Maximum likelihood estimation for survey data with informative interval censoring

Angel G. Angelov¹ · Magnus Ekström¹

Received: 24 May 2017 / Accepted: 22 May 2018 / Published online: 17 July 2018
© The Author(s) 2018

Abstract Interval-censored data may arise in questionnaire surveys when, instead of being asked to provide an exact value, respondents are free to answer with any interval without having pre-specified ranges. In this context, the assumption of noninformative censoring is violated, and thus, the standard methods for interval-censored data are not appropriate. This paper explores two schemes for data collection and deals with the problem of estimation of the underlying distribution function, assuming that it belongs to a parametric family. The consistency and asymptotic normality of a proposed maximum likelihood estimator are proven. A bootstrap procedure that can be used for constructing confidence intervals is considered, and its asymptotic validity is shown. A simulation study investigates the performance of the suggested methods.

Keywords Informative interval censoring · Maximum likelihood · Parametric estimation · Questionnaire surveys · Self-selected intervals

1 Introduction

In questionnaire surveys respondents are often allowed to give an answer in the form of an interval. For example, the respondent can be asked to select from several pre-specified intervals; this question format is known as range card. Another approach is called unfolding brackets, where the respondent is asked a sequence of yes-no

✉ Angel G. Angelov
agangelov@gmail.com

Magnus Ekström
magnus.ekstrom@umu.se

¹ Department of Statistics, Umeå School of Business, Economics and Statistics, Umeå University, Umeå, Sweden

questions that narrow down the range in which the respondent's true value is. These formats are suitable when asking questions that are difficult to answer with an exact value (e.g., recall questions) or when asking sensitive questions (e.g., asking about income) because they allow partial information to be elicited from respondents who are unable or unwilling to provide exact amounts. However, studies have found that the pre-specified intervals given to the respondents in a range-card question are likely to influence their answers. Such bias is known as bracketing effect (see, e.g., McFadden et al. 2005). Similarly, the unfolding brackets format is prone to the so-called anchoring effect, i.e., answers can be biased toward the starting value in the sequence of yes-no questions (see, e.g., Furnham and Boo 2011; Van Exel et al. 2006).

A format that does not involve any pre-specified values is the respondent-generated intervals approach, suggested by Press and Tanur (2004a, b), where the respondent is asked to provide both a point value (a best guess for the true value) and an interval. They employed Bayesian methods for estimating the parameters of the underlying distribution. Similar format, in which the respondent is free to answer with any interval containing his/her true value, was considered by Belyaev and Kriström (2010). They use the term self-selected interval (SSI). Estimating the underlying distribution using SSI data, however, requires some generally untestable assumptions related to how the respondent chooses the interval. To avoid such assumptions, Belyaev and Kriström (2012, 2015) introduced a novel two-stage approach. The idea is to ask the respondent first to provide an SSI and then to select from several sub-intervals of the SSI the one that most likely contains his/her true value. Data collected in a pilot stage are used for generating the sub-intervals in the second question. Belyaev and Kriström (2012, 2015) proposed a nonparametric maximum likelihood estimator of the underlying distribution for two-stage SSI data. Angelov and Ekström (2017) extended their work by exploring a sampling scheme where the number of sub-intervals in the second question is limited to two or three, which is motivated by the fact that a question with a large number of sub-intervals might be difficult to implement in practice, e.g., in a telephone interview.

Data consisting of self-selected intervals are a special case of interval-censored data. Let X be a random variable of interest. An observation on X is interval-censored if, instead of observing X exactly, only an interval $(L, R]$ is observed, where $L < X \leq R$ (see, e.g., Zhang and Sun 2010). Interval-censored data arise most commonly when the observed variable is the time to some event (known as survival data, failure time data, lifetime data, duration data, or time-to-event data). The problem of estimating the underlying distribution for interval-censored data has been approached through nonparametric methods by Peto (1973), Turnbull (1976), and Gentleman and Geyer (1994), among others. These estimators rely on the assumption of noninformative censoring, i.e., the observation process that generates the censoring is independent of the variable of interest (see, e.g., Sun 2006, p. 244). In the sampling schemes considered by Belyaev and Kriström (2010, 2012, 2015) and Angelov and Ekström (2017) this is not a reasonable assumption as it is the respondent who chooses the interval; thus, the standard methods are not appropriate. The existing methods for data with informative interval censoring (see Finkelstein et al. 2002; Shardell et al. 2007) are specific for time-to-event data and are not directly applicable in the context that we are discussing.

In this paper, we focus on parametric estimation of the underlying distribution function, i.e., we assume a particular functional form of the distribution. Compared to nonparametric methods, this approach usually leads to more efficient estimators, provided that the distributional assumption is true (see, e.g., Collett 1994, p. 107). The problem of choosing the right parametric model can be sidestepped by using a wide parametric family like the generalized gamma distribution (see, e.g., Cox et al. 2007) that includes most of the commonly used distributions as special cases (exponential, gamma, Weibull, and log-normal).

We suggest two modifications of the sampling scheme for SSI data studied in Angelov and Ekström (2017) and propose a parametric maximum likelihood estimator. In Sect. 2, we introduce the sampling schemes. In Sect. 3, the statistical model is defined and the corresponding likelihood function is derived. Asymptotic properties of the maximum likelihood estimator are established in Sect. 4. The results of a simulation study are presented in Sect. 5, and the paper is concluded in Sect. 6. In “Appendix” are given proofs and auxiliary results.

2 Sampling schemes

2.1 Scheme A

The rationale behind this scheme is that we need to have more information than just the self-selected intervals in order to estimate the underlying distribution. Therefore, we ask the respondent to select a sub-interval of the interval that he/she stated. The problem of deciding where to split the stated interval into sub-intervals can be resolved using some previously collected data (in a pilot stage) or based on other knowledge about the quantity of interest.

We consider the following two-stage scheme for collecting data. In the *pilot stage*, a random sample of n_0 individuals is selected and each individual is requested to give an answer in the form of an interval containing his/her value of the quantity of interest. It is assumed that the endpoints of the intervals are rounded, for example, to the nearest integer or to the nearest multiple of 10. Thus, instead of $(50.2, 78.7]$ respondents will answer with $(50, 79]$ or $(50, 80]$.

Let $d_0^* < d_1^* < \dots < d_{k'}^* < d_{k'}^*$ be the endpoints of all observed intervals. The set $\{d_j^*\} = \{d_0^*, \dots, d_{k'}^*\}$ can be seen as a set of typical endpoints. The data collected in the pilot stage are used only for constructing the set $\{d_j^*\}$, which is needed for the main stage. The set $\{d_j^*\}$ may also be constructed using data from a previous survey, or it can be determined by the researcher based on prior knowledge about the quantity of interest or other reasonable arguments. For instance, if it is known that the variable of interest ranges between 0 and 200 and that the respondents are rounding their endpoints to a multiple of 10, then a reasonable set of endpoints will be $\{0, 10, 20, \dots, 200\}$.

In the *main stage*, a new random sample of n individuals is selected and each individual is asked to state an interval containing his/her value of the quantity of interest. We refer to this question as Qu1. The stated interval is then split into two or three sub-intervals, and the respondent is asked to select one of these sub-intervals (the points of split are chosen in some random fashion among the points d_j^* that are

within the stated interval, e.g., equally likely or according to some other pre-specified probabilities). We refer to this question as Qu2. The respondent may refuse to answer Qu2, and this will be allowed for. If there are no points d_j^* within the stated interval, the second question is not asked.

Let $d_0 < d_1 < \dots < d_{k-1} < d_k$ be the union of $\{d_j^*\}$ and the endpoints of all intervals observed at the main stage. Note that k is unknown but, because of the rounding of endpoints, it can not be arbitrarily large. Let us define a set of intervals $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$, where $\mathbf{v}_j = (d_{j-1}, d_j]$, $j = 1, \dots, k$, and let $\mathcal{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ be the set of all intervals that can be expressed as a union of intervals from $\mathcal{V}$, i.e., $\mathcal{U} = \{(d_l, d_r] : d_l < d_r, l, r = 0, \dots, k\}$. For example, if $\mathcal{V} = \{(0, 5], (5, 10], (10, 20]\}$, then $\mathcal{U} = \{(0, 5], (5, 10], (10, 20], (0, 10], (5, 20], (0, 20]\}$. We denote $\mathcal{J}_h$ to be the set of indices of intervals from $\mathcal{V}$ contained in $\mathbf{u}_h$:

$$\mathcal{J}_h = \{j : \mathbf{v}_j \subseteq \mathbf{u}_h\}, \quad h = 1, \dots, m.$$

In the example with $\mathcal{V} = \{(0, 5], (5, 10], (10, 20]\}$, $\mathbf{u}_5 = (5, 20] = \mathbf{v}_2 \cup \mathbf{v}_3$, hence $\mathcal{J}_5 = \{2, 3\}$.

