Capture-recapture for population size estimation based upon zero-truncated count distributions with one-inflation

media title: Dice snakes in Graz, drink-driving in Britain, and the size of the Pleiades

Dankmar Böhning

Southampton Statistical Sciences Research Institute
University of Southampton
e-mail: d.a.bohning@soton.ac.uk

May 30, 2019

the idea of capture-recapture

- objective is to determine the size N of an elusive target population
- some mechanism (life trapping, register, surveillance system) identifies a unit repeatingly
- there is a count X informing about the number of identifications of each unit in the target population

sample

available: sample

$$X_1, X_2, ...X_N$$

leading to

Table: Frequency distribution of count X of repeated identifications

x	0	1	2	3	4	...	population size
f_x	f_0	f_1	f_2	f_3	f_4	...	N

problem

if $X_j = 0$ unit is not observed leading to a reduced observable sample

$$X_1, X_2, ...X_n$$

where – w.l.g. – we assume that

$$X_{n+1} = X_{n+2} = ... = X_N = 0$$

Table: Frequency distribution of count X of repeated identifications

x	0	1	2	3	4	...	observed size
f_x	-	f_1	f_2	f_3	f_4	...	n

hence

$$f_0 = N - n \text{ is unknown}$$

estimating the size of a dice snake population in Graz

- Tranninger and Friedl (2018) tried to estimate the size of a dice snake population in a closed area at the river Mur in Graz (Austria)
- work was motivated by resettlement project of the population due to the development of a water power plant in the vicinity
- how many dice snakes are there?
- was considered for several years but here we focus on 2014


Abbildung


eitenden Radweg wurden gezielt

dice snakes in Graz

- there were 31 capture occasions during the year
- X is the identification count per dice snake
- distribution is as follows:

Table: Frequencies of the number of times dice snakes have been identified in the target area in 2014

count of sightings	f_0	f_1	f_2	f_3	f_4	f_5	n
per dice snake		59	8	1	1	1	70

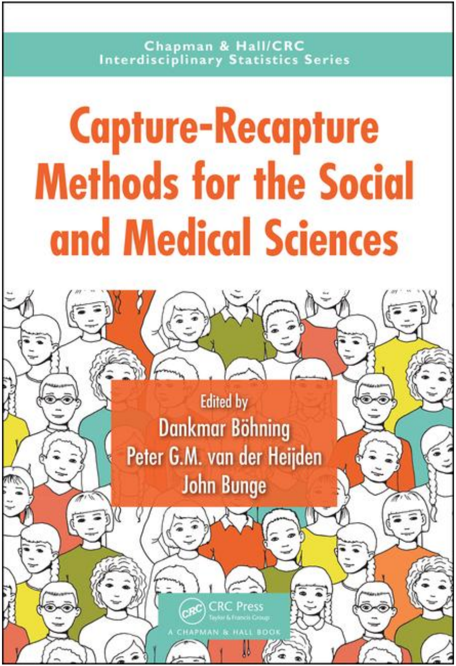
Figure: The Guardian 30 Dec 2016: "Thousands of drink-drivers offend again"

drink-driving in Britain

- drink-driving (DD) relates to driving (or attempting to drive) while being above the legal alcohol limit
- according to the Guardian (30/12/16): 219,000 motorist were caught once, 8,068 twice, etc. (see Table below)

Table: Frequency distribution of the count (per person) of DVLA reported drink-driving (DD) in the UK between 2011 and 2015 (figures are based on DR10 endorsements)

count of DD	f_0	f_1	f_2	f_3	f_4	f_5	f_6	n
frequency		219,008	8,068	449	46	5	2	227,578


Chapman & Hall/CRC
Interdisciplinary Statistics Series
Capture-Recapture Methods for the Social and Medical Sciences
Edited by
Dankmar Böhning
Peter G.M. van der Heijden
John Bunge
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

From: Ashot Hakobyan [mailto:aakopian57@gmail.com]
Sent: 20 July 2018 07:27
To: R.S.McCrea@kent.ac.uk<mailto:R.S.McCrea@kent.ac.uk>;
B.J.T.Morgan@kent.ac.uk<mailto:B.J.T.Morgan@kent.ac.uk>; Bohning D.A.;
P.G.M.vanderHeijden@uu.nl<mailto:P.G.M.vanderHeijden@uu.nl>;
jab18@cornell.edu<mailto:jab18@cornell.edu>

Subject: Astronomical estimator

Dear colleagues,

I thank you for your excellent books:
"Analysis of Capture-Recapture Data", Rachel S. McCrea and Byron J. T. Morgan
"Capture-recapture methods for the social and medical sciences".

