Inference and Density Estimation with Interval Statistics

Ximing Wu*

November, 2006

Abstract

Individual data from a continuous distribution are often partitioned into a collection of intervals defined by either fixed interval limits or sample quantiles. In this study, we derive asymptotic distribution of interval statistics for both cases, allowing multiple sample statistics for each interval. Under fixed intervals, the covariance matrix is singular. We identify a computationally simple non-singular generalized inverse for corresponding $\chi^2$ test and density estimator. For quantile intervals, the interval limits are stochastic, which complicates the asymptotic distribution. We use an influence function approach to derive the joint distribution of interval limits and other interval statistics. We propose minimum distance estimators of the underlying distribution based on interval statistics. Asymptotic properties of the proposed estimators are established. Monte Carlo simulations suggest that the proposed estimators provide good finite sample performance. An empirical example on income distribution estimation based on interval statistics is presented.

JEL Classification: C13, C16, C25

Keywords: Interval Statistics; Density Estimation; Minimum Distance Estimator

*Department of Agricultural Economics, Texas A&M University, College Station, TX, 77843. Email: xwu@ag.tamu.edu.
1 Introduction

It is a common practice in economics and statistics that individual data from a continuous distribution is summarized in the form of sample statistics by intervals. As a result, researchers often face the problem of undertaking estimation and inference based on interval statistics. The most common interval statistic is the grouped data, where the sample is partitioned into a collection of intervals, and individual observations are represented by a binary variable indicating the intervals into which they have fallen. For example, individual income data are typically grouped to preserve confidentiality. Very often the mean of the entire sample is reported, but not the dispersion. Subsequently, the estimation of dispersion or inequality measures from grouped income data has been a focus of the inequality measurement literature. See for example, Gastwirth (1972), Gastwirth and Glauberman (1976), Kawani and Podder (1976), Cowell and Mehta (1982), Gastwirth et al. (1986), and Davies and Shorrocks (1989). Maasoumi (1998) reviews inequality measurement with grouped data. Heitjan (1989) offers a general review of inference with grouped data.

Inequality or dispersion only captures one aspect of the underlying distribution. Often it is desirable to estimate the underlying continuous distribution from grouped data. Given a parametric distribution, the MLE is efficient when individual data are available. However with grouped data, one can only use likelihood function based on the probability of each interval. For example, McDonald (1984) fits grouped data of U.S. income distribution to generalized beta distributions using multinomial MLE. Other distributions considered in the literature include Pareto (Aigner and Goldberg, 1970), exponential (Chen, 1996) and log-normal (McLaren et al., 1991). Burridge (1982) discusses estimation of unimodal distributions from grouped data. Alternatively, Hogg and Klugman (1983) and Wu and Perloff (forthcoming) present minimum distance estimators, and Krieger and Gastwirth (1984) use interpolation to estimate the underlying distribution. Victoria-Fester and Ronchetti (1997) discuss robust estimation with grouped data.

In this study, we examine the estimation and inference with general interval statistics, which include not only interval frequency, but also other sample statistics, such as mean or variance of each interval. One can summarize a sample of individual data by interval statistics in a variety of ways. Depending on how the intervals are defined, we group them
into two broad categories:

- Fixed interval: The intervals are defined by fixed or predetermined values, often multiples of 10's, 50's, 100's and so on. The frequency and/or other statistics of each interval are reported.

- Quantile interval: The intervals are defined by sample quantiles such as quartiles, quintiles or deciles. The limits and/or other statistics of each interval are reported.

The main reason we suggest this classification is that the asymptotic distribution of interval statistics differs between these two grouping methods. To the best of our knowledge, almost all existing work focuses on fixed intervals with only interval frequency reported. The asymptotic distribution of multiple statistics on intervals of either type have not been systematically examined, nor has estimation of the underlying distribution based on them. The goal of this study is three-folded. For both types of intervals, (i) we derive the asymptotic distribution of interval statistics in a most general framework; (ii) we propose a goodness-of-fit test for interval statistics; (iii) we develop a minimum distance estimator of the underlying distribution based on interval statistics.

Given fixed intervals, suppose the frequency and other statistics, such as the mean or variance, of each interval are reported. We show that their joint distribution is asymptotically normal with a singular covariance matrix. A $\chi^2$ goodness-of-fit test for interval statistics is proposed, where an arbitrary generalized inverse of the asymptotic covariance matrix can be used as weighting matrix. We then propose a minimum distance estimator of the underlying distribution based on interval statistics. When only a frequency table is available, the MLE is traditionally used to estimate the underlying distribution. However, the likelihood function is usually rather complicated if multiple statistics for each interval are reported. For example, it is difficult to define the likelihood function for MLE if the Gini index of each interval is provided together with interval frequency. In contrast, the proposed minimum distance estimator can easily accommodate multiple interval statistics of various forms. Moreover, we present a non-singular generalized inverse of the covariance matrix. This computationally simple matrix is used as optimal weighting matrix for the proposed minimum distance estimator.
Compared to the case of fixed intervals, the asymptotic distribution of interval statistics is more complicated when the intervals are defined by sample quantiles. This is because interval limits are now themselves stochastic. We derive the joint distribution of interval limits and other interval statistics using an influence function approach. A \( \chi^2 \) goodness-of-fit test and a minimum distance estimator are presented. Often the sample quantiles used to define the intervals are not reported. Hence, we further extend the proposed estimator to accommodate this case.

We investigate the finite sample performance of the proposed estimators using a series of Monte Carlo simulations. The estimators perform well with small sample sizes. We show that in addition to interval frequency (for fixed intervals) or sample quantiles (for quantile intervals), incorporating interval means can substantially improve the estimate. Lastly, we present an illustrative example by applying the proposed method to estimating China’s income distribution for rural and urban areas, for which only interval statistics are available.

The rest of the paper is organized as follows. In section 2, we derive the asymptotic distribution of general interval statistics and corresponding goodness-of-fit test. In Section 3, we present a minimum distance estimator for distributions based on interval statistics and discuss its asymptotic properties. We present Monte Carlo simulations and an empirical application in Section 4 and 5 respectively. Some concluding remarks are given in Section 6.

2 Asymptotic Distribution of Interval Statistics

Suppose \( \{X_t\}_{t=1}^n \) is an i.i.d. sample from a distribution \( F(x) \) defined on \( A \in \mathcal{R} \). It is assumed that \( F(x) \) is continuous and differentiable to at least second order, and has a density function \( f(x) = F'(x) \). The support of the distribution is partitioned into \( J \) intervals \( A_1, \ldots, A_J \) with \( P(X_1 \in A_j) = s_j > 0 \) for \( j = 1, \ldots, J \). Let \( \theta(x) : A \to \mathcal{R} \) be a Borel measurable function with \( \int \theta^2(x) dF(x) < \infty \). In this section, we derive the asymptotic distribution of interval conditional moments of \( \theta(x) \) jointly with interval frequency or interval limits. For the ease of exposition, we focus on the case where interval statistics include interval limits, interval frequency and interval conditional mean of \( \theta(x) \). Extension to more general cases, such as higher order moments of \( \theta(x) \), is straightforward.
2.1 Fixed intervals

We first discuss the case where interval limits are fixed or pre-determined. The most common example is a frequency table with non-stochastic interval limits. Define the interval indicator

$$I_j (X_t) = \begin{cases} 
1, & x_t \in A_j \\
0, & x_t \not\in A_j
\end{cases}.$$ 

Let $s = [s_1, \ldots, s_J]'$ and $\hat{s} = [n_1/n, \ldots, n_J/n]'$, where $n_j = \sum_{t=1}^{n} I_j (X_t)$ for $j = 1, \ldots, J$. It is well-known that

$$\sqrt{n} (\hat{s} - s) \xrightarrow{d} N(0, \Omega_1),$$

where $\Omega_1 = [\omega_{1,ij}]_{J \times J}$ with

$$\omega_{1,ij} = \begin{cases} 
s_i (1 - s_i), & i = j \\
-s_i s_j, & i \neq j
\end{cases}.$$ (1)

Note that $\text{rank}(\Omega_1) = J - 1$. Suppose $s$ is given by a completely specified distribution $F(x)$ such that $s_j = \int I_j(x) dF(x)$ for $j = 1, \ldots, J$. The classical Neyman $\chi^2$ test is used to test the goodness-of-fit of $F(x)$ to $\hat{s}$. The test statistic

$$Q_n = n \sum_{j=1}^{J} \frac{(\hat{s}_j - s_j)^2}{s_j}$$

is distributed asymptotically according to $\chi^2_{J-1}$.

