problem of making statistical inferences about a vessel's monthly bycatch rate of halibut. The data, relevant to making these inferences, consist of total weights of the baskets pooled and total weights of halibut in the pooled baskets, when all hauls selected for sampling are basket sampled. When only some of the selected hauls are basket sampled, and the others are whole haul sampled, the information used to make bycatch rate inferences include total haul weight, total basket sample weight, and weight of halibut in the pooled basket samples, for basket sampled hauls. For whole haul sampled hauls, the information used is total haul weight and total weight of the prohibited species in the haul.

Described in this report are the statistical analyses of the data taken by observers for the vessel incentive program. The goal of these analyses is useful statistical inferences about the bycatch rate of halibut for a given vessel, when the vessel fishes during the course of a given month.

The first step in the analyses of the data for a vessel-month is the application to the data of a statistical procedure which produces robust estimates of monthly bycatch rates. The term robust, here, refers to the fact that unusual haul sampling results, obtained by an observer, do not have an inordinately large effect on the bycatch rate estimate produced by the procedure. For example, the procedure is designed to produce a reliable bycatch rate estimate, when relatively few basket samples contain extremely large amounts of halibut, but the majority of the basket samples contain very little halibut. This robust bycatch rate estimation procedure is developed in Section 2.

The second step is the use of observer data to calculate a lower 95% confidence limit for a vessel's monthly bycatch rate. This confidence limit is a number such that we are 95% confident that the vessel's actual monthly bycatch rate exceeds the number. The method used for calculating the confidence limit is described in Section 3.

The methods, used for finding a bycatch rate estimate and a lower 95% confidence limit for a bycatch rate, are based on certain statistical assumptions about the nature of the observed random variables, namely total basket sample weight, prohibited species weight in a basket sample, total weight of a haul, and weight of the prohibited species in a haul. In order to determine whether or not the procedures for finding these estimates and limits are valid the reasonableness of these assumptions must be examined.

The data analysis techniques used on observer data to check the validity of the assumptions, that need to be made for application of the inferential procedures, are discussed in Section 4. Application of these techniques is the third step in the analyses of vessel incentive program data.

The fourth step is the calculation of 95% confidence limits by the use of a technique other than the procedure used in the second step. This step is discussed in Section 5, and it is meant to serve as an additional check of the reasonableness of the results obtained thus far by the data analyses. The technique is a general purpose procedure, for producing confidence limits, which may be applied without making any statistical assumptions about the variables involved.

For those vessel-month situations in which all of the hauls selected for sampling are basket sampled, one more set of analyses of the data is performed. This set essentially consists of the repetition of the first four steps on information which consists of basket sample measurements as well as weights of the hauls from which the basket samples were taken. This step serves as a final check of the reasonableness of the results obtained by the previous analyses. It is also an attempt to alleviate concerns expressed by some that total haul weight is not used in making inferences about monthly bycatch rates, when all hauls selected for sampling are basket sampled.

2. Robust Estimation of Bycatch Rates

We will consider first the case where all observer samples from hauls are pooled basket samples. Let the random variable x represent the total weight of a pooled basket sample and the random variable y represent the weight of the halibut in this pooled basket sample. The halibut bycatch rate may be defined to be the unknown constant r such that the mean of the random variable y-rx is zero.

Let $(x_1,y_1),\ldots,(x_n,y_n)$ represent n pairs of observations of x and y. These pairs of observations are to be used to make statistical inferences about the unknown parameter r, and we will first consider the problem of finding an estimate of r. The estimate of r will be the value of q such that a reasonable, robust estimate of the mean of the random variable y-qx is zero.

