Abstract—We investigate the estimation of the size of a broadcast audience. We give the Maximum Likelihood (ML) estimators for different scenarios. Using Poisson approximation we obtain upper and lower bounds for the audience size. The ML estimators has closed-form expressions, so are the bounds. We treat the problem of a malfunctioning feedback flow control, where the sender is overwhelmed by feedback messages. The ML estimate of the audience size is obtained using the Expectation-Maximization (EM) algorithm, even under such condition of censored values due to feedback implosion. Quantifying the influence of feedback implosion the loss of information for the audience estimation, we further investigated the choice of the parameters in order to optimize estimation results.

Keywords—Broadcast, Multicast, Audience Estimation, Order Statistics, EM Algorithm

I. INTRODUCTION

The Internet now serves as an international mass medium connecting millions of people and the numbers are growing. Today traditional telephone service is being integrated with Internet data traffic. The increasing Internet access bandwidth via cable modems [1] and xDSL [2] fosters the logical next step: IP broadcast - the integration of traditional TV and radio with the Internet.

IP telephony emerges in three different setups: PC to PC, PC to gateway, and gateway to gateway [3]. We can foresee the same level of networking heterogeneity for IP broadcast. The broadcast signal can be delivered via satellite to network access points where it continues packetized within the Internet, it can be delivered via IP multicast, and it can be delivered via unicast through a hierarchy of broadcast hubs [4].

Rather than solely mapping an already existing broadcast service onto a new medium, IP broadcast can take advantage of the Internet’s capability to transport data to and from a receiver. The maybe most significant figure for the broadcast industry is the demographics and the size of its daily audience. These numbers are usually furnished by choosing a small representative subset of receivers with dedicated set-top boxes that send feedback once a day about the viewing behavior of the household members [5].

With the convergence of TV and the Internet it will be possible to provide audience measurements that are (i) real-time and (ii) precise.

In this paper we will examine a protocol and different audience estimators on the performance of:

- The accuracy of the estimate.
- The convergence time.

Given the range of heterogeneity of possible setups for IP broadcast we will restrict the protocol support for estimation to instances at the sender and the receivers. We only assume a forward broadcast, or multicast, channel from the sender to all receivers and a unicast feedback channel from the receivers to the sender. The left hand side of Figure 1 illustrates this setting for a satellite broadcast. Here the broadcast sender sends a Request for Feedback (RFB) message with a sequence number. In turn the receivers send feedback to the sender. In the meanwhile the broadcast sender emitted the next RFB(i + 1). This message stops the sending of any outstanding feedback from the receivers for round i. In the following we will concentrate on such a simple protocol to estimate the size of the audience, n, by the number of returned feedback messages.

Arising Issues with such a protocol are:

- Feedback implosion: The number of feedback mes-
sages may overwhelm the sender, dependent on how many responses are sent in the time interval \( c \), see Figure 1.

- **A dynamic size \( n_t \) of the audience:** The audience may dynamically vary by orders of magnitude, which imposes i) a problem for feedback implosion, and ii) a problem for the convergence of the estimate towards a desired precision.

In this paper we concentrate entirely on the quality of the estimator. We give upper and lower bounds for different maximum likelihood estimators \( \hat{n} \). We give expressions for the best estimator under feedback implosion. Feedback implosion is modeled by considering the loss of feedback messages above a predefined value. We show how it is possible to still obtain a maximum likelihood estimate of the audience size under feedback implosion using the EM-algorithm.

The exact estimation method, described in the following, is more sophisticated than shown in Figure 1. It contains two parameters that can be used to i) regulate the tradeoff between the amount of feedback and the feedback latency [6], and ii) to improve convergence of the estimate. We investigate the accuracy of the estimate dependent on the parameters.

The rest of the paper proceeds as follows. The basic protocol is presented in section II and related work is discussed. In section III we examine the maximum likelihood estimator for the case where only one, the first, response is used to conclude on the audience size. In section IV we expand maximum likelihood estimation to the case, where not only the first but all feedback packets arriving at the sender are used to estimate the size of the audience. We use Poisson approximation and give bounds. Section V shows how a maximum likelihood can be obtained even for the case of missing information due to lost feedback in the implosion. In section VI we aim to find the optimal distribution in order to maximize the information for the estimation. Finally, Section VII concludes the paper.

II. RELATED WORK

Statisticians used to treat the case of estimating a population by a joint estimation of \( n \) the population size and the distribution. An example is the requirement to estimate the fish population by a randomly selected sample from some places in the sea.

Fortunately, we have a more deterministic setting where the distribution is known. We even have influence on the distribution and can use appropriate parameterization to achieve optimal results. We thus have to concentrate on just one parameter to estimate: \( n \).