We define truncated first and second moment of $\theta(x)$ on the $j^{th}$ interval:

$$\mu_j = \int_{A_j} \theta(x) dF(x) = E [\theta(X_1) I_j (X_1)],$$

$$\mu_j^{(2)} = \int_{A_j} \theta^2(x) dF(x) = E [\theta^2(X_1) I_j (X_1)],$$

and their sample analogs $\hat{\mu}_j = \frac{1}{n} \sum_{t=1}^{n} \theta(X_t) I_j (X_t)$ and $\hat{\mu}_j^{(2)} = \frac{1}{n} \sum_{t=1}^{n} \theta^2(X_t) I_j (X_t)$. Next we derive the joint asymptotic distribution of $\hat{s}$ and $\hat{\mu}$. Since $\int \theta^2(x) dF(x) < \infty$, it follows
that \( \mu_j^{(2)} \leq \infty \) for all \( j \). One can then easily show that

\[
\text{var}(\theta(X_1)I_j(X_1)) = E[\theta^2(X_1)I_j(X_1)] - \mu_j^2 = \mu_j^{(2)} - \mu_j^2,
\]

\[
\text{cov}(Y_{j1}, \theta(X_1)I_j(X_1)) = E[(I_j(X_1) - s_j)(\theta(X_1)I_j(X_1) - \mu_j)] = (1 - s_j)\mu_j.
\]

For \( i \neq j \),

\[
\text{cov}(\theta(X_1)I_i(X_1), \theta(X_1)I_j(X_1)) = E[(\theta(X_1)I_i(X_1) - \mu_i)(\theta(X_1)I_j(X_1) - \mu_j)] = -\mu_i\mu_j,
\]

\[
\text{cov}(I_i(X_1), \theta(X_1)I_j(X_1)) = E[(I_i(X_1) - s_i)(\theta(X_1)I_j(X_1) - \mu_j)] = -s_i\mu_j.
\]

It follows that the covariance between \( \hat{s} \) and \( \hat{\mu} \) is given by \( \Omega_2/n = [\omega_{2,ij}/n]_{J \times J} \) with

\[
\omega_{2,ij} = \begin{cases} 
(1 - s_i)\mu_i, & i = j \\
-s_i\mu_j, & i \neq j 
\end{cases}
\]  

(2)

and the variance of \( \hat{\mu} \) is given by \( \Omega_3/n = [\omega_{3,ij}/n]_{J \times J} \) with

\[
\omega_{3,ij} = \begin{cases} 
\mu_i^{(2)} - \mu_i^2, & i = j \\
-\mu_i\mu_j, & i \neq j 
\end{cases}
\]  

(3)

We now establish the joint distribution of \( \hat{s} \) and \( \hat{\mu} \).

**Proposition 1.** Given \( \Omega_1, \Omega_2 \) and \( \Omega_3 \) defined by (1), (2) and (3) respectively,

\[
\hat{W} = \sqrt{n} \left( \begin{array}{c} \hat{s} - s \\ \hat{\mu} - \mu \end{array} \right) \overset{d}{\to} N \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \begin{bmatrix} \Omega_1 & \Omega_2 \\ \Omega_2' & \Omega_3 \end{bmatrix}
\]

[All proofs are placed in Appendix.]

Next, we derive a goodness-of-fit test of a complete specified distribution to \( \hat{W} \). We aim to construct a test based on a proper quadratic transform of the multivariate normal sample statistics, which is expected to distributed asymptotically as a \( \chi^2 \) distribution. We present the underlying theorem below for the ease of reference.

**Serfling’s Theorem 3.5** (Serfling (2001), p.128). Let \( W = [W_1, \ldots, W_k] \) be \( N(\mu, \Omega) \), and
let $C_{k \times k}$ be a symmetric matrix. Assume that, for $\eta = (\eta_1, \ldots, \eta_k)$,

$$\eta\Omega = 0 \Rightarrow \eta\mu' = 0.$$ 

Then $WCW'$ has a (possibly noncentral) $\chi^2$ distribution if and only if

$$\Omega C \Omega C \Omega = \Omega C \Omega,$$

in which case the degrees of freedom is $\text{trace}(C\Omega)$ and the noncentrality parameter is $\mu C \mu'$. 

If $\Omega$ is non-singular, $C = \Omega^{-1}$ is the obvious choice. However in Proposition 1, $\Omega$ is not full-ranked. Therefore, we use instead a generalized inverse of $\Omega$ to construct the test. Define $\text{diag}(z)$ as the diagonal vector of square matrix $z$. Let $D_s = \text{diag}(s), D_\mu = \text{diag}(\mu), \nu_j = E(X_1|I_j(X_1) = 1), D_\nu = \text{diag}(\nu)$, $j = 1, \ldots, J$. Since $\nu_j = \mu_j/s_j$, it follows that

$$\Omega_2 = \Omega_1 D_s^{-1}D_\mu = \Omega_1 D_\nu.$$ 

Hence $\text{rank}([\Omega_1, \Omega_2]) = \text{rank}(\Omega_1) = J - 1$.

Further denote $\sigma^2 = [\sigma_1^2, \ldots, \sigma_J^2]$ with $\sigma_j^2 = \text{var}(X_1|I_j(X_1) = 1), D_\sigma = \text{diag}(\sigma^2)$. Since $\mu_j^{(2)} = s_j(\sigma_j^2 + \nu_j^2)$, we have for $\Omega_3$,

$$\omega_{3,ij} = \begin{cases} s_i\nu_i^2 - s_i^2\nu_i^2 + s_i\sigma_i^2, & i = j \\ -\nu_i s_i s_j \nu_j, & i \neq j \end{cases}.$$ 

It then follows that

$$\Omega_3 = D_\nu \Omega_1 D_\nu + D_s D_\sigma.$$ 

Therefore, we can rewrite the covariance matrix for $\hat{W}$

$$\Omega = \begin{bmatrix} \Omega_1 & \Omega_2 \\ \Omega_2' & \Omega_3 \end{bmatrix} = \begin{bmatrix} \Omega_1 & \Omega_1 D_\nu \\ D_\nu \Omega_1 & D_\nu \Omega_1 D_\nu + D_s D_\sigma \end{bmatrix}.$$ 

Assuming $\sigma_j^2 > 0$ for all $j$, $\Omega_3$ has a full rank $J$ thanks to the “ridge” $D_s D_\sigma$. Therefore, the vector space of $\Omega$ is spanned by those of $\Omega_1$ and $\Omega_3$, which has a total rank of $2J - 1$. 

7
We further note that the $\chi^2$ test statistic for interval frequency $\hat{s}$, whose covariance matrix $\Omega_1$ is singular, uses the non-singular $D_s^{-1}$ as weighting matrix. We conjecture that one can construct the weighting matrix of the $\chi^2$ statistic for $\hat{W}$ by replacing the upper left $\Omega_1$ in (5) with $D_s$ then taking inverse. This heuristic argument is presented formally below.

