The Analytical Estimator for Sparse Data

Wen-Hui Lo, *Member, IAENG* , and Sin-Horng Chen

***Abstract*—In parameter estimation of normal distribution, the conventional truncated normal estimator worked well only if the sample size is greater than 20. In this study, we consider to extend its usage for sparse data cases with sample size under 20. We derive a wide-sense truncated normal joint probability distribution function, composing of coverage, range, the sample of the first order, and data samples themselves, to analyze the problem of truncated normal distribution in sparse data estimation. We successfully improve the traditional truncated normal estimation by simply finding the solution from quadric polynomials without complex computations. Besides, we also shapes the formulations to guarantee the convergence of the population mean estimation if the standard deviation of population is known.**

***Index Terms*—truncated normal distribution, coverage, coverage interval, sparse data**

ACRONYMS

pdf	probability density function
s-	statistical(ly)
cdf	cumulative distribution function
MLE	maximum likelihood estimation
MSE	mean square error
VTNJ *pdf*	variably truncated normal joint *pdf*
BLUE	the best linear unbiased estimation
CLT	central limit theorem
GLI	Gauss Legendre integration
MLL	marginal log likelihood
i.i.d	independent and identical distribution
DM	distribution mismatch

NOTATION

$p_?(.)$	*pdf* or conditional *pdf* for a certain variable
Pr(.)	probability
x	random variable of normal distribution
$f_?(.)$	*pdf* of a certain random variable
$f_x(.)$	*pdf* of the normal population of random variable x with mean u and standard deviation σ ; i.e., $f_x(x) = N(u, \sigma^2)$
$F_x(.)$	*cdf* of the normal population of random variable x ; $F_x(x) = \int_{-\infty}^{x} f_x(y)dy$
u	population mean
σ	standard deviation of population
$x_{i:n}, 1 \le i \le n$	the ranked random variable resulting from sorting the samples of x
n	sample size
$u[0,1]$	standard uniform distribution in $[0,1]$
ξ	random sequence of the standard normal distribution
$\xi_{i:n}, 1 \le i \le n$	order statistics random variable generated from the ranked random variable ξ of the standard normal *pdf*
X_n	random sequence of length n
$\mathrm{E}_?[.]$ or $\mathrm{E}_{(?)}[.]$	expectation operator
$Cov[\cdot,\cdot]$	covariance operator
$Min[.]$	take the minimum value in set
I	identity vector
B	covariance matrix
L	likelihood
r	range
c	coverage
$U(.)$	unit step function
$Z(Cc_t, n)$	normalized factor for the fixed coverage point Cc_t , t is the sampling index
η_j	root of the Hermite polynomials expanded coverage the order of Hermite polynomials
$[a,b]$	the interval for interval estimation of coverage
$w_{Hm}(\gamma_i)$	the roots of the i-th Hermite polynomial
$P_v(.)$	the v -th Legendre polynomial
r_s	the random variable of range on standard normal *pdf*
$\Phi(.)$	*cdf* of standard normal distribution
$\bar{x}$	if no emphasis, it is the sample mean or average of the truncated data
$\overline{x^2}$	mean of square
$w_{P_v}(\kappa_t)$	the weighting coefficient of the t-th root of the v th order Legendre polynomial

Manuscript received October 13, 2008.

Wen-Hui Lo is with the Communication Engineering Department, National Chiao Tung University, Hsinchu, Taiwan. (phone: 886-3-571-2121 ext 54555; fax: 886-3-590-7897; e-mail: hs3341.cm90g@nctu.edu.tw);

Prof. Sin-Horng Chen is with the Communication Engineering Department, National Chiao Tung University, Hsinchu, Taiwan (phone: 886-3-572-1691; fax: 886-3-571-0116; e-mail: schen@cc.nctu.edu.tw; address: 1001 University Road, Hsinchu, Taiwan, 30050, ROC)

I. Introduction

Robust parameter estimation in sparse data is generally applied to the tasks when data collection is time-consuming or of high sampling cost. This study focuses on the mean estimation of normally distributed random variables under the sparse data constraint. Since the truncation or censoring scheme is usually adopted in sparse data estimation, our major goal is to improve the truncated normal estimator

proposed by Cohen [1]. There are some shortcomings in his truncated normal estimator, including the need of looking-up tables for the positions of initial searching points, the need of a couple of endpoints to compute the standard deviation, the constraint that the expression of endpoints must be deterministic, and non-guarantee of convergence. Focusing on those problems, we successfully extend Cohen's algorithm to express it for sparse data and to release the solving criteria to require only one truncated point expressed by random variable.

In the beginning, we point out that there exists a distribution mismatch (DM) problem if the sample size is less than 20. Then we propose a new method to overcome the problems of needing a table to look up for finding the initial searching points and of requiring deterministic truncated points encountered in the roots-finding task of the truncated normal estimation. A macro view random variable of coverage interval [2], [3] which is the expression recommended for the uncertainty measurement [4], is then introduced. The *pdf* of coverage interval is firstly derived. Then a variably truncated normal joint (VTNJ) *pdf*, which considers coverage, coverage interval, the first order of ranked samples and the samples themselves, is created to compensate the DM effects. Lastly, we reduce the computations of VTNJ *pdf* by referring the suggestion of Chen [5], [6] about the parametric coverage interval to obtain a wide-sense parametric coverage estimator. It is a simplified result of the VTNJ *pdf*.

The remainder of the paper is organized as follows. Section 2 gives a brief review of previous studies. In Section 3, the proposed method is presented. The numerical implementation of the proposed method with the VTNJ *pdf* is discussed in Section 4. Section 5 derives the variably truncated normal joint distribution estimator (VTNJE). Experiments to evaluate the performance of the proposed method are described in Section 6. An application of using the results of realistic quantile mapping invariance (QMI) is presented in Section 7. Discussions and conclusions are given in the last section.

II. Paper Review

In parameter estimation of normally-distributed sparse data, there are two popular methods: the best linear unbiased estimation (BLUE) method and the maximum likelihood estimation (MLE) method. Balarkrishnan and Cohen [7], Lloyd [8], and Teichroew [9] have suggested the BLUE method for parameter estimation of normal random variables using order statistics. BLUE is a weighted least-square algorithm basing on the Gauss-Markov least-square theorem. It was popularly used for sparse data analysis. It is known that BLUE is unbiased and more efficient if it takes the censoring sampling scheme. We briefly discuss BLUE as follows.

Let x be a normal random variable with *pdf* $f_x(x) = N(u,\sigma^2)$. Assume that there are n independent observed samples $x_1,\cdots,x_n$ of x. Let $x_{1:n},\cdots,x_{n:n}$ be the ranked samples of $x_1,\cdots,x_n$ in increasing order. The BLUE estimator is formulated as the sum of products of the observed data and properly-chosen coefficients. By performing the standard normal transformation, $\xi_i=(x_i-u)/\sigma$, to the observed data and sorting them in increasing order, we have

$$
\begin{aligned}
&X_n = [x_1,\cdots,x_n]^T \\
&\xi = [\xi_1,\cdots,\xi_n]^T \\
&E\{\xi_{i:n}\} = \rho_{i:n} \\
&Cov\{\xi_{i:n},\xi_{j:n}\} = \beta_{i,j:n} \quad \text{for } 1\le i,j\le n \text{ and } i<j \\
&E\{x_{i:n}\} = u+\sigma\xi_{i:n} \\
&E\{X_n\} = u\mathrm{I}+\sigma\xi
\end{aligned}
\tag{1}
$$

$$
\mathrm{I}_n = [1,\cdots,1]^T_{n\times 1}
$$

$$
B = \sigma^2 I \tag{2}
$$

where I_n is an n-dimensional all-1 vector. Consider the generalized variance:

$$
(X_n - u\mathrm{I}_n - \sigma\xi)^T B^{-1}(X_n - u\mathrm{I}_n - \sigma\xi) \tag{3}
$$

Minimizing it with respect to u and σ, we obtain.