I am astronomer and very interested in "Capture-recapture methods". I hope that it will be interesting for you to know that such methods used in astronomy since 1968. At this year famous astronomer Ambartsumian had suggest and applied estimator, which now is known as Chao estimator.
In 1970, Ambartsumian (Astrophysics, 1970, Volume 6, Issue 1, pp.1-10) had prove that estimator gives only lower bound. Unfortunately this facts was missed on your books. Of course it is very explicable and understable.
I hope that this information can be usefull for you.

Sicerely yours
A. Akopian

ASTROPHYSICS 1
FLARE STARS IN T HE PLEIADES .

V. A. Ambartsumyan, L. V. Mirzoyan, E . S. Parsamyan, O. and L. K. E rastova
Ast rofizika, Vol. 6, o. 1, pp. 7-30, _1970

UDC 523. 841

We have collected data on 45 new flare stars in the Pleiades, discovered mainly during the observational season 1968-1969 at the Tonantzintla, Asiago, Byurakan, Budapest, and Alma-Ata Observatories (Table 1). To Tonant zintla Observatory the total number of flare stars discovered in the region of the Pleiades has now reached 146. One of them (H II 2411) belongs to the Hyades, Of the remaining 145 stars, 123 have shown one flare, 16 have shown two flares, and 6 more than two flares.
A special analysis of flare stars has been carried out and it was found that the total number of flare stars in the Pleiades should be greater than 600. The distribution of flare stars can be satisfactorily represented

flare stars in the Pleiades

- Pleiades is a star cluster about 444 light years away from planet Earth
- consists of 100s of stars only some are visible

Table: Frequency distribution of the count (per star) of flares (Ambartsumyan *et al.* 1970

count of flares	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_9	n
frequency		123	16	2	1	1	1	1	145

three case studies

- dice snakes in Graz
- DD in Britain
- flare stars in the Pleiades

what do they have in common?

- do not know the size
- many counts of ones (singletons)

predicting f_0

- find model for $P(X = x) = p_x = p_x(\theta)$
- find estimate $\hat{\theta}$ for θ leading to

$$\hat{p}_x = p_x(\hat{\theta})$$

- then use Horvitz-Thompson estimator for estimating f_0

$$\hat{f}_0 = n\frac{\hat{p}_0(\hat{\theta})}{1 - \hat{p}_0(\hat{\theta})}$$

as $E\left(n\frac{\hat{p}_0(\hat{\theta})}{1-\hat{p}_0(\hat{\theta})}\right) \rightarrow Np_0$ (if model is correct)

power series as model class

- consider

$$p_x(\theta) = a_x\theta^x/\eta(\theta), \tag{1}$$

where a_x are known coefficients and $\eta(\theta)$ is the normalizing constant

  - $a_x = 1/x!$ Poisson
  - $a_x = 1$ geometric
  - $a_x = \binom{T}{x}$ binomial

- note the property

$$\frac{a_x}{a_{x+1}}\frac{p_{x+1}}{p_x} = \frac{a_{x-1}}{a_x}\frac{p_x}{p_{x-1}} = \theta \tag{2}$$

power series as model class

- specifically for $x = 1$:

$$\frac{a_x}{a_{x+1}} \frac{p_{x+1}}{p_x} = \frac{a_{x-1}}{a_x} \frac{p_x}{p_{x-1}} = \theta$$

$$\frac{a_1}{a_2} \frac{p_2}{p_1} = \frac{a_0}{a_1} \frac{p_1}{p_0} = \theta$$

- so that Chao's estimator (SJoS 84, Biometrics 87, 89) arises:
-

$$p_0 = \frac{a_2 a_0}{a_1^2} \frac{p_1^2}{p_2} \to \hat{f}_0 = \frac{a_2 a_0}{a_1^2} \frac{f_1^2}{f_2}$$