**Theorem 2.** Define the weighting matrix $C = \begin{bmatrix} D_s & \Omega_2 \\ \Omega_2' & \Omega_3 \end{bmatrix}^{-1}$ and its sample analog $\hat{C}$. The goodness-of-fit statistic for $\hat{W}$ is given by

$$\hat{W}\hat{C}\hat{W}' \xrightarrow{d} \chi^2_{2J-1}.$$  

**Remark 1.** Note that there is an infinite number of generalized inverses of $\Omega$ that can be used in the construction of $\chi^2$ test above. However, Theorem 2 identifies a full-ranked computationally simple weighting matrix $C$. In next section, we use this non-singular weighting matrix as optimal weighting matrix for a minimum distance estimator of density based on interval statistics.

**Remark 2.** Section 3 discusses density estimation based on interval statistics. The estimates based on interval frequency alone and based on interval frequency and conditional expectation of $\theta(x)$ are both consistent, but the latter is generally more efficient. Denote $\Gamma_1 = \text{trace}(\Omega_1)^{-1}$ and $\Gamma = \text{trace}(\Omega)^{-1}$, where $\Omega_1$ and $\Omega$ are given by (1) and (5) respectively. The efficiency gain in terms of mean squared errors of density estimate is proportional to $\Gamma_1 - \Gamma$. Equation (4) suggests the larger is the within interval variation $\sigma^2$, the larger the efficiency gain. On the other hand, if all elements of $D_s$ are zero in (4), there shall be no efficiency gain from incorporating the interval conditional mean $\theta(x)$ for the density estimation. Intuitively, when all elements of $D_s$ are zero, the distribution degenerates to a discrete one. In this case, the interval frequency is the sufficient statistics, any additional information is redundant for the estimation of the underlying distribution.

### 2.2 Quantile intervals

This section derives the asymptotic distribution of interval statistics when intervals are defined by sample quantiles. Denote the cumulative probability for the $j^{th}$ interval $p_j =$
It is also defined \( p_0 = 0 \) and \( p_J = 1 \). Suppose a given random sample is partitioned into \( J \) intervals by sample quantiles \( \hat{\xi}_1, \ldots, \hat{\xi}_{J-1} \), where \( p_j = \frac{1}{n} \sum_{t=1}^{n} I \left( X_t \leq \hat{\xi}_j \right) \) for \( j = 1, \ldots, J \). We first present the distribution of \( \hat{\xi} = \left[ \hat{\xi}_1, \ldots, \hat{\xi}_{J-1} \right]' \), which serve as interval limits.

**Lemma 1.** Denote \( f_j = f \left( F^{-1} (p_j) \right) \) for \( j = 1, \ldots, J - 1 \). We have

\[
\sqrt{n} \left( \hat{\xi} - \xi \right) \overset{d}{\rightarrow} N \left( 0, \Omega \right),
\]

where \( \Omega = \{ \omega_{ij} \} \) with \( \omega_{ij} = p_i (1 - p_j) / f_j^2 \) for \( i = j \) and \( \omega_{ij} = p_i (1 - p_j) / (f_i f_j) \) for \( 1 < i < j < J - 1 \).

We then define conditional expectation of \( \theta (X) \) on the \( j \)th interval

\[
\nu_j = \frac{1}{p_j - p_{j-1}} \int_{\hat{\xi}_{j-1}}^{\hat{\xi}_j} \theta (x) \, dF (x),
\]

and its sample analog

\[
\hat{\nu}_j = \frac{\sum_{t=1}^{n} \theta (X_t) I \left( \hat{\xi}_{j-1} \leq X_t \leq \hat{\xi}_j \right)}{\sum_{t=1}^{n} I \left( \hat{\xi}_{j-1} \leq X_t \leq \hat{\xi}_j \right)}
\]

for \( j = 1, \ldots, J \), where \( \hat{\xi}_0 = \min (\{ X_t \}_{t=1}^{n}) \) and \( \hat{\xi}_J = \max (\{ X_t \}_{t=1}^{n}) \).

When intervals are defined by sample quantiles, the asymptotic distribution of interval statistics is more complicated compared to the case of fixed intervals. This is because interval limits are stochastic for quantile intervals. To derive the asymptotic distribution of \( \hat{\nu} = [\hat{\nu}_1, \ldots, \hat{\nu}_J]' \), we first present two lemmas that will be needed to prove the main result of this
The following notations are used:

\[
\begin{align*}
\tau_j &= p_{j-1} \theta (\xi_{j-1}) + (p_j - p_{j-1}) \nu_j + (1 - p_j) \theta (\xi_j), \\
\eta_j &= \nu_j - \tau_j, \\
\eta_j^- &= \theta (\xi_{j-1}) - \tau_j, \\
\eta_j^+ &= \theta (\xi_j) - \tau_j, \\
\nu_j^{(2)} &= \frac{1}{p_j - p_{j-1}} \int_{\xi_{j-1}}^{\xi_j} \theta^2 (x) dF (x).
\end{align*}
\]

The first lemma provides the asymptotic distribution of a single conditional mean of \( \theta (x) \) on an interval defined by sample quantiles.

**Lemma 2.** For \( j = 1, \ldots, J \), define

\[
\psi_j^2 = \frac{p_{j-1} (\eta_j^-)^2 + (p_j - p_{j-1}) \left( \nu_j^{(2)} - 2 \nu_j \tau_j + \tau_j^2 \right) + (1 - p_j) (\eta_j^+)^2}{(p_j - p_{j-1})^2}.
\]  

We then have

\[
\sqrt{n} (\hat{\nu}_j - \nu_j) \xrightarrow{d} N \left( 0, \psi_j^2 \right).
\]

To obtain joint distribution of \( \hat{\nu} \), we also need the covariance between \( \hat{\nu}_i \) and \( \hat{\nu}_j \) for \( i \neq j \).

The second lemma provides the needed result.

**Lemma 3.** Let integer \( 1 \leq i < j \leq J \). Define

\[
\psi_{ij} = \frac{1}{(p_i - p_{i-1}) (p_j - p_{j-1})} \left[ p_{i-1} \eta_i^- \eta_j^- + (p_i - p_{i-1}) \eta_i^- \eta_j^- \right. \\
+ (p_j - p_{j-1}) \eta_i^+ \eta_j^- + (p_j - p_{j-1}) \eta_i^+ \eta_j^+ + (1 - p_j) \eta_i^+ \eta_j^+ \right].
\]

The covariance between \( \hat{\nu}_i \) and \( \hat{\nu}_j \) is given by \( \frac{\psi_{ij}}{n} \).

Next, applying Lemma 1 and Lemma 2 to \( \hat{\nu} \) yields the asymptotic joint distribution.

**Proposition 3.** The asymptotic distribution of interval statistics \( \hat{\nu} \) is given by

\[
\sqrt{n} (\hat{\nu} - \nu) \xrightarrow{d} N \left( 0, \Omega \right),
\]
where the diagonal elements of $\Omega$ are given by (6), and the off-diagonal elements by (7).

Lastly, to derive the joint distribution of $\left(\hat{\xi}, \hat{\nu}\right)$, we need the covariance between sample quantiles and sample statistics on quantile intervals. The result is given by the lemma below.

**Lemma 4.** When $i \leq j - 1$,

$$
cov(\hat{\xi}_i, \hat{\nu}_j) = \frac{p_i \left[ (p_{j-1} - 1) \eta_j + (p_j - p_{j-1}) \eta_j + (1 - p_j) \eta_j^+ \right]}{nf_i};
$$

when $i \geq j$,

$$
cov(\hat{\xi}_i, \hat{\nu}_j) = \frac{(p_i - 1) \left[ p_{j-1} \eta_j - (p_j - p_{j-1}) \eta_j - p_j \eta_j^+ \right]}{nf_i}.
$$

Combining Proposition 3, Lemma 1 and 4, we obtain the following result.