$$
\begin{aligned}
&u\mathrm{I}_n^T B^{-1}\mathrm{I}_n + \sigma\mathrm{I}_n^T B^{-1}\xi = \mathrm{I}_n^T B^{-1}X_n \\
&u\xi^T B^{-1}\mathrm{I}_n + \sigma\xi^T B^{-1}\xi = \xi^T B^{-1}X_n
\end{aligned}
\tag{4}
$$

The solution of Eq.(4) is

$$
\begin{aligned}
u^* &= \left\{\frac{\xi^T B^{-1}\xi\mathrm{I}_n^T B^{-1} - \xi^T B^{-1}\mathrm{I}_n\xi^T B^{-1}}{(\xi^T B^{-1}\xi)(\mathrm{I}_n^T B^{-1}\mathrm{I}_n) - (\xi^T B^{-1}\mathrm{I}_n)^2}\right\}X_n \\
&= -\xi^T\Delta X_n = \sum_{i=1}^{n}\alpha_{1:i}x_{i:n}
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
\sigma^* &= \frac{\mathrm{I}_n^T B^{-1}\mathrm{I}_n\xi^T B^{-1} - \mathrm{I}_n^T B^{-1}\xi\mathrm{I}_n^T B^{-1}}{(\xi^T B^{-1}\xi)(\mathrm{I}_n^T B^{-1}\mathrm{I}_n) - (\xi^T B^{-1}\mathrm{I}_n)^2}X_n \\
&= \mathrm{I}_n^T\Delta X_n = \sum_{i=1}^{n}\alpha_{2:i}x_{i:n}
\end{aligned}
\tag{6}
$$

where u^* and σ^* are the estimated parameters, and $\alpha_{1:i}$ and $\alpha_{2:i}$ are weighting coefficients. These coefficients have been tabulated by Sarhan and Greenberg [10], [11] with entries in the 1956 tables being given for sample size up to 10 and in 1962 up to 20.

Generally speaking, BLUE performs well in small sample size. But it needs a table to look up, and this is a shortcoming. The other technique used is the MLE method which is often applied to the truncated normal distribution in sparse data condition. Cohen [1] derived the singly truncated and doubly truncated maximum likelihood estimators and found that they outperformed BLUE when the sample size was grater than 20. Cohen recognized the sparse data problem as a truncated normal *pdf* and defined its likelihood by

$$L=\left(\frac{U(x-x_{1:n})-U(x-x_{1:n}-r)}{\sqrt{2\pi}\sigma(F_x(x_{1:n}+r)-F_x(x_{1:n}))}\right)^n \exp(-\sum_{i=1}^{n}\frac{(x_i-u)^2}{2\sigma^2}) \quad (7)$$

If we take the transformations of $\xi_{1:n}=(x_{1:n}-u)/\sigma$ and $\xi_{n:n}=(x_{n:n}-u)/\sigma$ and differentiate the resulting log-likelihood function with respect to u and σ, we obtain the following two equations:

$$\begin{aligned}&\frac{n(\phi_\xi(\xi_{1:n})-\phi_x(\xi_{n:n}))}{\sigma(\Phi_\xi(\xi_{n:n})-\Phi_\xi(\xi_{1:n}))}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-u)\\&\sigma^2\left\{\frac{(\xi_{1:n}\phi_\xi(\xi_{1:n})-(\xi_{n:n})\phi_\xi(\xi_{n:n}))}{(\Phi_\xi(\xi_{n:n})-\Phi_\xi(\xi_{1:n}))}+1\right\}=\frac{1}{n}\sum_{i=1}^{n}(x_i-u)^2\end{aligned} \quad (8)$$

where ϕ and Φ are the standard normal *pdf* and *cdf*, respectively. By defining two new random variables

$$\Theta_L=\frac{\phi_\xi(\xi_{1:n})}{\Phi_\xi(\xi_{1:n}+r_s)-\Phi_\xi(\xi_{1:n})}$$

and

$$\Theta_R=\frac{\phi_\xi(\xi_{1:n}+r_s)}{\Phi_\xi(\xi_{1:n}+r_s)-\Phi_\xi(\xi_{1:n})},$$

we obtain the following two equations

$$H_1(\xi_1,\xi_2)=\frac{\bar{x}-x_{1:n}}{r}=\frac{\Theta_L-\Theta_R-\xi_1}{\xi_2-\xi_1} \quad (9)$$

$$H_2(\xi_1,\xi_2)\Rightarrow\frac{S^2}{r^2}=\frac{1+\xi_1\Theta_L-\xi_2\Theta_R-(\Theta_L-\Theta_R)^2}{(\xi_2-\xi_1)^2} \quad (10)$$

where r is the range, and $\bar{x}$ and S^2 are the sample mean and variance, respectively. Although the above two equations can be solved by applying the Newton and Raphson method to calculate the variables ξ_2 and ξ_1, it may be time-consuming. Cohen [1] therefore announced to use a chart consisting of intersecting graphs of the simultaneous equations. He proposed an approach to use a look-up table which provided a couple of endpoints, $[\xi_{1:n},\xi_{n:n}]$, to serve as the initial searching positions for the roots-finding task. After obtaining the estimates of $\xi_{1:n}$ and $\xi_{n:n}$ by using Eqs.(9) and (10), we then compute the estimated mean u^{**} by

$$u^{**}=x_{p:n}-\sigma^{**}\xi_{p:n}, \quad (11)$$

where p=1 or n and

$$\sigma^{**}=\frac{x_{n:n}-x_{1:n}}{\xi_{n:n}-\xi_{1:n}}. \quad (12)$$

The final solution is obtained via repeatedly applying the above procedure until a convergence is reach.

III. Proposed Method

Our task is to estimate the mean of a random variable x with unknown normal distribution $f_x(x)=N(\mu,\sigma^2)$ from a set of n observed samples $\{x_i,1\le i\le n\}$ with $n\le 20$. We first rank these n samples in increasing order and denote them by $\{x_{i:n},1\le i\le n\}$. The range and coverage of the sample set are then defined by $r=x_{n:n}-x_{1:n}$ and $c=F_x(x_{n:n})-F_x(x_{1:n})$, respectively. Coverage is a macro view of random variable to carry global information of all observed samples. The general relation among coverage c, range r, the minimum order $x_{1:n}$, and samples X_n is shown in Fig. 1. In our basic assumption, we think the macro view random variables should be consistent to the result of micro view random variable. The dash-line represents the interferences within the macro view random variables and the solid-line represents the interferences from the macro view to micro view random variables. A joint normal *pdf* of these four parameters will be built in the following basing on Fig. 1 to compensate the coverage mismatch. We treat the distribution as a variably truncated normal joint (VTNJ) *pdf* to represent the randomness of the truncated points of a truncated normal distribution depending on coverage and sample size.

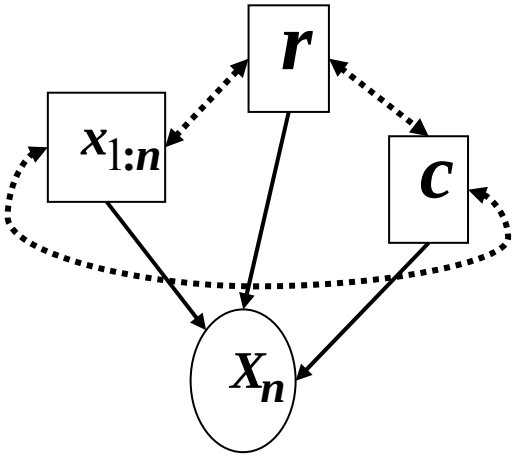

Fig. 1: Relation of variables' interference model

We first decompose $f_{x,x_{1:n},r,c;u,\sigma|n}(x,x_{1:n},r,c)$ into four conditional *pdf*s by

$$\begin{aligned}&f_{x,x_{1:n},r,c;u,\sigma|n}(x,x_{1:n},r,c)\\&=f_{x;u,\sigma|x_{1:n},r,c,n}(x)\cdot f_{x_{1:n}|r,n}(x_{1:n})\cdot f_{r|c,n}(r)\cdot f_{c|n}(c)\end{aligned} \quad (13)$$

where

$$f_{x;u,\sigma|x_{1:n},r,c,n}(x)=f_x(x)\frac{U(x-x_{1:n})-U(x-x_{1:n}-r)}{Q(x_{1:n},r)}$$

is the truncated normal *pdf* depending on the sample size, the truncated points and the sample's coverage; and $Q(x_{1:n},r)=F_x(x_{1:n}+r)-F_x(x_{1:n})$ is the sample coverage.

