Lecture 1: Introduction

MATH 667-01
Statistical Inference
University of Louisville

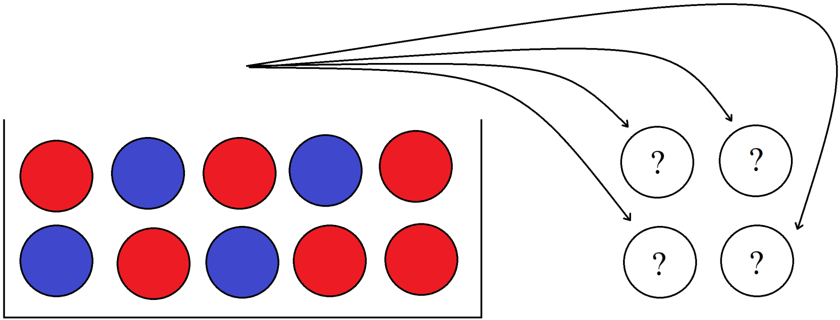
August 22, 2017

Introduction

- This lecture provides an introduction to important concepts on estimates and estimators through a basic example.
- First, an example of a question from probability is presented, and then a couple questions from statistics are discussed.
- In discussing the question from probability, we will review some basic probability theory in Chapter 1 of Casella and Berger (2001)[1] and a probability distribution discussed in Section 3.2 of Casella and Berger (2001).
- The statistical question involves estimation using concepts that will be presented in Section 7.2 and Section 5.2 of Casella and Berger (2001). This will include discussion of computing likelihood-based estimate based on observed data and of studying the sampling distribution of an estimator before the data is observed.

[1] Casella, G. and Berger, R. (2001). Statistical Inference, second edition. Duxbury Press.

A Probability Question



A Probability Question

- *Example L1.1*: Suppose that there are ten marbles in a box, 6 of which are red and 4 of which are blue. Four marbles are selected at random without replacement from the box. What is the probability that exactly one of the selected marbles is red?
- *Answer to Example L1.1*:
  Recall that if S is a finite sample space with equally likely outcomes, and we are interested in computing the probability of an event A, then

$$P(A) = \frac{\text{number of elements in } A}{\text{number of elements in } S} \qquad \text{(see p.16).}$$

A Probability Question

- *Answer to Example L1.1 continued*:
  Since we are selecting a sample of 4 objects without replacement from a population of 10 objects, the sample space is the set of all possible subsets of size 4 from the population, and the number of elements of S (that is, the number of possible samples) is

$$\binom{10}{4} = \frac{10!}{4! \cdot 6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} = 210.$$

- Next, we need to compute the number of ways of choosing 1 red marble and 3 blue marbles. The Fundamental Theorem of Counting (Thm 1.2.14 on p.13) states that if a job consists of k separate tasks, the ith of which can be done in n_i ways, $i = 1, \ldots, k$, then the entire job can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways.

- *Answer to Example L1.1 continued*:
  So, since there are $\binom{6}{1} = 6$ ways to choose the one red marble and $\binom{4}{3} = 4$ ways to choose three blue marbles, there are $6 \times 4 = 24$ ways to choose one red and 3 blue marbles from the box.
- So, the probability of the event A that exactly one of the selected marbles is red is

$$P(A) = \frac{\binom{6}{1}\binom{4}{3}}{\binom{10}{4}} = \frac{24}{210} = \frac{4}{35} \approx .114286.$$

- Alternately, the answer to *Example L1.1* can be obtained by letting X be the random variable representing the number of red marbles obtained when selecting 4 marbles at random without replacement from the box and computing $P(X = 1)$.
- See Section 1.4 (see p.27) for a review of random variables.
- In this experiment, X is a discrete random variable which follows a *hypergeometric distribution* (see p.87) that has probability mass function (pmf)

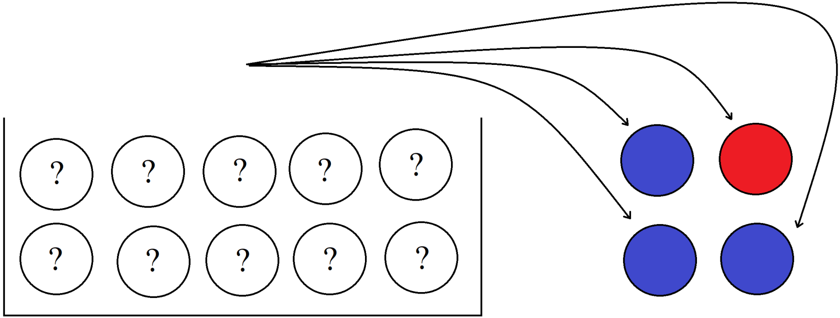
$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$$

for x in $\{\max\{0, M - N + K\}, \ldots, \min\{K, M\}\}$ where N is the number of objects in the population, M is the number of objects of the desired type in the population, and K is the number of objects selected in the sample.

