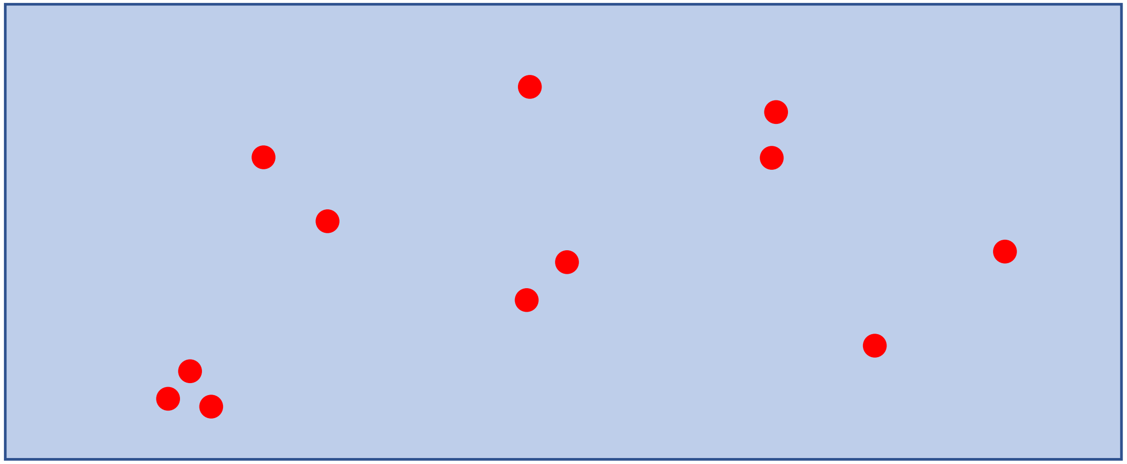
FISH 621
Estimation of Fish Abundance:
2: Sampling Theory cont.

Dr. Curry Cunningham: cjcunningham@alaska.edu

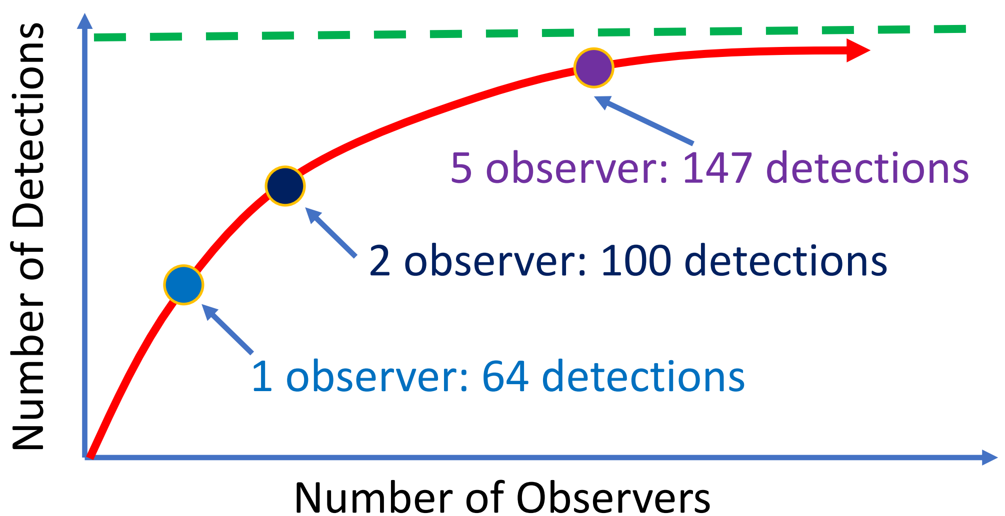
General Synthesis of Abundance Estimation Approaches

- Example
  - Let's consider a field containing flowers
    - And we want to estimate the ***total number*** of flowers
  - We have several general categories of approaches
    - Available to us
  - A single observer sees $n = 64$ unique plants



General Synthesis of Abundance Estimation Approaches

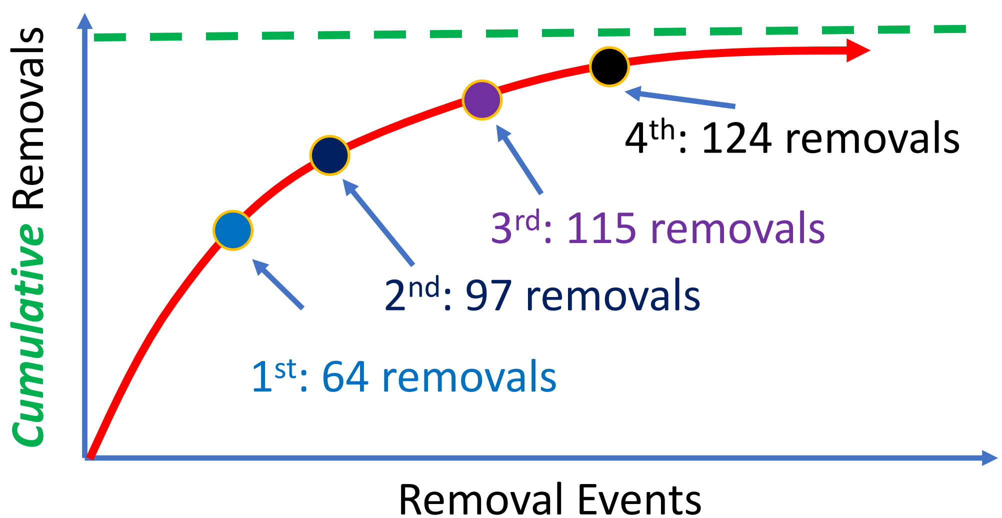
- Approach 1
  - We might assume that if enough observers survey the field
    - They must between them eventually see all the plants
  - Two observers count $n = 100$ plants
  - Five observers count $n = 147$ plants