Chao's estimator copes with heterogeneity

- model:

$$m_x(\theta) = \int_\theta a_x \theta^x / \eta(\theta) f(\theta) \; d\theta$$

  for some arbitrary heterogeneity distribution $f(\theta)$ as mixing distribution

- by means of the Cauchy-Schwarz inequality

$$E(XY)^2 \le E(X^2)E(Y^2)$$

- now with $X = \theta/\sqrt{\eta(\theta)}$ and $Y = 1/\sqrt{\eta(\theta)}$ we have

$$E[\theta/\eta(\theta)]^2 \le E[\theta^2/\eta(\theta)]E[1/\eta(\theta)]$$

- equivalently

$$m_1^2/a_1^2 \le m_2/a_2 \times m_0/a_0$$

- or

$$m_0 \ge \frac{a_2 a_0}{a_1^2}\frac{m_1^2}{m_2} \rightarrow \hat{f}_0 = \frac{a_2 a_0}{a_1^2}\frac{f_1^2}{f_2}$$

  and $\hat{f}_0$ a lower bound estimator

some illustrations

Chao's lower bound estimator: $\hat{f}_0 = \frac{a_2 a_0}{a_1^2} \frac{f_1^2}{f_2}$

- Poisson: $\hat{f}_0 = \frac{1}{2} \frac{f_1^2}{f_2}$
- binomial: $\hat{f}_0 = \frac{T(T-1)/2}{T^2} \frac{f_1^2}{f_2} = \frac{T-1}{2T} \frac{f_1^2}{f_2}$
- geometric: $\hat{f}_0 = \frac{f_1^2}{f_2}$

in the case studies

Chao's lower bound estimator: $\hat{f}_0 = \frac{1}{2} \frac{f_1^2}{f_2}$

- dice snakes: $\hat{f}_0 = \frac{1}{2} \frac{59^2}{8} = 218$
- drink-driving: $\hat{f}_0 = \frac{1}{2} \frac{219{,}008^2}{8069} = 2,972,515$
- flare stars: $\hat{f}_0 = \frac{1}{2} \frac{123^2}{16} = 473$

Finally, from Eq. (6) we have the following distribution law:

$$P_k = \frac{t^k}{k!}\int_0^\infty f_1(\nu')\, e^{-\nu' t}\nu'^k\, d\nu'. \tag{8}$$

This distribution is analogous to Eq. (2) with the difference that, in the present case, the expression for p_k includes the new frequency distribution function $f_1(\nu')$.

§6. **The effect of flare-frequency dispersion on the estimated total number of flare stars.** The distribution law given by Eq. (2) leads to a very important inequality for the mathematical expectation of the number of still undiscovered flare stars. Before we proceed to its derivation, let us consider an imaginary case where all the stars have the same frequency ν, and all the flares are equally accessible to observation with a given telescope. The mathematical expectation of the number of observed flares in a time t is then

$$\bar{n}_k = Ne^{-\nu t}\frac{(\nu t)^k}{k!}. \tag{9}$$

By writing this equation separately for k = 0, 1, 2, we immediately have

$$2\,\bar{n}_0\bar{n}_2 = \bar{n}_1^2, \tag{10}$$

and hence

$$\bar{n} = \frac{\bar{n}_1^2}{\ }$$

Let us now consider the general case of Eq. (2), when there are flares with different mean flare frequencies. We have already shown that this formula can be used even when the telescope does not record all the flares.

Table 4

k	n_k	n_k (obs.)
0	474	?
1	123	123
2	16	16
3	2	2
4	1	1
5	1	1
6	1	1
7	1	0
8	0.5	0
9	0.3	1

According to the Schwartz inequality, we have

$$\left(\int fg\,d\nu\right)^2 \leqslant \int f^2 d\nu \cdot \int g^2 d\nu. \tag{13}$$

It will be convenient to substitute

$$f = \nu\sqrt{e^{-\nu t}f(\nu)},$$
$$g = \sqrt{e^{-\nu t}f(\nu)}.$$

and hence

$$\bar{n}_0 = \frac{\bar{n}_1^2}{2\bar{n}_2}. \tag{11}$$

If we replace, approximately, the mathematical expectations with the numbers of stars which have flared once and twice, we can deduce from this formula the value of $\bar{n}_0$, i.e., the number of flare stars whose flares have not been detected. Thus, of the 145 flare stars known at present in the Pleiades, 123 have shown one flare and 16 have shown two flares. Substituting these numbers for $\bar{n}_1$ and $\bar{n}_2$, we obtain

$$\bar{n}_0 = 473,$$

and the total number of flare stars (recorded and unrecorded) should be close to N = 600, which is valid if Eq. (9) is valid.