Remark 1 The main difference between this scheme and the one explored in Angelov and Ekström (2017) is that with scheme A there is no exclusion of respondents, while with the former scheme respondents are excluded if they stated an interval with endpoints not belonging to $\{d_j^*\}$.

2.2 Scheme B

This scheme is a modification of scheme A with two follow-up questions after Qu1 aiming to extract more refined information from the respondents. The *pilot stage* is the same as in scheme A. The sets $\{d_0, \dots, d_k\}$, $\mathcal{V}$, $\mathcal{U}$, and $\mathcal{J}_h$ are also defined in the same way. In the *main stage*, a new random sample of n individuals is selected and each individual is asked to state an interval containing his/her value of the quantity of interest. We refer to this question as Qu1. The stated interval is then split into two sub-intervals, and the respondent is asked to select one of these sub-intervals. The point of split is the d_j^* that is the closest to the middle of the interval; if there are two points that are equally close to the middle, one of them is taken at random. This way of splitting the interval yields two sub-interval of similar length, which would be more natural for the respondent. We refer to this question as Qu2a. The interval selected at Qu2a is thereafter split similarly into two sub-intervals, and the respondent is asked to select one of them. We refer to this question as Qu2b. The respondent may refuse to answer the follow-up questions Qu2a and Qu2b. If there are no points d_j^* within the interval stated at Qu1 or Qu2a, the respective follow-up question is not asked. We assume that if a respondent has answered Qu2a, he/she has chosen the interval containing his/her true value, independent of how the interval stated at Qu1 was split. An analogous assumption is made about the response to Qu2b.

If we know the intervals stated at Qu1 and Qu2b, we can find out the answer to Qu2a. For this reason, if Qu2b is answered, the data from Qu2a can be omitted. Let Qu2 Δ denote the last follow-up question that was answered by the respondent. If the

respondent did not answer both Qu2a and Qu2b, we say that there is no answer at Qu2 Δ . We will distinguish three types of answers in the main stage:

- Type 1. ($\mathbf{u}_h$; NA), when the respondent stated interval $\mathbf{u}_h$ at Qu1 and did not answer Qu2 Δ ;
 Type 2. ($\mathbf{u}_h$; $\mathbf{v}_j$), when the respondent stated interval $\mathbf{u}_h$ at Qu1 and $\mathbf{v}_j$ at Qu2 Δ , where $\mathbf{v}_j \subseteq \mathbf{u}_h$;
 Type 3. ($\mathbf{u}_h$; $\mathbf{u}_s$), when the respondent stated interval $\mathbf{u}_h$ at Qu1 and $\mathbf{u}_s$ at Qu2 Δ , where $\mathbf{u}_s$ is a union of at least two intervals from $\mathcal{V}$ and $\mathbf{u}_s \subset \mathbf{u}_h$.

Similar types of answers can be considered for scheme A, as well. In what follows we will use these three types for both schemes (for scheme A, Qu2 Δ will denote Qu2).

3 Model and estimation

We consider the unobserved (interval-censored) values $x_1, \dots, x_n$ of the quantity of interest to be values of independent and identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with distribution function $F(x) = P(X_i \leq x)$. Our goal is to estimate $F(x)$ through a maximum likelihood approach. Let q_j be the probability mass placed on the interval $\mathbf{v}_j = (d_{j-1}, d_j]$:

$$q_j = P(X_i \in \mathbf{v}_j) = F(d_j) - F(d_{j-1}), \quad j = 1, \dots, k.$$

Because only intervals with endpoints from $\{d_0, \dots, d_k\}$ are observed, the likelihood function will depend on $F(x)$ through the probabilities q_j . In order to avoid complicated notation, we assume that $q_j > 0$ for all $j = 1, \dots, k$. The case when $q_j = 0$ for some j can be treated similarly (cf. Rao 1973, p. 356).

Let H_i , $i = 1, \dots, n$, be i.i.d. random variables such that $H_i = h$ if the i -th respondent has stated interval $\mathbf{u}_h$ at Qu1. The event $\{H_i = h\}$ implies $\{X_i \in \mathbf{u}_h\}$. Let us denote

$$w_{h|j} = P(H_i = h \mid X_i \in \mathbf{v}_j).$$

If $\mathbf{u}_h$ does not contain $\mathbf{v}_j$, then $w_{h|j} = 0$.

Hereafter we will need the following frequencies:

- $n_{h,NA}$ is the number of respondents who stated $\mathbf{u}_h$ at Qu1 and NA (no answer) at Qu2 Δ ;
 n_{hj} is the number of respondents who stated $\mathbf{u}_h$ at Qu1 and $\mathbf{v}_j$ at Qu2 Δ , where $\mathbf{v}_j \subseteq \mathbf{u}_h$;
 n_{h*s} is the number of respondents who stated $\mathbf{u}_h$ at Qu1 and $\mathbf{u}_s$ at Qu2 Δ , where $\mathbf{u}_s$ is a union of at least two intervals from $\mathcal{V}$ and $\mathbf{u}_s \subset \mathbf{u}_h$;
 $n_{h\bullet}$ is the number of respondents who stated $\mathbf{u}_h$ at Qu1 and any sub-interval at Qu2 Δ .

Now we will derive the likelihood for scheme B. If respondent i has given an answer of type 1, i.e., $\mathbf{u}_h$ at Qu1 and NA at Qu2 Δ , then the contribution to the likelihood can

be expressed using the law of total probability: $P(H_i = h) = \sum_{j \in \mathcal{J}_h} w_{h|j} q_j$. If an answer of type 2 is observed, i.e., $\mathbf{u}_h$ at Qu1 and $\mathbf{v}_j$ at Qu2 Δ , then the contribution to the likelihood is $w_{h|j} q_j$. And if a respondent has given an answer of type 3, i.e., $\mathbf{u}_h$ at Qu1 and $\mathbf{u}_s$ at Qu2 Δ , then the contribution to the likelihood is $\sum_{j \in \mathcal{J}_s} w_{h|j} q_j$. Thus, the log-likelihood function corresponding to the main-stage data is

$$\begin{aligned} \log L(\mathbf{q}) = & \sum_h n_{h,\text{NA}} \log \left(\sum_{j \in \mathcal{J}_h} w_{h|j} q_j \right) + \sum_{h,j} n_{hj} \log(w_{h|j} q_j) \\ & + \sum_{h,s} n_{h*s} \log \left(\sum_{j \in \mathcal{J}_s} w_{h|j} q_j \right) + c_1, \end{aligned} \quad (1)$$

where c_1 does not depend on $\mathbf{q} = (q_1, \dots, q_k)$. By similar arguments it can be shown that the log-likelihood for scheme A has essentially the same form as the log-likelihood (1), it differs by an additive constant (the pre-specified probabilities of choosing the points of split of the stated interval are incorporated in c_1).

If we want to estimate $F(x)$ without making any distributional assumptions, we can maximize the log-likelihood (1) with respect to $\mathbf{q}$ (for details see Angelov and Ekström 2017). Here we will assume that $F(x)$ belongs to a parametric family, i.e., $F(x)$ is a known function of some unknown parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$, and thus the probabilities q_j are functions of $\boldsymbol{\theta}$. Therefore, the log-likelihood will be a function of $\boldsymbol{\theta}$, i.e., $\log L(\boldsymbol{\theta}) = \log L(\mathbf{q}(\boldsymbol{\theta}))$, and in order to estimate $F(x)$ we need to estimate $\boldsymbol{\theta}$. For emphasizing that $F(x)$ depends on $\boldsymbol{\theta}$, we will sometimes write $F_{\boldsymbol{\theta}}(x)$.

