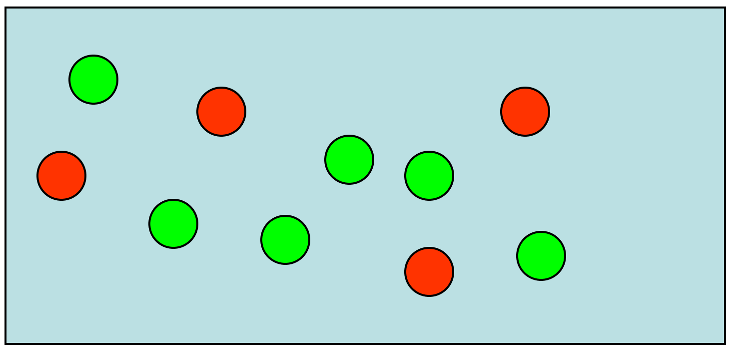
Take a Chance – Part 2

- The Hypergeometric Model
- Capture-Mark-Recapture Method
- Real World Applications
- Useful Practical Techniques
- The Maximum Likelihood Estimate
- Performing Experiments
- Quality of Estimation

The Hypergeometric Model

A box contains 4 red and 6 green marbles. You select 5 marbles. What is the probability that you will find 2 red marbles among the 5 you have selected?



GO TO THE WEB-SITE

http://socr.ucla.edu/htmls/exp/Ball_and_Urn_Experiment.html OR

http://www.ds.unifi.it/VL/VL_EN/applets/UrnExperiment.html

Interactive Fun

Your Notes Here

The Hypergeometric Model

A box contains *N* marbles, *K* red and the remaining *N-K* are green. You select *n* marbles. What is the probability that you will find *k* red marbles among the *n* you have selected?

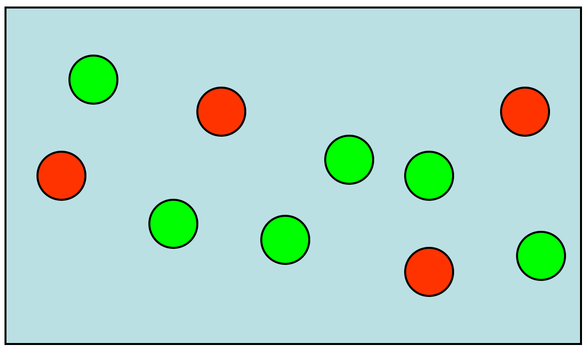
N = total

K = red

N –K = green

n = selected

k = red among *n* selected



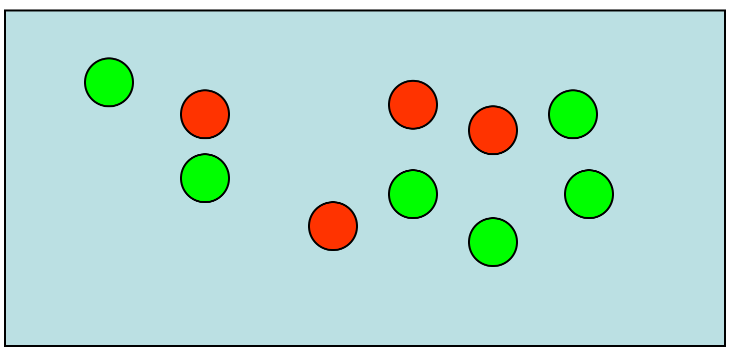
$$P(k\ red\ among\ n\ selected) = \frac{\binom{K}{k} \times \binom{N-K}{n-k}}{\binom{N}{n}}$$

Interactive Fun

Your Notes Here

The Most Likely Case

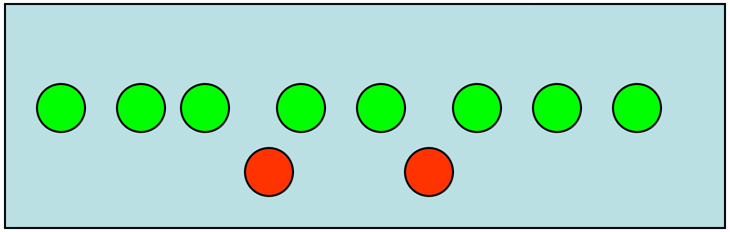
A box contains red and green marbles, a total of 10 marbles. You select 5 marbles and find that two of them are red. What is the most likely number of red marbles in the box? Is 4 red among 10 reasonable?