The potential of polling protocols had been early recognized by Ammar [8]. Bolot [9] was then the first one who evaluated audience estimation in the context of an important application - flow control for multicast. The audience were all the receivers having a classified reception quality. Dependent on the number of receivers with a certain reception class the sender reacts by increasing or decreasing its sending rate. Bolot used the intuitive method of starting with a high maximum value of receivers, sampled the receivers with a reply probability and divided the search space by two with every probing round until a response was received. The audience estimator was based upon this single response. In order to obtain an estimate within certain bounds Bolot’s experiment has to be repeated several times. We show that the number of trials is very high to obtain a reasonable confidence for the maximum likelihood estimator and give the lacking upper and lower bounds for this kind of estimation.

We will now proceed by presenting the estimation protocol we use throughout the paper. After the presentation we discuss a very recent contribution [10] in this context.

The NB scheme

For each round \( i \), the sender sends a request for feedback RFB(\( i \)) with the distribution \( F(\cdot | n) \), \( n \in (0, T) \) to all the \( n \) receivers. Upon receiving the packet from the sender, each receiver \( j \) sets a timer \( z_j(\cdot) \) by drawing \( z_j(\cdot) \) from \( F(\cdot | n) \) and then waits until either \( z_j(\cdot) \) is expired or the timer is cancelled (suppressed) by a new RFB(\( i + 1 \)) packet sent by the sender to start a new round (i.e., the \( (i + 1) \)-th round of probing). When \( z_j(\cdot) \) is expired, receiver \( j \) sends a packet to the sender containing the information \( z_j(\cdot) \). Denote by \( z_{1,n}(\cdot) \) the smallest timer at the \( i \)-th round. Once it receives the first packet, the sender starts the next round by sending \( F(\cdot | n + 1) \) (Z). However, the packets with timer in the time interval \([z_{1,n}(\cdot), z_{1,n}(\cdot) + c] \) are received by the sender, see Figure 1. We denote by \( z_{2,n}(\cdot), \ldots, z_{r_i,n}(\cdot) \) all the \( r_i \) packets arriving after \( z_{1,n}(\cdot) \) at the \( i \)-th round. The time from the expiration of the first timer, \( z_{1,n}(\cdot) \), to the arrival of RFB(\( i + 1 \)) is assumed to be a constant \( c \).

This scheme originally proposed in [11] was improved in [10] by removing a bias and by adding the means to cope with heterogeneous delays. Heterogeneous delays occur in the time, \( c \), from the expiration of the first timer, \( z_{1,n}(\cdot) \), to the arrival of RFB(\( i + 1 \)) due to differences in network delays between sender and different receivers. The idea in [10] is to correct the number of replies such that only the ones are counted that would, assuming a homogeneous delay, fall in the interval \([z_{1,n}(\cdot), z_{1,n}(\cdot) + c] \).

Further, Friedman [10] gives on the last page the maxi-
mum likelihood estimator. Unfortunately, no closed form was found and an iterative solution is proposed. We go one step further in this aspect. First, we use Poisson approximation and give a closed-form for \( \hat{n} \). Second, we provide upper and lower bound for the maximum likelihood estimator, which allows to converge to a desired quality.

III. Maximum likelihood estimation of the membership size from the first arrivals

First we discuss the NB method, when only the first arrival is taken into account. Let \( f_i(z) \) be the density function of the continuous distribution \( F_i(z) \). The likelihood function of the unknown parameter \( n \) given the observations \( \{x_{i,0} = z^{[i]}_{1:n}, i = 1, ..., k\} \) is

\[
L(n|Y) \propto \prod_{i=1}^{k} f_i(x_{i,0}) [1 - F_i(x_{i,0})]^n - 1.
\]

The maximum likelihood estimate of \( n \) is then

\[
\hat{n}_k = \frac{k}{\sum_{i=1}^{k} \ln(1 - F_i(x_{i,0}))}.
\]

(1)

It is easy to show that for any \( i \), \( -\ln(1 - F_i(x_{i,0})) \) is exponentially distributed. This means that these particular results are valid for any distribution \( F \). More precisely, we have the following result:

**Proposition 1:** For any increasing distribution \( F_i(x) \), \( Z = -\ln(1 - F_i(x_{i,0})) \) is exponentially distributed with \( E(-\ln(1 - F_i(x_{i,0}))) = 1/n \), that is, \( Z \) has the density function \( ne^{-nz} \), \( z > 0 \).