This large number of flare stars reinforces the earlier conclusion in [16] that all, or practically all, stars in the Pleiades which lie below a certain absolute magnitude are flare stars.

Next, substituting in turn k = 0 and k = 1 in Eq. (9) and dividing the resulting expression by $\bar{n}_0$, we obtain

$$\nu t = \frac{\bar{n}_1}{\bar{n}_0}. \tag{12}$$

For the aggregate in the Pleiades the above data taken in conjunction with Eq. (12) yield $\nu t \cong 0.26$.

If we assume that the total effective time of observations was approximately 750 hr (we do not know precisely the effective observational time in [20] and have assumed that it was ~100 hr), we find that the mean frequency of flares in the Pleiades is 0.00035 hr^{-1}

$$g = \sqrt{e^{-\nu t} f(\nu)}.$$

Instead of inequality (13) we then have

$$\left(\int_0^\infty \nu e^{-\nu t} f(\nu)\, d\nu\right)^2 \leqslant \int_0^\infty \nu^2 e^{-\nu t} f(\nu) \cdot \int_0^\infty e^{-\nu t} f(\nu)\, d\nu. \tag{14}$$

Multiplying both sides of this inequality by t^2, we obtain

$$p_1^2 \leqslant 2 p_0 p_2. \tag{15}$$

If we now multiply the last inequality by N^2, we obtain

$$\bar{n}_0 \geqslant \frac{\bar{n}_1^2}{2\bar{n}_2}. \tag{16}$$

Thus, by using the formula given by Eq. (11), which is valid for equal mean flare frequencies, we obtain in the general case the lower limit for the mathematical expectation of the number of stars for which the flares have not as yet been recorded. To obtain an idea about the change in $\bar{n}_0$ resulting from the presence of dispersion among the mean frequencies from the above lower limit, let us consider another imaginary case for which

$$f(\nu) = \frac{1}{b} e^{-b\nu}. \tag{17}$$

From Eq. (2) we can readily show that, in this case

$$\bar{n}_0 = \frac{\bar{n}_1^2}{\bar{n}_2}, \tag{18}$$

i.e., this value is larger by a factor of two than pre-

problems with Chao's estimator (or what is wrong when everything looks right)

- $\hat{f}_0 = \frac{a_2 a_0}{a_1^2} \frac{f_1^2}{f_2}$ builds heavily on f_1
- hence need to assume that f_1 is correct
- but what will happen if f_1 *overestimates* relative to m_x?

$$m'_x = \begin{cases} (1-\alpha) + \alpha m_x, & x = 1 \\ \alpha m_x, & x \neq 1 \end{cases}$$

- the lower bound estimator will loose its property and potentially largely *overestimate*
- fundamental difference to *zero-inflation* models

a synthetic example

- 500 counts sampled from $Po(1)$
- 500 extra-counts of 1 so that $N = 1,000$
- $\hat{f}_0 = \frac{1}{2}\frac{f_1^2}{f_2} = 2,434$

Table: one-inflated Poisson data

f_0	f_1	f_2	f_3	f_{4+}	n
186	690	95	32	7	814

need for modelling

- hence will focus on one-inflation modelling

$$p'_x(\theta) = \begin{cases} (1-\alpha) + \alpha p_x(\theta), & x = 1 \\ \alpha p_x(\theta), & x \neq 1 \end{cases}$$

  where $p_x(\theta) = b_x(\theta)/(1 - b_0(\theta))$ is a zero-truncated base distribution
- for example

$$p'_x = \begin{cases} (1-\alpha) + \alpha p_x(\theta), & x = 1 \\ \alpha p_x(\theta), & x \neq 1 \end{cases}$$

$$p_x(\theta) = \frac{\exp(-\theta)}{1 - \exp(-\theta)} \theta^x / x!$$

a general result

- consider an *arbitrary* inflation point x_1 and an *arbitrary* count density $p_x(\theta)$
- the associated x_1-inflation is

$$p'_x(\theta) = \begin{cases} (1-\alpha) + \alpha p_x(\theta), & x = x_1 \\ \alpha p_x(\theta), & x \neq x_1 \end{cases}$$