The conditional probabilities $w_{h|j}$ are nuisance parameters. If $w_{h|j}$ does not depend on j , the assumption of noninformative censoring will be satisfied. In our case, there are no grounds for making such assumptions about $w_{h|j}$, and therefore, we need the data from Qu2 Δ in order to estimate $w_{h|j}$. For this task we employ the procedure suggested in Angelov and Ekström (2017), which we outline here. The idea is first to estimate the probabilities $p_{j|h} = P(X_i \in \mathbf{v}_j | H_i = h)$, $j \in \mathcal{J}_h$. For a given h , a strongly consistent estimator $\tilde{p}_{j|h}$ of $p_{j|h}$, $j \in \mathcal{J}_h$ is obtained by maximizing the log-likelihood:

$$\sum_j n_{hj} \log p_{j|h} + \sum_s n_{h*s} \log \left(\sum_{j \in \mathcal{J}_s} p_{j|h} \right) + c_0,$$

where c_0 does not depend on $p_{j|h}$. Then, an estimator of $w_{h|j}$ is derived using the Bayes formula:

$$\tilde{w}_{h|j} = \frac{\tilde{p}_{j|h} \hat{w}_h}{\sum_s \tilde{p}_{j|s} \hat{w}_s},$$

where $\hat{w}_h = (n_{h\bullet} + n_{h,\text{NA}})/n$ is a strongly consistent estimator of $w_h = P(H_i = h)$.

To find the maximum likelihood estimate of the parameter $\boldsymbol{\theta}$, we insert the estimates of the probabilities $w_{h|j}$ into $\log L(\boldsymbol{\theta})$ and maximize with respect to $\boldsymbol{\theta}$. Alternatively, one may maximize the log-likelihood with respect to both $\boldsymbol{\theta}$ and the nuisance

parameters $w_{h|j}$ using standard numerical optimization methods. This is, however, a high-dimensional and computationally time-consuming optimization problem, which we avoid by simply plugging in the estimated nuisance parameters $\tilde{w}_{h|j}$ into the log-likelihood.

Remark 2 The proposed methodology for estimating $F_{\theta}(x)$ assumes that the respondents are selected according to simple random sampling. If this is not the case, extrapolating the results to the target population may be incorrect. For surveys with a complex design, parameter estimates can be obtained, for example, by using the pseudo-likelihood approach, in which the individual contribution to the log-likelihood is weighted by the reciprocal of the corresponding sample inclusion probability (see, e.g., Chambers et al. 2012, p. 60).

4 Asymptotic results

Let us consider q_j as a function of $\theta = (\theta_1, \dots, \theta_d)$, a multidimensional parameter belonging to a set $\Theta \subseteq \mathbb{R}^d$, and let the true value θ^0 be an interior point of Θ . In this section we prove the consistency and asymptotic normality of the proposed maximum likelihood estimator of θ . We also show the asymptotic validity of a bootstrap procedure which can be used for constructing confidence intervals.

Let $\theta^1, \theta^2 \in \Theta$ and $\|\theta^1 - \theta^2\|$ denote the Euclidean distance between θ^1 and θ^2 . Let the contribution of the i -th respondent to the log-likelihood be denoted $\text{lik}_i(\theta)$, whose precise definition is given by (13). We will consider the following assumptions:

A1 If $\theta^1 \neq \theta^2$, then $\mathbf{q}(\theta^1) \neq \mathbf{q}(\theta^2)$.

A2 For every $\delta > 0$, there exists $\varepsilon > 0$ such that

$$\inf_{\|\theta - \theta^0\| > \delta} \sum_j q_j(\theta^0) \log \frac{q_j(\theta^0)}{q_j(\theta)} \geq \varepsilon.$$

A3 The functions $q_j(\theta)$ are continuous.

A4 The functions $q_j(\theta)$ have first-order partial derivatives that are continuous.

A5 The set Θ is compact, and the functions $q_j(\theta)$ have first- and second-order partial derivatives that are continuous on Θ . Furthermore, $q_j(\theta) > 0$ on Θ .

A6 For each $\theta \in \Theta$, the Fisher information matrix $\mathbf{I}(\theta)$ with elements $I_{r\ell}(\theta) = -E_{\theta} \left(\frac{\partial^2 \text{lik}_i(\theta)}{\partial \theta_r \partial \theta_{\ell}} \right)$, $r, \ell = 1, \dots, d$, is nonsingular.

We say that $\tilde{\theta}$ is an approximate maximum likelihood estimator (cf. Rao 1973, p. 353) of θ if for some $c \in (0, 1)$,

$$L(\tilde{\theta}) \geq c \sup_{\theta \in \Theta} L(\theta). \quad (2)$$

Let γ_t be the probability that a respondent gives an answer of type t , for $t = 1, 2, 3$.

Theorem 1 Let $\tilde{\theta}$ be an approximate maximum likelihood estimator of θ .

- (i) If assumption A2 is satisfied, $\gamma_2 > 0$, and the conditional probabilities $w_{h|j}$ are known, then $\tilde{\theta} \xrightarrow{\text{a.s.}} \theta^0$ as $n \rightarrow \infty$.
- (ii) If assumption A2 is satisfied, $\gamma_2 > 0$, and a strongly consistent estimator of $w_{h|j}$ is inserted into the log-likelihood, then $\tilde{\theta} \xrightarrow{\text{a.s.}} \theta^0$ as $n \rightarrow \infty$.

Theorem 2 If assumptions A2 and A3 are satisfied, $\gamma_2 > 0$, and the conditional probabilities $w_{h|j}$ are known (or strongly consistently estimated), then the maximum likelihood estimator of θ exists and is strongly consistent.

Theorem 3 If assumptions A1 and A4 are satisfied and the conditional probabilities $w_{h|j}$ are known (or strongly consistently estimated), then there exists a root $\tilde{\theta}$ of the system of likelihood equations

$$\frac{\partial \log L(\theta)}{\partial \theta_r} = 0, \quad r = 1, \dots, d, \quad (3)$$

such that $\tilde{\theta} \xrightarrow{\text{a.s.}} \theta^0$ as $n \rightarrow \infty$.

In what follows, $\tilde{\theta}$ will denote the maximum likelihood estimator of θ , unless we state that it denotes an approximate maximum likelihood estimator.

For obtaining asymptotic distributional results about $\sqrt{n}(\tilde{\theta} - \theta^0)$ we will use the notion of weakly approaching sequences of distributions (Belyaev and Sjöstedt-de Luna 2000), which is a generalization of the well-known concept of weak convergence of distributions but without the need to have a limiting distribution. Two sequences of random variables, $\{X_n\}_{n \geq 1}$ and $\{Y_n\}_{n \geq 1}$, are said to have weakly approaching distribution laws, $\{\mathcal{L}(X_n)\}_{n \geq 1}$ and $\{\mathcal{L}(Y_n)\}_{n \geq 1}$, if for every bounded continuous function $\varphi(\cdot)$, $E \varphi(X_n) - E \varphi(Y_n) \rightarrow 0$ as $n \rightarrow \infty$. Further, we say that the sequence of conditional distribution laws $\{\mathcal{L}(X_n | Z_n)\}_{n \geq 1}$ weakly approaches $\{\mathcal{L}(Y_n)\}_{n \geq 1}$ in probability (along Z_n) if for every bounded continuous function $\varphi(\cdot)$, $E(\varphi(X_n) | Z_n) - E \varphi(Y_n) \rightarrow 0$ in probability as $n \rightarrow \infty$.

Theorem 4 Let assumptions A2, A4, and A6 be true, $\gamma_2 > 0$, and the conditional probabilities $w_{h|j}$ be known (or strongly consistently estimated). Then the maximum likelihood estimator $\tilde{\theta}$ exists and the distribution of $\sqrt{n}(\tilde{\theta} - \theta^0)$ weakly approaches $\mathcal{N}(\mathbf{0}, \mathbf{I}^{-1}(\theta^0))$ as $n \rightarrow \infty$.

The claim of Theorem 4 implies weak convergence, i.e., the limiting distribution of $\sqrt{n}(\tilde{\theta} - \theta^0)$ is multivariate normal with zero mean vector and covariance matrix $\mathbf{I}^{-1}(\theta^0)$.