A Statistics Question

- In statistics, we instead consider what we can infer about the unknown population based on a known sample.
- For example, suppose that we know that there are 10 red and blue marbles in a box but we do not know how many are red and how many are blue. We select 4 marbles at random without replacement and obtain 1 red marble and 3 blue marbles. What can we infer about the contents of the box? What is our best guess for the number of red marbles in the box?

A Statistics Question



- There are many ways to obtain an estimate of M, the number of red marbles in the box. In this class, we will consider a few of the possible methods of estimating this number.
- Now we will briefly introduce a popular estimate called the maximum likelihood estimate (MLE) to estimate a population "parameter" using observed data based on a parametric model.
- Suppose that $\boldsymbol{X}$ is an n-dimensional random variable with joint pmf (or in the continuous case, joint pdf) $f_\theta(\boldsymbol{x})$ where θ is a k-dimensional parameter in a parameter space Θ.

- *Definition L1.1*: The likelihood function $L : \Theta \to \mathbb{R}$ is defined by $L(\theta) = L(\theta; \boldsymbol{x}) = f_\theta(\boldsymbol{x})$.
- Note the difference between the pmf/pdf and the likelihood function. For the pmf/pdf, θ is considered to be known and the domain of f is the possible values of $\boldsymbol{X}$. For the likelihood function, $\boldsymbol{x}$ is considered to be known and the domain of L is Θ.
- *Definition L1.2* (see Def 7.2.4 on p.316): A value of θ in Θ which maximizes L is called a *maximum likelihood estimate* (MLE) of θ. Such a value is often denoted by $\hat{\theta}$ or $\hat{\theta}(\boldsymbol{x})$.
- The MLE finds the value(s) of the parameter which "makes the observed data most likely".

- *Example L1.2*: Suppose that there are ten marbles in a box, M of which are red and $10 - M$ of which are blue. We select four marbles at random without replacement and observe 1 red marble and 3 blue marbles. What is the maximum likelihood estimate of M, the number of red marbles in the box?
- *Answer to Example L1.2*: Let X be a hypergeometric random variable with parameters M, $N = 10$, and $K = 4$. The likelihood function for estimating M is

$$L(M) = L(M; X = 1) = \frac{\binom{M}{1}\binom{10-M}{3}}{\binom{10}{4}} \text{ for } M \in \{1, \ldots, 7\}\,.$$

- Now we compute the likelihood function for each value in the domain of L on the next slide.

A Statistics Question

- *Answer to Example L1.2 continued*:

$$\begin{aligned}
L(1) &= \binom{1}{1}\binom{9}{3}\Big/\binom{10}{4} = .400000 \\
L(2) &= \binom{2}{1}\binom{8}{3}\Big/\binom{10}{4} \approx .533333 \\
L(3) &= \binom{3}{1}\binom{7}{3}\Big/\binom{10}{4} = .500000 \\
L(4) &= \binom{4}{1}\binom{6}{3}\Big/\binom{10}{4} \approx .380952 \\
L(5) &= \binom{5}{1}\binom{5}{3}\Big/\binom{10}{4} \approx .238095 \\
L(6) &= \binom{6}{1}\binom{4}{3}\Big/\binom{10}{4} \approx .114285 \\
L(7) &= \binom{7}{1}\binom{3}{3}\Big/\binom{10}{4} \approx .033333
\end{aligned}$$

- Since L is maximized when $M = 2$, the MLE is $\hat{M} = 2$.
- The folllowing R command can be used to compute all of these values.

```
> dhyper(1,1:7,9:3,4)
[1] 0.40000000 0.53333333 0.50000000 0.38095238
[5] 0.23809524 0.11428571 0.03333333
```

- A list of some R commands with some probability distribution for computing cdf's, for inverting cdf's, for computing pdf's or pmf's, and for generating random values are available at `http://www.stat.umn.edu/geyer/old/5101/rlook.html`.

A Statistics Question

- Alternately, we can maximize L without checking all possible values of θ.
- Since the natural logarithm is an increasing function, maximizing L is equivalent to maximizing

$$\begin{aligned}\ell(M) &= \ln L(M) \\ &= \ln M + \sum_{i=0}^{2} \ln(10 - M - i) - \ln 6 - \ln 210.\end{aligned}$$

- Even though the domain of ℓ is $\{1, \ldots, 7\}$, let's temporarily consider all values in $[1, 7]$ and differentiate ℓ twice to obtain

$$\frac{d\ell}{dM} = \frac{1}{M} - \frac{1}{10 - M} - \frac{1}{9 - M} - \frac{1}{8 - M}$$

and

$$\frac{d^2\ell}{dM^2} = -\frac{1}{M^2} - \frac{1}{(10 - M)^2} - \frac{1}{(9 - M)^2} - \frac{1}{(8 - M)^2}.$$

- Since $\dfrac{d^2\ell}{dM^2} < 0$, ℓ is concave downward on $[1, 7]$ so $\dfrac{d\ell}{dM}$ is decreasing.
- Furthermore, since

$$\left.\frac{d\ell}{dM}\right|_{M=2} = \frac{11}{168} > 0$$

and

$$\left.\frac{d\ell}{dM}\right|_{M=3} = -\frac{37}{210} < 0,$$

ℓ is either maximized at $M = 2$ or $M = 3$.