We then derive the *pdf* of coverage. There were some previous studies concerning the issue of randomness of coverage. The early topic was called "the random division of an interval", which means the range may be cut as many small sub-ranges which can be added to calculate the

coverage [12], [13]. The *cdf* of coverage for small sample size can be expressed by [14]

$$\Pr_{c|n}(C > c) = \sum_{k=0}^{n-2} \binom{n}{k} c^k (1-c)^{n-k} \tag{14}$$

We now simplify the coverage *pdf* as a polynomial of *c*. The derivation is given as follows.

$$\begin{aligned}
\Pr_{c|n}(C > c) &= ((1-c)^n n!\{-(\frac{1}{1-c})^n c\,\Gamma(n) \\
&+ (-1)^n c(-\frac{c}{1-c})^n \Gamma(n) + (-1)^n (-\frac{c}{1-c})^n \Gamma(n+1) \\
&- (-1)^n c(-\frac{c}{1-c})^n \Gamma(n+1)\}) / (c\,\Gamma(n)\Gamma(n+1)) \\
&= \frac{n!(-c(-1+c^n)\Gamma(n) + (-1+c)c^n\Gamma(n+1))}{c\,\Gamma(n)\Gamma(n+1)} \\
&= \frac{n!}{\Gamma(n+1)} - \frac{c^{n-1}n!(c\,\Gamma(n) + \Gamma(n+1) - c\,\Gamma(n+1))}{\Gamma(n)\Gamma(n+1)} \\
&= 1 - nc^{n-1} + (n-1)c^n
\end{aligned} \tag{15}$$

where $\Gamma(.)$ denotes the Gamma function and $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$. Hence

$$\begin{aligned}
f_{c|n}(c) &= \frac{\partial}{\partial c}(1 - (\Pr_{c|n}(C > c))) = \frac{\partial}{\partial c}(nc^{n-1} - (n-1)c^n) \\
&= n(n-1)(c^{n-2} - c^{n-1}) \qquad \text{for } 0 \le c \le 1
\end{aligned} \tag{16}$$

It is worth to note that Eq.(16) is distribution-free because Pratt and Gibbsons [14] derived it without assuming the distribution of the sampled random variable so that it is appropriately applied to any kind of *pdf* . Fig. 2 displays the coverage *pdf* for some small values of *n*. It can be found from the figure that the coverage distribution deviates away from 1 progressively and spreads wider as the sample size decreases from 20. We call this special phenomenon as distribution mismatch (DM) because it implicitly indicates that there exists a serious mismatch between the distributions of observed samples and the random variable when the sample size is small. The DM phenomenon reveals an important cue to the modeling of sparse data: coverage may serve as a confidence factor to indicate the appropriateness of observed data for robust parameter estimation. A higher value of coverage means a better match of the samples to its original normal distribution. To exploit the DM phenomenon, we treat coverage as a random variable and add it to the VTNJ *pdf*.

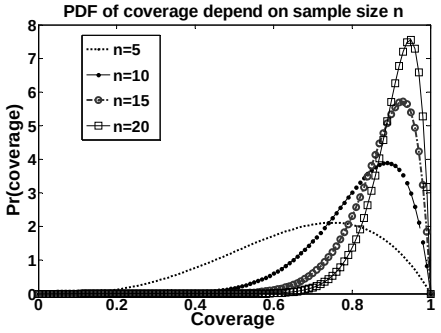

Fig. 2: The *pdf* of coverage for different sample size n

The other two terms in Eq.(13), $f_{x_{1:n}|r,n}(x_{1:n})$ and $f_{r|c,n}(r)$, can be derived from the nonparametric *pdf* of range of the ranked observed samples expressed by:

$$\begin{aligned}
&f_{r|n}(r) \\
&= \int_{dx_{1:n}} n(n-1) f_x(x_{1:n}) f_x(x_{1:n}+r)(F_x(x_{1:n}+r) - F_x(x_{1:n}+r))^{n-2}
\end{aligned} \tag{17}$$

The conditional *pdf* of the minimum sample given with range and sample size can be expressed by

$$\begin{aligned}
f_{x_{1:n}|r,n}(x_{1:n}) &= \frac{f_{x_{1:n},r|n}(r, x_{1:n})}{\int_{dx_{1:n}} f_{x_{1:n},r|n}(r, x_{1:n})} \\
&= \frac{f_x(x_{1:n})\, f_x(x_{1:n}+r)\{F_x(x_{1:n}+r) - F_x(x_{1:n})\}^{n-2}}{\int_{dx_{1:n}} f_x(x_{1:n})\, f_x(x_{1:n}+r)\{F_x(x_{1:n}+r) - F_x(x_{1:n})\}^{n-2}}
\end{aligned} \tag{18}$$

It can be derived easily by using the Bayes' theorem.

$f_{r|c,n}(r)$ is the range *pdf* given with coverage and sample size. Range is formally called coverage interval if we consider it as joint *pdf* in association with coverage. Coverage interval is a very important random variable in the field of measurement and its usage can be divided into the parametric coverage interval and non-parametric coverage interval. Parametric approach considers it as a *pdf* with parameters and non-parametric approach takes the general empirical distribution to compute the total coverage interval. In this paper, we consider the parametric coverage interval. If we think that the coverage interval is range constrained with coverage, it can be derived by the Jacobian determinant transform.

Now we take the variable transformation, shown in Fig. 3, to transform $x_{1:n}$ to $c\ (= F_x(x_{1:n}+r) - F_x(x_{1:n}))$ with the variable r being preserved.

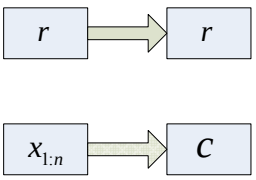

Fig. 3: The transformation rule for variables

Suppose that there are k roots η_j, $1 \le j \le k$, satisfying the coverage sampling point, $Q(x_{1:n}, r) = F_x(x_{1:n} + r) - F_x(x_{1:n}) = Cc$. The joint distribution of r and c can be expressed by

$$f_{r,c|n}(r, c = Cc) = \sum_{j=1}^{k} \{ f_{r,x_{1:n}|n}(r, x_{1:n}) \times \frac{1}{\left| \begin{matrix} \frac{\partial r}{\partial r} & \frac{\partial r}{\partial x_{1:n}} \\ \frac{\partial c}{\partial r} & \frac{\partial c}{\partial x_{1:n}} \end{matrix} \right|_+} \}_{x_{1:n}=\eta_j}$$

$$= \sum_{j=1}^{k} \{ f_{r,\eta_j|n}(r, \eta_j) \times \frac{1}{\left| \begin{matrix} 1 & 0 \\ f_x(\eta_j + r) & f_x(\eta_j + r) - f_x(\eta_j) \end{matrix} \right|_+} \}$$

$$= \sum_{j=1}^{k} \left(f_{r,\eta_j|n}(r, \eta_j) \frac{1}{\left| f_x(\eta_j + r) - f_x(\eta_j) \right|} \right) \tag{19}$$

But, since $Q(x_{1:n}, r) = F_x(x_{1:n} + r) - F_x(x_{1:n})$ is a transcendental function, we can not get an explicit transformation from the first order sample $x_{1:n}$ to coverage c. Alternatively, we can solve the problem by the sampling point method to express $f_{r|c,n}(r)$ by using the sampling points of $Q(x_{1:n}, r)$:

$$f_{r|c=Cc,n}(r) = \frac{f_{r,c|n}(r, c = Cc)}{\int_{dr} f_{r,c|n}(r, c = Cc)}$$

$$= \sum_{j=1}^{k} \left(\frac{f_x(\eta_j) f_x(\eta_j + r) \left\{ F_x(\eta_j + r) - F_x(\eta_j) \right\}^{n-2}}{\left| f_x(\eta_j + r) - f_x(\eta_j) \right|} \right) \cdot \frac{1}{Z(Cc, n)} \tag{20}$$

where $Z(Cc, n)$ is a normalization factor expressed by

$$Z(Cc, n) = \int_{dr} \left\{ \sum_{j=1}^{k} \left[\frac{n(n-1) f_x(\eta_j) f_x(\eta_j + r) \left\{ F_x(\eta_j + r) - F_x(\eta_j) \right\}^{n-2}}{\left| f_x(\eta_j + r) - f_x(\eta_j) \right|} \right] \right\};$$

$Cc = F_x(\eta_j + r) - F_x(\eta_j)$; η_j, $1 \le j \le k$ are the roots of $F_x(\eta_j + r) - F_x(\eta_j) = Cc$; and k is the number of roots. If $k > 1$, it may be approximated by expanding $F_x(.)$ with Hermite polynomials. If we consider to delete the outlier solutions of η_j, we may construct the following empirical constraints to select the solutions: $-4\sigma + u \le \eta_j \le 4\sigma + u$, $-4\sigma + u \le \eta_j + r \le 4\sigma + u$ and $0 < r < 8\sigma$.