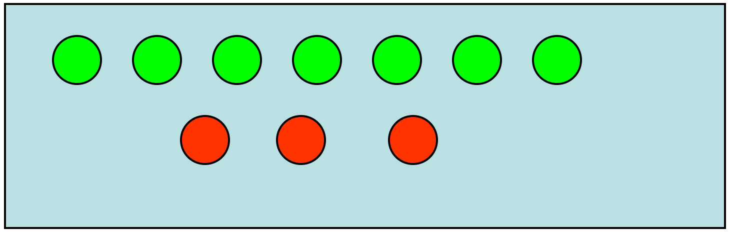
P(2 RED IN A SAMPLE OF 5) = 0.476

The Most Likely Case

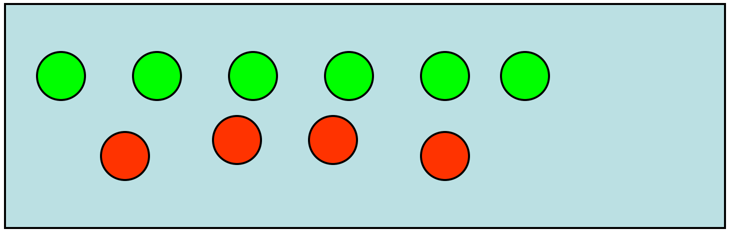
P(2 red in a sample of 5)



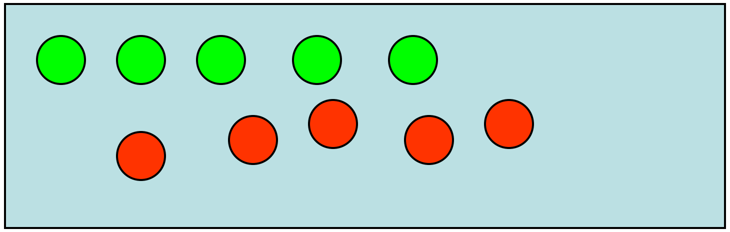
0.222



0.417



0.476



0.4

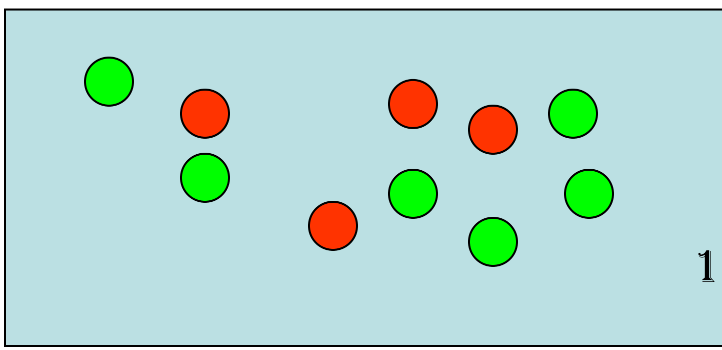
Estimation – Our Best Guess

Since 4 red and 6 green gives the highest probability to the event we have observed, 2 red in a sample of 5, we should assume that there are 4 red and 6 green marbles in the container.

This is called the Maximum Likelihood Estimate

The Most Likely Case

A box contains red and green marbles. The number of red marbles is 4. You select 5 marbles and find that 1 of them is red. What is the most likely total number of marbles in the box? Is 10 reasonable (4 red and 6 green)?