  where $\alpha \in [0, 1]$
- the associated *likelihood* is

$$L = [(1-\alpha) + \alpha p_1(\theta)]^{f_1} \prod_{x \neq x_1} [\alpha p_x(\theta)]^{f_x}$$

  where $p_1(\theta) = p_{x_1}(\theta)$ and $f_1 = f_{x_1}$

a general result

- the associated *log-likelihood* is

$$\log L = f_1 \log[1 - \alpha + \alpha p_1(\theta)] + \sum_{x \neq x_1} f_x \log p_x(\theta) + (n - f_1) \log \alpha$$

  where n is the sample size

- the *profile*-log-likelihood in θ is

$$\log PL(\theta) = \sup_{\alpha} \log L(\theta, \alpha)$$

- and

$$\hat{\alpha} = \frac{1 - f_1/n}{1 - p_1(\theta)}$$

  maximizes $\log L(\alpha, \theta)$ for fixed θ

a general result

- so that

$$1-\hat{\alpha}+\hat{\alpha}p_1(\theta)=1-\frac{1-f_1/n}{1-p_1(\theta)}+\frac{1-f_1/n}{1-p_1(\theta)}p_1(\theta)=f_1/n$$

- the *profile log-likelihood* is

$$\log L(\theta,\hat{\alpha})=f_1\log[1-\hat{\alpha}+\hat{\alpha}p_1(\theta)]+\sum_{x\neq x_1}f_x\log p_x(\theta)+(n-f_1)\log\hat{\alpha}$$

$$=f_1\log(f_1/n)+(n-f_1)\log\frac{1-f_1/n}{1-p_1(\theta)}+\sum_{x\neq x_1}f_x\log p_x(\theta)$$

$$=f_1\log(f_1/n)+(n-f_1)\log(1-f_1/n)+\sum_{x\neq x_1}f_x\log\left(\frac{p_x(\theta)}{1-p_1(\theta)}\right)$$

as $\sum_{x\neq x_1}f_x=n-f_1$

a general result

- fitted x_1-inflated log-likelihood

$$f_1 \log(f_1/n) + (n - f_1)\log(1 - f_1/n) + \sum_{x \neq x_1} f_x \log\left(\frac{p_x(\theta)}{1 - p_1(\theta)}\right)$$

- equals the x_1-truncated log-likelihood

$$\sum_{x \neq x_1} f_x \log\left(\frac{p_x(\theta)}{1 - p_1(\theta)}\right)$$

plus

$$f_1 \log(f_1/n) + (n - f_1)\log(1 - f_1/n)$$

which is independent of θ

x_1-inflation diagnostics

- fit the x_1-truncated likelihood

$$\log T_1 = \sum_{x \neq x_1} f_x \log\left(\frac{p_x(\hat{\theta})}{1 - p_1(\hat{\theta})}\right)$$

- get the fitted x_1-inflated log-likelihood

$$\log L_1 = f_1 \log(f_1/n) + (n - f_1) \log(1 - f_1/n) + \log T_1$$

- form the likelihood ratio statistic $\lambda = 2 \log \frac{L_1}{L_0}$ where

$$\log L_0 = \sum_x f_x \log p_x(\theta)$$

  is the non-inflated log-likelihood using all data

- note that $\lambda \sim 0.5\chi_0^2 + 0.5\chi_1^2$ because of the boundary problem

application to zero-truncated distributions

- for an arbitrary count density $b_x(\theta)$, the *base density*, consider the associated zero-truncated count density

$$p_x(\theta) = b_x(\theta)/(1 - b_0(\theta)), x = 0, 1, \cdots$$

- then the one-inflated density is

$$p'_x = \begin{cases} (1-\alpha) + \alpha p_x(\theta), & x = 1 \\ \alpha p_x(\theta), & x \neq 1 \end{cases}$$