From Proposition 1, we have

**Proposition 2:** Let \( \hat{n}_k \) be the maximum likelihood estimate of \( n \) given in (1), then (i)

\[
\hat{n}_k = \frac{d}{\sum_{i=1}^{k} \ln(1 - F_i(x_{i,0}))},
\]

where \( \chi^2_{2k} \) is the Chi-square deviate with \( 2k \) degrees of freedom; and (ii) \( 1/\hat{n}_k \) is an unbiased estimate of \( 1/n \), that is,

\[
E(1/\hat{n}_k) = 1/n.
\]

Proposition 2 can be used to compute the needed number of rounds for given precisions. We present two such examples: (i) upper bounds

\[
P((n - \hat{n}_k)/n < \delta) = P(n > (1 - \delta)^{-1}\hat{n}_k)
\]

(3)

\[
= P(\chi^2_{2k} > 2k/(1 - \delta)) < \alpha,
\]

and (ii) lower bounds

\[
P((n - \hat{n}_k)/n < -\delta) = P(n < (1 + \delta)^{-1}\hat{n}_k)
\]

(4)

\[
= P(\chi^2_{2k} < 2k/(1 + \delta)) < \alpha,
\]

where \( \delta \geq 0 \) and \( 0 < \alpha < 1 \). Again, note that these bounds on the estimator’s precision are valid for any increasing distribution \( F \) used in the NB scheme.

Table I and Table II show that a large number of rounds are required to obtain a good quality for the audience estimate. As an example consider the upper bound given in Table 1: 230 polling rounds are required to guarantee that in less than \( \alpha = 5\% \) of all the cases the actual audience \( n \) is more than 11\% higher than the estimate \( \hat{n} \) \( (\delta = 10\%) \).

Similarly, a very high number of polling rounds can be observed for the lower bound. For both, lower and upper bound, tightening the bound (decreasing \( \delta \)) leads to an explosion in the number of polling rounds required and therefore to a slow convergence.

The slow convergence here is due to the fact that only the minimal timer is considered for estimation purpose.

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>( (1 - \delta)^{-1} )</th>
<th>( 0.01 )</th>
<th>( 0.025 )</th>
<th>( 0.05 )</th>
<th>( 0.10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>1.11</td>
<td>465</td>
<td>328</td>
<td>230</td>
<td>137</td>
</tr>
<tr>
<td>0.20</td>
<td>1.25</td>
<td>98</td>
<td>69</td>
<td>48</td>
<td>28</td>
</tr>
<tr>
<td>0.30</td>
<td>1.43</td>
<td>36</td>
<td>25</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>0.40</td>
<td>1.67</td>
<td>17</td>
<td>12</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>0.50</td>
<td>2.00</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>0.60</td>
<td>2.50</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0.67</td>
<td>3.00</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>4.00</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0.80</td>
<td>5.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

In the next section we will consider the minimal timer and the next \( r \) timers to improve convergence of the audience estimation. For this case we will obtain a general result by considering an arbitrary continuous timer distribution.

IV. A common continuous timer distribution

\( F^{(r)}(z) = F(z) \): No Feedback Implosion

In the NB scheme, the number of responses received by the sender is \( 1 + r \); we now also consider the \( r \) responses in addition to the response with the minimal timer. The NB scheme is executed \( k \) times. In each of the \( k \) rounds we observe \( 1 + r_i \) responses, \( i = 1, ..., k \). We derive bounds for the maximum likelihood estimation of the size
of the audience using all responses. Since we have more information the precision of the estimator improves.

The maximum likelihood estimator of \( n \) is not treated easily. Friedman [10] used an iterative solution to obtain \( \hat{n} \). We give a closed-form using Poisson approximation.

Let \( f(z) \) be the density function of \( F(z) \), that is, \( f(z) = dF(z)/dz \). The likelihood function of the unknown parameter \( n \) given the observations \( Y = \{(x_{i,0} = z_{1}^{(i)}, \ldots, x_{i,\tau+1} = z_{\tau+1}^{(i)}): i = 1, \ldots, k\} \) is

\[
L(n|Y) \propto \prod_{i=1}^{k} n f(x_{i,0}) [1 - F(x_{i,0})]^{|n-1|}
\]

\[
\cdot \left( n - 1 \right) \sum_{i=1}^{k} \frac{r_{i}^{-1} p_{i}^{-1} (1 - p_{i})^{-1 - r_{i}}}{F(x_{i,0} + c) - F(x_{i,0})}
\]

\[
= \sum_{i=1}^{k} \frac{r_{i}^{-1} p_{i}^{-1} (1 - p_{i})^{-1 - r_{i}}}{F(x_{i,0} + c) - F(x_{i,0})}
\]

\[
\cdot (r_{i} + 1)! \prod_{j=0}^{r_{i}} f(x_{i,j})
\]

\[\] for \( i = 1, \ldots, k \). We see that the set \( S = \{(x_{i,0}, r_{i}): i = 1, \ldots, k\} \) is a set of sufficient statistics for the unknown parameter \( n \in [1, \infty) \).