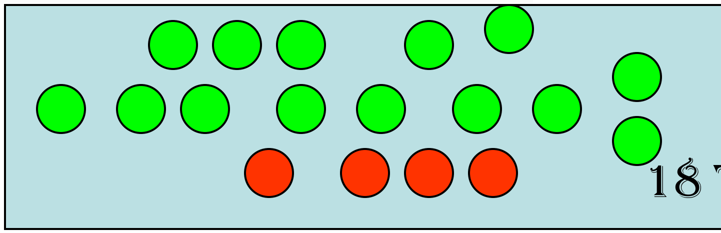


P(1 RED IN A SAMPLE OF 5) = 0.238

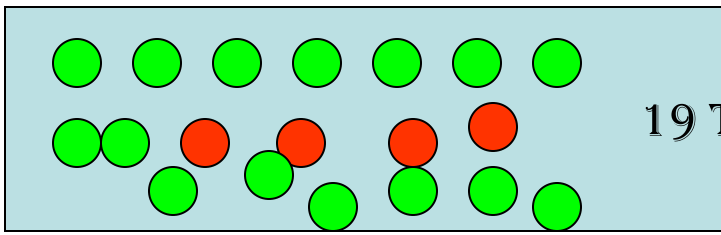
10 TOTAL

The Most Likely Case

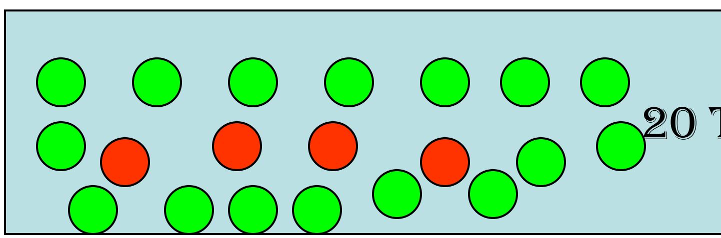
P(3 RED IN A SAMPLE OF 5)