To be sure, one possible estimate of the mean of y-qx, for any value of q, is

$$\frac{1}{n}\sum_{i=1}^{n}(y_i-qx_i) = \bar{y}-q\bar{x} \,, \tag{1}$$

where $\bar{x}$ and $\bar{y}$ represent the averages of the n x_i's and the n y_i's, respectively. The value of q which makes this estimate zero is $q=\bar{y}/\bar{x}$, and this would be the corresponding estimate of the bycatch rate. The problem with using (1) as an estimate of the mean of y-qx is that it is not at all necessarily robust. That is, a relatively few inordinately large values for some of the y_i-qx_i's can have a profound influence on this estimate. Our analyses of vessel incentive program data have suggested that, in some cases, the distribution of the random variable y-qx is highly skewed to the right with a heavy right tail. When this is the case, one would be leery of using the average of a set of observations of y-qx to estimate the mean of this random variable. So we seek a procedure for estimating the mean of the random variable y-qx which does not have this drawback.

For any given value of q, let $z_1,\ldots,z_n$ represent the ordered values of $y_1-qx_1,\ldots,y_n-qx_n$. If μ_q represents the mean of y-qx, consider an estimator of μ_q of the form

$$\hat{\mu}_q = \frac{\sum_{i=1}^{s} z_i + (n-s)\,t}{n}, \tag{2}$$

where t is a number which does not exceed z_n and s is the number of z_i's which do not exceed t. Essentially, this estimator is formed by replacing all of the z_i's greater than t by the value of t itself, and averaging the resulting set of z_i's.

Suppose that f(z) represents the p.d.f. of the random variable y-qx. It follows from the work of Searls (1966) that the value of t, which makes (2) the estimator of μ_q with minimum mean squared error, among all such estimators, is the solution, for t, to the equation

$$\frac{1}{n}p_t(t-m_t) - (1-p_t)(m_t'-t) = 0, \tag{3}$$

where

$$p_t=\int_a^t f(z)\ dz,\ p_t\, m_t=\int_a^t z\ f(z)\ dz,\ (1-p_t)\,m_t'=\int_t^\infty z\ f(z)\ dz,$$

and a is the parameter such that $f(z)>0$, if $z \geq a$, and $f(z)=0$, if $z<a$.

The p.d.f. f(z) is unknown, but many analyses of vessel incentive program data have indicated that a reasonable model for f(z) is a generalized gamma p.d.f. of the form

$$
\begin{aligned}
f(z) &= \frac{d}{\Gamma(c)\,b^{cd}}(x-a)^{cd-1}e^{-(\frac{x-a}{b})^{d}}, x>a, -\infty<a<\infty, b, c>0, d>0 \\
&= \frac{1}{\sqrt{2\pi}\,(x-a)\,c}e^{-\frac{1}{2c^2}(\ln\frac{x-a}{b})^2}, x>a, -\infty<a<\infty, b, c>0, d=0 .
\end{aligned}
\tag{4}
$$

The appropriateness of the generalized gamma distribution as a model for the distribution of y-rx will be discussed in Section 4.

If f(z) is given by (4), then p_t, $p_t m_t$, and $(1-p_t)m_t'$ in (3) are

$$
\begin{aligned}
p_t &= \frac{1}{\Gamma(c)}\int_0^{(\frac{t-a}{b})^d} u^{c-1}e^{-u}du,\ d>0 \\
&= \Phi(\frac{1}{c}\ \ln\ \frac{t-a}{b}),\ d=0,
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
p_t m_t &= \frac{b}{\Gamma(c)}\int_0^{(\frac{t-a}{b})^d} u^{c+\frac{1}{d}-1}e^{-u}\,du,\ d>0 \\
&= be^{\frac{c^2}{2}}\ \Phi[\frac{1}{c}\ln\ \frac{t-a}{b}-c]
\end{aligned}
\tag{6}
$$

and

$$
\begin{aligned}
(1-p_t)\,m_t' &= \frac{b\Gamma(c+\frac{1}{d})}{\Gamma(c)} - p_t m_t,\ d>0 \\
&= be^{\frac{c^2}{2}} - p_t m_t,\ d=0,
\end{aligned}
\tag{7}
$$

where Φ denotes the standard normal distribution function. The integrals in (5) and (6) may be evaluated by use of the algorithm of Lau (1985).