Let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be the observed main-stage data. Each data point $\mathbf{y}_i$ is a vector of size four, where the first two elements represent the endpoints of the interval stated at Qu1 and the last two elements represent the endpoints of the interval stated at Qu2 Δ . We consider $\mathbf{y}_1, \dots, \mathbf{y}_n$ to be values of i.i.d. random variables $\mathbf{Y}_1, \dots, \mathbf{Y}_n$. We denote $\mathbf{Y}_{1:n} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)$. Let $\mathbf{Y}_1^*, \dots, \mathbf{Y}_n^*$ be i.i.d. random variables taking on the values $\mathbf{y}_1, \dots, \mathbf{y}_n$ with probability $1/n$, i.e., $\mathbf{Y}_1^*, \dots, \mathbf{Y}_n^*$ is a random sample with replacement from the original data set $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$. We say that $\mathbf{Y}_1^*, \dots, \mathbf{Y}_n^*$ is a bootstrap sample. Let $\tilde{\theta}^*$ be the maximum likelihood estimator of θ from the bootstrap sample $\mathbf{Y}_1^*, \dots, \mathbf{Y}_n^*$.

Theorem 5 *Let assumptions A2, A5, and A6 be true, $\gamma_2 > 0$, and the conditional probabilities $w_{h|j}$ be known (or strongly consistently estimated). Then the distribution of $\sqrt{n}(\tilde{\theta}^* - \tilde{\theta}) \mid \mathbf{Y}_{1:n}$ weakly approaches the distribution of $\sqrt{n}(\tilde{\theta} - \theta^0)$ in probability as $n \rightarrow \infty$.*

This result can be applied for constructing confidence intervals for θ_r , $r = 1, \dots, d$. Let $G_{\text{boot}}(x) = P(n^{1/2}(\tilde{\theta}_r^* - \tilde{\theta}_r) \leq x \mid \mathbf{Y}_{1:n})$. The interval

$$\left[\tilde{\theta}_r - n^{-1/2} G_{\text{boot}}^{-1}(1 - \alpha/2), \quad \tilde{\theta}_r - n^{-1/2} G_{\text{boot}}^{-1}(\alpha/2) \right] \quad (4)$$

is an approximate $1 - \alpha$ confidence interval for θ_r (hybrid bootstrap confidence interval; see Shao and Tu 1995, p. 140).

5 Simulation study

We have conducted a simulation study to examine the performance of the proposed methods. The data for the pilot stage and for Qu1 at the main stage are generated in the same way. We describe it for Qu1 to avoid unnecessary notation. In all simulations, the random variables $X_1, \dots, X_n$ are independent and have a Weibull distribution:

$$F(x) = P(X_i \leq x) = 1 - \exp(-(x/\sigma)^\nu), \quad \text{for } x > 0,$$

where $\nu = 1.5$ and $\sigma = 80$. The Weibull distribution has a flexible shape and is used in various contexts, for example, in contingent valuation studies where people are asked how much they would be willing to pay for a certain nonmarket good (see, e.g., Alberini et al. 2005). Contingent valuation is a natural application area for the sampling schemes considered here because they account for respondent uncertainty.

Let $U_1^L, \dots, U_n^L$ and $U_1^R, \dots, U_n^R$ be sequences of i.i.d. random variables defined below:

$$\begin{aligned} U_i^L &= M_i U_i^{(1)} + (1 - M_i) U_i^{(2)}, \\ U_i^R &= M_i U_i^{(2)} + (1 - M_i) U_i^{(1)}, \end{aligned} \quad (5)$$

where $M_i \sim \text{Bernoulli}(1/2)$, $U_i^{(1)} \sim \text{Uniform}(0, 20)$, and $U_i^{(2)} \sim \text{Uniform}(20, 50)$. Let $(L_{1i}, R_{1i}]$ be the interval stated by the i -th respondent at Qu1. The left endpoints are generated as $L_{1i} = (X_i - U_i^L) \mathbb{1}\{X_i - U_i^L > 0\}$ rounded downwards to the nearest multiple of 10. The right endpoints are generated as $R_{1i} = X_i + U_i^R$ rounded upwards to the nearest multiple of 10. The data for the follow-up questions Qu2a and Qu2b are generated according to scheme B. The probability that a respondent gives no answer to Qu2 Δ is 1/6. All computations were performed in R (R Core Team 2016). The R code can be obtained from the first author upon request.

It is of interest to investigate to what extent the set of endpoints $\{d_j^*\}$ influences the properties of the estimator of $\theta = (\nu, \sigma)$. For this purpose, we explore three different ways of obtaining the set $\{d_j^*\}$, i.e., three variations of scheme B, specified below:

- (i) pilot stage with sample size $n_0 = 20$;
- (ii) pilot stage with sample size which is the same as in the main stage, $n_0 = n$;
- (iii) skipping the pilot stage and using instead a predetermined set of endpoints $\{d_j^*\} = \{0, 10, 20, \dots, 300, 320, 340, \dots, 400\}$, which is a reasonable set given the rounding to a multiple of 10 and the likely values of X_i .

Under the settings of our simulations, the set $\{d_j^*\}$ will on average be smallest in scenario (i) and largest in scenario (iii).

First, we compare the suggested estimator of θ under the three variations of scheme B and the maximum likelihood estimator when $X_1, \dots, X_n$ are observed without censoring (uncensored observation scheme). For each scheme, 40000 samples of different sizes are generated. Table 1 presents the relative bias and the root mean square error over the simulations. If $\tilde{\nu}$ is an estimator of ν , the relative bias of $\tilde{\nu}$ is defined as $\text{rb}(\tilde{\nu}) = 100 \text{bias}(\tilde{\nu})/\nu$. The root mean square error is of more or less the same magnitude in each of the three scenarios for obtaining the set of endpoints $\{d_j^*\}$. However, if we look at the results for $n = 1000$, the bias is smallest when the set of endpoints is largest. This indicates that the set $\{d_j^*\}$ should not be too small (ideally one would like the set to contain all endpoints that future respondents will give). As we can expect, the error with the uncensored scheme is lower; however, the difference is pretty small. The bias is fairly close to zero with all schemes. Analogous simulations for scheme A displayed comparable results with a slightly higher root mean square error. We also conducted similar simulations with the scheme suggested in Angelov and Ekström (2017) which showed a larger bias, e.g., for $n_0 = n = 100$, $\text{rb}(\tilde{\nu}) = 6.7$, for $n_0 = n = 1000$, $\text{rb}(\tilde{\nu}) = 1.7$, while with schemes A and B of the current paper, $\text{rb}(\tilde{\nu}) < 1$ in each of the cases studied. This bias can be attributed to the exclusion of respondents in the former scheme. For the sake of brevity, the detailed simulation results for scheme A and the scheme of Angelov and Ekström (2017) are omitted.

In addition, we have performed simulations to examine potential bias due to wrongly assuming that $w_{h|j}$ does not depend on j . This assumption implies noninformative censoring, and in this case the likelihood will be proportional to

$$\prod_{i=1}^n [F_{\theta}(b_i) - F_{\theta}(a_i)], \quad (6)$$

where $(a_i, b_i]$ is the last interval stated by respondent i at the series of questions Qu1, Qu2Δ (cf. Sun 2006, p. 28). We compare the estimator suggested in this paper with an estimator assuming noninformative censoring, obtained by maximizing the likelihood (6). For generating data we use the model stated above with $M_i \sim \text{Bernoulli}(1/100)$ in (5). This model corresponds to a specific behavior of the respondents, that is, at Qu1 they tend to choose an interval in which the true value is located in the right half of the interval. The estimator assuming noninformative censoring has been applied both to the full data (Qu1 and Qu2Δ) and to the data only from Qu1. Table 2 displays the relative bias and the root mean square error of the estimators based on 40000 simulated samples of sizes $n = 100$ and $n = 1000$, with scheme variation B(ii). For $n = 1000$ when using the full data, the bias of the estimator assuming noninformative censoring is substantially greater than the bias of our estimator. Similar thing is observed for the root mean square error. The results with $n = 100$ indicate that when using the