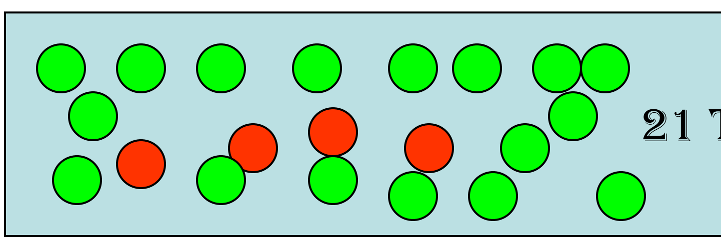
18 TOTAL — 0.4673



19 TOTAL — 0.4696



20 TOTAL — 0.4696



21 TOTAL — 0.4678

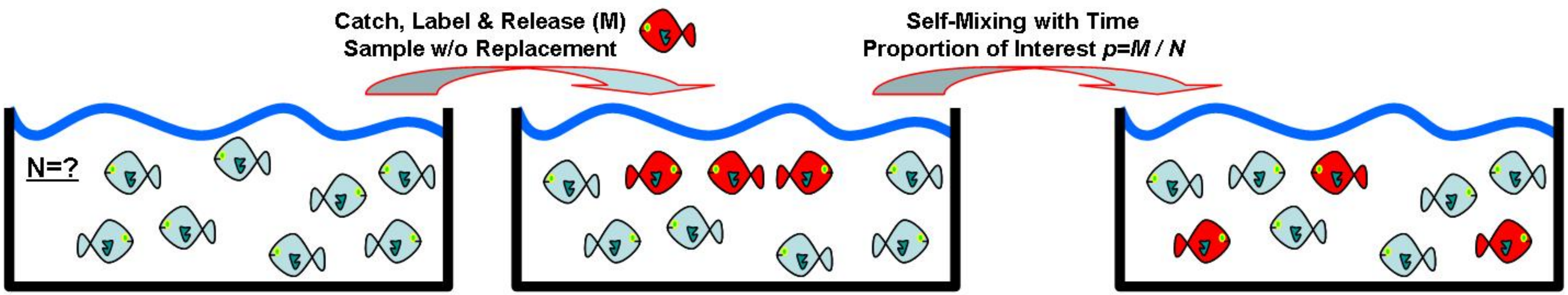
Estimation – Our Best Guess

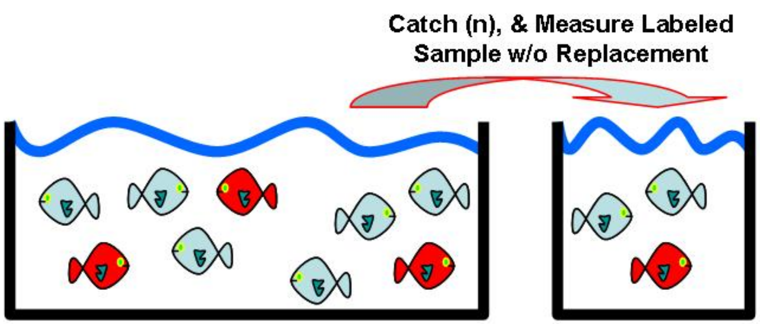
Since 4 red and 15 green or 16 green, with the total of 19 or 20 marbles gives the highest probability to the event we have observed, 1 red in a sample of 5, we should assume that there are 19 or 20 marbles in the container.

This is called the Maximum Likelihood Estimate

Estimation of the Size of Population From Recapture Data

Suppose 1000 fish caught in a lake are tagged and released. After some time allowed for mixing, a new catch of 1000 fish is made and 100 among them are found tagged. Can we estimate the number of fish in the lake? Here is an intuitive solution:





$$\hat{p} = \frac{\textit{Labeled in Sample}}{n} \approx p = \frac{M}{N}$$

$$\Rightarrow \hat{N} \approx \frac{M}{\hat{p}} = \frac{1{,}000}{\left(\frac{100}{1{,}000}\right)} = 1{,}000 \times 10 = 10{,}000$$

Children on the streets of São Paulo

Applications to Social Problems

Capture-recapture to estimate the number of street children in a city in Brazil

Authors: ***R Q Gurgel***, ***J D C da Fonseca***, ***D Neyra-Castañeda***, ***G V Gill***, ***L E Cuevas***

Background: Street children are an increasing problem in Latin America. It is however difficult to estimate the number of children in the street as this is a highly mobile population.

Aims: To estimate the number of street children in Aracaju, northeast Brazil, and describe the characteristics of this population.

Methods: Three independent lists of street children were constructed from a non-governmental organisation and cross-sectional surveys. The number of street children was estimated using the capture-recapture method. The characteristics of the children were recorded during the surveys.

Results: The estimated number of street children was 1456. The estimated number of street children before these surveys was 526, although non-official estimates suggested that there was a much larger population. Most street children are male, maintain contact with their families, and are attending school. Children contribute to the family budget a weekly average of R$21.2 (£4.25, €6.0, US$7.5) for boys and R$17.7 (£3.55, €5.0, US$6.3) for girls.

Conclusion: Street children of Aracaju have similar characteristics to street children from other cities in Brazil. The capture-recapture method could be a useful method to estimate the size of this highly mobile population. The major advantage of the method is its reproducibility, which makes it more acceptable than estimates from interested parties.