As number of observers increases, the maximum should be the population size

General Synthesis of Abundance Estimation Approaches

- Approach 2: Depletion, catch-effort, change-in-ratio
  - We might assume that the rate at which the ***number of new detections*** falls off, after ***removal***
    - Tells us something about the ***proportion*** of the population removed
  - Each observer removes all plants they detect
    - If large proportion removed by first observer
    - Then a second observer using equal effort should detect fewer plants
  - Number removed by four observers: 64, 33, 18, 9



Can fit a line to the observed series to estimate when no more plants will be left (*i.e. the point at which additional removal occasions will yield no more removals*)

General Synthesis of Abundance Estimation Approaches

- Approach 3: Mark-recapture
  - Simplest conceptually
    - Yet has spawned the most complicated variety of sub-models
  - If we keep track of which plants were detected by the first observer
    - That are also detected by the second observer
    - We now know something about ***detection probability***
- If the first observer detected $n = 64$ plants
  - And the second observed $n = 67$ plants
    - 34 of which were the ***same*** as the 1st observer (i.e. recaptures)
  - We know the second observer ***detected*** 34/64-ths of plants ***known*** to be present
    - $P(observe) = \sim 53\%$
    - Abundance estimate: $\widehat{N} = \frac{n}{p} = \frac{67}{0.53} = \sim 126$

Estimation

- (Recap) Goal of estimation:
  - Estimate a set of parameters $\Theta = \{\theta_i, i = 1, \ldots, p\}$, by collecting data $X = \{x_i, i = 1, \ldots, n\}$
- There are several methods of estimation
  - All sharing goal of getting as close to the true parameter value as possible
    - Although we will never know what it is!
- One method of estimation comes from decision theory
  - And involves the concept of loss
    - Denoted $\ell(\hat{\theta}|\theta)$

Common Loss Functions

- Practically we want to minimize the difference between our estimated $\hat{\theta}$
  - And ***true*** but unknown value θ for our parameters
- Squared error loss
  - $\ell(\hat{\theta}|\theta) = (\hat{\theta} - \theta)^2$
- Relative squared error loss
  - $\ell(\hat{\theta}|\theta) = (\hat{\theta} - \theta)^2/\theta^2$
- Absolute loss
  - $\ell(\hat{\theta}|\theta) = |\hat{\theta} - \theta|$

Risk Functions

- Closely related to ***loss*** is the concept of ***risk***
  - Where risk is the expected value of our loss function
    - $R(\hat{\theta}|\theta) = E[\ell(\hat{\theta}|\theta)]$
- In practice the we can write the risk
  - Or expected loss with respect to the data, as
    - $R(\hat{\theta}|\theta) = \int_X \ \ell(\hat{\theta}|\theta) f(X|\theta) dX$
  - The integral of the loss function multiplied by the joint probability density function $f(X|\theta)$
    - With respect to the data (this is our likelihood function!)
- The idea here is to use mathematical theory
  - To choose an estimator $\hat{\theta}$
    - That makes the risk small

Introducing the Likelihood

- As an alternative to all that risky business…
- We can use the pdf (probability density function)
  - To construct the likelihood that the ***data*** $\boldsymbol{X}$ occurred
    - Given the ***unknown*** parameters Θ
- If we assume the data were collected ***independently*** and ***identically*** (i.e. having the same pdf)
  - Then the ***joint pdf***, called the ***likelihood***, is the product of individual pdf's
    - $\mathcal{L}(\Theta|X) = \prod_{i=1}^{n} f(x_i) = f(x_1)f(x_2)\cdots f(x_n)$

