Estimating the Complete Sample Size from an Incomplete Poisson Sample

May 23, 2021

Outline

1. Our Problem
   - Estimate the Complete Sample Size N
   - Background Information
2. Estimators of N
   - Three kinds of MLE Based on Different Likelihoods
3. Asymptotic Properties of the Estimators
   - Asymptotic Bias and Variance
4. A Monte Carlo Simulation
   - Simulation Details
   - Simulation Results and Analysis

1. Our Problem

- Target

  Given a sample of n observations from $Poi(\lambda)$ where all observations of which $X = 0$ are missing, estimate the complete sample size N.

- Background

  Dahiya and Gross proposed the conditional maximum likelihood estimator of N in 1973 and discussed its asymptotic properties. The new estimators offer some modification.

2. Estimators of N

- Conditional ML estimator: $\hat{N}_c$, $\hat{\lambda}_c$
  - Let $p = 1 - e^{-\lambda}$, $q = 1 - p$ and

$$L_1(\lambda, N) = \binom{N}{n} p^n q^{N-n}, \quad L_2(\lambda) = \left(\frac{q}{p}\right)^n \prod_{x=1}^{R} \left(\frac{\lambda^x}{x!}\right)^{n_x}$$

  - $\hat{N}_c$ and $\hat{\lambda}_c$ maximize L_1 and L_2 simultaneously and satisfy

$$\bar{X}^* = \hat{\lambda}_c/(1 - e^{-\hat{\lambda}_c}), \quad \hat{N}_c = \left[\sum_{x=1}^{R} xn_x/\hat{\lambda}_c\right] = [n/\hat{p}]$$

where $\bar{X}^* = S/n$ if $n \geq 1$ else 0, $\hat{p} = 1 - e^{-\hat{\lambda}_c}$, $S = \sum_{x \geq 1} xn_x$.

  - Define $G(V) = -V\ln(1 - n/V)$, $\hat{V}_c = n/\hat{p}$ is the unique solution to $S = G(V)$, then $\hat{N}_c = [\hat{V}_c]$.

2. Estimators of N

- Unconditional ML estimator: $\hat{N}_u$, $\hat{\lambda}_u$
  - Logarithmic unconditional likelihood:

$$\mathcal{L} = \ln L = \ln\binom{N}{n} - \lambda N + S \ln \lambda$$

For fixed N, $\hat{\lambda}_u = S/N$. Define

$$\mathcal{L}_N^* = \ln\binom{N}{n} - S + S(\ln S - \ln N)$$

$$D(V) = \ln(1 - n/V)/\ln(1 - 1/V), V > n$$

$\hat{N}_u = [\hat{V}_u]$ with $\hat{V}_u$ being the unique solution to $S = D(V)$.

- Limitations of $\hat{N}_c$ and $\hat{N}_u$
  - If S nears n, $\hat{N}_c$ and $\hat{N}_u$ are large.
  - If $S = n$, $\hat{N}_c$ and $\hat{N}_u$ are not defined.

2. Estimators of N

- Modified ML estimator: $\hat{N}_m$, $\hat{\lambda}_m$
  - To modify the unconditional likelihood L, multiply it by

$$b(\lambda) = r^{\rho+1}\lambda^{\rho}e^{-r\lambda}/\Gamma(\rho+1), \quad P(N=j) = 1/N_0, \quad j = 1, \dots, N_0$$

    where N_0 is a bound on N. Then the modified likelihood is given by

$$h(\lambda, N) = C(r, \rho, N_0) \cdot \binom{N}{n} e^{-\lambda(N+r)}\lambda^{S+\rho}$$

  - Finally we obtain $\hat{\lambda}_m = \dfrac{S+\rho}{N+r}$ and $\hat{N}_m = [\hat{V}_m]$, where $\hat{V}_m$ is the unique solution of the equation

$$S + \rho = D(V, r), \quad D(V, r) = \ln(1 - n/V)/\ln(1 - 1/(V + r)).$$

3. Asymptotic Properties of the Estimators

Let $\hat{V}_i = N + a_i\sqrt{N} + b_i + O_p(1/\sqrt{N}), i = c, u, m$

- Asymptotic properties of $\hat{V}_i, i = c, u, m$

  $$(\hat{V}_i - N)/\sqrt{N} \,\dot{\sim}\, N(0, \sigma^2), \quad \sigma^2 = q/(p - \lambda q)$$

