Basu's elephants

Justin Sheen

July 16, 2024

1 Horvitz-Thompson estimator

The following is adapted from the Wikipedia page on the Horvitz-Thompson estimator.

Let Y_i for $i = 1, 2, \ldots, n$ be independent samples across $N \geq n$ distinct strata with mean across all strata μ. Suppose that π_i is the inclusion probability that a randomly sampled individual belongs to the ith stratum. The Horvitz-Thompson estimator of the total, Y, is given by:

$$\widehat{Y}_{HT} = \sum_{i=1}^{n} \pi_i^{-1} Y_i$$

2 Basu's elephants

The following is a funny example of when the Horvitz-Thompson estimator does not behave pleasantly, adapted from Ghosh and Haziza (2017).[1]

A circus owner needs a rough estimate of the total weight of his 50 adult elephants. Because weighing elephants is difficult, the owners wants to estimate the total weight by weighing just one elephant.

The owner looks back at his records and finds that previously the average-sized elephant was an elephant named Sambo. After checking with a trainer, the owner learns that Sambo is still the average-sized elephant of the herd. The owner plans to weigh Sambo, let y be the weight of Sambo, and multiply it by 50, so that the estimate of the total weight of the elephants is $50y$.

The circus statistician is not impressed with the plan.

"How can you get an unbiased estimate of the total weight of 50 elephants with this plan?"

"Hm... I supposed you're right to raise these concerns," replied the circus owner.

Together the circus owner and circus statistician work out a new plan. In their new plan, they give Sambo a selection probability of 99/100 and divide the remaining 1/100 selection probability of 1/4900 equally to the other 49 elephants. They put their plan in motion and randomly select an elephant from the 50. Naturally, Sambo is selected.

"How are you going to estimate Y?" asked the circus statistician.

"The estimate ought to be $50y$," replied the circus owner.

"Oh no! That can't possibly be right," said the circus statistician. "I recently read an article in the Annals of Mathematical Statistics where it is proven that the Horvitz-Thompson estimator is the unique hyperadmissible estimator in the class of all generalized polynomial unbiased estimators."

"All right," said the circus owner, "then what is the Horvitz-Thompson estimate in this case?"

"Since the selection probability for Sambo in our plan was 99/100," says the circus statistician, "the proper estimate is $100y/99$."

"I see," says the circus owner, clearly pleased at the reasonableness of this estimate, "and how would you have estimated the total if our sampling plan made us select the biggest elephant, Jumbo?"

"According to the Horvitz-Thompson estimation method," says the circus statistician, "the proper estimate would be $4900y$," where y is Jumbo's weight."

The cirus statistician is promptly fired.

References

[1] Malay Ghosh and David Haziza. Revisiting basu's circus example: Another look at the horvitz-thompson estimator. *Calcutta Statistical Association Bulletin*, 68(1-2):33–37, 2016.