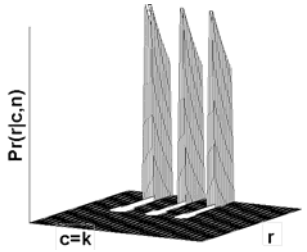

Fig. 4 Profile-conditional *pdf* by the sampling strategy, where k represents a certain constant.

IV. Numerical Implementation with the Variably Truncated Normal Joint *pdf*

A consequence of applying the sampling point method to compute $f_{r|c,n}(r)$ is that the interval estimation based on $f_{r|c,n}(r)$ has to be realized by using the sampling point method also. This can be solved by using the Gauss Legendre Integration (GLI) [15]. GLI is a well-known numerical method to give a good approximation to the integration of a function. It represents the integration of a function in the standard interval $[-1,1]$ by a weighted sum. By using GLI, we have

$$\int_a^b f_{c|n}(c)dc = \int_{-1}^{1} f_{c|n}\left(\frac{b-a}{2}\varsigma + \frac{b+a}{2}\right)\frac{(b-a)}{2}d\varsigma$$
$$= \frac{b-a}{2}\sum_{\tau=1}^{v} w_v(\varsigma_\tau) \cdot f_{c|n}\left(\frac{b-a}{2}\varsigma_\tau + \frac{b+a}{2}\right) + R_v(\varsigma) \tag{21}$$

where a and b are the endpoints of the interval estimation for coverage; $-1 < \varsigma_\tau < 1$ and ς_τ is the τth root of the Legendre polynomials $P_v(\varsigma)$ of order v; $1 \le \tau \le v$ is the root index;

$$w_v(\varsigma_\tau) = \frac{(b-a)}{(1-\varsigma_\tau^2)(P_v'(\varsigma_\tau))^2} \tag{22}$$

is the weighting function;

$$R_v(\varsigma) = \frac{2^{(2v+1)}(v!)^4}{(2v+1)((2v)!)^3} p_{c|n}^{(2v)}(\varsigma) \tag{23}$$

is the error of the approximation; and

$$P_v(\varsigma) = \frac{1}{2^v v!}\frac{\partial^v}{\partial \varsigma^v}(\varsigma^2 - 1)^v, \text{ for } v = 0, 1, 2, \cdots.$$

Then, the VTNJ *pdf* can be implemented by the numerical technique. Its result is also an interval estimation for the coverage fluctuation. In this study, the default settings are $a = E_{c|n}[c] - 0.005$ and $b = E_{c|n}[c] + 0.005$ to consider a coverage interval of 0.01.

It will be perfect if we can use a fixed sampling number for GLI to reduce the error so as to make it approach to its minimum. As shown in Eq.(23), the error of GLI is related to the differential order of the integrated function. Obviously, its differential order is finite. From Eq.(23), if the GLI

sampling number v meets the condition of $2v \geq n-1$, the estimation error $R_v(\xi)$ will be reduced to zero. In this case, GLI will approach to the theoretical optimal solution of no errors. Besides, Eq.(16) shows another important fact that the coverage *pdf* is independent of the distribution of the sampled random variable. So, we can claim that the *pdf* of coverage is distribution free. This property makes $f_{c|n}(c)$ freely connect to any kind of $f_{r|c,n}(r)$ by Chain rule.

If we want to directly calculate the VTNJ *pdf* in $p_{r|c,n}(r)$, we will face the problem that the mean and standard deviation of the population must be known in advance. But this is unrealistic in our mission. We therefore adopt an alternative approach to construct a new bridge to conjoint with these variables. The idea is to transform the observed data into the standard normal domain. The suggestion is shown in Fig. 5. As shown in the figure, we transform the observed ranked samples into the domain of standard normal by $\xi_{i:n} = (x_{i:n} - u)/\sigma$. Each transform pair is marked with the same digit number. The range is also transformed by $r_s = \xi_{n:n} - \xi_{1:n}$. Notice that the transform is quantile mapping invariance (QMI) for the macro view random variables.

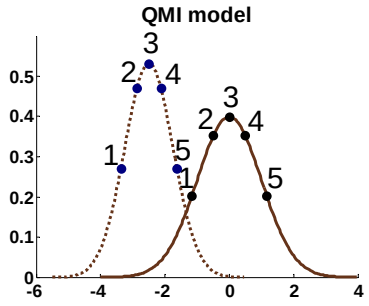

Fig. 5: Relative quantile mapping invariance based on their percentiles. Dash-line represents the original normal *pdf* and solid-line represents the standard normal *pdf*.

V. Derive the Variably Truncated Normal Joint Distribution Estimator (VTNJE)

We then apply GLI to the VTNJ *pdf* to obtain the marginal log likelihood by:

$$MLL(\cdot) \equiv \sum_{t=1}^{v}\left\{\frac{b-a}{2}n(n-1)(Cc_t^{\,n-2}-Cc_t^{\,n-1})w_{P_v}(\kappa_t)\int_{dr_s}\int_{d\xi_{1:n}} G\right\} \quad (24)$$

where

$$G = \log\left\{\left(\frac{1}{\sqrt{2\pi}\sigma\left(\Phi_\xi(\xi_{1:n}+r_s)-\Phi_\xi(\xi_{1:n})\right)}\right)^n \cdot\exp\left\{-\sum_{i=1}^{n}\frac{(x_i-u)^2}{2\sigma^2}\right\}\right\}\cdot p_{\xi_{1:n}|r_s,n}(\xi_{1:n})\cdot p_{r_s|c=Cc_t,n}(r_s)$$

A. Simplify the MLL through Coverage Interval

The marginal log likelihood is complicated and computionally time-consuming. We suggested an idea to reduce its computation basing on the coverage interval. An example of the profile-conditional *pdf*, $f_{r|c,n}(r)$, is plotted in Fig. 6. It is to demonstrate the fact that if we would like to guarantee the coverage of the estimation to be large enough to greater than a lower bound, then there will be much more tolerance intervals qualified for solutions to reside. Let us return to Eq.(16) to inspect the *pdf* of coverage which is distribution-free. We find that its form is inconvenient for parameter estimation due to the no use of derivative operator. Fortunately, Chen [6] gave a good suggestion to the computation of coverage. According to the conclusion of Chen, the *pdf* of coverage can be parametric if we constrain the coverage interval (range) to be the minimum of all possible values. The plot shown in Fig. 6 demonstrates that $f_{r|c,n}(r)$ looks like an impulse with its distribution concentrating near the minimum-case. It is hence reasonable to take $Min[r_s]$ to substitute all other possible values of r_s.