- In order to assess the quality of a method of estimating a parameter, we need to study the *sampling distribution* of the estimator of the parameter.
- To do this, we consider the experiment before the data is observed and consider all possible samples and their probabilities. For each possible sample, we determine the estimate of the parameter if the sample is observed. This allows us to compute the pmf/pdf of the estimator.
- In general, it is important to note that the distribution of the estimator depends on the true value of the parameter.
- Sampling distributions are introduced more formally in Section 5.2 (Def 5.2.1 on p.211), but here the concept is presented for our example.

- Now suppose that there are ten marbles in a box, M of which are red and $10 - M$ of which are blue. We select four marbles at random without replacement and observe x red marbles and $4 - x$ blue marbles. For each possible value of x in $\{0, 1, 2, 3, 4\}$, we find the maximum likelihood estimate of M.
- Using the same method discussed in the answer to *Example L1.1*, we find that
  $\hat{M} = 0$ if $x = 0$,
  $\hat{M} = 2$ if $x = 1$,
  $\hat{M} = 5$ if $x = 2$,
  $\hat{M} = 8$ if $x = 3$,
  and $\hat{M} = 10$ if $x = 4$,
  as shown in the R code on the next slide.

```
> dhyper(0,0:6,10:4,4)
[1] 1.000000000 0.600000000 0.333333333
[4] 0.166666667 0.071428571 0.023809524
[7] 0.004761905
> dhyper(1,1:7,9:3,4)
[1] 0.40000000 0.53333333 0.50000000 0.38095238
[5] 0.23809524 0.11428571 0.03333333
> dhyper(2,2:8,8:2,4)
[1] 0.1333333 0.3000000 0.4285714 0.4761905
[5] 0.4285714 0.3000000 0.1333333
> dhyper(3,3:9,7:1,4)
[1] 0.03333333 0.11428571 0.23809524 0.38095238
[5] 0.50000000 0.53333333 0.40000000
> dhyper(4,4:10,6:0,4)
[1] 0.004761905 0.023809524 0.071428571
[4] 0.166666667 0.333333333 0.600000000
[7] 1.000000000
```

- So we can write

$$\hat{M}(x) = \begin{cases} 0 & \text{if } x = 0 \\ 2 & \text{if } x = 1 \\ 5 & \text{if } x = 2 \\ 8 & \text{if } x = 3 \\ 10 & \text{if } x = 4 \end{cases};$$

  given a sample x, $\hat{M}(x)$ is the *maximum likelihood estimate* of M.

- Before we observe the data, X is a hypergeometric random variable with parameters M, $N = 10$ and $K = 4$, and we want to find the distribution of the *maximum likelihood estimator* $\hat{M}(X)$.

- *Example L1.3*: Suppose that there are ten marbles in a box, M of which are red and $10 - M$ of which are blue where M is unknown. If the true value of M is 6, find the sampling distribution for the maximum likelihood estimator of M based on a sample of four marbles selected at random without replacement.

- *Answer to Example L1.3*: When $M = 6$, the pmf of $\hat{M}(X)$ is given below.

$$
\begin{aligned}
P(\hat{M}(X) = 0) &= P(X = 0) = \binom{6}{0}\binom{4}{4} \Big/ \binom{10}{4} = \frac{1}{210} \approx .005 \\
P(\hat{M}(X) = 2) &= P(X = 1) = \binom{6}{1}\binom{4}{1} \Big/ \binom{10}{4} = \frac{24}{210} \approx .114 \\
P(\hat{M}(X) = 5) &= P(X = 2) = \binom{6}{2}\binom{4}{2} \Big/ \binom{10}{4} = \frac{90}{210} \approx .429 \\
P(\hat{M}(X) = 8) &= P(X = 3) = \binom{6}{3}\binom{4}{3} \Big/ \binom{10}{4} = \frac{80}{210} \approx .381 \\
P(\hat{M}(X) = 10) &= P(X = 4) = \binom{6}{4}\binom{4}{0} \Big/ \binom{10}{4} = \frac{15}{210} \approx .071 \\
P(\hat{M}(X) = m) &= 0 \text{ if } m \text{ is not in } \{0, 2, 5, 8, 10\}
\end{aligned}
$$