**Proposition 4.** Let $\Omega_1$ be given by Lemma 1, $\Omega_2$ by Lemma 4, and $\Omega_3$ by Proposition 3. The asymptotic joint distribution of sample quantiles and sample statistics on quantile intervals is given by

$$
\hat{W} = \sqrt{n} \left( \begin{array}{c} \hat{\xi} - \xi \\ \hat{\nu} - \nu \end{array} \right) \overset{d}{\rightarrow} N \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left[ \begin{array}{cc} \Omega_1 & \Omega_2 \\ \Omega_2' & \Omega_3 \end{array} \right].
$$

The goodness-of-fit test for $\hat{W}$ is then immediate.

**Theorem 5.** Let $\Omega$ be the covariance matrix for $\hat{W}$ as defined in Proposition 4 and $\hat{\Omega}$ its sample analog. Under the assumption that $\Omega$ is full-ranked,

$$
\hat{W} \hat{\Omega}^{-1} \hat{W} \overset{d}{\rightarrow} \chi^2_{2J-1}.
$$

**Remark 3.** Given $J$ intervals, the degrees of freedom of the proposed $\chi^2$ tests for fixed interval case and quantile interval case are both $2J - 1$, but for different reasons. For the former, $\hat{W}$ is a vector of length $2J$, but its covariance matrix is singular with reduced rank $2J - 1$. For the latter, the covariance matrix is full-ranked, but $\hat{W}$ is a vector of length $2J - 1$ because the sample quantile $\hat{\xi}$ has only $J - 1$ elements when the sample is partitioned into $J$ intervals.
3 Minimum distance estimator

In this section, we present a minimum distance estimator of the underlying distribution based on interval statistics, whose asymptotic distributions are discussed above.

Consider an i.i.d. sample \( \{X_i\}_{i=1}^n \) from a continuous distribution \( F(x; \lambda_0) \) with density \( f(x; \lambda_0) \). Let \( f(x; \lambda) \) be an estimate of \( f(x; \lambda_0) \), and \( \xi_j(\lambda) \) is the \( p_j^{th} \) quantile of \( F(x; \lambda) \).

When interval limits are fixed, define

\[
m_1(\lambda) = \begin{bmatrix}
\int_{A_1} f(x; \lambda) \, dx - \hat{s}_1 \\
\vdots \\
\int_{A_J} f(x; \lambda) \, dx - \hat{s}_J \\
\int_{A_1} \theta(x) f(x; \lambda) \, dx - \hat{\mu}_1 \\
\vdots \\
\int_{A_J} \theta(x) f(x; \lambda) \, dx - \hat{\mu}_J
\end{bmatrix}.
\]

When intervals are defined by sample quantiles, define

\[
m_2(\lambda) = \begin{bmatrix}
F^{-1}(p_1; \lambda) - \hat{\xi}_1 \\
\vdots \\
F^{-1}(p_{J-1}; \lambda) - \hat{\xi}_{J-1} \\
\frac{1}{p_1-p_0} \int_{F^{-1}(p_1; \lambda)}^{F^{-1}(p_0; \lambda)} \theta(x) f(x; \lambda) \, dx - \hat{\nu}_1 \\
\vdots \\
\frac{1}{1-p_J} \int_{F^{-1}(1; \lambda)}^{F^{-1}(p_J; \lambda)} \theta(x) f(x; \lambda) \, dx - \hat{\nu}_J
\end{bmatrix}.
\]

The following conditions are assumed.

**Assumptions**

For \( k = 1, 2 \),

1. \( \lambda_0 \in \Lambda \), which is compact, is the unique solution to the moment condition \( Em_k(\lambda) = 0 \);

2. \( \sup_{\lambda \in \Lambda} \int \theta^2(x) \, dF(x; \lambda) < \infty \);

3. \( m_k(\lambda) \) is continuous on \( \Lambda \) and for all \( \lambda \in \Lambda \) it is twice continuously differentiable in a neighborhood \( \mathcal{N} \) of \( \lambda_0 \);
4. \sup_{\lambda \in A} \left\| \nabla_{\lambda} m_k (\hat{\lambda}) - G_k (\lambda_0) \right\| \xrightarrow{P} 0;

5. Let \( \Psi_1 \) and \( \Psi_2 \) be the asymptotic covariance matrix given respectively in Proposition 1 and 4. \( \Psi_k (\hat{\lambda}) \xrightarrow{P} \Psi_k (\lambda_0) \), which is positive-definite;

6. \( G_k (\lambda_0)^t \Psi_k^{-1} (\lambda_0) G_k (\lambda_0) \) is non-singular.

**Proposition 6.** Define the minimum distance estimator

\[
\hat{\lambda} = \arg \min_{\lambda} m_k (\lambda) \Psi_k^{-1} (\lambda) m_k (\lambda), k = 1, 2.
\]

Under assumptions 1-6,

\[
\sqrt{n} \left( \hat{\lambda} - \lambda_0 \right) \xrightarrow{d} N \left( 0, [G_k (\lambda_0)^t \Psi_k^{-1} (\lambda_0) G_k (\lambda_0)]^{-1} \right).
\]

Under assumption 2, \( G_1 (\lambda) \), the gradient for \( m_1 (x; \lambda) \), is well defined and can be easily derived. The gradient for \( m_2 (x; \lambda) \) is more complicated as we need to take into account the stochastic nature of quantile intervals. To calculate the gradient of \( m_2 (x; \lambda) \), we first define the derivative of the quantile function with respect to a scalar parameter in the following proposition.

**Proposition 7.** Suppose the length of \( \lambda \) is \( Q \). Let \( \delta \) be a vector of zeros with the same length and its \( q \)th element replaced by \( \delta_q \), \( 1 \leq q \leq Q \). Under the assumption that \( F (x; \lambda) \) is continuously differentiable,

\[
\frac{\partial F^{-1} (p; \lambda)}{\partial \lambda_q} = \lim_{\delta_q \to 0} \frac{p - F (F^{-1} (p; \lambda); \lambda + \delta) / \delta_q}{f (F^{-1} (p; \lambda); \lambda)}.
\]

The derivative for the quantile function with respect to \( \lambda \) is then calculated elementwisely using (8). In particular, the \( q \)th element of the gradient for \( m_2 (\lambda) \) is given by, according to the Leibniz’ formula,

\[
\frac{\partial}{\partial \lambda_q} \int_{F^{-1}(p_{j-1}; \lambda)}^{F^{-1}(p_j; \lambda)} \theta (x) f (x; \lambda) \, dx = \int_{F^{-1}(p_{j-1}; \lambda)}^{F^{-1}(p_j; \lambda)} \theta (x) \frac{\partial f (x; \lambda)}{\partial \lambda_q} \, dx \\
+ \theta (x) f (x; \lambda) \frac{\partial F^{-1} (p_j; \lambda)}{\partial \lambda_q} - \theta (x) f (x; \lambda) \frac{\partial F^{-1} (p_{j-1}; \lambda)}{\partial \lambda_q}.
\]
Remark 4. When only a subset of interval statistics is reported, one can still use the minimum distance estimator defined in Theorem 5. For example, if only a frequency table is given under fixed intervals, we can use the proposed estimator with moment conditions being the top half of $m_1(\lambda)$. When the intervals are defined by sample quantiles, very often interval limits are not reported. In this case, we only need to use the bottom half of $m_2(\lambda)$ as the moment condition in the estimation.

Remark 5. To ensure identification, the dimension of $\lambda$ cannot be larger than the number of moment conditions used in the estimation. This condition is actually applied by the identification of Assumption 1.

4 Simulations

In this section, we present Monte Carlo simulations of the proposed estimator based on interval statistics. We consider the standard normal distribution with sample sizes of 200 and 500 respectively.