Table 1 Simulation results for different sampling schemes

Scheme	n	$m(\tilde{\nu})$	$rb(\tilde{\nu})$	$rmse(\tilde{\nu})$	$m(\tilde{\sigma})$	$rb(\tilde{\sigma})$	$rmse(\tilde{\sigma})$
B, $n_0 = 20$	100	1.510	0.675	0.127	80.111	0.139	5.747
B, $n_0 = n$	100	1.514	0.937	0.127	80.077	0.097	5.754
B, no pilot	100	1.514	0.902	0.127	80.051	0.064	5.707
Uncensored	100	1.521	1.385	0.123	80.028	0.034	5.614
B, $n_0 = 20$	200	1.501	0.088	0.088	80.088	0.110	4.082
B, $n_0 = n$	200	1.504	0.282	0.089	80.023	0.028	4.026
B, no pilot	200	1.504	0.295	0.088	80.048	0.060	4.072
Uncensored	200	1.510	0.693	0.085	79.997	-0.004	3.969
B, $n_0 = 20$	500	1.495	-0.335	0.056	80.126	0.157	2.583
B, $n_0 = n$	500	1.501	0.042	0.055	80.018	0.023	2.562
B, no pilot	500	1.501	0.049	0.056	80.045	0.056	2.555
Uncensored	500	1.504	0.277	0.053	79.996	-0.005	2.516
B, $n_0 = 20$	1000	1.493	-0.454	0.040	80.097	0.121	1.824
B, $n_0 = n$	1000	1.500	0.011	0.039	80.018	0.023	1.795
B, no pilot	1000	1.500	-0.003	0.039	80.001	0.001	1.810
Uncensored	1000	1.502	0.140	0.037	79.993	-0.009	1.770

m mean, rb relative bias, $rmse$ root mean square error

Table 2 Comparison of our estimator (I) with an estimator assuming noninformative censoring (II)

Estimator	n	Scheme	$m(\tilde{\nu})$	$rb(\tilde{\nu})$	$rmse(\tilde{\nu})$	$m(\tilde{\sigma})$	$rb(\tilde{\sigma})$	$rmse(\tilde{\sigma})$
I	100	B, $n_0 = n$	1.524	1.567	0.126	79.356	-0.806	5.649
II	100	B, $n_0 = n$	1.492	-0.555	0.123	77.577	-3.029	6.122
II	100	Qu1 data	1.316	-12.248	0.224	65.129	-18.588	15.959
I	1000	B, $n_0 = n$	1.503	0.229	0.038	79.899	-0.126	1.780
II	1000	B, $n_0 = n$	1.469	-2.091	0.049	77.705	-2.869	2.901
II	1000	Qu1 data	1.296	-13.616	0.208	65.107	-18.616	15.004

m mean, rb relative bias, $rmse$ root mean square error

full data, for ν the bias of our estimator is a bit greater than the bias of the other estimator, while for σ the bias of our estimator is smaller. Yet, with our estimator the estimated distribution function more closely resembles the true distribution. If only the data from Qu1 are used, the bias under the assumption of noninformative censoring is considerably larger for both sample sizes.

Finally, we compare the performance of the bootstrap confidence intervals (4) and the confidence intervals constructed using normal approximation (see Theorem 4). Table 3 shows results based on 1500 simulated samples of sizes $n = 100$ and $n = 1000$ using scheme variation B(iii); the confidence level is 0.95. One bootstrap confidence interval is calculated using 1000 bootstrap samples. For both sample sizes we see that the bootstrap confidence intervals have similar coverage and length as the confidence intervals based on normal approximation.

Table 3 Confidence intervals: coverage proportion and average length

Method	n	ν		σ	
		CP	AL	CP	AL
Normal approximation	100	0.941	0.484	0.941	22.199
Bootstrap	100	0.933	0.494	0.935	21.979
Normal approximation	1000	0.945	0.152	0.947	7.047
Bootstrap	1000	0.953	0.152	0.941	7.045

CP coverage proportion, AL average length, $\alpha = 0.05$

6 Conclusion

We considered two schemes (A and B) for collecting self-selected interval data that extend sampling schemes studied before in the literature. Under general assumptions, we proved the existence, consistency, and asymptotic normality of a proposed parametric maximum likelihood estimator. In comparison with the scheme used in a previous paper (Angelov and Ekström 2017), the new schemes do not involve exclusion of respondents and this leads to a smaller bias of the estimator as indicated by our simulation study. Furthermore, the simulations showed a good performance of the estimator compared to the maximum likelihood estimator for uncensored observations. It should be noted that the censoring in this case is imposed by the design of the question. A design allowing uncensored observations might introduce bias in the estimation if respondents are asked a question that is difficult to answer with an exact amount (e.g., number of hours spent on the internet) and they give a rough best guess. We also demonstrated via simulations that ignoring the informative censoring can lead to bias. We presented a bootstrap procedure for constructing confidence intervals that is easier to apply compared to the confidence intervals based on asymptotic normality where, e.g., the derivatives of the log-likelihood need to be calculated. According to our simulations, the two approaches yield similar results in terms of coverage and length of the confidence intervals. Finally, it would be of interest in future research to develop a test for assessing the goodness of fit of a parametric model.

Acknowledgements The authors would like to thank Maria Karlsson, Marie Wiberg, Philip Fowler, and two anonymous reviewers for their valuable comments which helped to improve this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

We first introduce some notation and expressions for the log-likelihood that will be used henceforth. Let us denote by n' , n'' , and n''' the number of respondents who gave an answer of type 1, 2, and 3, respectively, and let $n_{\bullet j}$ be the number of respondents

who stated $\mathbf{v}_j$ at $\text{Qu}2\Delta$. The following are satisfied:

$$n' = \sum_h n_{h,\text{NA}}, \quad n'' = \sum_j n_{\bullet j}, \quad n''' = \sum_{h,s} n_{h* s}, \quad n' + n'' + n''' = n.$$

From (1) we have

$$\begin{aligned} \frac{\log L(\mathbf{q})}{n} &= \frac{n'}{n} \sum_h \frac{n_{h,\text{NA}}}{n'} \log \left(\sum_{j \in \mathcal{J}_h} w_{h|j} q_j \right) + \frac{n''}{n} \sum_j \frac{n_{\bullet j}}{n''} \log q_j \\ &\quad + \frac{n'''}{n} \sum_{h,s} \frac{n_{h* s}}{n'''} \log \left(\sum_{j \in \mathcal{J}_s} w_{h|j} q_j \right) + c_2, \end{aligned} \quad (7)$$

where $c_2 = (1/n)(c_1 + \sum_{h,j} n_{hj} \log w_{h|j})$. Using the notations

$$\begin{aligned} \hat{\gamma}_1 &= \frac{n'}{n}, \quad \hat{\gamma}_2 = \frac{n''}{n}, \quad \hat{\gamma}_3 = \frac{n'''}{n}, \\ \hat{w}_{h,\text{NA}} &= \frac{n_{h,\text{NA}}}{n'}, \quad \hat{q}_j = \frac{n_{\bullet j}}{n''}, \quad \hat{w}_{h* s} = \frac{n_{h* s}}{n'''}, \\ w_h &= \sum_{j \in \mathcal{J}_h} w_{h|j} q_j, \quad w_{h* s} = \sum_{j \in \mathcal{J}_s} w_{h|j} q_j, \end{aligned}$$

we can write the log-likelihood (7) in a more compact way:

$$\frac{\log L(\mathbf{q})}{n} = \hat{\gamma}_1 \sum_h \hat{w}_{h,\text{NA}} \log w_h + \hat{\gamma}_2 \sum_j \hat{q}_j \log q_j + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h* s} \log w_{h* s} + c_2. \quad (8)$$

Taking into account that $q_j = q_j(\boldsymbol{\theta})$, the log-likelihood (8) may also be written as follows

$$\begin{aligned} \frac{\log L(\boldsymbol{\theta})}{n} &= \hat{\gamma}_1 \sum_h \hat{w}_{h,\text{NA}} \log w_h(\boldsymbol{\theta}) + \hat{\gamma}_2 \sum_j \hat{q}_j \log q_j(\boldsymbol{\theta}) \\ &\quad + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h* s} \log w_{h* s}(\boldsymbol{\theta}) + c_2, \end{aligned} \quad (9)$$

where $w_h(\boldsymbol{\theta}) = \sum_{j \in \mathcal{J}_h} w_{h|j} q_j(\boldsymbol{\theta})$ and $w_{h* s}(\boldsymbol{\theta}) = \sum_{j \in \mathcal{J}_s} w_{h|j} q_j(\boldsymbol{\theta})$.