- according to the previous result we can *restrict inference* on the *zero*-*one*-truncated density

$$p_x^{++}(\theta) = b_x(\theta)/[1 - b_0(\theta) - b_1(\theta)]$$

for $x = 2, 3, \cdots$

finding the base distributions in the case studies

Table: comparative distributional analysis for the three case studies based on the 0-1 truncated likelihood

case study	distribution	0-1 log-L	AIC	BIC
dice snakes	Poisson	-11.41	24.82	25.22
	geometric	-11.04	24.07	24.47
	NB	-11.04	26.07	26.87
	NB dispersion: 0.9999 (0.9995 – 1.0005)			
DD	Poisson	-2127.90	4257.80	4264.85
	geometric	-2116.79	4235.58	4242.64
	NB	-2116.79	4237.58	4251.69
	NB dispersion: 0.9999 (0.9997 – 1.0001)			
flare stars	Poisson	-31.50	65.00	66.09
	geometric	-27.53	57.05	58.14
	NB	-26.61	57.23	59.41
	log - NB dispersion: 11.16 (-85.66 – 107.98)			

negative-binomial

for completeness the density function of the negative-binomial with $\theta = (\mu, \delta)$:

$$b(x, \theta) = \frac{\Gamma(x + \frac{1}{\delta})}{\Gamma(x+1)\Gamma(\frac{1}{\delta})} \left(\frac{1/\delta}{\mu + 1/\delta}\right)^{1/\delta} \left(\frac{\mu}{\mu + 1/\delta}\right)^{x}$$

for $x = 0, 1, 2, \cdots$ using the mean parameterization, so that

- $E(X) = \mu$ and $Var(X) = (1 + \delta\mu)\mu$
- $\mu > 0$ is the mean and
- $\delta > 0$ is the dispersion parameter
- geometric: $\delta = 1$ and
- Poisson: $\delta \to 0$

is there evidence of one-inflation?

Table: zero-truncated-one-inflated and zero-truncated geometric log-likelihood with likelihood ratio statistic

case study	0-trunc.-1-infl. Log L	0-trunc. Log-L	$2 \log \lambda$ (P-val)
dice snakes	-41.48	-42.97	2.98 (0.042)
DD	$-38,626.33$	$-38,685.17$	117.70 (0.000)
flare stars	-89.25	-96.58	14.66 (0.000)

modified Horvitz-Thompson estimation

- conventional Horvitz-Thompson estimator

$$\hat{f}_0 = n \frac{b_0(\theta)}{1 - b_0(\theta)}$$

  has property $E(\hat{f}_0) = Np_0(\theta)$ (in the case of no inflation)

- needs to be *modified* here as n contains the one-inflated part

$$\hat{f}_0 = (n - f_1) \frac{b_0(\theta)}{1 - b_0(\theta) - b_1(\theta)}$$

- has (again) property $E(\hat{f}_0) = Np_0(\theta)$ and, ultimately, the modified Horvitz-Thompson estimator

$$\hat{N} = n + (n - f_1) \frac{b_0(\theta)}{1 - b_0(\theta) - b_1(\theta)}$$

  with $E(\hat{N}) = N$

population size estimates for the three case studies

- as θ is unknown, a plug-in estimate is used based on the 0-1-truncated geometric as evidenced in the previous analysis
- the conventional HTE would use the 0-truncated geometric
- the modified HTE would use the 0-1-truncated geometric

Table: zero-truncated-one-inflated and zero-truncated geometric log-likelihood with likelihood ratio statistic

case study	n	$\hat{N}$ (mHTE)	$\hat{N}$ (cHTE)
dice snakes	70	127	358
DD	$227,578$	$2,336,517$	$5,897,792$
flare stars	145	205	671

uncertainty assessment

the conventional, nonparametric bootstrap is as follows:

1. Draw a sample of size N from the observed distribution defined by the probabilities $\frac{f_0}{N}, \frac{f_1}{N}, \frac{f_2}{N}, \cdots, \frac{f_m}{N}$.
2. Derive $\hat{\theta}$ and $\hat{N}$ for the bootstrap sample in 1).
3. Repeat step 1) and 2) B times, leading to a sample of estimates $N^{(1)}, \cdots, N^{(B)}$
4. Calculate the bootstrap standard error as

$$SE^* = \frac{1}{B}\sum_{b=1}^{B}(N^{(b)} - \bar{N}^*)^2,$$

where $\bar{N}^* = \frac{1}{B}\sum_{b=1}^{B} N^{(b)}$.