<table>
<thead>
<tr>
<th>Table 1: Simulation design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>interval limits</td>
</tr>
<tr>
<td>limits</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>(1) population quintiles</td>
</tr>
<tr>
<td>(2) population quintiles</td>
</tr>
<tr>
<td>(3) sample quintiles</td>
</tr>
<tr>
<td>(4) sample quintiles</td>
</tr>
<tr>
<td>(5) sample quintiles</td>
</tr>
</tbody>
</table>

We consider five commonly used formats of interval statistics as listed in Table 1. The random $i.i.d.$ sample is divided into five groups. We first use the population quintiles as interval limits. In case one, we report interval frequency and interval limits; in case two, we also report the conditional mean of each interval. We then use sample quintiles as interval limits for case three to five, where interval frequencies are known by construction. In case three, the sample quintiles used to define intervals are reported. In case four, instead of sample quintiles, we provide conditional mean of each interval. Case five reports both sample quintiles and conditional means.
The minimum distance estimator proposed in previous section is used to estimate the underlying distribution. The simulation results are summarized in Table 2, where the mean squared errors (MSE) of the estimated standard deviation and the mean integrated squared errors (MISE) of the estimated density are reported. All numbers are multiplied by 10,000 to improve readability.

<table>
<thead>
<tr>
<th></th>
<th>n=200</th>
<th>n=500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>MSE(σ)</td>
<td>55</td>
<td>36</td>
</tr>
<tr>
<td>MISE(f)</td>
<td>5.0</td>
<td>3.9</td>
</tr>
</tbody>
</table>

When sample size is 200, the estimates based on interval frequency for fixed intervals (case 1) have a MSE of 55 for the standard deviation. As expected, when interval means are also incorporated (case 2), the MSE improves to 36 by 35%. Under quantile intervals, the MSE for estimates based on sample quantiles is 50 (case 3), comparable to that of case one. When the estimates are based on conditional means but not sample quantiles (case 4), the performance improves by 50%. This is a considerable improvement given that the amount of information input is roughly the same for case 3 and 4: case 3 uses only sample quintiles, and case 4 uses only conditional means. When we use both sample quintiles and conditional means for our estimates (case 5), the performance is essentially identical to that of case 4.

Similar pattern is suggested by the MISE of the density estimates. For fixed interval case, the MISE improves by 22% when conditional means are incorporated. When the intervals are defined by sample quintiles, the performance improves by 37% (from case 3 to case 5). The estimate based on interval means alone (case 4) is considerably better than that base on sample quintiles alone (case 3). In case 5, where both sample quintiles and interval means are used, the performance is only marginally better than that of case 4.

Since the difference in grouping method has been explicitly addressed in the calculation of asymptotic covariance, the performance of the proposed estimator does not appear to be affected by the method of grouping. When only interval frequency or limits are reported, the performances for case 1 and case 3 are virtually identical. With interval means incorporated (case 2 and case 5), the results are also comparable, with case 5 performing slightly better.
The simulations with a larger sample size (n=500) provide improved results, whose pattern are qualitatively similar to those from smaller sample size. In terms of the MISE, the performances under two different grouping methods are nearly indistinguishable. Our results clearly demonstrate the usefulness of the proposed estimator, especially for the case when interval limits are not observable to the researchers.

5 Empirical example

In this section, we apply the proposed method to estimating China’s income distribution from interval statistics. During the past two decades, China has averaged a two-digit growth. At the same time, the income inequality has risen considerably. Wu and Perloff (2005) documented this growth in inequality. A major obstacle of studying China’s income distribution is the lack of microdata. Only interval statistics are available. The proposed method allows efficient estimation of the underlying distribution based on interval statistics.

The interval statistics for annual household income per capita are reported by the *Statistics Yearbook of China* each year. For rural area, a fixed income grid in multiples of one hundred yuan is used. The sample is partitioned into 20 intervals, ranging from “below 100 yuan” to “over 5,000 yuan” as listed in Table 3. The frequency of each interval is reported. For urban area, the sample is divided into eight intervals by the $5^{th}$, $10^{th}$, $20^{th}$, $40^{th}$, $60^{th}$, $80^{th}$ and $90^{th}$ sample quantile. The interval averages are reported, but not those sample quantiles used to define the intervals. Hence, the formats of interval statistics for rural and urban area correspond respectively to case 1 and 4 in the previous section.

We use reported income information for 2001 to estimate the income distribution of rural and urban area separately. The density function is specified as

$$f(x) = \exp \left\{ -\lambda_0 - \sum_{i=1}^{4} \lambda_i \log^i (c + x^2) \right\}, \ x \geq 0,$$

where $c$ is a small positive constant,\(^1\) and $\lambda_0 = \int_0^\infty \exp \left\{ -\sum_{i=1}^{4} \lambda_i \log^i (c + x^2) \right\} dx$ such that $f(x)$ integrates to one over its domain. This density function has a relatively simple

---

\(^1\)One can view $c$ as the minimum income or the subsistence level of income. In our estimation, $c$ is set to be equal to 100 yuan.
functional form and at the same time is flexible enough to allow for skewed, fat-tailed or even multi-modal distribution. Note that when $c = 0$ and $\lambda_3 = \lambda_4 = 0$, the density is reduced to the log-normal distribution, which is commonly used to model income distribution. For a detailed discussion of this density specification as a maximum entropy density, see Wu and Perloff (2005).

The interval statistics and corresponding estimates are reported in Table 3 for both areas. The proposed simple density function accurately predicts sample interval statistics. The estimated densities are plotted in Figure 1, where solid line denotes urban income density and dashed line is rural density. The results clearly suggest a substantial rural-urban income gap. The estimated densities allow us to compare various aspects of the two distributions. Average urban income (7,287 Yuan) is almost three times of average rural income (2,519 Yuan). The proportion of people with annual income below 1,000 Yuan is estimated to be 12.54% in rural area, while it is essentially zero (0.002%) in urban area. At the other end of the distribution, 17.88% of the urban population has annual income over 10,000 Yuan, while only 0.43% of the rural population reaches this income level. Because of its large proportion of low income families, income inequality is actually higher in rural area. According to our calculation, the Gini index is 0.35 for rural area, considerably higher than 0.26 for urban area.

<table>
<thead>
<tr>
<th>Range</th>
<th>Rural income</th>
<th>Urban income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(s)</td>
<td>(s(\lambda))</td>
</tr>
<tr>
<td>&lt;100</td>
<td>0.48</td>
<td>0.46</td>
</tr>
<tr>
<td>100-200</td>
<td>0.21</td>
<td>0.45</td>
</tr>
<tr>
<td>200-300</td>
<td>0.33</td>
<td>0.53</td>
</tr>
<tr>
<td>300-400</td>
<td>0.60</td>
<td>0.74</td>
</tr>
<tr>
<td>400-500</td>
<td>0.89</td>
<td>0.85</td>
</tr>
<tr>
<td>500-600</td>
<td>1.36</td>
<td>1.09</td>
</tr>
<tr>
<td>600-800</td>
<td>3.88</td>
<td>3.37</td>
</tr>
<tr>
<td>800-1000</td>
<td>5.47</td>
<td>5.04</td>
</tr>
<tr>
<td>1000-1200</td>
<td>6.30</td>
<td>6.15</td>
</tr>
<tr>
<td>1200-1300</td>
<td>3.45</td>
<td>3.89</td>
</tr>
</tbody>
</table>

*: \(\hat{s}\) and \(s(\lambda)\) denote respectively sample and estimated interval frequency.
**: \(\hat{\mu}\) and \(\mu(\lambda)\) denote respectively sample and estimated interval average.
6 Conclusion

Individual data from continuous distributions are often represented by sample statistics defined on a collection of sample space intervals. In this study, we group interval statistics into two broad categories according to the method of interval grouping: those defined by fixed values and those defined by sample quantiles. For either type, we allow multiple statistics on each interval. The asymptotic distributions of for both types of interval statistics are derived. For the case of quantile intervals, the asymptotic distribution is more complicated as the intervals are themselves stochastic. We use the influence function approach to derive the joint distribution of interval limits and other interval statistics. Goodness-of-fit tests for interval statistics are constructed based on their corresponding asymptotic distributions.

