FISH 621 Laboratory #1: Sampling Theory

Curry Cunningham 2021

Instructions

The purpose of this lab is to:

- Explore foundational concepts in probability theory that we will leverage throughout the semester.
- Gain experience with the pdf's for the Normal, Binomial, and Poisson distributions, and their associated functions in R.
- Explore likelihood profiling and our simple estimators.
-

If you have a question during the lab, please un-mute yourself and ask, or type it into the chat box. There is a high likelihood that someone else has the same question. It is more fun if we all learn together in our distance-learning world.

I have posted the lecture slides to the ***Canvas site***, so you can reference this material as you work through the lab.

This and all other labs will be graded based on your attendance and participation.

Lab Contents

- **621_Lab 1_Sampling Theory.pdf** (this file)
- **621_Lab 1_Sampling Theory.R** R script with exercises.

Exercise 2: Probability Distributions in R

In this exercise we will explore some of the standard probability distributions we will utilize repeatedly throughout the semester:

- Normal distribution
- Binomial distribution: Number of "successes" (n) out of some number of trials (N), conditional on probability (p) of success.
- Poisson distribution: Counts of things per unit time or space.

Although we often come up with rather mundane examples for these distributions (i.e. coin flips for the binomial), we will utilize the emergent properties of these distributions to ask more interesting questions in the context of abundance estimation, and they will form a critical component of the likelihoods that allow us to fit our models to data we or others collect.

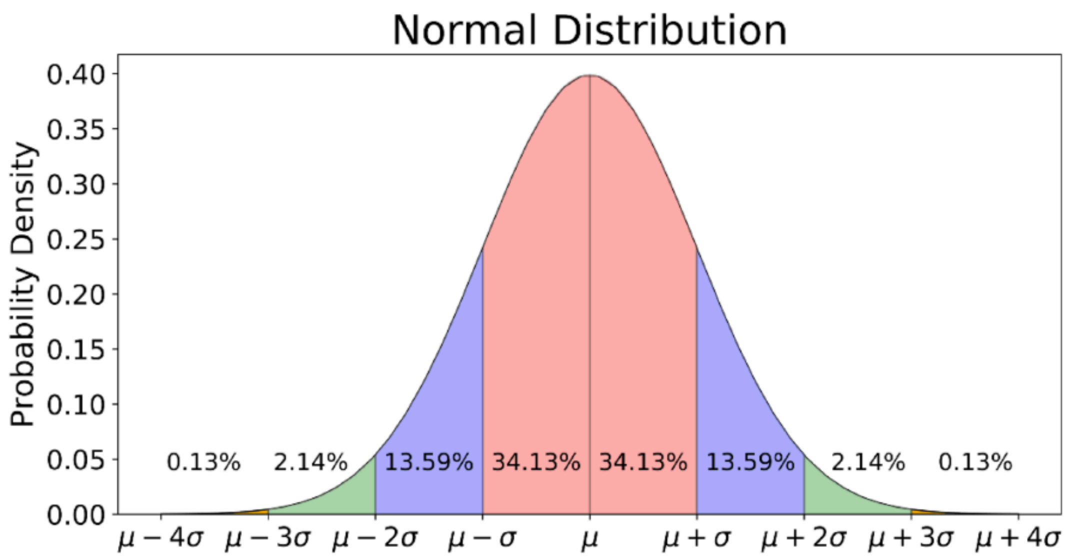
The standard probability functions in R will be very useful in the context of statistical analyses and simulation. Probability functions starting with “r”, like `rnorm(n, …)` will simulate random draws from a probability distribution of number `n=`.

Probability functions starting with “d”, like `dnorm(x, …)` will return the probability density for a value `x=`, given the distribution.

Probability functions starting with “p”, like `pnorm(q, …)` will return the cumulative probability density (area under the curve) for all values less `q=`, given the distribution.

- Normal
  - $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

Expected value: $E(x) = \mu$
Variance: $var(x) = \sigma^2$



- Binomial
  - $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
  - $f(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$

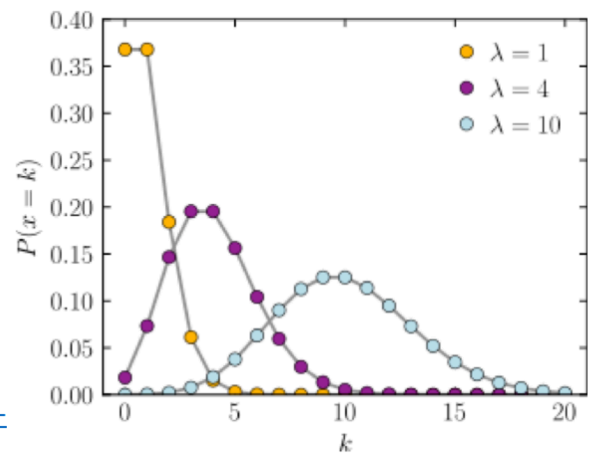
Factorial derivation

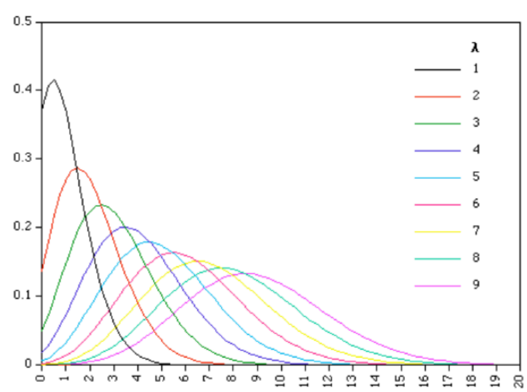
$5! = 5 * 4 * 3 * 2 * 1$

- Results from sampling with ***replacement***, totaling outcomes that have an attribute
  - Where p is the ***probability*** that each sampled item has the attribute
  - n is the number sampled
  - x is the number that have the attribute
- Summary
  - Expected value: $E(x) = np$
  - Variance: $var(x) = np(1-p)$
    - Note that the variance is ***less than the mean***, because $p < 1$