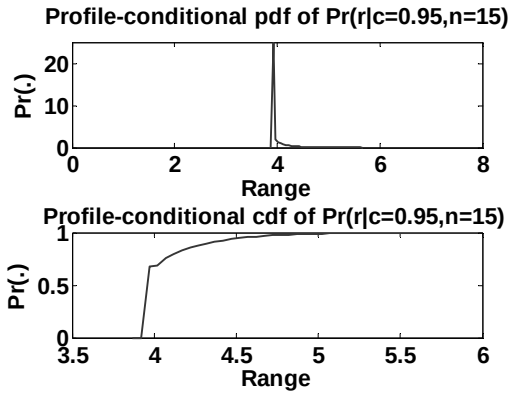

Fig. 6: Exemplified profile-conditional *pdf* to show the impulse properties for the sample size $n = 15$ and coverage=0.95 of standard normal *pdf*.

B. Algebraic Closed From for Parameter Estimation

Let we apply the result of Fig. 6 to simplify Eq.(24). It can then be optimized and expressed as two quadric equations of variables σ and u. Take the roots of these two quadric equations will result in the following solutions:

$$\sigma^* = \frac{B_\sigma \pm \sqrt{(B_\sigma)^2 + 4\left(\sum_{t=1}^{v} nD_t\right)C_\sigma}}{2\left(\sum_{t=1}^{v} nD_t\right)}, \quad (25)$$

where

$$B_\sigma = \left(\sum_{t=1}^{v} D_t\left(E_{\xi_{1:n}|c=Cc_t,Min\{r_s\},n}\{\xi_{1:n}\}\right)\left(\sum_{i=1}^{n}(x_i - x_{1:n})\right)\right),$$

$$C_\sigma = \left(\sum_{t=1}^{v}\left(D_t\sum_{i=1}^{n}(x_i - x_{1:n})^2\right)\right),$$

$$D_t = \left(\frac{b-a}{2}n(n-1)(Cc_t^{\,n-2} - Cc_t^{\,n-1})\left(w_{P_v}(\kappa_t)\right)\right);$$

and

$$u^* = \frac{-B_u \pm \sqrt{B_u^2 - 4(\sum_{t=1}^{v} D_t)C_u}}{2(\sum_{t=1}^{v} D_t)} \quad (26)$$

where

$$B_u=\sum_{t=1}^{v}\left\{D_t\left[(\bar{x}-x_{1:n})\left(E_{\xi_{1:n}|c=Cc_t,Min\{r_s\},n}\left\{\xi_{1:n}^2\right\}\right)-2x_{1:n}\right]\right\},$$

$$C_u=\sum_{t=1}^{v}\left\{D_t\left[x_{1:n}^2+\left(\bar{x}x_{1:n}-\overline{x^2}\right)\left(E_{\xi_{1:n}|c=Cc_t,Min\{r_s\},n}\left\{\xi_{1:n}^2\right\}\right)\right]\right\}.$$

Here, $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ is the sample mean, $\overline{x^2}=\frac{1}{n}\sum_{i=1}^{n}x_i^2$ is the mean of sample square, $w_{P_v}(\kappa_t)=\frac{(b-a)}{(1-\kappa_t^2)(P_v'(\kappa_t))^2}$ is the weighting coefficient of the t-th root of the v-th order Legendre polynomial, $[a,b]$ is the coverage estimation interval, and $\Phi_{\xi}(\xi)$ is the *cdf* of the standard normal distribution. The same strategy can be applied to the other endpoint $\xi_{n:n}$ via replacing $\xi_{1:n}$ by $\xi_{n:n}$. Then, the VTNJ *pdf* can be implemented by the numerical technique. Its result is also an interval estimation for the coverage fluctuation. From Fig. 2, it clearly shows that the coverage is a random variable if the sample size is less than 20. Hence, we had better to set the most observed interval to inspect its randomness. Define the following $\beta\%$-inspection interval (β-II):

> $\beta\%$-inspection interval is an interval estimation for the coverage random variable over the interval [a,b] with $a=E_{c|n}[c]-\beta/2$, and $b=E_{c|n}[c]+\beta/2$.

We then aim at calculating the most possible happening probability.

VI. Experiments

By checking Eqs.(25) and (26), we find that they are mainly affected by the sample mean, $\bar{x}$, and the individual ranked samples, $x_{i:n}, 1\le i\le n$. Our strategy is to adjust the coverage to make it approach to the real coverage, generated from $\bar{x}$ and $x_{i:n}, 1\le i\le n$, in order to compensate the DM effects. We examine two methods. One is to view the joint effect of $\bar{x}$ and $x_{i:n}$ under our suggestion of QMI (see Fig. 5). The other is to realize the QMI basing only on the real coverage. Its purpose is to see the effect of sample mean without coverage estimation.

A. Test the results with consistency to sample mean under the QMI principle—case of the default percentile

We first formed an interval estimation for coverage by performing a coverage estimation from the expectation of order statistics by $E_{c|n}[c]$ and adding fluctuation of ± 0.005.

The VTNJE might work normally without the tasks of looking up the tables so that it supported more conveniences for the computer programs. We compared it to the best estimator, sample mean. If the performance is not far off too much, then we admitted its goodness. Since truncated data represents part of the data are lost but sample mean represents the complete data condition. We then examined the accuracy of the conventional sample mean estimator. Two different conditions for sample mean were considered. One was to constrain the sample means in the interval of $-0.3\sigma+u\le\bar{x}\le 0.3\sigma+u$. It was referred to as the good sample mean case. The other was to constrain the sample means in the interval of $-2.3\sigma+u\le\bar{x}\le-1.3\sigma+u$ or $1.3\sigma+u\le\bar{x}\le 2.3\sigma+u$, and was referred to as the bad sample mean case. Three estimators were compared: Scheme A represented the conventional sample mean estimator; Scheme B was the coverage-based estimator defined below

$$u^*=u_p=x_{p:n}-\frac{\sum_{t=1}^{v}\left\{D_t\left[E_{\xi_{p:n}|c=Cc_t,Min\{r_s\},n}\left\{\xi_{p:n}\right\}\right]\right\}}{\sum_{t=1}^{v}D_t}\sigma_p \qquad (27)$$

where p was constrained to be either 1 or n which corresponded to the endpoints of the range; and Scheme C was the estimator defined in Eq. (26). If $p=1$, then the term $E_{\xi_{p:n}|c=Cc_t,Min\{r_s\},n}\left\{\xi_{p:n}\right\}$ can be computed by $(-1)E_{\xi_{1:n}|c=Cc_t,Min\{r_s\},n}\left\{\xi_{1:n}\right\}$. The results are displayed in Fig. 7. It can be found from the figure that MSEs were very small for the case of good sample mean for all three estimators; while the MSEs were all large for the case of bad sample mean. This shows that the performance of VTNJE will follow that of the sample mean which is a uniform minimum variance unbiased estimator (UMVUE). In other words, the performances of the two VTNJ estimators follow the best one. Those results also imply that very low MSE can be reached provided that the sample mean is near the population mean. In other words, if we want to obtain a guaranteed coverage, then the difference between the estimated mean and the sample mean should be small.

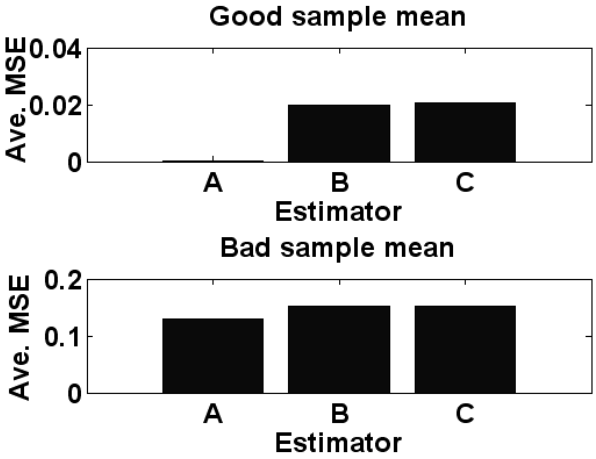

Fig. 7: Comparison of the conventional sample mean estimator and two coverage-based mean estimators.