The procedure used for finding a value for t in (2) is the following one. A search is used to find the value of t which satisfies (3). For any given value of t, the quantities p_t, m_t, and

m_t' in (3) are determined by use of (5), (6), and (7) after a, b, c, and d in these equations are replaced by estimates of these parameters.

The parameter estimates are based upon $z_1,\ldots,z_n$, and a is estimated first by

$$\hat{a} = \sum_{I=1}^{n} a_i z_i, \tag{8}$$

where $a_1=1+[1-(1/n)]^n$, and $a_i=[1-(i/n)]^n-[1-(i-1)/n]^n$, for $i=2,\ldots,n$. The estimate (8) is the estimate of the lower bound of a random variable proposed by Cooke (1979).

The parameters b, c, and d are estimated by using $v_1,\ldots,v_n$, where $v_i=z_i-\hat{a}$, for $i=1,\ldots,n$, and a maximum likelihood estimation argument. The estimate of d is the solution, for d, to the equation

$$\frac{1}{d}+c\frac{1}{n}\sum_{i=1}^{n}\ln v_i-c\frac{\sum_{i=1}^{n} v_i^d \ln v_i}{\sum_{i=1}^{n} v_i^d} = 0, \tag{9}$$

provided a positive value of d satisfies (9). For any given value of d, c in (9) is the solution to the equation

$$\Psi(c)-\ln c+\ln \frac{1}{n}\sum_{i=1}^{n} v_i^d-\frac{1}{n}\sum_{i=1}^{n}\ln v_i^d = 0. \tag{10}$$

The solution to (10), for c, after d in (10) is replaced by its estimate, is the estimate of c. The estimate of b is

$$\hat{b} = [\frac{1}{n\hat{c}}\sum_{i=1}^{n} v_i^{\hat{d}}]^{1/\hat{d}},$$

where $\hat{c}$ and $\hat{d}$ are the estimates of c and d, respectively.

If no positive value of d satisfies (9), the estimate of d is zero, and the estimates of b and c are, respectively,

$$\hat{b} = (\prod_{i=1}^{n} v_i)^{\frac{1}{n}} \text{ and } \hat{c} = [\frac{1}{n}\sum_{i=1}^{n} (\ln v_i - \ln \hat{b})^2]^{1/2}.$$

Thus, for any value of q, $\hat{\mu}_q$, given by (2), may be calculated. $\hat{\mu}_q$ is an estimate of the mean of y-qx. Further, a search procedure may be used to find the value of q such that $\hat{\mu}_q$ is zero. This value of q is the estimate of the bycatch rate r.

For the case where some of a vessel's hauls, selected for sampling by an observer, are basket sampled and some are whole haul sampled, a modification of the procedure described above may be used to estimate a bycatch rate. Let x_i represent the total haul weight for the i-th sampled haul, and y_i represent x_i times the ratio of the weight of halibut in the sample to the total weight of the sample from the i-th sampled haul, for i=1,...,n. Then proceed as before.

3. Confidence Limits for Bycatch Rates

The second statistical inference that is made about a bycatch rate r is a lower 95% confidence limit for this parameter. This limit is also based on the pairs $(x_1,y_1),\ldots(x_n,y_n)$. Here x_i represents pooled basket sample weight or total haul weight, for the i-th selected haul, depending upon whether observer sampling was strictly basket sampling or a mixture of basket sampling and whole haul sampling. Also, for the i-th selected haul, y_i represents either halibut weight in the pooled basket sample, for the case of strictly pooled basket sampling, or total haul weight times the ratio of halibut weight in the sample to the weight of the sample, for the mixed sampling case.

The lower 95% confidence limit that is used is essentially due to Fieller (1940). The derivation of it starts with the assumption that $\bar{y}$-r$\bar{x}$ has, approximately at least, a normal distribution with mean zero. Methods for examining the appropriateness of this assumption will be discussed in the next section.