Common pdfs

- Poisson
  - $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ lambda
  - Counts per unit time or space
    - French mathematician [Siméon-Denis Poisson] developed in 1830 to describe the number of times a gambler would win in a large number of tries
- Summary
  - Expected value: $E(x) = \lambda$
  - Variance: $var(x) = \lambda$





Exercise 3: Binomial Likelihood: Known Detection Probability

In this exercise we will leverage our new-found familiarity with the binomial distribution and utilize it in the context of our simple estimator of abundance: $\widehat{N} = \frac{n}{p}$. In this context we will assume we know (or have estimated) the detection probability p.

With this knowledge we can ask (among other things):

- What is the probability of observing n animals, given the observation probability and a true abundance of N?
- What is the probability that our true abundance is $\widehat{N}$, given the number of animals we observed n and our known detection probability p?

Unknown

- Standard notation
  - N: population size (abundance)
  - $\widehat{N}$: estimator of population size

Our best guess

  - n: number of animals detected (sample size)
  - p: probability of detecting an animal

In order to estimate $\widehat{\mathbf{N}}$ we must learn about $\boldsymbol{p}$

- Basic estimator
  - Population size estimate is the number detected divided by the detection probability
    - $\widehat{N} = \frac{n}{p}$

Exercise 4: Simple Random Sampling

Simple random sampling is one of the most basic ways we can sample an area for abundance, or a population for some attribute like average length, age, weight, ect.

The underlying idea is that we sample n units out of a possible N total units in our sample frame, and from this we should be able to get unbiased estimates of key population parameters:

- The ***sample mean*** is the average of the y-values in the whole sample n
  - $\bar{y} = \frac{1}{n}(y_1 + y_2 + \cdots + y_n) = \frac{1}{n}\sum_{i=1}^{n} y_i$
- The ***sample variance*** is defined as
  - $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
- An ***unbiased estimator*** of this variance is
  - $\widehat{var}(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{s^2}{n}$ *Sample variance*

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

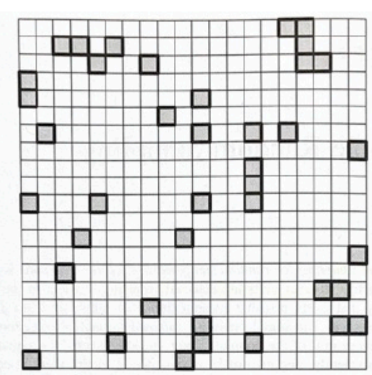
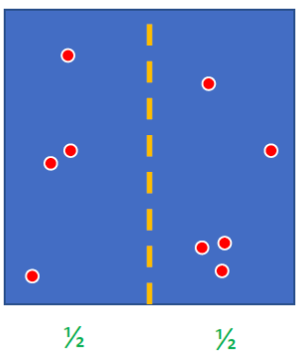


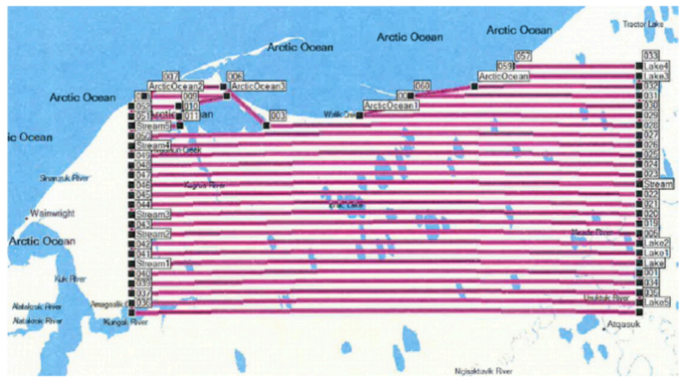
Figure 2.1. Simple random sample of 40 units from a population of 400 units.

Exercise 5: SRS – Alaskan Caribou

In this exercise we will explore the caribou aerial survey example we discussed in lecture. As a reminder, transects 1 mile in width were flown across the Arctic Coastal Plain, covering $n = 15$ out of a possible $N = 286$ miles.

- 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)





North Slope Burough

Assuming our transects were indeed sampled randomly and without replacement, and all caribou are detectable within each transect, we can use our estimators to tell us something about the population of caribou on the Coastal Plain and how certain we are about this estimate.

- 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
    - $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
    - $s^2 = \frac{(1-25.93)^2+(50-25.93)^2+\cdots+(4-25.93)^2}{15-1} = 919.0667$

- 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: $\sqrt{4{,}748{,}879} = 2{,}179$