### A. Maximum likelihood estimation using Poisson approximation

We assume that \( p_{i} \) is small and \( n \) is large. Using Poisson approximation (see, for example, [12, ch. 10])

\[
\frac{X_{i}}{r_{i}!} e^{-\lambda_{i}} \approx (n - 1) \left( \frac{1}{r_{i}} \right) p_{i}^{-1} (1 - p_{i})^{n-1 - r_{i}}
\]

where \( \lambda_{i} = (n - 1)p_{i} \), in (1), we have

\[
\ell(n|Y) \approx k \ln(n) + (n - 1) \sum_{i=1}^{k} \ln(1 - F(x_{i,0}))
\]

\[
+ \ln(n - 1) \sum_{i=1}^{k} r_{i} - (n - 1) \sum_{i=1}^{k} p_{i}
\]

\[
\approx k \ln(n) + n \sum_{i=1}^{k} \ln(1 - F(x_{i,0}))
\]

\[
+ \ln(n) \sum_{i=1}^{k} r_{i} - n \sum_{i=1}^{k} p_{i}
\]

Thus, we have the following approximation to the maximum likelihood estimate of \( n \):

\[
\frac{1}{\hat{n}_{k}} = \frac{\sum_{i=1}^{k} \left( x_{i,0} - \ln(1 - F(x_{i,0})) \right)}{k + \sum_{i=1}^{k} r_{i}}
\]

that is,

\[
\hat{n}_{k} = \frac{k + \sum_{i=1}^{k} r_{i}}{\sum_{i=1}^{k} \ln(1 - F(x_{i,0})) + \sum_{i=1}^{k} p_{i}}
\]

**Proposition 3:** \( \hat{n}_{k} \to n \) in probability 1, as \( k \to \infty \).

For comparison of the estimator (7) with other estimators suggested in the literature, we write (7) as

\[
\hat{n}_{k} = \frac{1}{S_{0} + \frac{k}{S_{p}} + \frac{k}{S_{r}}}
\]

where

\[
S_{0} = - \sum_{i=1}^{k} \ln(1 - F(x_{i,0})), \quad S_{r} = \sum_{i=1}^{k} r_{i}
\]

and

\[
S_{p} = \sum_{i=1}^{k} p_{i}
\]

Given \( \{x_{i,0}: i = 1, \ldots, k\} \), the \( r_{i}s \) are conditionally distributed and

\[
\mathbb{E} \left( \frac{r_{i}}{p_{i}} \right) \approx n \quad \text{and} \quad \text{Var} \left( \frac{r_{i}}{p_{i}} \right) \approx \frac{n}{p_{i}}
\]

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>((1 + \delta)^{-1})</th>
<th>(0.01)</th>
<th>(0.025)</th>
<th>(0.05)</th>
<th>(0.10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1111</td>
<td>0.9</td>
<td>512</td>
<td>365</td>
<td>259</td>
<td>160</td>
</tr>
<tr>
<td>0.2500</td>
<td>0.8</td>
<td>121</td>
<td>87</td>
<td>62</td>
<td>39</td>
</tr>
<tr>
<td>0.4286</td>
<td>0.7</td>
<td>50</td>
<td>36</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>0.6667</td>
<td>0.6</td>
<td>26</td>
<td>19</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>1.0000</td>
<td>0.5</td>
<td>16</td>
<td>12</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>1.5000</td>
<td>0.4</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>2.3333</td>
<td>0.3</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>4.0000</td>
<td>0.2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>9.0000</td>
<td>0.1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
This suggests that the weighted unbiased estimate, with the weights proportional to the inverse of the variances,

\[
\frac{S_c}{S_p}
\]

is the minimum-variance-unbiased linear estimate, and thereby better than, for example, the unweighted average

\[
\frac{1}{k} \sum_{i=1}^{k} \frac{r_i}{p_i},
\]

which is analogous to the estimator proposed in ([10], [11]).

The Poisson approximation provides not only a simple way to estimate the audience size \( n \) but also a simple way of accommodating a dynamic audience in which the audience size \( n_t \) changes over time \( t \). At this point in time we lack real data of audience dynamics, nevertheless, suppose that the mean audience size in a session is \( n \) and that \( n_t \) are independently and identically distributed as

\[
n_t | n \sim \text{Poisson}(n),
\]

that is, \( pr(n_t = m | n) = \frac{n^m e^{-n}}{m!} \). Suppose also that given \( n_t \), \( r_i \) follows the binomial distribution \( \text{Binomial}(p_i, n_t) \). Then the observations \( r_i \) are independently distributed with

\[
r_i | n \sim \text{Poisson}(p_i; n),
\]

\[B. \text{The upper and lower bounds: A Bayesian Approach}\]