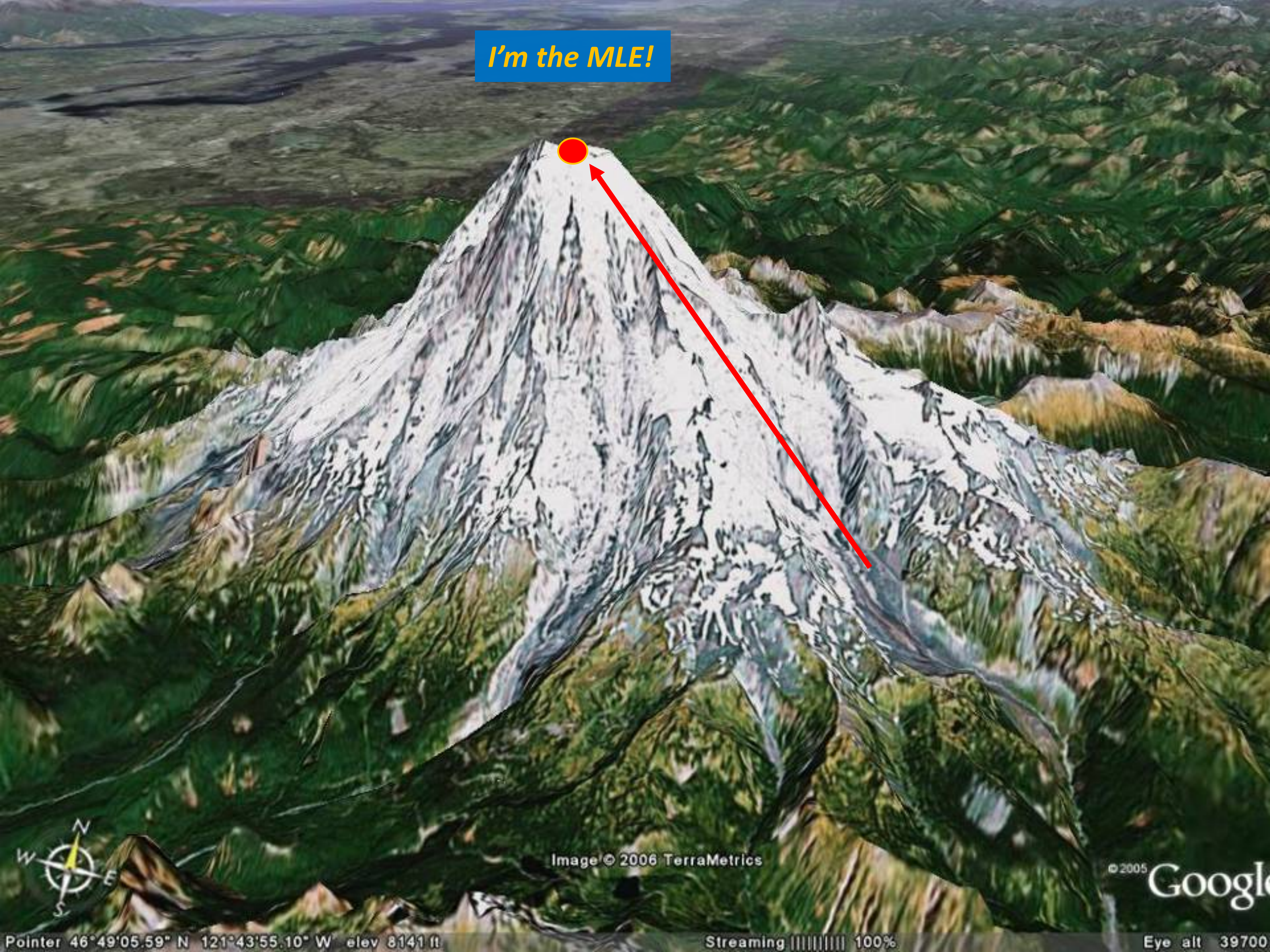
Introducing the Likelihood

- From the perspective of ***estimation***
  - The likelihood is viewed as a function of the ***unknown*** parameters
    - Given the data: $\mathcal{L}(\Theta|X)$
- The likelihood is maximized as a function of the ***unknown*** parameters
  - To obtain the maximum likelihood estimators (MLEs)
  - In order to get MLEs it is frequently more convenient to work with the natural log of the likelihood
    - $ln\mathcal{L}(\Theta|X) = \sum_{i=1}^{n} ln[f(x_i)]$