Now

$$Var\ (\bar{y}-r\bar{x}) = Var\ \bar{y} + r^2\ Var\ \bar{x} - 2r\ Cov(\bar{x},\bar{y})\ , \tag{11}$$

and, if the basic assumption is true,

$$P[\frac{\bar{y}-r\bar{x}}{\sqrt{VAR(\bar{y}-r\bar{x})}} > -1.645] = 0.95\ . \tag{12}$$

If the parameters in (11) are replaced by estimates, the result is substituted in the left hand side of the inequality in (12), and the resulting inequality is operated on algebraically, it turns out that the inequality is equivalent to

$$r > \frac{\bar{y}}{\bar{x}}\,\frac{1-2.706\ C_{yx}-1.645\sqrt{C_{yy}+C_{xx}-2C_{yx}-2.706\ (C_{yy}C_{xx}-C_{yx}^2)}}{1-2.706\ C_{xx}}\ , \tag{13}$$

where

$$C_{yy}=\frac{\Sigma(y_i-\bar{y})^2}{(\Sigma y_i)^2},\quad C_{xx}=\frac{\Sigma(x_i-\bar{x})^2}{(\Sigma x_i)^2},\quad and\ C_{yx}=\frac{\Sigma(x_i-\bar{x})\ (y_i-\bar{y})}{(\Sigma y_i)\ (\Sigma x_i)}\ .$$

Thus the right hand side of (13) is a lower 95% confidence limit for r.

4. Examining the Appropriateness of Assumptions

The procedure being used for estimating the bycatch rate, r, uses as a model, for the distribution of the random variable z=y-rx, the generalized gamma distribution with p.d.f. given by Equation (4). In order to check the reasonableness of this model, a nonparametric estimate of the p.d.f. of z is found and compared with a member of the generalized gamma family of p.d.f.'s. The generalized gamma family member used is given by (4), after the parameters are replaced by estimates.

Both the nonparametric estimate of the p.d.f. of z and the generalized gamma parameter estimates are obtained by using as observations $z_1,\ldots,z_n$, where $z_i=y_i-\hat{r}x_i$, for $i=1,\ldots,n$, and $\hat{r}$ is the estimate of r described in Section 2. The method for obtaining generalized gamma parameter estimates is the one discussed in Section 2.

The procedure for finding the nonparametric estimate of the p.d.f. of z begins with estimating the location parameter, a, of the p.d.f. of z by (8). Then a number λ is determined. λ is such that $w_1,\ldots,w_n$ can be regarded as a set of observations from a symmetric distribution, where $w_i=k(z_i)$, for $i=1,\ldots,n$ and

$$\begin{aligned} k(z) &= \frac{(z-\hat{a})^{\lambda}-1}{\lambda},\ \lambda\neq 0 \\ &= \ln\,(z-\hat{a}),\ \lambda=0\ . \end{aligned}$$

The method for determining λ is described in the appendix.

After λ is determined, a kernel estimate of the p.d.f. of $w=k(z)$ may be found. This estimate is of the form

$$\hat{g}(w) = \frac{1}{\sqrt{2\pi}\,nh}\sum_{i=1}^{n} e^{-\frac{(w-w_i)}{2h^2}}\ , \tag{14}$$

where h is a smoothing parameter. Silverman (1986) provided a detailed account of nonparametric p.d.f. estimates of the form (14). The observations $w_1,\ldots,w_n$ and the method of Sheather and Jones (1991) are used to get a value for the smoothing parameter h.

Finally, the estimate of the p.d.f. of z is obtained by use of (14) and a transformation of variables. This estimate is

$$\frac{(z-\hat{a})^{\lambda-1}}{\sqrt{2\pi}\,nh}\sum_{i=1}^{n} e^{-\frac{[k(z)-k(z_i)]^2}{2h^2}}\ . \tag{15}$$

The method for finding the nonparametric p.d.f. estimate (15) was essentially suggested by Wand, Marron, and Ruppert (1991). But the techniques used for finding values for $\hat{a}$, λ, and h are different from the methods suggested by Wand, Marron, and Ruppert.

The basic assumption needed for determining a lower 95% confidence limit for a bycatch rate is that

$$P[\frac{\overline{y-rx}}{\hat{\sigma}}>-1.645] = 0.95 , \quad (16)$$

where $\overline{y-rx}$ is the average of $y_1-rx_1,\ldots,y_n-rx_n$ and $\hat{\sigma}$ is an estimate of the standard deviation of $\overline{y-rx}$.