A simple way of making inference about the audience size \( n \), such as upper and lower bounds, is to take the following Bayesian approach (see, for example, [13]) More specifically, we specify our prior belief about \( n \) in terms of a prior distribution \( pr(n) \), and make inference about \( n \) using the posterior distribution

\[
pr(n | S) \propto pr(n) f(S | n),
\]

where \( f(S | n) \) is the likelihood function of \( n \) given the observed data

\[
S \equiv \{ (x_{i,0}, r_i) : i = 1, \ldots, k \}.
\]

We use the uniform distribution as the prior distribution of \( n \), that is,

\[
pr(n) \propto \text{constant} \quad (n > 0).
\]

From the results in Section III and the Poisson approximation for larger \( n \) in Section IV-A, the likelihood, as a function of \( n \), is then

\[
f(S | n) \propto n^{k+S_r} e^{-(S_0+S_p)n},
\]

where \( S_0 \), \( S_r \), and \( S_p \) are given in equations (8) and (9). This leads to the posterior distribution

\[
pr(n | S) = \text{Gamma}(k + S_r + 1, S_0 + S_p),
\]

(10)

where \( \text{Gamma}(\alpha, \beta) \) denotes the Gamma distribution with the density function

\[
\frac{\beta^n n^{n-1}}{\Gamma(n)} e^{-\beta n},
\]

where \( \Gamma(\cdot) \) is the Gamma function. Thus, upper and lower bounds of \( n \) can be obtained by calculating the incomplete Gamma function.

We conclude this section with the following remarks:

(a) It is easy to sequentially update the posterior distribution.

(b) When the prior distribution is uniform, the posterior mode is the maximum likelihood estimator \( \hat{n} \) given in (7).

\[V. \text{A COMMON CONTINUOUS TIMER DISTRIBUTION } F^{(i)}(z) = F(z) \text{ and Feedback Implosion}\]

We model feedback implosion by the exceedance of a limit \( m \) of number of feedback packets per round. We assume that the first \( m \) packets are received in this round and that other packets are lost. Therefore \( m \) is the number of feedback packets that are available to the sender for estimation during the time period \( (x_{i,0}, x_{i,0} + c) \) for any \( i \). If the actual number of responses \( r + 1 \) received in any round is larger than \( m \) we assume this additional information is lost and do not consider it for the estimation.

\[A. \text{Maximum likelihood estimation using the EM algorithm}\]

The Poisson approximation of Section IV leads to the following the complete-data log-likelihood function:

\[
\ell(n | Y_{\text{COM}}) = n \sum_{i=1}^{k} [\ln(1 - F(x_{i,0})) - p_i] + \ln(n) \left[ k + \sum_{j=1}^{k} r_j \right].
\]

Thus, \( \left\{ \sum_{i=1}^{k} [\ln(1 - F(x_{i,0})) - p_i], k + \sum_{i=1}^{k} r_i \right\} \) provides a set of the sufficient statistics for the unknown parameter \( n \). The complete-data maximum likelihood estimate of \( n \) has a closed-form, as given in Section 3. To find the maximum likelihood estimate of \( n \) from the observed data \( Y_{\text{OBS}} \) with censored values (greater than or equal \( m \)), we use the EM algorithm [14], which is an iterative algorithm. With a reasonable starting value of \( n \) (\( n \in (0, +\infty) \)), each iteration of the EM algorithm for
The EM algorithm:

**E-step:** Compute \( r_i^{(t)} = \mathbb{E}(r_i | Y_{\text{OBS}}, n = \hat{n}_k^{(t-1)}) \), where
\[
\hat{n}_k^{(t)} = \frac{k + \sum_{i=1}^{k} r_i^{(t)}}{\sum_{i=1}^{k} [p_i - \ln(1 - F(x_{i,0}))]}.
\]

and thereby
\[
\lambda_i^{(t)} = \hat{n}_k^{(t)} p_i \quad \text{for} \quad i = 1, \ldots, k.
\]

For simulation study, one can use the maximum likelihood estimate of \( n \) from the first arrivals as the starting values.