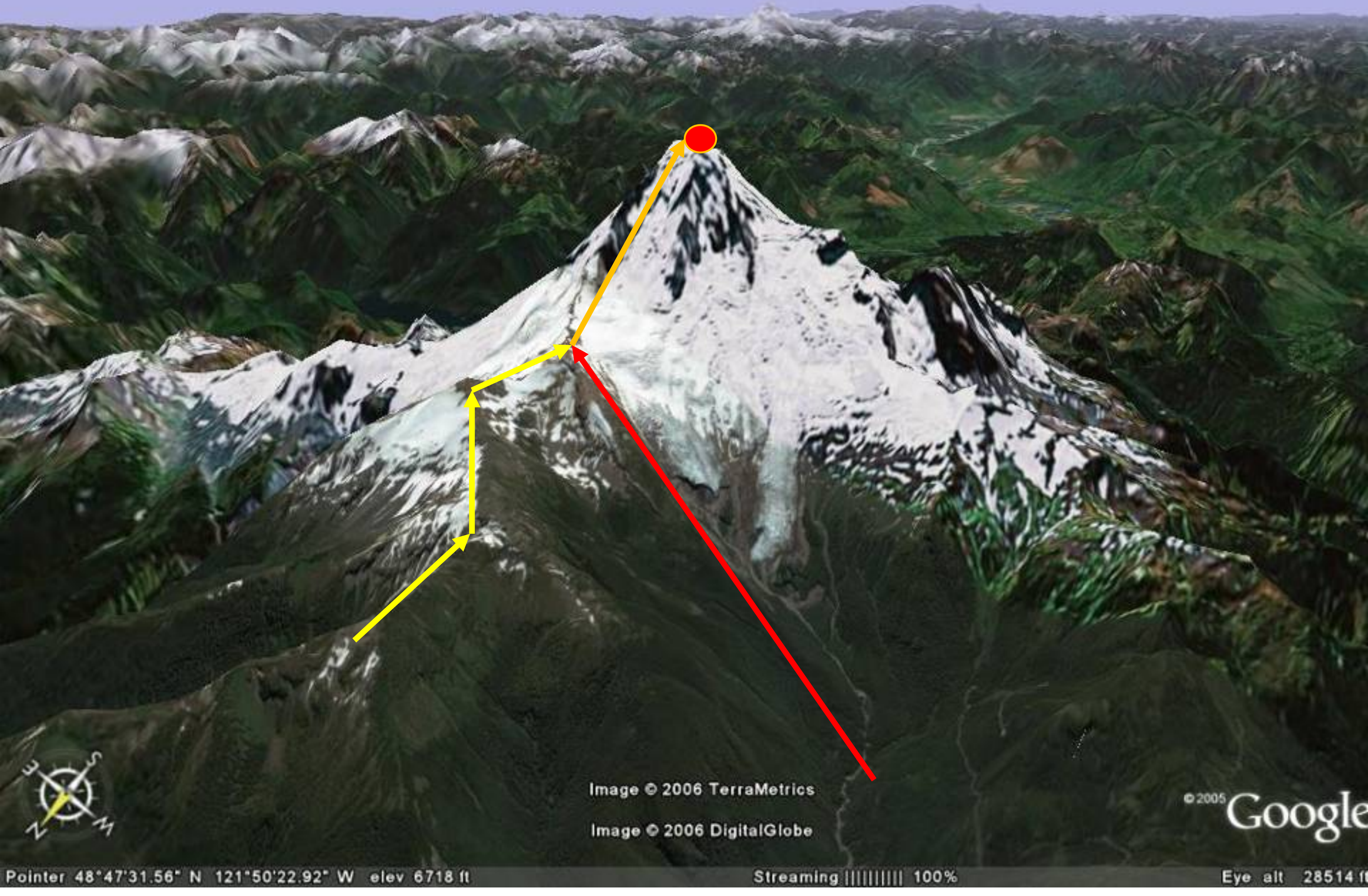
Likelihood Optimization

- To find the values for our parameters Θ
  - That maximize the log likelihood
    - Or minimize the negative log likelihood
- "Hill climbing in ***reverse***"
  - We want to walk down the hill, until every direction is up
    - R. Hilborn
- We want to search across values for our model parameters
  - Find the values that minimize our $\mathrm{NLL} = -\,ln\left(L(Parameters|Data)\right)$
    - Or maximize our likelihood


Maximizing the Total Likelihood
Minimizing the $-ln(L)$
Image © 2006 TerraMetrics
© 2005 Google
Pointer 46°49'05.59" N 121°43'55.10" W elev 8141 ft
Streaming 100%
Eye alt 39700


I'm the MLE!
Image © 2006 TerraMetrics
© 2005 Google
Pointer 46°49'05.59" N 121°43'55.10" W elev 8141 ft
Streaming 100%
Eye alt 39700

There are multiple ways to climb the hill depending on starting values!



Negative Log Likelihood Curvature

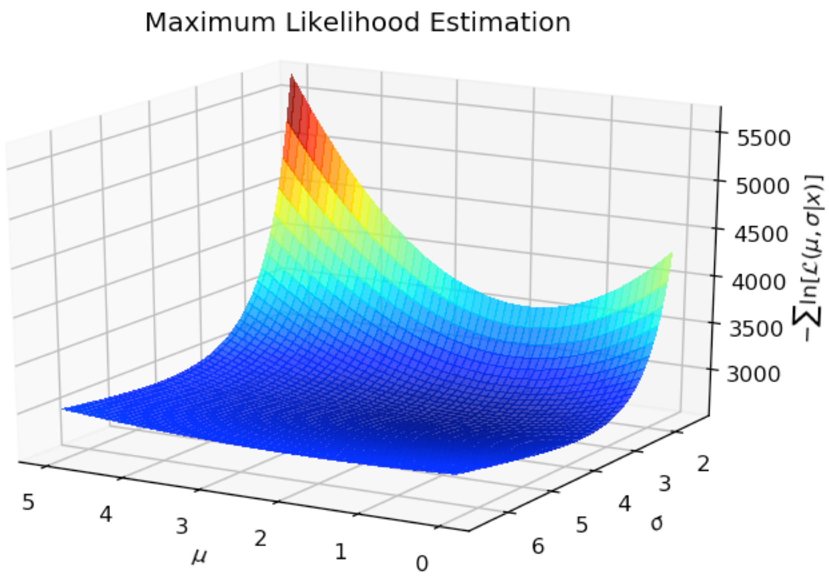
- Estimator and hessian matrix are often found by ***numerical methods***

- Delta method: general method for deriving the variance of asymptotically normal random variables
  - For variables that are asymptotically lognormal
    - Compute variance in log space, then transform Confidence Interval estimates

Slide adapted from: M. Rudd

Challenging Likelihood Surfaces

Flat negative log likelihood surface



Source: Math Stack Exchange

Multiple local minima (maxima)

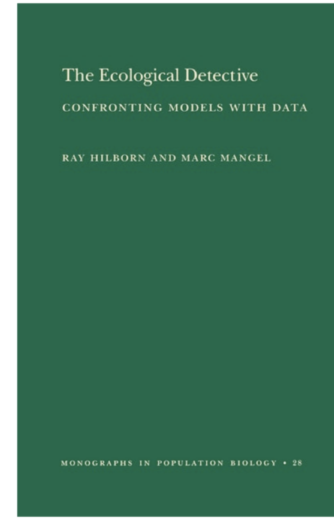
Basic Approach

- Start with a guess about parameter value x
- See which direction is down (calculate the slope (dy/dx)
  - Rate of change in objective function value (y)
    - With respect to parameter (x)
- Look at the slope and the change in slope (1^{st} and 2^{nd} derivatives) to guess about how far we have to go until we reach the bottom (lowest –loglikelihood)
- Move that direction
- Start over again
  - Repeat!