Applications to Ecology (California)

Estimating Bobcat Population Sizes and Densities in a Fragmented Urban Landscape Using Noninvasive Capture–Recapture Sampling

Authors: Emily W. Ruell, Seth P. D. Riley, Marlis R. Douglas, John P. Pollinger, and Kevin R. Crooks

Bobcats (*Lynx rufus*) are valuable indicators of connectivity in the highly fragmented landscape of coastal southern California, yet their population sizes and densities are largely unknown. Using noninvasive scat sampling in a capture–recapture framework, we estimated population sizes for 2 similar areas of natural habitat with differing levels of isolation by human development in Santa Monica Mountains National Recreation Area, California. We used scat transects with geographic information system land-use layers and home-range sizes of bobcats to estimate effective sampling area and population densities. Estimates of population size in the study area connected to a much larger habitat area (26–31 individuals) were similar to estimates for the area that was completely surrounded by development (25–28 individuals). Bobcat densities for the 2 study areas also were similar (ranging from 0.25 to 0.42 bobcat/km2) and likely represent recent population declines because of notoedric mange likely interacting with toxicants. These methods proved effective despite particularly low densities of bobcats and may be especially useful when study areas are geographically isolated, reducing the uncertainty in size of the sampling area.

Applications to Ecology (India)

Estimation of tiger densities in India using photographic captures and recaptures
Authors: K. Ullas Karanth, James D. Nichols

The tiger (Panthera tigris) is an endangered big cat whose demographic status is uncertain across its entire distributional range, spanning 13 Asian countries. Because of their large body size and carnivorous diet (Eisenberg 1981), tigers naturally occur at low population densities. Further, wild tiger populations are now being affected by adverse factors such as prey depletion due to overhunting, tiger poaching, and habitat shrinkage and fragmentation. Although recent field surveys, combined with forest cover maps, have generated more accurate distribution maps for Tigers, their utility for assessing the status and viability of tiger populations is limited by the absence of reliable data on population densities.

Most prevailing methods of counting wild tigers appear to fail, because they are unable to dent with three important ecological characteristics of the species: scarcity, extensive range, and secretiveness. Because of their secretive behavior, tigers cannot be visually counted under usual field conditions. Consequently, most methods depend on counting tiger tracks. In India, the official "censuses" of tiger populations are based on the assumptions that each individual tiger can be identified by its unique track shape, and that track prints of every tiger can be simultaneously found and recorded.

The application of radiotelemetry to estimate tiger densities is constrained by the small number of animals that can be tagged simultaneously, uncertainty about numbers of untagged tigers, and the high costs and effort involved.

Based on the fact that tigers are individually identifiable from their stripe patterns, there is potential for estimating their population size using photographic "captures," within the theoretical framework of formal capture-recapture theory.

Capture and Marking Methods

Capture methods are as varied as are animals, and some are listed below. Marking is often a very difficult problem, as you want to mark an individual, but you do not want to alter the individual beyond your ability to detect the mark. Specifically, you do not want the capturing and marking to affect individual mortality or fecundity, either through the trauma of marking or by making the animal more visible to predators (if prey) or to prey (if a predator). In addition, you do not want to alter the marked individuals behavior so that it is either more or less easy to capture a second time. The mark should be permanent, or at least last as long as the study. Some capture and marking techniques are listed below with the animals upon which the technique is used.

- **Animal**
  Capture Technique
  Marking Technique
- **Birds**
  netting, nest invasion
  leg banding, recording color patterns, feather clipping
- **Mammals**
  live trapping, sedation
  hair clipping, collars, dyeing hair, recording color pattern, clipping nails, leg banding, subdermal radio transponders
- **Amphibians**
  Pit traps, hand collection
  toe clipping, paint injection, recording color pattern
- **Insects**
  Sweep Netting, Pit trapping, Aspirating, Pheromone Trapping, Light Trapping
  Fluorescent dust, Tags (glued on), Recording Wing Damage Patterns, Radioactive Tracers
- **Mollusks** (invertebrate animals., like squid)
  Hand collection
  Paint, Glue-on Tags

Maximum Likelihood Estimation

Suppose 1000 fish caught in a lake are tagged and released. After some time allowed for mixing, a new catch of 1000 fish is made and 100 among them are found tagged. It is conceivable that the total number of fish in the lake is 1,900 (1,000 tagged and 900 untagged).