For the E-step, we use the result that if \( r \sim \text{Poisson}(\lambda) \), then for \( m > 0 \),
\[
\mathbb{E}(r | r \geq m, \lambda) = \frac{\sum_{k=m}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}}{\sum_{k=m}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}} = \lambda \left( 1 + \frac{P(r = m - 1)}{1 - P(r \leq m - 1)} \right).
\]

**M-step:** Update the estimate of \( n \):
\[
\hat{n}_k^{(t)} = \frac{k + \sum_{i=1}^{k} r_i^{(t)}}{\sum_{i=1}^{k} [p_i - \ln(1 - F(x_{i,0}))]}.
\]

The observed-data likelihood function corresponding to the Poisson approximation to the complete-data likelihood function has the form of
\[
\ell(\theta | Y_{\text{OBS}}) = \left( k + \sum_{r_i < m} r_i \right) \theta + \left( \sum_{i=1}^{k} \ln(1 - F(x_{i,0})) - \sum_{r_i < m} p_i \right) e^{\theta} + \sum_{r_i \geq m} \ln \left( 1 - \sum_{j=0}^{m-1} \frac{p_i^j e^{\theta}}{j! e^{-\theta}} \right)
\]

The observed Fisher information [15] is then
\[
I(\theta) = -n \sum_{i=1}^{k} \ln(1 - F(x_{i,0})) + \sum_{r_i < m} \lambda_i
\]
\[
+ \sum_{r_i \geq m} \left( \frac{(\lambda_i - m) \lambda_i P(r_i = m - 1)}{1 - P(r_i = m - 1)} + \frac{\lambda_i^2 P^2(r_i = m - 1) \lambda_i}{(1 - P(r_i = m - 1))^2} \right)
\]

and thereby
\[
I(\hat{\theta}) = -\hat{n}_k \sum_{i=1}^{k} \ln(1 - F(x_{i,0})) + \sum_{r_i < m} \hat{\lambda}_i
\]
\[
+ \sum_{r_i \geq m} (r_i^{(\infty)} - m)(r_i^{(\infty)} - \hat{\lambda}_i),
\]

where \( \hat{\lambda}_i = \hat{n}_k p_i \). We use
\[
\ln(\hat{n}_k) \sim N(\ln(n), I^{-1}(\hat{\theta}))
\]
to obtain the confidence intervals.

**Remark 1:** When \( m = \infty \), these results can be used for the case with unlimited number of feedback packets in Section 3.

VI. OPTIMAL TIMER DISTRIBUTION

We consider two cases due to implosion: (i) incomplete data with censored values; and (2) fully observed data with time delay.

A. Incomplete data with censored values and no time delay

In [11] it was shown that among three truncated distributions on \((0, T]\), uniform, beta, and exponential, the exponential distribution performs best in the trade-off of feedback delay and implosion.

We therefore consider the class of the timer distributions
\[
F_\beta(x) = \frac{e^{\beta x} - 1}{e^{\beta T} - 1} \quad (0 \leq x \leq T)
\]
indexed by \( \beta \ (> 0) \). The objective is to maximize the expected Fisher information
\[
I^*(\theta) = \frac{1}{k} \mathbb{E}(I(\theta)) = 1 + \mathbb{E} \left[ \lambda_i P(r_i < m | \lambda_i) + (\lambda_i - m) \lambda_i P(r_i = m - 1 | \lambda_i) \right]
\]
\[
+ \frac{\lambda_i^2 P^2(r_i = m - 1 | \lambda_i)}{1 - P(r_i \leq m - 1 | \lambda_i)}
\]
\[
= 1 + \mathbb{E} \left[ \lambda_i P(r_i < m | \lambda_i) + (\lambda_i - m) m P(r_i = m | \lambda_i) \right]
\]
\[
+ \frac{m^2 P^2(r_i = m | \lambda_i)}{1 - P(r_i < m | \lambda_i)} \quad (11)
\]

over \((\beta, c)\) with the fixed session size \( n \), implosion limit \( m \), and \( T = 1 \). Note that this is the same as to maximize \( I^*(\theta) \) over \((\beta, T)\) with the fixed session size \( n \), implosion limit \( m \), and \( c \) the \((\beta, c, T)\) and \((\beta^*, c^*, T^*)\) are equivalent.
in the sense of maximization of \(I^*(\theta)\) with the mapping such that
\[
\frac{\beta}{\beta^*} = \frac{c^*}{c} = \frac{T^*}{T}.
\]

It is interesting to see the two extreme cases: (1) using only the first arrivals (i.e., \(m = 0\)), the expected Fisher information is constant, which is consistent with the results of Section III; and (2) implosion never happens (i.e., \(m = \infty\)), the expected Fisher information is
\[
1 + E(\lambda_i) = 1 + E(np_i) \approx 1 + \left(\frac{n + e^{\beta T}}{e^{\beta T} - 1}\right)
\]
(12)

For any fixed \(T\) and \(c\), this information is infinity at \(\beta = \infty\), i.e., when \(F_{\beta}(x)\) is degenerated to the point mass at \(x = T\). This is consistent with fact that the estimate of \(n\) is exact when \(F_{\beta}(x)\) is degenerated to the point mass at \(x = T\) and all receivers send feedback.