Slide adapted from: R. Hilborn

Newton’s method

Yes that is Sir Isaac Newton
Ecological Detective pp 267 (Hilborn and Mangel)



$$Let$$

Parameter value

Objective function value

$$y = f(x)$$

$$h(x) = \frac{dy}{dx} \qquad h'(x) = \frac{dh(x)}{dx}$$

begin with starting estimate x_1

$$x_{i+1} = x_i - \lambda \frac{h(x_i)}{h'(x_i)}$$

continue until x stops changing

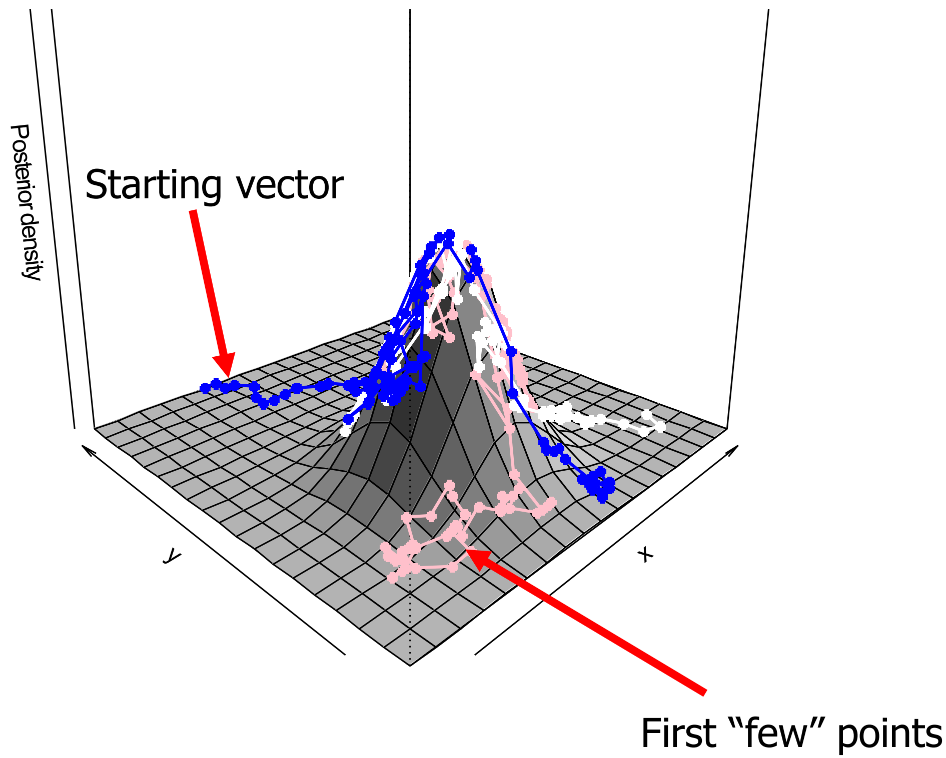
Slide adapted from: R. Hilborn

Other Minimization Approaches

- Derivative free - the SIMPLEX method
    - This is a very sophisticated form of hill climbing
- Simulated annealing
    - Randomly jump to a new spot, if it is better then stay there, if it is worse, go back to initial jump
- MCMC
    - Propose new parameter value, calculate posterior probability
        - Accept conditional on acceptance probability

Slide adapted from: R. Hilborn

Markov Chain Monte Carlo (MCMC)