When it is determined that the generalized gamma distribution is a reasonable model for the distribution of y-rx, the validity of the basic assumption (16) may be examined by use of simulation.

This is done by use of the following algorithm:

1. Generate a random sample of observations, say $u_1,\ldots,u_n$, from a generalized gamma distribution.
2. Calculate $\overline{u}/s_{\overline{u}}$, where $\overline{u}=\frac{1}{n}\sum_{i=1}^{n} u_i$ and $s_{\overline{u}} = \frac{[\sum_{i=1}^{n}(u_i-\overline{u})^2]^{1/2}}{n}$.
3. Repeat steps 1 and 2 a large number, M, times obtaining the ratios $r_1^*,\ldots,r_M^*$, where r_j^* is the ratio $\overline{u}/s_{\overline{u}}$ obtained by the j-th repetition of steps 1 and 2, for j=1,...,M.
4. Determine $p=(\#r_j^*>-1.645)/M$.

If p, in step 4, is close to 0.95, the conclusion would be that the 95% lower confidence limit for the bycatch rate is reasonably valid.

Step 1 of the algorithm may be carried out by noting that if the random variable z has a generalized gamma distribution with parameters a, b, c, and d, then $(z-a)^d$ has a two parameter gamma distribution with parameters b and c. The algorithms of Cheng (1977) and Ahrens and Dieter (1974) may be used to generate random samples from a gamma distribution, and these samples are easily transformed into samples from a generalized gamma distribution by adding the value of the parameter a to the 1/d power of each observation.

The values of the generalized gamma distribution parameters used in step 1 of the simulation study algorithm would be close to those parameter estimates obtained when checking the reasonableness of the generalized gamma distribution as a model for the distribution of y-rx. However, these estimates may have to be adjusted somewhat to insure that the equation

$$\hat{a}+\frac{\hat{b}\,\Gamma(\hat{c}+\frac{1}{\hat{d}})}{\Gamma(\hat{c})}=0 \qquad (17)$$

holds. This is necessary because, by definition of bycatch rate, the mean of the distribution of y-rx is zero. The left hand side of (17) is the mean of a generalized gamma distribution whose parameter values are the numbers $\hat{a}$, $\hat{b}$, $\hat{c}$, *and* $\hat{d}$.

5. Other Confidence Limits for Bycatch Rates

There is another method that may be used to find a lower confidence limit for a bycatch rate. It is an application of a statistical technique commonly referred to as the bootstrap.

The data gathered for making statistical inferences about a bycatch rate r can be summarized by the n pairs $(x_1,y_1),\ldots,(x_n,y_n)$. These pairs may be referred to as data points, and we denote the j-th of these by d_j, for j=1,...,n.

The procedure for finding a bootstrap lower 95% confidence limit for the bycatch rate r may be described as follows:

1. Calculate an estimate, $\hat{r}$, of r using all of the data points $\{d_1,\ldots,d_n\}$.
2. Select at random and with replacement d_i's from the set $\{d_1,\ldots,d_n\}$, one at a time, until n data points are

selected. Denote the selected data points by $d_1^*, \ldots, d_n^*$.

3. Calculate an estimate of r using the data points $d_1^*, \ldots, d_n^*$.
4. Repeat steps 2 and 3 a large number, B, of times obtaining the estimates, of r, $\hat{r}_1^*, \ldots, \hat{r}_B^*$.
5. Let $p=[\#(\hat{r}_i^*<\hat{r})]/B$, t be such that

$$\int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}]\ dx = p\ ,$$

and

$$q = \int_{-\infty}^{2t-1.645} \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}]\ dx\ .$$

6. The bootstrap lower 95% confidence limit for r is the value of s such that

$$\frac{\#(\hat{r}_i^* \le s)}{B} = q\ .$$

6. The Final Step

One final analysis of the data is performed for the case where all observer samples are pooled basket samples. That is, the analyses described above are repeated using the points $(x_1,y_1), \ldots, (x_n,y_n)$, where here x_i and y_i represent, respectively, the total haul weight and the product of the total haul weight and the ratio of the weight of the halibut in the sample to the weight of the sample, for the i-th haul selected for sampling. This final analysis serves as a final check of the reasonableness of the results obtained by the previous analyses, and it is an attempt to alleviate concerns expressed by some that total haul weight is not used in making inferences about bycatch rates.