We propose minimum distance estimators to estimate the underlying continuous distribution based on interval statistics. Under fixed intervals, the covariance matrix for interval statistics are singular. We present a computationally-simple non-singular generalized inverse of the covariance matrix, which is used as optimal weighting matrix of the proposed estimator. When the intervals are defined by sample quantiles, we further generalize the estimator...
to allow for the case that the sample quantiles used to define the intervals are not reported.

Monte Carlo simulations suggest that the proposed estimators have good small sample performance. Conventional distribution estimation based on grouped data only use the frequency of each interval. We show that by incorporating additional information such as conditional mean of each interval, one can improve the estimates substantially.
References


Burridge, J. “Some Unimodality Properties of Likelihoods Derived from Grouped Data.” 

Chen, Z. M., and J. M. “Confidence Interval for the Mean of the Exponential Distribution, 


Gastwirt, J. L. “Estimation of Lorenz-Curve and Gini-Index.” *Review of Economics and 


Gastwirth, J. L., T. K. Nayak, and A. M. Krieger. “Large Sample Theory for the Bounds on 
the Gini and Related Indexes of Inequality Estimated from Grouped Data.” *Journal of 

164-179.


Kakwani, N. C., and N. Podder. “Efficient Estimation of the Lorenz Curve and Associated 
Inequality Measures from Grouped Observations.” *Econometrica*, 44, no. 1(1976): 137- 
148.


Appendix

Proof of Proposition 1. Because \( \int \theta^2 (x) \, dF(x) < \infty \), it follows that \( \mu_j^{(2)} = \int_{A_j} \theta^2 (x) \, dF(x) < \infty \) for \( j = 1, \ldots, J \). By multivariate Lindenberg-Lévy CLT, \( \sqrt{n} (\hat{\mu} - \mu) \to N(0, \Omega_3) \). Combining the asymptotic distributions for \( \hat{s} \) and \( \hat{\mu} \) yields the desired result.

Proof of Proposition 2. Define

\[
B = \begin{bmatrix} D_s & \Omega_2 \\ \Omega_2' & \Omega_3 \end{bmatrix},
\]

and

\[
Z = \begin{bmatrix} I \_J & -\Omega_2 \Omega_3^{-1} \\ 0 & I \_J \end{bmatrix},
\]

where \( I \_J \) is a \( J \times J \) identity matrix. We can block-diagonalize \( B \) by

\[
ZBZ' = \begin{bmatrix} D_s - \Omega_2^{-1} \Omega_3^{-1} \Omega_2' & 0 \\ 0 & \Omega_3 \end{bmatrix}.
\]

Since \( \Omega_3 \) has a full rank and \( \text{rank}(D_s - \Omega_2 \Omega_3^{-1} \Omega_2') = J \), \( \text{rank}(B) = 2J \). Therefore, \( C = B^{-1} \) is full-ranked. We then have

\[
C = \begin{bmatrix} \Omega_* & -\Omega_* \Omega_2 \Omega_3^{-1} \\ -\Omega_3^{-1} \Omega_2' \Omega_* & \Omega_3^{-1} \Omega_2' \Omega_* \Omega_2 \Omega_3^{-1} + \Omega_3^{-1} \end{bmatrix},
\]

where

\[
\Omega_* = D_s^{-1} - D_s^{-1} \Omega_2 \left( \Omega_2' D_s^{-1} \Omega_2 - \Omega_3 \right)^{-1} \Omega_2' D_s^{-1}.
\]

Using the fact that \( \Omega_2 = \Omega_1 D_\nu \) and \( \Omega_3 = D_\nu \Omega_1 D_\nu + D_\nu D_\sigma \), one can then show, after some tedious algebraic manipulations, that

\[
\Omega C \Omega = \Omega.
\]
We note that $C\Omega$ is idempotent since 

$$C (\Omega C\Omega) = C\Omega.$$ 

$C$ is not a Moore-Penrose generalized inverse of $\Omega$ as $C\Omega C \neq C$; or, it is a non-reflective generalized inverse of $\Omega$. Therefore, trace$(C\Omega)$ = rank$(C\Omega)$ by the fact that the trace of an idempotent matrix is equal to its rank. Since the rank of the product is no greater than the minimum rank of its components, we have rank$(C\Omega) = \min (\text{rank } (C), \text{rank } (\Omega)) = 2J - 1$. It then follows that $W_n CW_n' \rightarrow \chi_{2J-1}^2$ by Theorem 3.5 of Serfling (2001, p. 128).

**Proof of Lemma 1.** Let $F_{\varepsilon}(x) = (1 - \varepsilon) F(x) + \varepsilon H_{x_0}(x)$ where $H_{x_0}(x)$ is a CDF with mass 1 at $x = x_0$. Because 

$$F_{\varepsilon}(F_{\varepsilon}^{-1}(p)) = p,$$

we have 

$$(1 - \varepsilon) F(F_{\varepsilon}^{-1}(p)) + \varepsilon H_{x_0}(F_{\varepsilon}^{-1}(p)) = p.$$ 

Taking derivative with respect to $\varepsilon$ yields 

$$- F(F_{\varepsilon}^{-1}(p)) \bigg|_{\varepsilon=0} + (1 - \varepsilon) f(F_{\varepsilon}^{-1}(p)) \bigg|_{\varepsilon=0} \frac{d}{d\varepsilon} F_{\varepsilon}^{-1}(p) \bigg|_{\varepsilon=0} + H_{x_0}(F_{\varepsilon}^{-1}(p)) \bigg|_{\varepsilon=0} = 0.$$ 

Rearranging terms, we have 

$$\frac{d}{d\varepsilon} F_{\varepsilon}^{-1}(p) \bigg|_{\varepsilon=0} = \frac{p - H_{x_0}(F_{\varepsilon}^{-1}(p))}{f(F_{\varepsilon}^{-1}(p))} = \frac{p - I(F(x_0) \leq p)}{f(F_{\varepsilon}^{-1}(p))}.$$ 

Therefore, we obtain the influence function of $\xi_j = F^{-1}(p_j)$: 

$$\text{IF}_j(x) = \frac{p_j - I(x \leq F^{-1}(p_j))}{f(F^{-1}(p_j))}.$$
It then follows that

\[
Var \left( \sqrt{n} \hat{F}^{-1}(p_j) \right) = E \left[ IF_{F^{-1}(p_j)}^2(x) \right]
\]

\[
= p_j \left( p_j - 1 \right) f_j + (1 - p_j) \left( \frac{p_j}{f_j} \right)^2
\]

\[
= p_j \left( 1 - p_j \right) f_j^2.
\]

Similarly, define

\[
IF_{i,j} = \frac{p_i - I(x \leq F^{-1}(p_i)) p_j - I(x \leq F^{-1}(p_j))}{f(F^{-1}(p_i)) / f(F^{-1}(p_j))}.
\]

Then

\[
cov \left( \sqrt{n} \hat{F}^{-1}(p_i) , \sqrt{n} \hat{F}^{-1}(p_j) \right) = E \left[ IF_{i,j}(x) \right]
\]

\[
= p_i \left( p_i - 1 \right) (p_j - 1) f_i f_j + (p_j - p_i) \frac{p_i (p_{j-1})}{f_i f_j} + (1 - p_j) \frac{p_i p_j}{f_i f_j}
\]

\[
= p_i \left( 1 - p_j \right) f_i f_j.
\]

**Proof of Lemma 2.** Let \( F_n \) be the empirical CDF. An asymptotically equivalent version of \( \nu_j \) is defined as

\[
T(F_n) = \nu_j,
\]

with

\[
T(F) = \frac{1}{p_j - p_{j-1}} \int_{F^{-1}(p_{j-1})}^{F^{-1}(p_j)} \theta(x) dF(x) = \frac{1}{p_j - p_{j-1}} \int_{p_{j-1}}^{p_j} \int_{F^{-1}(t)} \theta(F^{-1}(t)) dt.
\]