B. Test the results with consistency to sample mean under the QMI principle—case of realistic percentile

In the test phase, we eliminated the effects caused by the QMI mapping mismatch for $\xi_{1:n}$ to $x_{1:n}$ or $\xi_{n:n}$ to $x_{n:n}$. In such a case, $\xi_{1:n}=(x_{1:n}-u)/\sigma$ and $\xi_{n:n}=(x_{n:n}-u)/\sigma$

were known. But, we pretended that we did not know u and σ. The fluctuation assumption for coverage was therefore not needed. So, the previous formulation could be simplified and expressed by

$$\sigma^* = \frac{\xi_{p:n}\left(\sum_{i=1}^{n}(x_i - x_{p:n})\right)}{2n} \pm \frac{\sqrt{\left(\xi_{p:n}\left(\sum_{i=1}^{n}(x_i - x_{p:n})\right)\right)^2 + 4n^2\left(\sum_{i=1}^{n}(x_i - x_{p:n})^2\right)}}{2n}, \quad (28)$$

for $\sigma^* > 0$ and

$$u^* = \frac{-\left(\left(\overline{x} - x_{p:n}\right)\left(\xi_{p:n}^2\right) - 2x_{p:n}\right)}{2} \pm \frac{\sqrt{\left[\left(\overline{x} - x_{p:n}\right)\left(\xi_{p:n}^2\right) - 2x_{p:n}\right]^2 - 4\left[x_{p:n}^2 + \left(\overline{x}x_{p:n} - \left(\overline{x^2}\right)\right)\left(\xi_{p:n}^2\right)\right]}}{2} \quad (29)$$

where p was constrained to be either 1 or n. Actually, Eq. (28) is equivalent to Eq.(29) because $u^* = x_{p:n} - \xi_{p:n}\sigma^*$. We generated 1,000 trials to examine the new estimator and used MSE as the score of comparison. The results are listed in Table 1.

Table 1: Performance of realistic QMI analysis

Item	Sample mean	Realistic QMI
MSE	0.0765	0.0252

Notice that the MSE of realistic QMI was defined by $\frac{1}{1000}\sum\left(\frac{(u_1 + u_n)}{2} - u\right)^2$, where u_1 and u_n were the estimated results for $x_{1:n}$ and $x_{n:n}$, respectively. It can be found from Table 1 that the realistic QMI mean estimator performed better than the sample mean estimator.

VII. APPLICATION OF USING THE RESULTS OF REALISTIC QMI

The above testing results of realistic QMI show us that if we are able to take the relative coverage for the range, then we can probably reduce the bias of the sample mean. Now we utilize the above results to analyze the problem in more depth. The transform $u = x_{p:n} - \xi_{p:n}\sigma$ has only two degree of freedom. Remember that all coefficients in Eqs.(28) and (29) are simple scalars or polynomials. So, if σ is known, the degree of freedom is reduced to 1. We therefore have an opportunity to approach the real value by recursive estimations.

A. Test the convergence

We use three types of normal distribution to test the robustness of Eq. (28). They are $N(10,1^2)$, $N(12,1.3^2)$ and $N(8,0.6^2)$, respectively. In each test, 1000 trials with 13 samples in each trial were tested. In each trial, we first sorted the 13 samples to find the two endpoints $x_{1:n}$ and $x_{n:n}$. We then take the QMI transform using a pre-assumption pseudo mean, u_s, to obtain $\xi_{1:n}$ and $\xi_{n:n}$. Then, the estimate σ^* was calculated by Eq. (28). We denoted it as σ_p^*. The final estimated mean was obtained by $u_p^* = x_{p:n} - \sigma_p^*\xi_{p:n}$. Let us denote $Tr(i) = \left((u_1^*(i) - u_s)^2 + (u_n^*(i) - u_s)^2\right)/2$ as the mean square error of pseudo mean for the *i*-th trial. Then the average MSE of pseudo mean for the 1000 trials was $1/1000\sum_{i=1}^{1000} Tr(i)$. We took the error between the pseudo mean and real mean, $(u_s - u)$, as a reference. It is noted that the inspection interval for $(u_s - u)$ was $[(-1.5\sigma/\sqrt{n}) + u, (1.5\sigma/\sqrt{n}) + u]$. Fig. 8 displays the average MSE of pseudo mean versus $(u_s - u)$. It can be clearly found from the figure that, for all the three tests using different normal distributions, the average mean square error of pseudo mean became smaller as the absolute value of $(u_s - u)$ decreased. This shows that the estimation will converge by recursive estimation.

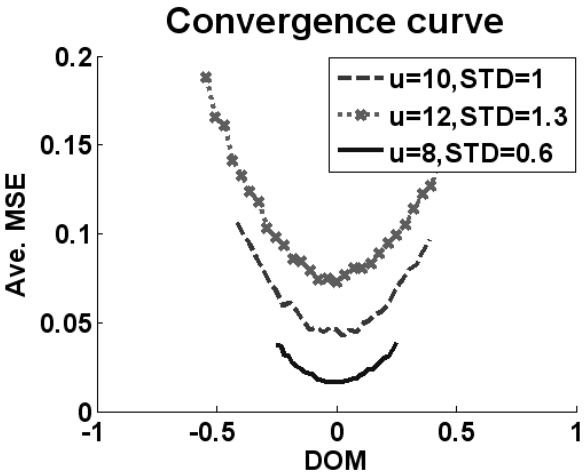

Fig. 8: Convergence curve for *VTNJE* depending on different mean and standard deviation and STD is known, DOM= $(u_s - u)$, u is population mean and STD is population's standard deviation

B. Comparison of the different estimators

We compared three different estimators in term of their stabilities and efficiencies. These three types of estimators are respectively derived by Cohen from Eq.(8) to Eq.(12), another is our VTNJE and the other is the complete data estimator which is the sample mean, the average of total samples. We generated 13 samples each trial which is submitted to the standard normal distribution, $N(0,1^2)$ and totally accounted to 5000 trials

In this scheme we apply three types of non-truncated intervals to force truncating the data outside the non-truncated intervals which represents the truncated percent from low, [-2, 3], to high, [-1.5, 1.75]. In such planning, we may easily to realize the results between the incomplete data, Cohen, VTNJE and complete data, sample mean.

The formulation derived by Cohen request of the initial searching points so that we divided the initial searching condition into two classes, bad and good. The bad condition indicating the initial searching position for mean, u, is outside the interval, $[-2\sigma/\sqrt{n}+u, 2\sigma/\sqrt{n}+u]$ and good condition representing the initial searching position is inside the interval, $[-0.5\sigma/\sqrt{n}+u, 0.5\sigma/\sqrt{n}+u]$. Table 1 is the average of the 5000 MSEs which is the square of the estimator subtracts the real population mean conditional on the bad initial and Table 2 is the same but conditional on the good initial. In these two tables, it is obvious to conclude that our VTNJE is stable and outperforms the Cohen. Furthermore, it is little efficient to the sample mean.

TABLE 1: COMPARISON WITH DIFFERENT ESTIMATORS IN ASSOCIATION WITH BAD INITIAL SEARCHING POINTS

Estimator	Non-truncated Interval		
	[-2,3]	[-1.8,2.5]	[-1.5,1.75]
Cohen	1.439	1.420	0.991
VTNJE	0.061	0.059	0.059
Sample mean	0.078	0.075	0.076

TABLE 2: COMPARISON WITH DIFFERENT ESTIMATORS IN ASSOCIATION WITH GOOD INITIAL SEARCHING POINTS

Estimator	Non-truncated Interval		
	[-2,3]	[-1.8,2.5]	[-1.5,1.75]
Cohen	0.610	0.582	0.731
VTNJE	0.061	0.060	0.059
Sample mean	0.078	0.075	0.076

VIII. DISCUSSIONS AND CONCLUSIONS

This study develops the variably truncated normal joint *pdf* to attack the DM problem. We have discussed the DM problem to demonstrate the weakness of the classical truncated normal distribution when the sample size is less than 20. For coverage interval which is a macro view random variable, we address the *pdf* of coverage interval and show the merit of its shape (see Fig. 6). This proofs the conclusion of Chen [6], who suggested taking the minimum value of coverage interval instead of the other possible ones. Moreover, it is worthy noting that the *pdf* of coverage interval, shown in Eq.(20), is expressed as a result of general form by order statistics which is a non-parametric statistical method and the *pdf* of coverage, shown in Eq.(16), is distribution-free. Hence the VTNJE is appropriate to formulate both the parametric and non-parametric coverage intervals. That is to say, we unify the framework for the parametric and non-parametric coverage intervals.