Lemma 1 (Information inequality) *Let $\sum_i a_i$ and $\sum_i b_i$ be convergent series of positive numbers such that $\sum_i a_i \geq \sum_i b_i$. Then*

$$\sum_i a_i \log \frac{b_i}{a_i} \leq 0,$$

with strict inequality if $a_i \neq b_i$ for at least one i .

A proof of Lemma 1 can be found in Rao (1973, p. 58).

Proof of Theorem 1 Let us consider the case when the conditional probabilities $w_{h|j}$ are known. By convention, we define $0 \log 0 = 0$ and $0 \log(a/0) = 0$ on the basis that $\lim_{x \downarrow 0} x \log x = 0$ and $\lim_{x \downarrow 0} x \log(a/x) = 0$ for $a > 0$. Using (2) and Lemma 1, we get

$$\begin{aligned} & \frac{\log c}{n} + \hat{\gamma}_1 \sum_h \hat{w}_{h,NA} \log \frac{w_h(\theta^0)}{\hat{w}_{h,NA}} + \hat{\gamma}_2 \sum_j \hat{q}_j \log \frac{q_j(\theta^0)}{\hat{q}_j} + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h*s} \log \frac{w_{h*s}(\theta^0)}{\hat{w}_{h*s}} \\ & \leq \hat{\gamma}_1 \sum_h \hat{w}_{h,NA} \log \frac{w_h(\tilde{\theta})}{\hat{w}_{h,NA}} + \hat{\gamma}_2 \sum_j \hat{q}_j \log \frac{q_j(\tilde{\theta})}{\hat{q}_j} + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h*s} \log \frac{w_{h*s}(\tilde{\theta})}{\hat{w}_{h*s}} \leq 0. \end{aligned} \quad (10)$$

From the strong law of large numbers (SLLN) it follows that

$$\begin{aligned} \hat{\gamma}_t & \xrightarrow{\text{a.s.}} \gamma_t, \quad t = 1, 2, 3, \\ \hat{w}_{h,NA} & \xrightarrow{\text{a.s.}} w_h(\theta^0), \\ \hat{q}_j & \xrightarrow{\text{a.s.}} q_j(\theta^0), \\ \hat{w}_{h*s} & \xrightarrow{\text{a.s.}} w_{h*s}(\theta^0), \end{aligned} \quad (11)$$

as $n \rightarrow \infty$. Combining (10) and (11) yields

$$\hat{\gamma}_1 \sum_h \hat{w}_{h,NA} \log \frac{w_h(\tilde{\theta})}{\hat{w}_{h,NA}} + \hat{\gamma}_2 \sum_j \hat{q}_j \log \frac{q_j(\tilde{\theta})}{\hat{q}_j} + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h*s} \log \frac{w_{h*s}(\tilde{\theta})}{\hat{w}_{h*s}} \xrightarrow{\text{a.s.}} 0$$

as $n \rightarrow \infty$. From this and Lemma 1 it follows that

$$\hat{\gamma}_2 \sum_j \hat{q}_j \log \frac{q_j(\tilde{\theta})}{\hat{q}_j} \xrightarrow{\text{a.s.}} 0.$$

Using (11) and the assumption $\gamma_2 > 0$, we get

$$\sum_j q_j(\theta^0) \log \frac{q_j(\tilde{\theta})}{q_j(\theta^0)} \xrightarrow{\text{a.s.}} 0.$$

This, together with assumption A2, implies that $\tilde{\theta} \xrightarrow{\text{a.s.}} \theta^0$, which is what we had to prove.

In the case, when a strongly consistent estimator of $w_{h|j}$ is inserted into the log-likelihood, the proof follows the same lines. $\square$

Proof of Theorem 2 We give a proof for the case when $w_{h|j}$ are known; the proof for the case when $w_{h|j}$ are strongly consistently estimated is similar and thus omitted. From Lemma 1 we have

$$\begin{aligned}\sum_h w_h(\boldsymbol{\theta}^0) \log w_h(\boldsymbol{\theta}^0) &\geq \sum_h w_h(\boldsymbol{\theta}^0) \log w_h(\boldsymbol{\theta}), \\ \sum_{h,s} w_{h*s}(\boldsymbol{\theta}^0) \log w_{h*s}(\boldsymbol{\theta}^0) &\geq \sum_{h,s} w_{h*s}(\boldsymbol{\theta}^0) \log w_{h*s}(\boldsymbol{\theta}),\end{aligned}$$

and from assumption A2 we deduce that for every $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| > \delta\}$,

$$\sum_j q_j(\boldsymbol{\theta}^0) \log q_j(\boldsymbol{\theta}^0) > \sum_j q_j(\boldsymbol{\theta}^0) \log q_j(\boldsymbol{\theta}).$$

Combining the above inequalities and using that $\gamma_2 > 0$, we obtain that for every $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| > \delta\}$,

$$\begin{aligned}\gamma_1 \sum_h w_h(\boldsymbol{\theta}^0) \log w_h(\boldsymbol{\theta}^0) + \gamma_2 \sum_j q_j(\boldsymbol{\theta}^0) \log q_j(\boldsymbol{\theta}^0) + \gamma_3 \sum_{h,s} w_{h*s}(\boldsymbol{\theta}^0) \log w_{h*s}(\boldsymbol{\theta}^0) \\ > \gamma_1 \sum_h w_h(\boldsymbol{\theta}^0) \log w_h(\boldsymbol{\theta}) + \gamma_2 \sum_j q_j(\boldsymbol{\theta}^0) \log q_j(\boldsymbol{\theta}) + \gamma_3 \sum_{h,s} w_{h*s}(\boldsymbol{\theta}^0) \log w_{h*s}(\boldsymbol{\theta}).\end{aligned}$$

It then follows, using (11), that for every $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| > \delta\}$ and for large enough n ,

$$\begin{aligned}\hat{\gamma}_1 \sum_h \hat{w}_{h,NA} \log w_h(\boldsymbol{\theta}^0) + \hat{\gamma}_2 \sum_j \hat{q}_j \log q_j(\boldsymbol{\theta}^0) + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h*s} \log w_{h*s}(\boldsymbol{\theta}^0) \\ > \hat{\gamma}_1 \sum_h \hat{w}_{h,NA} \log w_h(\boldsymbol{\theta}) + \hat{\gamma}_2 \sum_j \hat{q}_j \log q_j(\boldsymbol{\theta}) + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h*s} \log w_{h*s}(\boldsymbol{\theta}),\end{aligned}$$

or equivalently, $\log L(\boldsymbol{\theta}^0) > \log L(\boldsymbol{\theta})$. Therefore,

$$\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}) = \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| \leq \delta} L(\boldsymbol{\theta}).$$

From this and assumption A3, it follows that $L(\boldsymbol{\theta})$ is continuous and its supremum over Θ is attained at some point $\tilde{\boldsymbol{\theta}}$ within the set $\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| \leq \delta\}$. Because δ is arbitrary, $\tilde{\boldsymbol{\theta}} \xrightarrow{\text{a.s.}} \boldsymbol{\theta}^0$ as $n \rightarrow \infty$. $\square$

Proof of Theorem 3 We give a proof for the case when $w_{h|j}$ are known; the proof for the case when $w_{h|j}$ are strongly consistently estimated follows the same lines. Similarly to Rao (1973, p. 361), let us consider the function

$$\gamma_1 \sum_h w_h(\theta^0) \log \frac{w_h(\theta^0)}{w_h(\theta)} + \gamma_2 \sum_j q_j(\theta^0) \log \frac{q_j(\theta^0)}{q_j(\theta)} + \gamma_3 \sum_{h,s} w_{h*s}(\theta^0) \log \frac{w_{h*s}(\theta^0)}{w_{h*s}(\theta)} \quad (12)$$

on the set $\{\theta : \|\theta - \theta^0\| \leq \delta\}$, which is a neighborhood of θ^0 . Note that for δ small enough, $\{\theta : \|\theta - \theta^0\| \leq \delta\} \subseteq \Theta$. By the continuity assumption A4, the infimum of (12) over the set $\{\theta : \|\theta - \theta^0\| = \delta\}$ is attained, and then by assumption A1 and Lemma 1 there exists $\varepsilon > 0$ such that

$$\inf_{\|\theta - \theta^0\| = \delta} \left(\gamma_1 \sum_h w_h(\theta^0) \log \frac{w_h(\theta^0)}{w_h(\theta)} + \gamma_2 \sum_j q_j(\theta^0) \log \frac{q_j(\theta^0)}{q_j(\theta)} + \gamma_3 \sum_{h,s} w_{h*s}(\theta^0) \log \frac{w_{h*s}(\theta^0)}{w_{h*s}(\theta)} \right) > \varepsilon.$$