Using the influence for \( F^{-1}(p) \) derived in the proof of Lemma 1, we obtain the influence
function of $T(F)$ as

$$
\text{IF}_T(x) = \frac{d}{d\varepsilon} \left. p_j \int_{p_{j-1}}^{p_j} \theta' \left( F^{-1} (t) \right) F^{-1}_\varepsilon (t) \, dt \right|_{\varepsilon = 0}
$$

$$
= \frac{1}{p_j - p_{j-1}} \int_{p_{j-1}}^{p_j} \theta' \left( F^{-1} (t) \right) \frac{\text{I}_x (F^{-1} (t))}{f (F^{-1} (t))} \, dt
$$

$$
= \frac{1}{p_j - p_{j-1}} \int_{p_{j-1}}^{p_j} \theta' \left( F^{-1} (t) \right) [t - \text{I}_x (F^{-1} (t)) \, dF^{-1} (t)
$$

$$
= \frac{1}{p_j - p_{j-1}} \int_{p_{j-1}}^{p_j} [t - \text{I}_x (F^{-1} (t)) \, d\theta (F^{-1} (t))
$$

For $x < F^{-1}(p_{j-1})$,

$$
\text{IF}_T(x) = \frac{1}{p_j - p_{j-1}} \int_{p_{j-1}}^{p_j} (t - 1) \, d\theta (F^{-1} (t))
$$

$$
\text{IF}_T(x) = \frac{1}{p_j - p_{j-1}} \left\{ p_j \theta (\xi_j) - p_{j-1} \theta (\xi_{j-1}) - \int_{p_{j-1}}^{p_j} \theta (F^{-1} (t)) \, dt - \theta (\xi_j) + \theta (\xi_{j-1}) \right\}
$$

For $x > F^{-1}(p_j)$,

$$
\text{IF}_T(x) = \frac{1}{p_j - p_{j-1}} \int_{p_{j-1}}^{p_j} t \, d\theta (F^{-1} (t))
$$

$$
\text{IF}_T(x) = \frac{1}{p_j - p_{j-1}} \left\{ p_j \theta (\xi_j) - p_{j-1} \theta (\xi_{j-1}) - \int_{p_{j-1}}^{p_j} \theta (F^{-1} (t)) \, dt \right\}
$$

For $F^{-1}(p_{j-1}) \leq x \leq F^{-1}(p_j)$,

$$
\text{IF}_T(x) = \frac{1}{p_j - p_{j-1}} \int_{p_{j-1}}^{p_j} [t - \text{I}_x (F^{-1} (t)) \, d\theta (F^{-1} (t))
$$

$$
\text{IF}_T(x) = \frac{1}{p_j - p_{j-1}} \left\{ \int_{p_{j-1}}^{p_j} t \, d\theta (F^{-1} (t)) - \int_{F(x)}^{p_j} \, d\theta (F^{-1} (t)) \right\}
$$

$$
\text{IF}_T(x) = \frac{1}{p_j - p_{j-1}} \left\{ p_j \theta (\xi_j) - p_{j-1} \theta (\xi_{j-1}) - \int_{p_{j-1}}^{p_j} \theta (F^{-1} (t)) \, dt - \theta (\xi_j) + \theta (x) \right\}
$$

25
We then have
\[ IF_T (x) = \begin{cases} 
\frac{1}{p_j - p_{j-1}} \left[ \theta (\xi_{j-1}) - \tau_j \right], & x < \xi_{j-1}; \\
\frac{1}{p_j - p_{j-1}} \left[ \theta (x) - \tau_j \right], & \xi_{j-1} \leq x \leq \xi_j; \\
\frac{1}{p_j - p_{j-1}} \left[ \theta (\xi_j) - \tau_j \right], & x > \xi_j.
\end{cases} \]

Note that
\[
\int IF_T (x) \, dF(x) = \frac{1}{p_j - p_{j-1}} \left\{ p_{j-1} \left[ \theta (\xi_{j-1}) - \tau_j \right] + \int_{p_{j-1}}^{p_j} \theta \left( F^{-1}(t) \right) \, dt - (p_j - p_{j-1}) \tau_j + (1 - p_j) \left[ \theta (\xi_j) - \tau_j \right] \right\}
\]
\[= \frac{1}{p_j - p_{j-1}} \left\{ \int_{p_{j-1}}^{p_j} \theta \left( F^{-1}(t) \right) \, dt + p_{j-1} \theta (\xi_{j-1}) + (1 - p_j) \theta (\xi_j) - \tau_j \right\}
\]
\[= 0.\]

It then follows that
\[ Var \left( T \left( F_n \right) \right) = E \left( IF_T^2 (x) \right) \]
\[= \frac{p_{j-1} \left( \eta_j^- \right)^2 + (p_j - p_{j-1}) \left( \nu_j^{(2)} - 2 \nu_j \tau_j + \tau_j^2 \right) + (1 - p_j) \left( \eta_j^+ \right)^2}{n (p_j - p_{j-1})^2}.\]

The asymptotic result follows immediately by the CLT.

**Proof of Lemma 3.** Define
\[ T_i (F) = \frac{1}{p_i - p_{i-1}} \int_{p_{i-1}}^{p_i} F^{-1}(t) \, dt, \]
\[ IF_{T_i} (x) = \frac{1}{p_j - p_{j-1}} \int_{p_{j-1}}^{p_j} [t - I \left( F (x) \leq t \right)] \, d\theta_k \left( F^{-1}(t) \right), \]
and \( T_j (F) \) and \( IF_{T_j} (x) \) similarly. Define the product of \( IF_{T_i} (x) \) and \( IF_{T_j} (x) \)
\[ IF (x) = a \int_{p_{i-1}}^{p_i} [t - I \left( F (x) \leq t \right)] \, d\theta \left( F^{-1}(t) \right) \int_{p_{j-1}}^{p_j} [t - I \left( F (x) \leq t \right)] \, d\theta \left( F^{-1}(t) \right). \]
where \( a = \frac{1}{(p_i - p_{i-1})(p_j - p_{j-1})}. \)

Depending on the value of \( x \), there are five possibilities:
Case 1. $x < F^{-1}(p_{i-1})$, 

$$\text{IF} (x) = a \int_{p_{i-1}}^{p_i} (t - 1) d\theta \left( F^{-1}(t) \right) \int_{p_{j-1}}^{p_j} (t - 1) d\theta \left( F^{-1}(t) \right)$$

$$= a \eta_i^- \eta_j^-.$$

Case 2. $F^{-1}(p_{i-1}) \leq x \leq F^{-1}(p_i)$,

$$\text{IF} (x) = a \int_{p_{i-1}}^{p_i} [t - I(F(x) \leq t)] d\theta \left( F^{-1}(t) \right) \int_{p_{j-1}}^{p_j} (t - 1) d\theta \left( F^{-1}(t) \right)$$

$$= a \left[ \theta(x) - \tau_i \right] \eta_j^-.$$

Case 3. $F^{-1}(p_i) < x < F^{-1}(p_{j-1})$,

$$\text{IF} (x) = a \int_{p_{i-1}}^{p_i} t d\theta \left( F^{-1}(t) \right) \int_{p_{j-1}}^{p_j} (t - 1) d\theta \left( F^{-1}(t) \right)$$

$$= a \eta_i^+ \eta_j^-.$$

Case 4. $F^{-1}(p_{j-1}) \leq x \leq F^{-1}(p_j)$,

$$\text{IF} (x) = a \int_{p_{i-1}}^{p_i} t d\theta \left( F^{-1}(t) \right) \int_{p_{j-1}}^{p_j} [t - I(F(x) \leq t)] d\theta \left( F^{-1}(t) \right)$$

$$= a \eta_i^+ \left[ \theta(x) - \tau_j \right].$$

Case 5. $x > F^{-1}(p_j)$,

$$\text{IF} (x) = a \int_{p_{i-1}}^{p_i} t d\theta \left( F^{-1}(t) \right) \int_{p_{j-1}}^{p_j} t d\theta \left( F^{-1}(t) \right)$$

$$= a \eta_i^+ \eta_j^+.$$

It then follows that

$$E[\text{IF} (x)] = \frac{1}{(p_i - p_{i-1})(p_j - p_{j-1})} \left[ p_{i-1} \eta_i^- \eta_j^- + (p_i - p_{i-1}) \eta_i \eta_j^- ight]$$

$$+ (p_{j-1} - p_i) \eta_i^+ \eta_j^- + (p_j - p_{j-1}) \eta_i^+ \eta_j + (1 - p_j) \eta_i^+ \eta_j^+ \right].$$
Proof of Proposition 3. Using Lemma 1, 2 and the multivariate Lindeberg-Lévy CLT, we obtain the asymptotic normality immediately.