We use Hermite polynomials to expand the coverage function accurately. It not only uses the high order polynomials to approach the real curve, but also guarantees the convergence for the condition when σ is known in advance (see Eqs.(28) and (29)).

The new truncated normal estimator needs to know only one truncated point for estimation (see Eqs. (25) and (26)); thus it is superior to the old truncated normal estimator which needs a couple of endpoints to do iterations (see Eqs. (9) and (10)).

The third goodness of the VTNJ *pdf* is that it does not need any looking-up table for root-finding and it is expressed in an analytical closed form (see Eqs. (25) and (26)). This may save time for computation. Furthermore, in the default QMI test, we have showed that our coverage-based mean estimator follows the sample mean so that the VTNJ *pdf* also solves the truncated normal problems with knowing only the possible information of the truncated points (see Eq. (21)).

Lastly, we reformulate the equations for the case when σ is known. It works well if σ is known in our estimation process. In the original MLE formulation derived by Cohen, the solving process often encounters the underflow problems. Since the coefficients of the variables are probability or cumulative probability of normal distribution. That is inconvenient for the inverse function representation. We have proofed the convergence if σ is known in advance and show the corresponding convergence curve as Fig. 8. So, our truncated normal estimator outperforms the old one obviously.

IX. APPENDIX : DERIVED THE ALGEBRAIC CLOSED FORM FOR VTNJE

Our principal goal is to establish an analytical form of estimator for the truncated normal distribution so that some special skills will be applied to the whole schemes including marginal likelihood, withdraw certain terms in the derived equations and external adding a certain factor in the equation. If the posterior analysis takes the good performance, then these schemes are right.

First of all, we take integration for the total macro view random variables $x_{1:n}$, r, c accumulating to the log likelihood . The complete marginal log likelihood can be expressed by

$$\begin{aligned} MLL(.) = & \\ & \sum_{t=1}^{v}(D_t\cdot(\int_{dr_s}\int_{d\xi_{1:n}}((-\frac{n}{2}\log 2\pi - n\log\sigma)\times Q))) \\ & +\sum_{t=1}^{v}(D_t\cdot(\int_{dr_s}\int_{d\xi_{1:n}}((-n\log(\Phi(\xi_{1:n}+r_s)-\Phi(\xi_{1:n})) \\ & -\sum_{i=1}^{n}\frac{(x_i-u)^2}{2\sigma^2})\times Q))) \end{aligned} \tag{30}$$

where,

$$D_t = (\frac{b-a}{2}n(n-1)(Cc_t^{\,n-2}-Cc_t^{\,n-1})\times\left(w_{P_v}(\kappa_t)\right)$$

$$Q = p_{\xi_{1:n}|r_s,n}(\xi_{1:n}\mid r_s,n)\cdot p_{r_s|c,n}(r_s\mid c=Cc_t,n)$$

$p_{r_s|c,n}(r_s\mid c=Cc_t,n)$: profile-conditional *pdf* of r_s

We would like to use only one truncated point to process the estimation. Thus, we take the equation $u = x_{1:n} - \sigma\xi_{1:n}$

and Eq.(30) to make a simultaneous equation.

$$\begin{aligned}&\Rightarrow MLL(.)\\&=\sum_{t=1}^{v}(D_t\cdot(\int_{dr_s}\int_{d\xi_{1:n}}((-\frac{n}{2}\log 2\pi-n\log\sigma)Q)))\\&+\sum_{t=1}^{v}(D_t\cdot(\int_{dr_s}\int_{d\xi_{1:n}}(-n\log(\Phi(\xi_{1:n}+r_s)-\Phi(\xi_{1:n}))Q)))\\&+\sum_{t=1}^{v}(D_t\cdot(\int_{dr_s}\int_{d\xi_{1:n}}((-\sum_{i=1}^{n}\frac{(x_i-x_{1:n}+\sigma\xi_{1:n})^2}{2\sigma^2})Q)))\end{aligned}\tag{31}$$

Here we prepared to withdraw the partial expression in Eq.(31) and it is the second term, $\sum_{t=1}^{v}\left\{D_t\cdot\left(\int_{dr_s}\int_{d\xi_{1:n}}(-n\log(\Phi(\xi_{1:n}+r_s)-\Phi(\xi_{1:n}))Q)\right)\right\}$.

We have two reasons to make that decision. Once it is a transcendental function which is difficult to obtain an explicit expression for the variables. The second reason is that we have found $\Phi(\xi_{1:n}+r_s)-\Phi(\xi_{1:n})$ to be a coverage variable. Remember that we considered all the macro view random variables, $x_{1:n}$, r, c combined as a joint *pdf* and if we referred the Eq.(16), we would find the maximum power for coverage is $n-1$ in the *pdf* of coverage. That is, the dominant term has been present in the *pdf* of coverage and no care whether the coverage variable, $\Phi(\xi_{1:n}+r_s)-\Phi(\xi_{1:n})$, is existing.

Take the expansion for the third term:

$$\begin{aligned}&\sum_{t=1}^{v}(D_t\cdot(\int_{dr_s}\int_{d\xi_{1:n}}((-\sum_{i=1}^{n}\frac{(x_i-x_{1:n}+\sigma\xi_{1:n})^2}{2\sigma^2})Q)))\\&=\sum_{t=1}^{v}(D_t\cdot(\int_{dr_s}\int_{d\xi_{1:n}}((-\frac{1}{2}(\sum_{i=1}^{n}(\frac{(x_i-x_{1:n})}{\sigma})^2\\&\qquad+2\frac{(x_i-x_{1:n})}{\sigma}\xi_{1:n}+{\xi_{1:n}}^2))Q)))\end{aligned}\tag{32}$$

$$\begin{aligned}&=\sum_{t=1}^{v}(D_t\cdot(-\frac{1}{2}(\sum_{i=1}^{n}(\frac{(x_i-x_{1:n})}{\sigma})^2\\&+2\frac{(x_i-x_{1:n})}{\sigma}\cdot E_{\xi_{1:n},r_s|c=Cc_t,n}[\xi_{1:n}]+E_{\xi_{1:n},r_s|c=Cc_t,n}[\xi_{1:n}^2])))\end{aligned}\tag{33}$$

$E_{\vartheta}[.]$: take the expectation operator

When we withdraw the coverage term, the equation will become Eq.(33) and the other problem generated. If we inspect Eq.(33), it will be found that there is going to no any coverage interval term, r_s, to be left after integrating the variable r_s. This result violates Eq.(12) derived by Cohen. Since our VTNJ estimator is the extending work of his truncated normal estimator so that we should preserve the information for r_s. Thus we take the suggestion by Chen [6] to select the minimum coverage interval , $Min[r_s]$, representing the information for coverage interval, r_s. The new simplified equation is Eq.(34).

$$\begin{aligned}&\Rightarrow\sum_{t=1}^{v}D_t(-\frac{n}{2}\log 2\pi-n\log\sigma)\\&\quad+\sum_{t=1}^{v}D_t(-\frac{1}{2}\sum_{i=1}^{n}(\frac{(x_i-x_{1:n})}{\sigma})^2\\&\quad-\sum_{i=1}^{n}\frac{(x_i-x_{1:n})}{\sigma}\cdot E_{\xi_{1:n}|c=Cc_t,Min[r_s],n}[\xi_{1:n}]\\&\quad-\frac{n}{2}E_{\xi_{1:n}|c=Cc_t,Min[r_s],n}[\xi_{1:n}^2])\end{aligned}\tag{34}$$