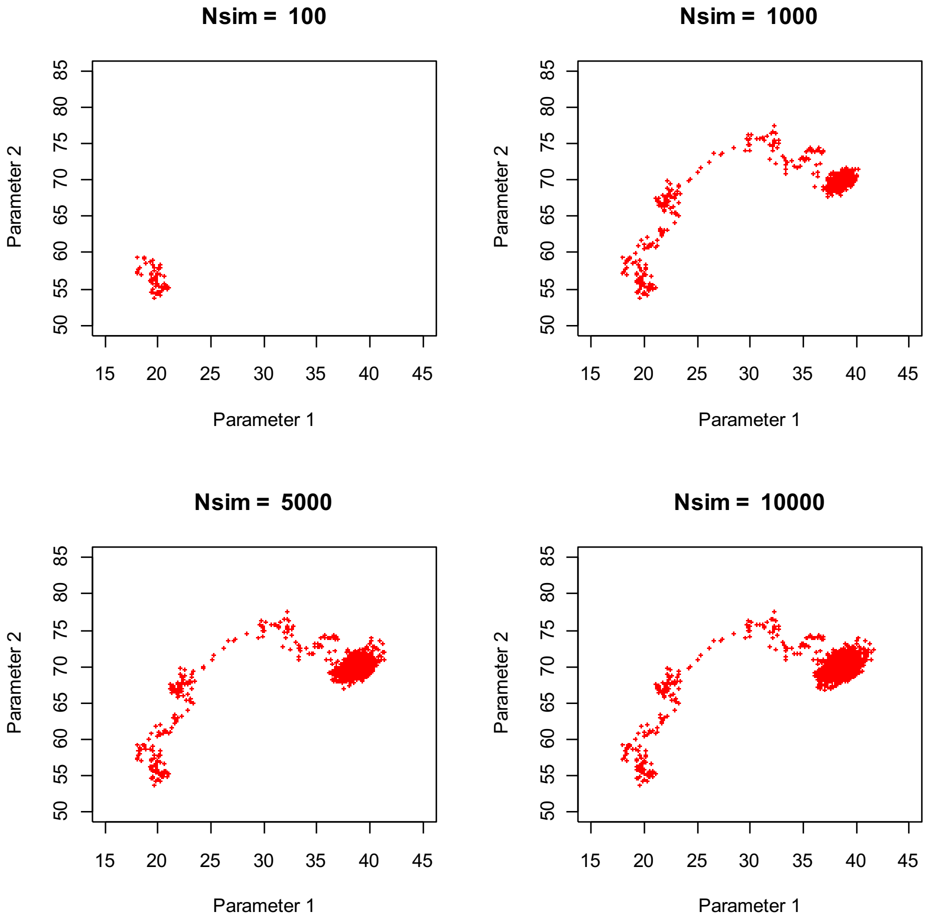
Similar to maximum likelihood estimation, but instead of only caring about the end point (MLE)

We care about our path (MCMC chain), and want to quantify how much time we spend at different places on our surface

Slide adapted from: M. Rudd

Markov Chain Monte Carlo (MCMC)

Impact of increasing the number of simulations from the posterior (i.e. MCMC iterations)



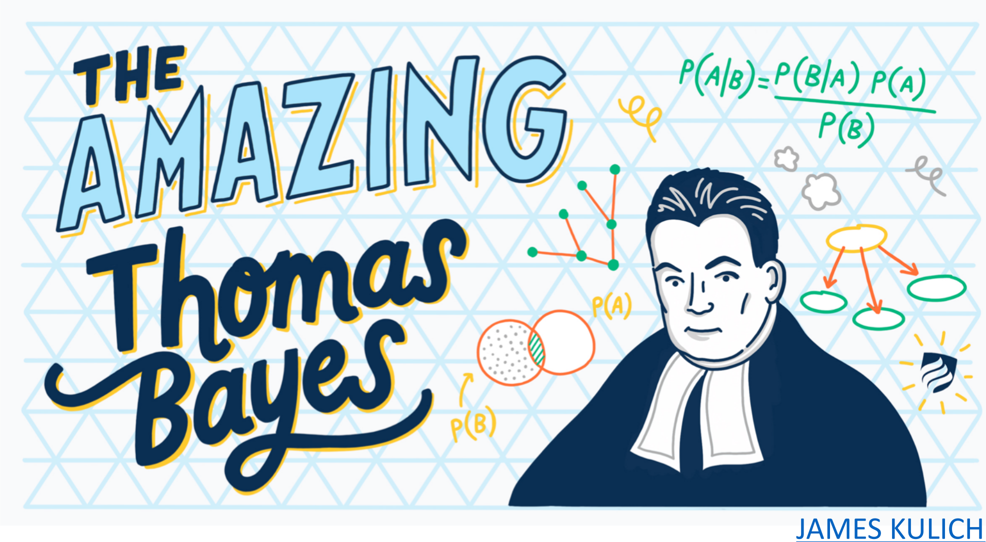
Slide adapted from: M. Rudd

The Bayesian Perspective

- Takes the likelihood principle one step further
  - The parameter set Θ is viewed as a set of random variables
    - With a pdf called the prior distribution $\pi(\Theta)$
- Using Bayes Theorem
  - The ***posterior distribution*** can be calculated from the prior and likelihood

$$p(\Theta|X) = \frac{\pi(\Theta)p(X|\Theta)}{\int_{\Theta} \pi(\Theta)p(X|\Theta)d\Theta}$$

$$f(\Theta|X) = \frac{\pi(\Theta)\mathcal{L}(X|\Theta)}{\int_{\Theta} \pi(\Theta)\mathcal{L}(X|\Theta)d\Theta}$$



JAMES KULICH

But, what about the *Variance*?

- It is clearly important to determine how well our estimator performs
  - At a minimum we should know something about the variance of an estimator $var(\widehat{\Theta})$
    - However, we can never really know the variance, which is itself a parameter of the population.
- But fear not!
  - We can estimate the variance, $\widehat{var}(\widehat{\Theta})$, using one of several methods

Methods for estimating the variance, $\widehat{var}(\widehat{\Theta})$:

- Theoretical
  - Statistical gyrations lead to a formula
    - I'm usually not smart enough to derive this directly ☹
- Approximation
  - Delta Method (Seber 1982, p. 7-9)
    - A Taylor series expansion is applied to the estimator, which leads to an approximate formula
- Jackknife
  - One data point is removed and the estimation is performed
    - Resulting in a pseudo-value
  - Repeat this for all data and $\widehat{var}(\widehat{\Theta})$ is an empirical function of the pseudo-values
- Bootstrap
  - Become one of the most common approaches to estimating variance
  - Data sampled with replacement with sample size equal to original dataset
    - Estimation procedure performed
  - After many replications, empirical variance of the bootstrap estimates is used as $\widehat{var}(\widehat{\Theta})$
  - Estimate of bias is also obtained
    - Difference between bootstrap mean and original estimate

But, what about the *Variance*?