Proof of Lemma 4. Define

\[
IF_{\xi,\nu_j}(x) = \frac{p_i - I(x \leq F^{-1}(p_i))}{F^{-1}(p_i)} \int_{p_{j-1}}^{p_j} \frac{[t - I(x \leq F^{-1}(t))]}{p_j - p_{j-1}} d\theta(F^{-1}(t)).
\]

For \( i \leq j - 1 \),

\[
E[IF_{\xi,\nu_j}(x)] = p_i \frac{p_i - 1}{f_i} \eta_j^- + (p_j - p_{j-1}) \frac{p_i}{f_i} \eta_j^- + (p_j - p_j - 1) \frac{p_i}{f_i} \eta_j^+ + (1 - p_j) \frac{p_i}{f_i} \eta_j^+
\]

\[
= p_i \left[(p_j - 1) \eta_j^- + (p_j - p_j - 1) \eta_j^- + (1 - p_j) \eta_j^+ \right].
\]

On the other hand, for \( i \geq j \),

\[
E[IF_{\xi,\nu_j}(x)] = p_{j-1} \frac{p_i - 1}{f_i} \eta_j^- + (p_j - p_j - 1) \frac{p_i}{f_i} \eta_j^- + (p_j - p_j - 1) \frac{p_i - 1}{f_i} \eta_j^+ + (1 - p_j) \frac{p_i}{f_i} \eta_j^+
\]

\[
= p_{j-1} \left[(p_j - 1) \eta_j^- + (p_j - p_j - 1) \eta_j^- + (1 - p_j) \eta_j^+ \right].
\]

Proof of Proposition 4. Using Lemma 1, 4 and Theorem 3, the asymptotic normality follows directly according to the multivariate Lindeberg-Lévy CLT.

Proof of Theorem 5. Under the assumption of full-ranked \( \Omega \), the result is immediate by Theorem 3.5 of Serfling (2001, p. 128).

Proof of Proposition 6. Without loss of generality, denote \( \mu \) the interval statistics used in the estimation, which may include the interval frequency, sample quantiles and interval moments. Under the assumption \( \sup_{\lambda \in \Lambda} \int \theta(x)^2 dF(x; \lambda) < \infty \), we have

\[
\sqrt{n} (\hat{\mu} - \mu) \xrightarrow{d} N(0, \Psi)
\]

by Theorem 1 and \( \hat{\mu}_j^{(2)} = \mu_j^{(2)} + o_p(1) \) by the LLN for \( i = 1, \ldots, K \). The first-order condition
for $\hat{\lambda}$ is

$$G\left(\hat{\lambda}\right)\Psi^{-1}\left(\hat{\lambda}\right)\{E\hat{\lambda}\mu - \hat{\mu}\} = 0.$$ 

Since $E_{\lambda_0}\mu = \mu(\lambda_0)$ and $\mu(\lambda_0) = \hat{\mu} + o_p(1)$, we have

$$E\hat{\lambda}\mu - \hat{\mu} = E\hat{\lambda}\mu - \mu(\lambda_0) + o_p(1) = G(\lambda_0)\left(\hat{\lambda} - \lambda_0\right) + o_p(1).$$

Thus we can rewrite the first order condition

$$G\left(\hat{\lambda}\right)\Psi^{-1}\left(\hat{\lambda}\right)\{\sqrt{n}\left[\mu\left(\hat{\lambda}\right) - \mu(\lambda_0)\right] - G(\lambda_0)\sqrt{n}\left(\hat{\lambda} - \lambda_0\right)\} + o_p(1) = 0.$$ 

Because $G(\cdot)$ is continuous and $\hat{\lambda} \xrightarrow{p} \lambda_0$, $G\left(\hat{\lambda}\right) = G(\lambda_0) + o_p(1)$; also, by assumption $\psi\left(\hat{\lambda}\right) = \psi(\lambda_0) + o_p(1)$. Then

$$G(\lambda_0)'\psi^{-1}(\lambda_0)G(\lambda_0)\sqrt{n}\left(\hat{\lambda} - \lambda_0\right) = G(\lambda_0)'\psi^{-1}(\lambda_0)\sqrt{n}\left[\mu\left(\hat{\lambda}\right) - \mu(\lambda_0)\right] + o_p(1).$$

Since $\sqrt{n}\left[\mu\left(\hat{\lambda}\right) - \mu(\lambda_0)\right] \xrightarrow{d} N(0, \psi(\lambda_0))$, by Slusky theorem,

$$G(\lambda_0)'\psi^{-1}(\lambda_0)G(\lambda_0)\sqrt{n}\left(\hat{\lambda} - \lambda_0\right) \xrightarrow{d} N\left(0, G(\lambda_0)'\psi^{-1}(\lambda_0)G(\lambda_0)\right).$$

Therefore, under assumption that $G(\lambda_0)'\psi^{-1}(\lambda_0)G(\lambda_0)$ is non-singular,

$$\sqrt{n}\left(\hat{\lambda} - \lambda_0\right) \xrightarrow{d} N\left(0, [G(\lambda_0)'\psi^{-1}(\lambda_0)G(\lambda_0)]^{-1}\right).$$

**Proof of Proposition 7.** Define for $\delta_\eta$ in the neighborhood of zero,

$$F_\varepsilon = (1 - \varepsilon) F(x; \lambda) + \varepsilon F(x; \lambda + \delta)$$

It follows that

$$F_\varepsilon (F_\varepsilon^{-1}(p)) = p,$$

$$(1 - \varepsilon) F\left(F_\varepsilon^{-1}(p); \lambda\right) + \varepsilon F\left(F_\varepsilon^{-1}(p); \lambda + \delta\right) = p.$$
Taking derivative with respect to $\varepsilon$ yields

$$-F_{\varepsilon} \left( F_{\varepsilon}^{-1} (p) \right) |_{\varepsilon=0} + (1 - \varepsilon) f \left( F_{\varepsilon}^{-1} (p; \lambda); \lambda \right) \frac{d F_{\varepsilon}^{-1} (p)}{d\varepsilon} |_{\varepsilon=0} + F \left( F_{\varepsilon}^{-1} (p; \lambda + \delta) \right) |_{\varepsilon=0} = 0.$$ 

Rearranging terms gives

$$\frac{d F_{\varepsilon}^{-1} (p)}{d\varepsilon} |_{\varepsilon=0} = \frac{p - F \left( F_{\varepsilon}^{-1} (p; \lambda); \lambda + \delta \right)}{f \left( F_{\varepsilon}^{-1} (p; \lambda); \lambda \right)}.$$ 

We then have

$$\frac{d F^{-1} (p; \lambda)}{d\lambda_q} = \lim_{\delta_q \to 0} \frac{p - F \left( F^{-1} (p; \lambda); \lambda + \delta \right)}{f \left( F^{-1} (p; \lambda); \lambda \right)} / \delta_q.$$