Taking the partial derivative of Eq.(34) with respect to σ and setting it to zero, i.e.,

$$\begin{aligned}&\frac{\partial}{\partial\sigma}MLL(.)=\\&\Rightarrow\sum_{t=1}^{v}(D_t(-\frac{n}{\sigma}+\sum_{i=1}^{n}\frac{(x_i-x_{1:n})^2}{\sigma^3}\\&\qquad+\sum_{i=1}^{n}\frac{(x_i-x_{1:n})}{\sigma^2}\cdot E_{\xi_{1:n}|c=Cc_t,Min[r_s],n}[\xi_{1:n}]))=0\\&=(\sum_{t=1}^{v}nD_t)\sigma^2\\&\quad-(\sum_{t=1}^{v}D_t(E_{\xi_{1:n}|c=Cc_t,Min[r_s],n}[\xi_{1:n}]\sum_{i=1}^{n}(x_i-x_{1:n})))\sigma\\&\quad-\sum_{t=1}^{v}(D_t\sum_{i=1}^{n}(x_i-x_{1:n})^2)=0\end{aligned}\tag{35}$$

Solving Eq.(35), we obtain an estimate of the standard deviation of the population:

$$\sigma^*=\frac{B_\sigma\pm\sqrt{(B_\sigma)^2+4\left(\sum_{t=1}^{v}nD_t\right)C_\sigma}}{2\left(\sum_{t=1}^{v}nD_t\right)},\sigma^*>0\tag{36}$$

$$B_\sigma=\left(\sum_{t=1}^{v}D_t\left(\left(E_{\xi_{1:n}|c=Cc_t,Min[r_s],n}[\xi_{1:n}]\right)\sum_{i=1}^{n}(x_i-x_{1:n})\right)\right)$$

$$C_\sigma=\left(\sum_{t=1}^{v}\left(D_t\sum_{i=1}^{n}(x_i-x_{1:n})^2\right)\right)$$

$$D_t=(\frac{b-a}{2}n(n-1)(Cc_t^{\,n-2}-Cc_t^{\,n-1})\times\left(w_{P_v}(\kappa_t)\right)$$

Then, we consider to substitute σ to u in order to get a new quadratic equation of u.

By substituting $\sigma=(x_{1:n}-u)/\xi_{1:n}$ in Eq.(30), we obtain

$$\begin{aligned}&MLL(.)\\&=\sum_{t=1}^{v}(D_t\cdot(\int_{dr_s}\int_{d\xi_{1:n}}((-\frac{n}{2}\log 2\pi-n\log(\frac{x_{1:n}-u}{\xi_{1:n}}))\times Q)))\\&\quad+\sum_{t=1}^{v}(D_t\cdot(\int_{dr_s}\int_{d\xi_{1:n}}((-n\log(\Phi(\xi_{1:n}+r_s)-\Phi(\xi_{1:n}))\\&\qquad-\sum_{i=1}^{n}\frac{(x_i-u)^2}{2(x_{1:n}-u)^2}\xi_{1:n}^2)\times Q)))\end{aligned}\tag{37}$$

Taking the partial derivative of Eq.(34) with respect to

u and set it to zero, i.e.

$$\frac{\partial}{\partial u} MLL(.) = 0$$

$$\Rightarrow \sum_{t=1}^{v}(D_t \frac{n}{x_{1:n}-u}) - \sum_{t=1}^{v}(D_t \sum_{i=1}^{n} \frac{(x_i - x_{1:n})(x_i - u)}{(x_{1:n}-u)^3} E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}]) = 0$$

$$= \sum_{t=1}^{v}(nD_t(x_{1:n}-u)^2) - \sum_{t=1}^{v}(D_t \sum_{i=1}^{n}((x_i^2 - x_i u - x_{1:n}x_i + ux_{1:n})E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}])) = 0$$

$$= \sum_{t=1}^{v}(nD_t(x_{1:n}-u)^2) - \sum_{t=1}^{v}(D_t((\sum_{i=1}^{n} x_i^2 - n\bar{x}u - n\bar{x}x_{1:n} + nux_{1:n})E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}])) = 0$$

$$= \sum_{t=1}^{v}(D_t(x_{1:n}-u)^2) - \sum_{t=1}^{v}(D_t((\overline{x^2} - (\bar{x} - x_{1:n})u - \bar{x}x_{1:n})E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}])) = 0 \quad (38)$$

.

$$\Rightarrow \sum_{t=1}^{v}(D_t(x_{1:n}^2 - 2x_{1:n}u + u^2)) - \sum_{t=1}^{v}(D_t((\overline{x^2})(E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}]))) + (\sum_{t=1}^{v}(D_t((\bar{x} - x_{1:n})(E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}])))u + \sum_{t=1}^{v}(D_t((\bar{x}x_{1:n})(E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}]))) = 0 \quad (39)$$

With simple mathematical manipulations, the above equation can be simplified and expressed by

$$\Rightarrow \sum_{t=1}^{v}(D_t)u^2 + (\sum_{t=1}^{v}(D_t((\bar{x} - x_{1:n})(E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}]) - 2x_{1:n})))u + \sum_{t=1}^{v}(D_t(x_{1:n}^2 + (\bar{x}x_{1:n} - (\overline{x^2}))(E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}]))) = 0 \quad (40)$$

Solving Eq.(40), we obtain an estimator of u:

$$\Rightarrow u^* = \frac{-B_u \pm \sqrt{B_u^2 - 4(\sum_{t=1}^{v} D_t)C_u}}{2(\sum_{t=1}^{v} D_t)} \quad (41)$$

where

$$B_u = \sum_{t=1}^{v}\left(D_t\left((\bar{x} - x_{1:n})\left(E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}]\right) - 2x_{1:n}\right)\right)$$

$$C_u = \sum_{t=1}^{v}\left(D_t\left(x_{1:n}^2 + \left(\bar{x}x_{1:n} - \left(\overline{x^2}\right)\right)\left(E_{\xi_{1:n}|c=Cc_t, Min[r_s],n}[\xi^2_{1:n}]\right)\right)\right)$$

REFERENCES

[1] A. Clifford Cohen. "Truncated and censored samples — theory and applications," New York, Marcel Dekker, pp.31-43, 1991.

[2] Paweł Fotowicz, "An analytical method for calculating a coverage interval," Metrologia, vol. 43, pp.42-45, 2006.

[3] Shuo-Huei Lin, Wenyew Chan and Lin-An Chen, "A non-parametric coverage interval," Metrologia, vol. 45, pp. L1-L4, 2008.

[4] "Guide to the expression of uncertainty in measurement supplement 1 — numerical methods for the propagation of distributions draft of JCGM document," JCGM, p.38, 2004..

[5] Lin-An Chen and Hui-Nien Hung, "Extending the discussion on coverage intervals and statistical coverage intervals," Metrologia, vol. 43, pp.L43-L44, 2006.

[6] Lin-An Chen, Jing-Ye Huang and Hung-Chia Chen, "Parametric coverage interval," Metrologia, vol. 44, pp.L7-L9, 2007.

[7] N. Balarkrishnan and A. Clifford Cohen , "Order statistics and inference estimation methods," Academic Press, Inc., 1991.

[8] E. H. Lloyd, "Least-square estimation of location and scale parameters using order statistics," Biometrika, vol. 39, pp.88-95, 1952.

[9] D. Teichroew, "Tables of expected values of order statistics for samples of size twenty and less from the normal distribution," The Ann. of Math. Stat., vol. 27, pp.410-426, 1956.

[10] Ahmed E. Sarhan and Bernard G. Greenberg, "Estimation of location and scale parameters by order statistics from singly and doubly censored samples, part one. The normal distribution up to size 10," The Ann. of Math. Stat., vol. 27, pp.427-451, 1956, (correction , vol. 40, p.325)

[11] Ahmed E. Sarhan and Bernard G. Greenberg. eds., "Contributions to order statistics," Wiley, New York, 1962.

[12] Wilks, S. S., "Statistical prediction with special reference to the problem of tolerance limits," Ann. Math. Statist. vol. 13, pp. 400-409, 1948

[13] Wilks, S. S., "Mathematical Statistics," Wiley, New York, 1962.

[14] John W. Pratt and Jean D. Gibbsons, "Concepts of nonparametric theory," Springer series in Statistics, Spring Verlag, 1981.

[15] P. Abbott, "Tricks of the trade: Legendre-Gauss quadrature," Mathematica Journal, vol. 9, pp.689-691, 2005.