The maximization of \(I^*(\theta)\) over \((\beta, c)\) is generally difficult. The complication also involves the computational burden in finding the maximum likelihood estimate of \(n\) using the EM algorithm: The larger the number of rounds of implosion, the more time consuming the EM algorithm. The probability that the number of feedback packets exceeds the implosion limit is
\[
P(r_i \geq m) = 1 - E[P(r_i < m|\lambda_i)].
\]

A.1 Uniform timer distributions: A simulation study

Consider \(\beta = 0\) and maximize the expected Fisher information \(I^*(\theta)\) over \(c\) (0 < \(c\) < 1) with the fixed \(n\), \(m\), and \(T = 1\). We use the Monte Carlo method to provide empirical results. The \(\lambda_i\) is generated as follows: take \(u\) from the uniform distribution \(U(0, 1)\); solve
\[
-\ln(u) = -n \ln(1 - F(x_o))
\]
for \(x_o\), that is,
\[
x_o = T(1 - u^{1/n}) = (1 - u^{1/n});
\]
and let
\[
\lambda_i = np_i = \frac{ne}{1 - x_o} = \frac{nue^{-1/n}}{n}, \quad (u^{1/n} \geq c); \quad (u^{1/n} < c).
\]
(13)
The results of the optimal \(c\) and the corresponding expected information and implosion probability are given in Table III and Table IV for \(m = 2^i - 1\) and \(n = 10^j\) with \(i = 1, \ldots, 6\) and \(j = 2, \ldots, 6\). (The sample size for Monte Carlo is 1,000,000). For fixed \(m\), the optimal \(nc\) is constant so are the expected information and the implosion probability. This implies that the \(c\) can be easily adjusted in practice for the next round according the current estimates (See Section 5). For fixed \(n\), the optimal \(nc\) increases. The information gained from the packets arrived after the first one, as a function of \(m\), is increasing and convex (maybe quadratic), as indicated, for example, by the values of \((I^*(\theta) - 1)/m\) for \(n = 1,000: (0.65, 0.69, 0.73, 0.78, 0.81, 0.85)\), which has approximately the constant first-order difference 0.04. The vector of the ratios of \((I^*(\theta) - 1)/nc\), (0.41, 0.63, 0.76, 0.84, 0.90, 0.93) corresponding to \(m = 1, 3, 7, 15, 31,\) and 63, indicates that large implosion limits, compared to the case of \(m = 0\), reduce network traffic substantially for estimating the session size and allow satisfactory precision. The corresponding implosion probability also decreases as \(m\) increases.

### Table III

<table>
<thead>
<tr>
<th>(n = 10^2)</th>
<th>(m)</th>
<th>(n \times c)</th>
<th>(I^*(\theta))</th>
<th>(P(r \geq m))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 = 2^1 - 1)</td>
<td>1.6</td>
<td>1.65</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>(3 = 2^2 - 1)</td>
<td>3.3</td>
<td>3.07</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>(7 = 2^3 - 1)</td>
<td>6.7</td>
<td>6.13</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>(15 = 2^4 - 1)</td>
<td>13.6</td>
<td>12.63</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>(32 = 2^5 - 1)</td>
<td>27.8</td>
<td>26.25</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>(63 = 2^6 - 1)</td>
<td>56.2</td>
<td>54.48</td>
<td>0.22</td>
<td></td>
</tr>
</tbody>
</table>

### Table IV

<table>
<thead>
<tr>
<th>(n = 10^3, 10^4, 10^5, ) and (10^6)</th>
<th>(m)</th>
<th>(n \times c)</th>
<th>(I^*(\theta))</th>
<th>(P(r \geq m))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 = 2^1 - 1)</td>
<td>1.6</td>
<td>1.65</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>(3 = 2^2 - 1)</td>
<td>3.3</td>
<td>3.07</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>(7 = 2^3 - 1)</td>
<td>6.8</td>
<td>6.14</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>(15 = 2^4 - 1)</td>
<td>13.8</td>
<td>12.64</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>(32 = 2^5 - 1)</td>
<td>28.1</td>
<td>26.26</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>(63 = 2^6 - 1)</td>
<td>57.3</td>
<td>54.54</td>
<td>0.24</td>
<td></td>
</tr>
</tbody>
</table>
The implosion probability corresponding to the optimal $c$ is large, especially for small implosion limit $m$. The EM algorithm can be slow. Thus, smaller values of $c$ can be considered.