It then follows, using (11), that for large enough n ,

$$\inf_{\|\theta - \theta^0\| = \delta} \left(\hat{\gamma}_1 \sum_h \hat{w}_{h,NA} \log \frac{w_h(\theta^0)}{w_h(\theta)} + \hat{\gamma}_2 \sum_j \hat{q}_j \log \frac{q_j(\theta^0)}{q_j(\theta)} + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h*s} \log \frac{w_{h*s}(\theta^0)}{w_{h*s}(\theta)} \right) > 0.$$

Hence, for every $\theta \in \{\theta : \|\theta - \theta^0\| = \delta\}$ and as $n \rightarrow \infty$, we have

$$\begin{aligned} & \hat{\gamma}_1 \sum_h \hat{w}_{h,NA} \log w_h(\theta^0) + \hat{\gamma}_2 \sum_j \hat{q}_j \log q_j(\theta^0) + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h*s} \log w_{h*s}(\theta^0) \\ & > \hat{\gamma}_1 \sum_h \hat{w}_{h,NA} \log w_h(\theta) + \hat{\gamma}_2 \sum_j \hat{q}_j \log q_j(\theta) + \hat{\gamma}_3 \sum_{h,s} \hat{w}_{h*s} \log w_{h*s}(\theta), \end{aligned}$$

or equivalently, $\log \bar{L}(\theta^0) > \log L(\theta)$, which implies that $\log L(\theta)$ has a local maximum at some point $\bar{\theta}$ within the set $\{\theta : \|\theta - \theta^0\| < \delta\}$. Assumption A4 implies that $\log L(\theta)$ has partial derivatives and therefore $\bar{\theta}$ is a root of the system of likelihood equations (3). Because δ is arbitrary, $\bar{\theta} \xrightarrow{\text{a.s.}} \theta^0$ as $n \rightarrow \infty$. $\square$

Proof of Theorem 4 Let us consider the case when $w_{h|j}$ are known; the case when $w_{h|j}$ are strongly consistently estimated can be treated similarly. The existence of the maximum likelihood estimator $\tilde{\theta}$ follows from Theorem 2. The log-likelihood (9) can be expressed as follows

$$\frac{\log L(\boldsymbol{\theta})}{n} = \sum_a \hat{\pi}_a \log \pi_a(\boldsymbol{\theta}) + c_2,$$

where $\hat{\pi}_a$ is a relative frequency, $\pi_a(\boldsymbol{\theta})$ is an expression of the form $\sum_{j \in \mathcal{J}^{(a)}} w_{h|j} q_j(\boldsymbol{\theta})$, and $\mathcal{J}^{(a)}$ is an index set. The continuity of the derivatives of $q_j(\boldsymbol{\theta})$ implies the continuity of the derivatives of $\pi_a(\boldsymbol{\theta})$. Thus, the proof of asymptotic normality follows the same lines as that of proposition (iv) in Rao (1973, p. 361). $\square$

For proving Theorem 5 we will consider assumptions B1–B4 stated below. Recall that the contribution of the i -th respondent to the log-likelihood is denoted $\text{lik}_i(\boldsymbol{\theta})$, given by

$$\begin{aligned} \text{lik}_i(\boldsymbol{\theta}) = & \mathbb{1}\{\mathbf{Y}_i = (\mathbf{u}_h; \text{NA})\} \log \sum_{j \in \mathcal{J}_h} w_{h|j} q_j(\boldsymbol{\theta}) \\ & + \mathbb{1}\{\mathbf{Y}_i = (\mathbf{u}_h; \mathbf{v}_j)\} \log(w_{h|j} q_j(\boldsymbol{\theta})) \\ & + \mathbb{1}\{\mathbf{Y}_i = (\mathbf{u}_h; \mathbf{u}_s)\} \log \sum_{j \in \mathcal{J}_s} w_{h|j} q_j(\boldsymbol{\theta}). \end{aligned} \quad (13)$$

B1 The partial derivatives

$$\frac{\partial \text{lik}_i(\boldsymbol{\theta})}{\partial \theta_r}, \quad \frac{\partial^2 \text{lik}_i(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_\ell}, \quad r, \ell = 1, \dots, d,$$

exist and are continuous functions of $\boldsymbol{\theta} \in \Theta$.

B2 For each $\boldsymbol{\theta} \in \Theta$, there exist $K_1(\boldsymbol{\theta}), K_2(\boldsymbol{\theta}) \in \mathbb{R}$ such that

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}} \left| \frac{\partial \text{lik}_i(\boldsymbol{\theta})}{\partial \theta_r} \right|^3 & \leq K_1(\boldsymbol{\theta}), \quad r = 1, \dots, d, \\ \mathbb{E}_{\boldsymbol{\theta}} \left| \frac{\partial^2 \text{lik}_i(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_\ell} \right|^3 & \leq K_2(\boldsymbol{\theta}), \quad r, \ell = 1, \dots, d. \end{aligned}$$

B3 For each $\boldsymbol{\theta} \in \Theta$ and for every $\delta > 0$, there exists $\varepsilon(\delta, \boldsymbol{\theta})$ such that for $\varepsilon \leq \varepsilon(\delta, \boldsymbol{\theta})$,

$$\mathbb{E}_{\boldsymbol{\theta}} \left(\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}'\| < \varepsilon} \left| \frac{\partial^2 \text{lik}_i(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_\ell} - \frac{\partial^2 \text{lik}_i(\boldsymbol{\theta}')}{\partial \theta_r \partial \theta_\ell} \right| \right) \leq \delta.$$

B4 For each $\boldsymbol{\theta} \in \Theta$,

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}} \left(\frac{\partial \text{lik}_i(\boldsymbol{\theta})}{\partial \theta_r} \right) & = 0, \quad r = 1, \dots, d, \\ \mathbb{E}_{\boldsymbol{\theta}} \left(\frac{\partial \text{lik}_i(\boldsymbol{\theta})}{\partial \theta_r} \frac{\partial \text{lik}_i(\boldsymbol{\theta})}{\partial \theta_\ell} \right) & = -\mathbb{E}_{\boldsymbol{\theta}} \left(\frac{\partial^2 \text{lik}_i(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_\ell} \right), \quad r, \ell = 1, \dots, d. \end{aligned}$$

Lemma 2 *If assumption A5 is satisfied and the conditional probabilities $w_{h|j}$ are known (or strongly consistently estimated), then assumptions B1–B4 hold true.*

Proof of Lemma 2 Assumption B1. From the continuity of the first- and second-order partial derivatives of $q_j(\theta)$, it follows that $\frac{\partial \text{lilik}_i(\theta)}{\partial \theta_r}$ and $\frac{\partial^2 \text{lilik}_i(\theta)}{\partial \theta_r \partial \theta_\ell}$ are also continuous.

Assumption B2. Because we have an experiment with a finite number of outcomes, the respective expectations can be expressed as finite sums and are therefore finite.

Assumption B3. From assumption B1, we have that $\frac{\partial^2 \text{lilik}_i(\theta)}{\partial \theta_r \partial \theta_\ell}$ is continuous on the compact set Θ , implying that it is uniformly continuous on Θ . Therefore

$$\sup_{\|\theta - \theta'\| < \varepsilon} \left| \frac{\partial^2 \text{lilik}_i(\theta)}{\partial \theta_r \partial \theta_\ell} - \frac{\partial^2 \text{lilik}_i(\theta')}{\partial \theta_r \partial \theta_\ell} \right| \xrightarrow{\text{a.s.}} 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Using Lebesgue's dominated convergence theorem (see, e.g., Roussas 2014, p. 75), we get

$$\mathbb{E}_\theta \left(\sup_{\|\theta - \theta'\| < \varepsilon} \left| \frac{\partial^2 \text{lilik}_i(\theta)}{\partial \theta_r \partial \theta_\ell} - \frac{\partial^2 \text{lilik}_i(\theta')}{\partial \theta_r \partial \theta_\ell} \right| \right) \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0,$$

which is what we had to prove.