APPENDIX

Suppose that v is a positive random variable and λ is a constant such that $w=k(v)$, where

$$k(v) = \frac{v^{\lambda}-1}{\lambda}, \ \lambda\neq 0$$
$$= \ln v \ , \ \lambda=0 \ ,$$

has, approximately at least, a symmetric distribution. To get a value for λ, we use, as a model for the distribution of w, the distribution with p.d.f.

$$f(w;a,b,c) = \frac{\Gamma(\frac{c}{2c-1})}{\sqrt{\frac{\pi}{2c-1}}\,\Gamma(\frac{1}{4c-2})\,b}\left[1+(2c-1)\left(\frac{w-a}{b}\right)^2\right]^{-\frac{c}{2c-1}}, \qquad c>\frac{1}{2}$$
$$= \frac{1}{2\sqrt{2}\,b\,\Gamma(C+1)}\, e^{\left[-\left|\frac{w-a}{\sqrt{2}b}\right|^{1/c}\right]}, \qquad 0<c\leq\frac{1}{2}. \qquad (I)$$

If $c=1/2$, $f(w;a,b,c)$ is a normal p.d.f.. If $c>1/2$, it is essentially the p.d.f. for a three parameter t distribution. The weights of the tails of this distribution increase as c increases, and they become very heavy if $c>1$. If $c<1/2$, the distribution, whose p.d.f. is (I), has lighter than normal tails, and the tail weights decrease as the value of c decreases. The tails are extremely light, much like those of a uniform distribution, if c is close to zero. Thus the distributional model, given by (I), is a very flexible one which yields symmetric distributions with tail weights ranging from very light to very heavy.

If (I) is the p.d.f. of k(v) and $v_1,\ldots,v_n$ represent the order statistics for a random sample of observations of v, the logarithm of the likelihood function of the parameters a, b, c, and λ is

$$(\lambda-1)\sum_{i=1}^{n}\ln v_i - \sum_{i=1}^{n}\ln f[k(v_i);a,b,c] \ . \qquad (II)$$

To find an estimate of λ in (I), we find the value of λ which maximizes (II). This value is found by a search, and at each step in the search a, b, and c, in (II), are replaced by suitable functions of λ and the v_i's.

For each value of λ, the value of c in (II) is determined as follows. Let

$$r(\lambda) = \frac{A_3(\lambda)-A_2(\lambda)}{A_4(\lambda)-A_1(\lambda)} ,$$

where $A_4(\lambda)$ and $A_1(\lambda)$ represent, respectively, the averages of the last and first 0.05n values of the ordered set $S=\{k(v_1),...,k(v_n)\}$, and $A_2(\lambda)$ and $A_3(\lambda)$ represent, respectively, the averages of the first and last 0.25n values of the set S which remain after the first and last 0.25n values are discarded. If $r(\lambda)<0.1574$, c is the solution to the equation

$$r(\lambda) = 0.20\frac{1-(1+c[q(c)]^2)^{\frac{c-1}{2c}}}{(1+c[p(c)]^2)^{\frac{c-1}{2c}}} , \qquad \text{(III)}$$

where

$$q(c)=0.6745+0.2454c+0.0795c^2-0.0054c^3-0.0105c^4$$

and

$$p(c)=1.6448+1.5238c+1.4202c^2+0.9830c^3+0.4339c^4 .$$

If $r(\lambda)\geq 0.1574$, c is the solution to the equation

$$h(\lambda)=\frac{Q_c(0.81)}{Q_c(0.69)}+3\frac{Q_c(0.95)}{Q_c(0.81)} , \qquad \text{(IV)}$$

where

$$h(\lambda)=\frac{\hat{Q}(0.81)-\hat{Q}(0.19)}{\hat{Q}(0.69)-\hat{Q}(0.31)}+3\frac{\hat{Q}(0.95)-\hat{Q}(0.06)}{\hat{Q}(0.81)-\hat{Q}(0.19)} ,$$