- Two quantities related to the variance
  - Standard error (SE)
    - Square root of variance
    - $SE(\widehat{\Theta}) = \sqrt{\widehat{var}(\widehat{\Theta})}$
  - Coefficient of variation (CV)
    - Conversion of the SE to a relative scale or percentage
    - $CV(\widehat{\Theta}) = SE(\widehat{\Theta})/\widehat{\Theta}$
- Some people mistakenly refer to the standard error as the standard deviation (SD)
  - SD is equivalent to the amount of variation expected among n observations
    - $SD = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)}$
  - In contrasted, the estimated variance of the mean is ***much smaller***
    - $SE(\bar{x}) = SD/\sqrt{n}$

Simple Random Sampling

- Equivalent to “random sampling without replacement”
  - Sampling design in which n distinct units are selected from N possible units in the population
    - In such a way that every possible combination of n units is ***equally likely*** to be selected
- Selecting n
  - Sample 1: N with equal probability
    - Repeat until n units are selected, discarding repeats
  - Easy to do in R
    - `S <- sample(1:400, 40, replace=FALSE)`
- Each sample $1{:}\,n$, could be
  - Length of a fish in a pool
  - Number of fish in a bottom trawl haul at a specific location in space

$n = 40$ **from** $N = 400$

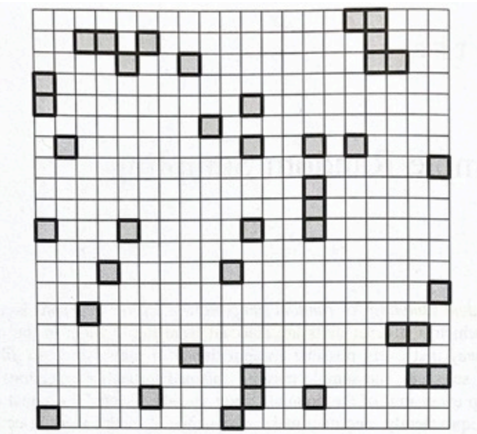


Figure 2.1. Simple random sample of 40 units from a population of 400 units.

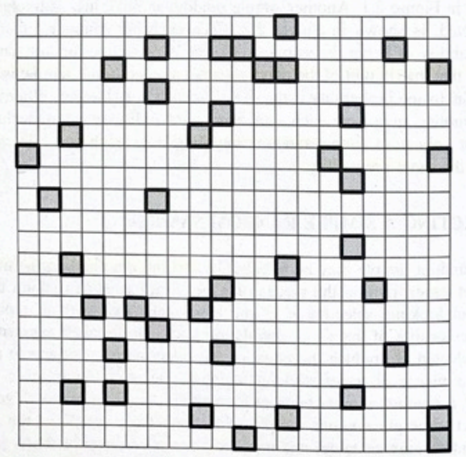


Figure 2.2. Another simple random sample of 40 units.

Thompson

SRS: Estimating the Population Mean

- With SRS the sample mean $\bar{y}$ is an ***unbiased estimator*** of the population mean μ
  - The ***population mean*** is the average of the y-values in the whole population N
    - $\mu = \frac{1}{N}(y_1 + y_2 + \cdots + y_N) = \frac{1}{N}\sum_{i=1}^{N} y_i$
  - The ***sample mean*** is the average of the y-values in the whole sample n
    - $\bar{y} = \frac{1}{n}(y_1 + y_2 + \cdots + y_n) = \frac{1}{n}\sum_{i=1}^{n} y_i$

SRS: Estimating the Population Variance

- Also with SRS the sample variance s^2 is an ***unbiased estimator*** of the finite-population variance σ^2
  - The ***finite-population variance*** is defined as
    - $\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \mu)^2$
  - The ***sample variance*** is defined as
    - $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$

SRS: Estimating the Population Variance

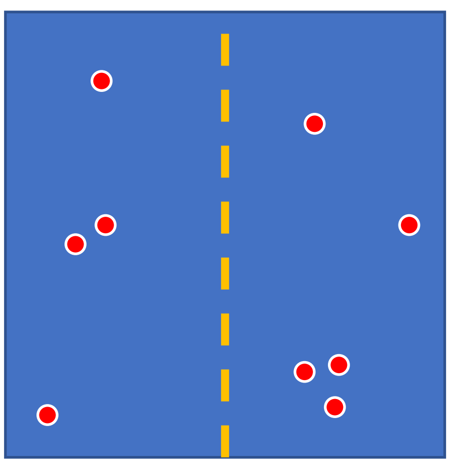
- The ***variance of the estimator*** $\bar{y}$ with SRS is
  - $var(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{\sigma^2}{n}$ ← ***Finite-population variance***
- An ***unbiased estimator*** of this variance is
  - $\widehat{var}(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{s^2}{n}$ ← ***Sample variance***
- The square roof of the variance of the estimator is its standard error
  - The estimated standard error ***is generally not*** an unbiased estimator of the actual standard error
- The quantity $(N-n)/N$ is the finite-population correction factor
  - If the population is large relative to the sample size
    - i.e. n/N is small
    - Finite population correction will be close to 1
    - Variance of the sample mean $\bar{y}$ will be approximately equal to σ^2/n