Assumption B4. The identities in this assumption follow from the fact that we have an experiment with a finite number of outcomes and thus the respective expectations can be expressed as finite sums. Indeed, if Y is a random variable that can take a finite number of possible values and $P(Y = y) = p_\theta(y)$, then

$$\begin{aligned} \mathbb{E}_\theta \left(\frac{\partial \log p_\theta(Y)}{\partial \theta_r} \right) &= \mathbb{E}_\theta \left(\frac{1}{p_\theta(Y)} \frac{\partial p_\theta(Y)}{\partial \theta_r} \right) \\ &= \sum_y \frac{1}{p_\theta(y)} \frac{\partial p_\theta(y)}{\partial \theta_r} p_\theta(y) = \frac{\partial}{\partial \theta_r} \sum_y p_\theta(y) = 0. \end{aligned}$$

The same argument leads to the second identity. $\square$

Lemma 3 *If assumptions A2 and A3 are satisfied, $\gamma_2 > 0$, and the conditional probabilities $w_{h|j}$ are known (or strongly consistently estimated), then $\tilde{\theta}^*$ exists and $\tilde{\theta}^* \xrightarrow{\text{a.s.}} \theta^0$ as $n \rightarrow \infty$.*

Proof of Lemma 3 The proof follows the same arguments as that of Theorem 2 but, instead of the classical SLLN, the strong law of large numbers for the bootstrapped mean (see, e.g., Athreya et al. 1984) is used. $\square$

We will present a general result about bootstrapping maximum likelihood estimators that is used in the proof of Theorem 5. Let $z_1, \dots, z_n$ be observed values of i.i.d. random variables $Z_1, \dots, Z_n$ whose distribution depends on some unknown parameter $\theta = (\theta_1, \dots, \theta_d) \in \Theta \subseteq \mathbb{R}^d$. The contribution of the i -th observation to the log-likelihood is denoted $\text{lilik}_i(\theta)$. Let $Z_{1:n} = (Z_1, \dots, Z_n)$ and $\tilde{\theta}$ be the maximum

likelihood estimator of θ . We define $Z_1^*, \dots, Z_n^*$ to be i.i.d. random variables taking on the values $z_1, \dots, z_n$ with probability $1/n$, i.e., $Z_1^*, \dots, Z_n^*$ is a bootstrap sample. Let $\tilde{\theta}^*$ be the maximum likelihood estimator of θ from the bootstrap sample $Z_1^*, \dots, Z_n^*$.

Lemma 4 *Suppose that*

- (i) *assumptions A6 and B1–B4 are true;*
- (ii) *the estimator $\tilde{\theta}$ exists and $\tilde{\theta} \xrightarrow{\text{a.s.}} \theta^0$ as $n \rightarrow \infty$;*
- (iii) *the estimator $\tilde{\theta}^*$ exists and $\tilde{\theta}^* \xrightarrow{\text{a.s.}} \theta^0$ as $n \rightarrow \infty$.*

Then the distribution of $\sqrt{n}(\tilde{\theta}^ - \tilde{\theta}) \mid Z_{1:n}$ weakly approaches the distribution of $\sqrt{n}(\tilde{\theta} - \theta^0)$ in probability as $n \rightarrow \infty$.*

For a proof of Lemma 4, see Belyaev and Nilsson (1997, Corollary 3).

Proof of Theorem 5 The idea of the proof is to show that the conditions of Lemma 4 are fulfilled. By using the fact that assumption A5 implies assumption A3 and combining the results obtained in Theorem 2, Lemma 2, and Lemma 3, we see that these conditions are satisfied. Thus, the assertion of Theorem 5 follows directly. $\square$

References

- Alberini, A., Rosato, P., Longo, A., Zanatta, V.: Information and willingness to pay in a contingent valuation study: the value of S. Erasmo in the Lagoon of Venice. *J. Environ. Plan. Manag.* **48**(2), 155–175 (2005)
- Angelov, A.G., Ekström, M.: Nonparametric estimation for self-selected interval data collected through a two-stage approach. *Metrika* **80**(4), 377–399 (2017)
- Athreya, K.B., Ghosh, M., Low, L.Y., Sen, P.K.: Laws of large numbers for bootstrapped U-statistics. *J. Stat. Plan. Inference* **9**(2), 185–194 (1984)
- Belyaev, Y., Kriström, B.: Approach to analysis of self-selected interval data. Working Paper 2010:2, CERE, Umeå University and the Swedish University of Agricultural Sciences (2010). <https://doi.org/10.2139/ssrn.1582853>
- Belyaev, Y., Kriström, B.: Two-step approach to self-selected interval data in elicitation surveys. Working Paper 2012:10, CERE, Umeå University and the Swedish University of Agricultural Sciences (2012). <https://doi.org/10.2139/ssrn.2071077>
- Belyaev, Y., Kriström, B.: Analysis of survey data containing rounded censoring intervals. *Inf. Appl.* **9**(3), 2–16 (2015)
- Belyaev, Y., Nilsson, L.: Parametric maximum likelihood estimators and resampling. Research Report 1997-15, Department of Mathematical Statistics, Umeå University (1997). <http://www.diva-portal.org/smash/get/diva2:709550/FULLTEXT01.pdf>
- Belyaev, Y., Sjöstedt-de Luna, S.: Weakly approaching sequences of random distributions. *J. Appl. Probab.* **37**(3), 807–822 (2000)
- Chambers, R.L., Steel, D.G., Wang, S., Welsh, A.: Maximum Likelihood Estimation for Sample Surveys. CRC Press, Boca Raton (2012)
- Collett, D.: Modelling Survival Data in Medical Research. Chapman & Hall, London (1994)
- Cox, C., Chu, H., Schneider, M.F., Muñoz, A.: Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat. Med.* **26**(23), 4352–4374 (2007)
- Finkelstein, D.M., Goggins, W.B., Schoenfeld, D.A.: Analysis of failure time data with dependent interval censoring. *Biometrics* **58**(2), 298–304 (2002)
- Furnham, A., Boo, H.C.: A literature review of the anchoring effect. *J. Socio-Econ.* **40**(1), 35–42 (2011)
- Gentleman, R., Geyer, C.J.: Maximum likelihood for interval censored data: consistency and computation. *Biometrika* **81**(3), 618–623 (1994)
- McFadden, D.L., Bemmaor, A.C., Caro, F.G., Dominitz, J., Jun, B.H., Lewbel, A., Matzkin, R.L., Molinari, F., Schwarz, N., Willis, R.J., Winter, J.K.: Statistical analysis of choice experiments and surveys. *Mark. Lett.* **16**(3–4), 183–196 (2005)

- Peto, R.: Experimental survival curves for interval-censored data. *J. R. Stat. Soc. C Appl. Stat.* **22**(1), 86–91 (1973)
- Press, S.J., Tanur, J.M.: An overview of the respondent-generated intervals (RGI) approach to sample surveys. *J. Mod. Appl. Stat. Methods* **3**(2), 288–304 (2004a)
- Press, S.J., Tanur, J.M.: Relating respondent-generated intervals questionnaire design to survey accuracy and response rate. *J. Off. Stat.* **20**(2), 265–287 (2004b)
- R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (2016)
- Rao, C.R.: *Linear Statistical Inference and Its Applications*, 2nd edn. Wiley, New York (1973)
- Roussas, G.: *An Introduction to Measure-Theoretic Probability*, 2nd edn. Academic Press, Boston (2014)
- Shao, J., Tu, D.: *The Jackknife and Bootstrap*. Springer, New York (1995)
- Shardell, M., Scharfstein, D.O., Bozzette, S.A.: Survival curve estimation for informatively coarsened discrete event-time data. *Stat. Med.* **26**(10), 2184–2202 (2007)
- Sun, J.: *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer, New York (2006)
- Turnbull, B.W.: The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Stat. Soc. B Stat. Methodol.* **38**(3), 290–295 (1976)
- Van Exel, N., Brouwer, W., Van Den Berg, B., Koopmanschap, M.: With a little help from an anchor: discussion and evidence of anchoring effects in contingent valuation. *J. Socio-Econ.* **35**(5), 836–853 (2006)
- Zhang, Z., Sun, J.: Interval censoring. *Stat. Methods Med. Res.* **19**(1), 53–70 (2010)