$\hat{Q}(t)$ represents an estimate of the 100t quantile based on the set S, and $Q_c(t)$ represents the solution, for x, to the equation

$$t=0.5[1+\frac{1}{\Gamma(c)}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\frac{x^{1+\frac{i}{c}}}{c+i}] .$$

This approach to obtaining a value for c in (II) is, essentially, that of Kappenman (1988). $r(\lambda)$ is an estimate of the quantity

$$\frac{\frac{1}{0.25}[\int_{Q(0.5)}^{Q(0.75)} w\ f(w)\ dw-\int_{Q(0.25)}^{Q(0.5)} w\ f(w)\ dw]}{\frac{1}{0.05}[\int_{Q(0.95)}^{\infty} w\ f(w)\ dw-\int_{-\infty}^{Q(0.05)} w\ f(w)\ dw]} \tag{V}$$

and (V) is the right hand side of (III), if f(w) is given by (I) and c>0.5. $h(\lambda)$ is an estimate of

$$\frac{Q(0.81)-Q(0.19)}{Q(0.69)-Q(0.31)}+3\frac{Q(0.95)-Q(0.05)}{Q(0.81)-Q(0.19)}\ , \tag{VI}$$

where Q(t) represents the 100t quantile of the distribution of w, and (VI) is the right hand side of (IV), if the distribution's p.d.f. is (I).

For any fixed values of c and λ, the maximum likelihood estimators of a and b are given by the solution, for a and b, to the equations

$$\sum_{i=1}^{n}[k(v_i)-a]\ |k(v_i)-a|^{(1/c)-2}=0\ ,\ b=(\frac{1}{nc}\sum_{i=1}^{n}|[k(v_i)-a]/\sqrt{2}|^{1/c})^{c}\ ,$$

if c≤1/2, or by the solution, for a and b, to the equations

$$\sum_{i=1}^{n}\frac{k(v_i)-a}{b^2+(2c-1)[k(v_i)-a]^2}=0\ ,\ n-2(c-2)\sum_{i=1}^{n}\frac{[k(v_i)-a]^2}{b^2+(2c-1)[k(v_i)-a]^2}=0\ ,$$

if c>1/2. These estimators are the functions of λ and the v_i's which replace a and b in (II), when (II) is searched for the value of λ which maximizes it.

REFERENCES

Ahrens, J.H. and Dieter, U. (1974). Computer methods for sampling from gamma, beta, Poisson, and binomial distributions. Computing, 12, 223-246.

Cheng, R.C.H. (1977). The generation of gamma variables with non-integral shape parameter. Applied Statistics, 26, 71-75.

Cooke, P. (1979). Statistical inference for the bounds of random variables. Biometrika, 66, 367-374.

Fieller, E.C. (1940). The biological standardization of insulin. Journal of the Royal Statistical Society, Series B, 16, 186-194.

Kappenman, R.F. (1988). Robust symmetric distribution location estimation and regression. Journal of Statistical Planning and Inference, 19, 55-72.

Lau, C. (1985). A simple series for the incomplete gamma integral. In Applied Statistics Algorithms, P. Griffiths and I.D. Hill editors, Royal Statistical Society/Ellis Horwood Limited.

Searls, D.T. (1966). An estimator for a population mean which reduces the effect of large true observations. Journal of the American Statistical Association, 61, 1200-1204.

Sheather, S.J. and Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society, Series B, 53, 683-690.

Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall Monographs on Statistics and Applied Probability.

Wand, M.P., Marron, J.S., and Ruppert, D. (1991). Transformations in density estimation. Journal of the American Statistical Association, 86, 343-353.