Math Matters

Apply It.

The math behind . . . Estimating Population Sizes

Technical terms used:

Capture-recapture, mark and recapture, Lincoln Petersen Method, the German Tank Problem, serial number analysis.

Uses and applications:

Estimating wild life population size, such as for the black bear; estimating the prevalence of unreported cases of a disease, such as childhood diabetes and heart attacks; estimating the number of iPhones sold by Apple.

How it works:

Suppose you want to estimate the size of a population but you can't directly count its members. This problem is common for wild life biologists who need to estimate the population of an endangered species. Statistics provide a surprisingly simple yet effective strategy. First capture a sample of the animals, mark them, and then release them back into the wild. Let those animals mix back in with the rest of the population and then capture a second sample of animals.

The fraction of animals marked in the second sample gives us a good idea of how big the total population is. If a large fraction of the second sample are marked then the total population is small since you must have captured a big fraction of the total population in the first sample. On the other hand, if the fraction of marked animals in the second sample is small then the total population is large because you must have captured only a small fraction of the total population in the first sample. The method relies on the assumption that the ratio of marked to unmarked animals in the second sample is very close to the ratio of the number of animals in the first sample to the total population of animals.

Interesting fact:

Serial numbers on objects in a sample from a large batch of objects can be used in a similar way to estimate the total number of objects in the batch. During World War II, Allied Forces used serial numbers on parts from captured German tanks to estimate the rate at which German factories were manufacturing tanks. At the end of the war, the serial number estimates and estimates provided by Allied Force spies were compared to captured production records. Serial number estimates of monthly tank production turned out to be much better estimates. (Table compares values for three months.)

Date	Serial Number Estimate	Allied Spies Estimate	Production Record
June, 1940	169	1000	122
June, 1941	244	1550	271
August, 1942	327	1550	342

References:

[1] The Estimation of Animal Abundance and Related Parameters, G.A.F. Seber (The Blackburn Press, 2002).

[2] An Empirical Approach to Economic Intelligence in World War II, Richard Ruggles and Henry Brodie, *Journal of the American Statistical Association*, Vol. 42, No. 237 (March 1947), pp. 72-91.

Submitted by Eric Chi, Rice University, second place, *Math Matters, Apply It!* contest January 2010.

Society for Industrial and Applied Mathematics
www.siam.org
www.siam.org/careers/matters.php