SRS: Estimating the Population Total

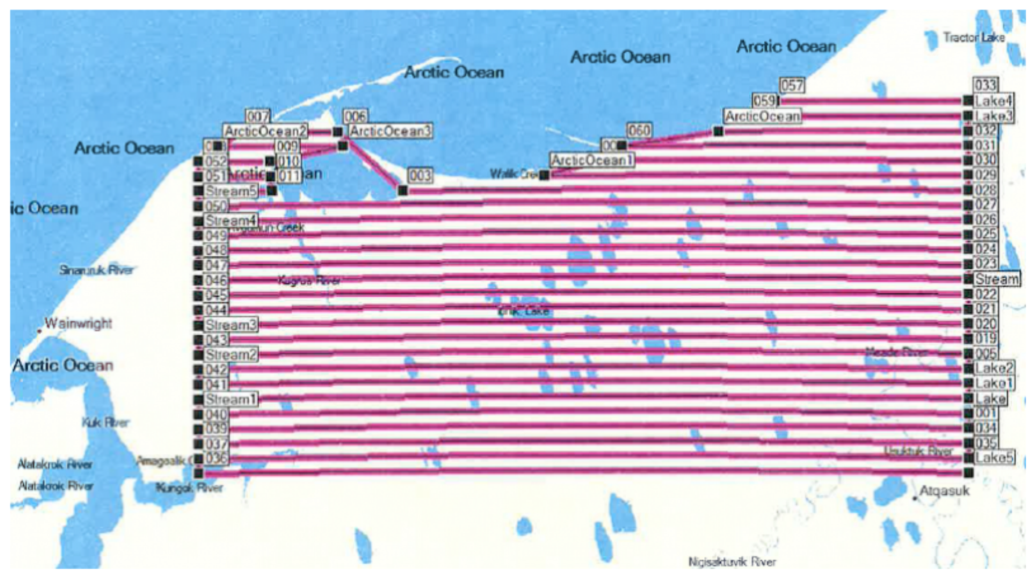
- To estimate the population total τ, where
  - $\tau = \sum_{i=1}^{N} y_i = N\mu$
- The sample mean is multiplied by N
  - An ***unbiased estimator*** of the ***population total*** is
    - $\hat{\tau} = N\bar{y} = \frac{N}{n}\sum_{i=1}^{n} y_i$
- Since the estimator $\hat{\tau}$ is N times the estimator $\bar{y}$
  - The variance of $\hat{\tau}$ is N^2 times the variance of $\bar{y}$
    - $var(\hat{\tau}) = N^2 var(\bar{y}) = N(N-n)\frac{\sigma^2}{n}$ ***Finite-population variance***
  - An ***unbiased estimator*** of the variance of the ***population total*** is
    - $\widehat{var}(\hat{\tau}) = N^2 \widehat{var}(\bar{y}) = N(N-n)\frac{s^2}{n}$ ***Sample variance***

SRS In Practice: Alaskan Caribou

- Arctic Coastal Plain of Alaska
  - Caribou counted from an aircraft flying over selected lines across the study region
    - Davis et al. (1979), Valkenburg (1990)
- All caribou with ½ mile to either side of each line
  - 1 mile wide strip(s)



½ ½



North Slope Burough

SRS In Practice: Alaskan Caribou

- A random sample of $n = 15$ North-South strips
  - Were selected from the $N = 286$-mile wide study region
- Number of caribou in the $n = 15$ sample units
  - $y_i = 1, 50, 21, 98, 2, 36, 4, 29, 7, 15, 86, 10, 21, 5, 4$
  - Sample mean
    - $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1+50+\cdots+4}{15} = 25.9333$
  - Sample variance
    - $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
    - $s^2 = \frac{(1-25.93)^2+(50-25.93)^2+\cdots+(4-25.93)^2}{15-1} = 919.0667$

SRS In Practice: Alaskan Caribou

- Number of caribou in the $n = 15$ sample units
  - $y_i = 1, 50, 21, 98, 2, 36, 4, 29, 7, 15, 86, 10, 21, 5, 4$
    - $\bar{y} = 25.9333$, and $s^2 = 919.0667$
- Estimated ***variance*** of the sample mean
  - $\widehat{var}(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{s^2}{n} = \left(\frac{286-15}{286}\right)\frac{919.07}{15} = 58.0576$
    - So that the estimated standard error is: $\sqrt{58.06} = 7.62$
- Estimate of ***the total number of caribou*** in the sample region
  - $\hat{\tau} = \frac{N}{n}\sum_{i=1}^{n} y_i = N\bar{y} = 286(25.9333) = 7{,}417$
- Estimated ***variance associated with the estimate*** of the total
  - $\widehat{var}(\hat{\tau}) = N(N-n)\frac{s^2}{n} = N^2\widehat{var}(\bar{y}) = 286^2(58.06) = 4{,}748{,}879$
    - Giving an estimated standard error of: $\sqrt{4{,}748{,}879} = 2{,}179$