problem : neither f_0 nor N are *known*

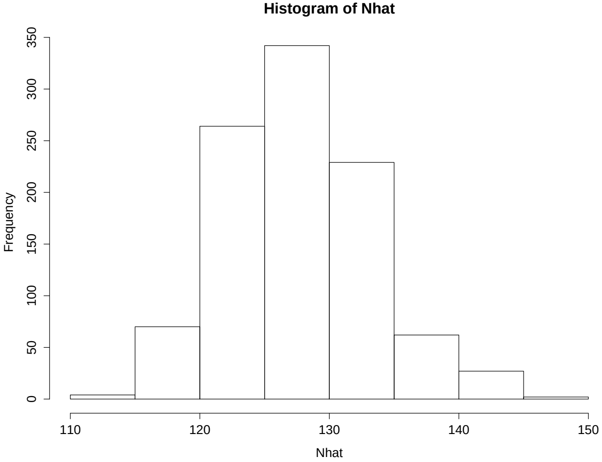
uncertainty assessment

we suggest a *semi-parametric* bootstrap as follows:

1. Draw a sample of size $||\hat{N}||$ from the observed distribution defined by the probabilities $\frac{\hat{f}_0}{\hat{N}}, \frac{f_1}{\hat{N}}, \frac{f_2}{\hat{N}}, \cdots, \frac{f_m}{\hat{N}}$. (Here $||x||$ denotes the rounding of x to the nearest integer.)
2. Derive $\hat{\theta}$ and $\hat{N}$ for the bootstrap sample in 1).
3. Repeat step 1) and 2) B times, leading to a sample of estimates $N^{(1)}, \cdots, N^{(B)}$
4. Calculate the bootstrap standard error as

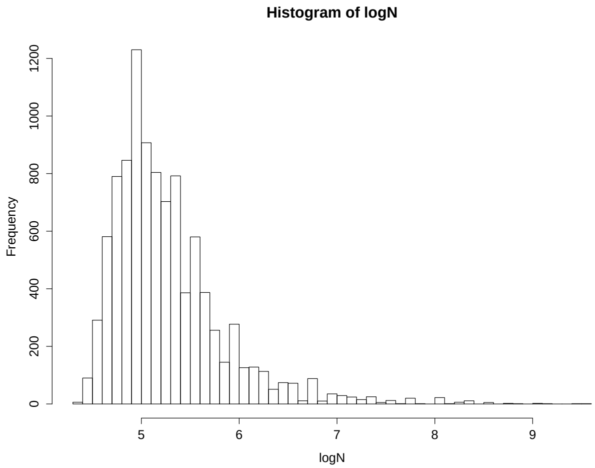
$$SE^* = \frac{1}{B} \sum_{b=1}^{B} (N^{(b)} - \bar{N}^*)^2,$$

where $\bar{N}^* = \frac{1}{B} \sum_{b=1}^{B} N^{(b)}$.



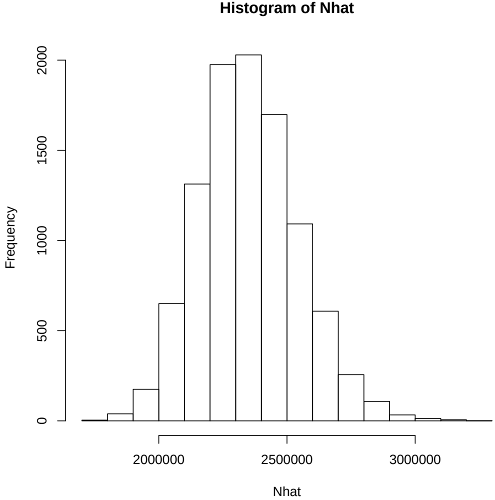
95% percentile bootstrap interval
for dice snake example

```
    2.5%       50%    97.5%
117.8974 127.0198 140.3741
```



95% robust bootstrap interval
for dice snake example

robust CI: 96.25 -- 299.78



95% percentile bootstrap interval for DD example

2.5%	50%	97.5%	
2,008,895	2,333,519	2,756,244	
8.8%	9.7%	12.1%	observed drink-driving

relevant works

- Böhning and van der Heijden (2019, AoAS)
- Böhning, Kaskasamkul, and van der Heijden (2019, Metrika)
- Anan, Böhning, and Maruotti (2017, JSCS)
- Böhning *et al.* (2013, 2016, Biometrics)
- Böhning (2016, Statistical Science)
- Böhning, Bunge, and van der Heijden (2018, Chapman&Hall/CRC)