- *N* = total = 1,900
- *K* = red = 1,000 (tagged the first time)
- *N* –*K* = green = 900 (not caught the first time)
- *n* = selected = 1,000 (caught the second time)
- *k* = red among *n* selected = 100 (tagged found in the second catch)

1900 TOTAL

1000 RED, 900 GREEN

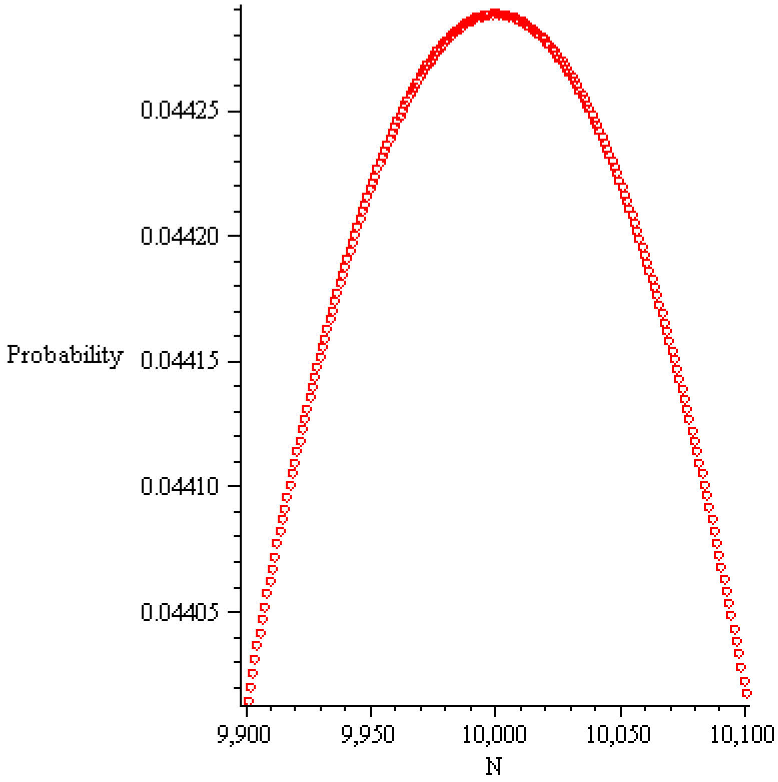
1000 SELECTED

100 RED FOUND

$$P(100\ red\ among\ 1000\ selected) = \frac{\binom{1000}{100} \times \binom{1900-1000}{1000-100}}{\binom{1900}{1000}} = 10^{-430}$$

TOO UNLIKELY TO HAVE OCCURRED!

Other Population Sizes



THE MAXIMUM PROBABILITY OF THE OBSERVED EVENT HAPPENS FOR N = 9,999 AND N=10,000, AND EQUALS 0.04428917172

THE M.L.E. FOR N IS THEREFORE 9,999 OR 10,000.

Performing the Experiment

No. Experiment	First Catch	Second Catch	Tagged	M.L.E. of N
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
Proportion of experiments with the M.L.E. different from N by more than 5%			Average:	

Quality of Estimation

The estimate of N varies from experiment to experiment, however, most of the time it stays close to the true population size (is consistent).

Test

1. Explain the statement: the probability of rain tomorrow is 25%.
2. How many combinations are possible in a lock using four rings, each ring has 10 possible positions.

3. A team of three students is to be selected from this class. In how many ways can this be done?

Test

4. A group of 5 students is to be selected from this class. What is the probability that this group will have 2 girls and three boys?
5. In order to estimate the number of illegal drug users in a certain location, 100 such individual were identified. After some time, a total of 100 illegal drug users were arrested and 20 of them were those already identified. Estimate the population size of illegal drug users in this location.

Your Answers

Your Notes Here

Your Answers

Your Notes Here