- Theoretical asymptotic bias of $\hat{V}_i$
  - $E(b_c) = \lambda pq/2(p - \lambda q)^2$
  - $E(b_u) = (\lambda q)^2/2(p - \lambda q)^2$
  - $E(b_m) = \lambda q\{\lambda q + 2(p - \lambda q)(r - \rho/\lambda)\}/2(p - \lambda q)^2$

- Best choice of (ρ, r) for $\hat{V}_m$

  For small λ, $E(b_m) = \frac{2}{\lambda^2}(1 - \rho) - \frac{2}{\lambda}(\frac{2}{3} - r - \frac{\rho}{3}) + \frac{1}{3} - \frac{2}{3}r - \frac{\rho}{18}$
  $(\rho, r) = (1, \frac{1}{3})$ is the best choice

3. Asymptotic Properties of the Estimators

Table 1: Summary

	Equation	Asymptotic Bias	Asymptotic Dis.
$\hat{V}_c$	$S = G(V)$	$\lambda pq/2(p-\lambda q)^2$	$(\hat{V}_i - N)/\sqrt{N}$
$\hat{V}_u$	$S = D(V)$	$(\lambda q)^2/2(p-\lambda q)^2$	$\dot{\sim}\ N(0, \sigma^2)$,
$\hat{V}_m$	$S + \rho = D(V, r)$	$\dfrac{\lambda q\{\lambda q + 2(p-\lambda q)(r - \rho/\lambda)\}}{2(p-\lambda q)^2}$	$\sigma^2 = q/(p-\lambda q)$

* $p = 1 - e^{-\lambda}$, $q = 1 - p$

4. A Monte Carlo Simulation

- Simulation Details

1. Generate random Poisson samples for each pair (N, λ),

   $N = 25$, 50, 100; $\lambda =$0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 3.00, 5.00.

2. Compute $\hat{V}_u, \hat{V}_c, \hat{V}_m$ and obtain $\hat{N}_u, \hat{N}_c, \hat{N}_m$

3. Repeat 1000 times for each pair (N, λ), compute the simulated

   average bias $\bar{B}$ and its MSE/N

$$\bar{B} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{N}_i - N)$$

$$MSE/N = \frac{1}{N \cdot 1000} \sum_{i=1}^{1000} (\hat{N}_i - N)^2$$

4. A Monte Carlo Simulation

- Simulation Results

Table 2: Estimated Biases and Mean Squared Errors, $\lambda \geq 0.75$

	Bias			*MSE/N*			
λ	$\hat{N}_u$	$\hat{N}_c$	$\hat{N}_m$	$\hat{N}_u$	$\hat{N}_c$	$\hat{N}_m$	σ^2
				$N = 50$			
0.75	2.081	3.208	-0.649	4.27670	4.64956	2.96614	2.72480
1.00	0.642	1.392	-0.496	1.87656	2.00512	1.61084	1.39221
1.25	-0.215	0.320	-0.762	0.93430	0.96484	0.88772	0.80623
1.50	-0.436	0.001	-0.688	0.53876	0.54810	0.52088	0.50462
1.75	-0.584	-0.278	-0.711	0.33368	0.33140	0.32726	0.33282
2.00	-0.361	-0.093	-0.436	0.23522	0.23686	0.23080	0.22784
3.00	-0.543	-0.459	-0.545	0.07254	0.07162	0.07226	0.06217
5.00	-0.313	-0.313	-0.313	0.00838	0.00838	0.00838	0.00702

See full table

4. A Monte Carlo Simulation

- Simulation Results

Table 3: Estimated Bias and Mean Squared Error, $\lambda < 0.\ 75$

	Bias		MSE/N	
	N			
λ	50	100	50	100
0.1	-39.421	-73.042	31.69	54.89
0.2	-28.292	-38.746	17.97	20.45
0.3	-17.081	-13.289	8.97	10.70
0.4	-7.553	-1.000	5.78	10.65
0.5	-3.156	0.181	5.41	8.41
0.6	-0.842	-0.085	4.89	5.58

4. A Monte Carlo Simulation

- Result Analysis

  - For $\lambda \geq 0.75$ and fixed N, the three estimators perform better as λ increases.
  - For $\lambda < 0.75$, the simulated bias of $\hat{N}_m$ is much smaller than the asymptotic bias of the other two estimators, but it is still worse than its own asymptotic bias.
  - In general, the overall trends are consistent with the corresponding results in the selected paper.