B. Fully observed data with reduced sampling frequency

Let $m$ be the implosion limit and let $r_i$ be the number of feedback packets. For a simple case, we assume that the only complication due to implosion is the time delay, which is proportional to the number of feedback packets. For clarity, we define the time unit as the server time for feedback packets that are less than or equal the implosion limit, that is, for $r_i \leq m - 1$, where $m \geq 2$. Since information on $\theta$ obtained by the $i$-th round is $1 + \lambda_i$, the information per unit time is approximately

$$I_i^\gamma(\theta) = E[(1 + \lambda_i)P(r_i < m|\lambda_i)]$$

$$+ \gamma E\left[\frac{1}{\sum_{r_i=m}^{m} \frac{m}{r_i+1} P(r_i|\lambda_i)}\right]$$

$$= E[(1 + \lambda_i)P(r_i < m|\lambda_i)]$$

$$+ \gamma E\left[\frac{1 + \lambda_i}{\lambda_i} P(r_i \geq m + 1|\lambda_i)\right],$$

(14)

which $\gamma$ is the positive proportionality constant due to the time delay. The coefficient of $P(r_i > m|\lambda_i)$, $\frac{m}{\lambda_i}$, can be viewed as the penalty for allowing for excessive amount of feedbacks. More generally, for complicated implosion effects on the time delay of the server we consider the objective function of the form

$$I_i^\gamma(\theta) = E[(1 + \lambda_i)P(r_i < m|\lambda_i)]$$

$$+ \gamma E\left[\frac{1}{\lambda_i} P(r_i \geq m + 1|\lambda_i)\right],$$

(15)

In particular, we consider in the sequel the following objective function

$$I_\infty^\gamma(\theta) = E[(1 + \lambda_i)P(r_i < m|\lambda_i)],$$

(16)

which is most conservative among all $I_i^\gamma(\theta)$ in the sense that $I_\infty^\gamma(\theta) < I_i^\gamma(\theta)$ for all $p$ and $\gamma$.

C. Optimal timer distribution for the objective function $I_\infty^\gamma(\theta)$

We are considering the objective function $I_\infty^\gamma(\theta)$ for the optimal parametrization of a set of functions $F$.

Let $g(\lambda m) = (1 + \lambda_i)P(r < m\lambda)$, then (1) $I_\infty^\gamma(\theta) = E(g(np_i|m))$; and (2) for all $m \geq 2$, $g(\lambda|m)$ is unimodal with the maximum $\lambda_0$ satisfying

$$\sum_{r=0}^{m-2} \frac{\lambda^r}{r!} - \frac{\lambda^m}{(m-1)!} = 0.$$  

(17)

In general, it is difficult to find the optimal timer distribution. For simplicity, we consider the quadratic approximation to $I_\infty^\gamma(\theta)$ with respect to $z_i = \ln(np_i)$ so that we can find a good choice of the timer distribution based on the expectation and variance of $z_i$.

Denote by $\lambda_0$ the solution to Equation (17). Suppose that $F$ is a class of timer distributions with non-empty subset $F_\gamma \equiv \{F : F \in F, E_{F_0}(z_i|S) = \ln(\lambda_0)\} \neq \emptyset$. A timer distribution $F_0 \in F$ is said to be optimal among $F$ if

1. $E_{F_0}(z_i|S) = \ln(\lambda_0)$, that is, $F_0 \in F_\gamma$; and
2. $V_{F_0}(z_i|S) = \min_{F \in F_\gamma} V_F(z_i|S)$.

As an example, consider the class of the exponential timer distribution. Let $n_0$ and $v_0$ be the current posterior mean and variance of $n$ (or the current maximum likelihood estimate of $n$ and its associated asymptotic variance), respectively. We can get the optimal exponential timer distribution, which is given by

$$\beta_{opt} \approx \max \left\{ \ln \left( \frac{n_0}{v_0} \right), -2 \right\}$$

and

$$\gamma_{opt} \approx \left\{ \begin{array}{ll}
\frac{1}{n_0} \ln \left( 1 + \frac{\lambda_0(e^\beta_{opt} - 1)}{n_0 + 1} \right) & \text{if } \beta_{opt} > 0; \\
\frac{\lambda_0}{n_0 + 1} & \text{if } \beta_{opt} = 0.
\end{array} \right.$$  

Intuitively, when $v_0$ is large, for instance, at the first few rounds, steep distributions are recommended to avoid feedback implosion.

VII. Conclusion

We investigated estimators of a broadcast audience. We give upper and lower bounds for two different maximum likelihood estimators: (i) the estimator based only on the broadcast audience can be estimated via maximum likelihood and (ii) the estimator accounting for all feedback received.

The beauty of the results come with the closed-form expressions we obtain for, first, the maximum likelihood estimator and, second, the upper and lower bound.

We also treat the case of feedback implosion and its impact on the estimation. We show how the size of the broadcast audience can be estimated via maximum likelihood using the EM-algorithm. We further give first results for the optimality of a timer distribution for the estimation purpose.
REFERENCES


