STAT 88: Lecture 15

Contents

Last time

Unbiased estimators

Probability distributions often have parameters that we wish to estimate. An estimator is a random variable and there is uncertainty what you will get. With an unbiased estimator, on average the estimator will be correct.

Ex Suppose the population has population mean μ, i.e. any sample X from the population has mean $E(X) = \mu$. Let $X_1, \ldots, X_n$ be a SRS from the population. We use the sample mean $\bar{X}$ as an estimator of the population mean μ. Sample mean is always unbiased since

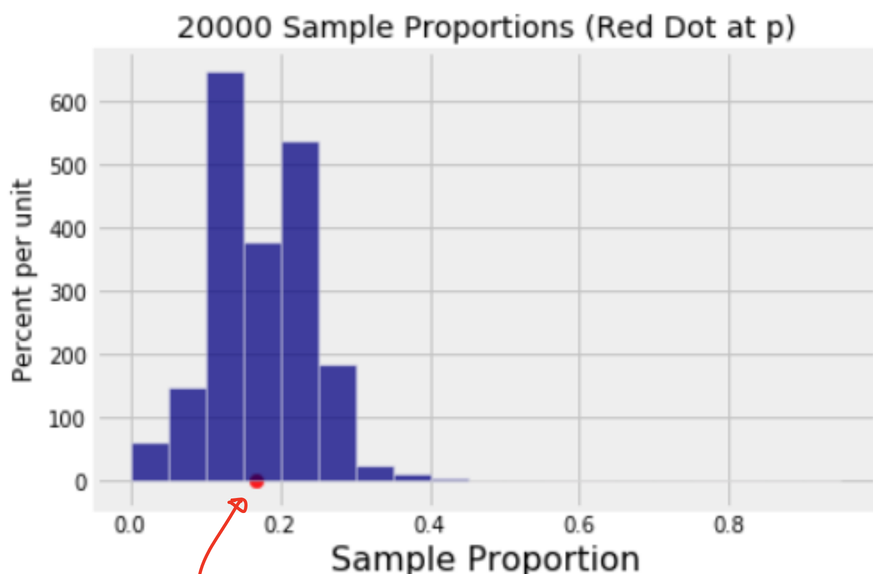
$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu.$$

Ex Suppose the population consists of zeros and ones. Then the population mean p is the population proportion of ones. If $X_1, \ldots, X_n$ are i.i.d. samples from the population, the sample mean $\bar{X}$ is the sample proportion of ones in your sample. By unbiasedness of the sample mean, we have

$$E(\bar{X}) = p.$$

The sampling distribution of sample proportions from the 20,000 repeated experiments:

```
n = 30
p = 0.1667
Average of observed sample proportions = 0.1664
```



p = population proportion

$= E(\overline{x})$

Warm up: (Exercise 5.7.11) Let X be the number of cars owned by a Cal student. Here is the distribution of X.

number of cars	0	1	2
probability	2θ	θ	$1-3\theta$

(a) Find $E(X)$ (as a function of θ).

(b) Let $X_1, \ldots, X_n$ be the number of cars owned by n randomly picked students. Use $\bar{X}$ to find an unbiased estimator of θ.

5.4. Unbiased Estimators (Continued)

Estimating the largest possible value

parameter

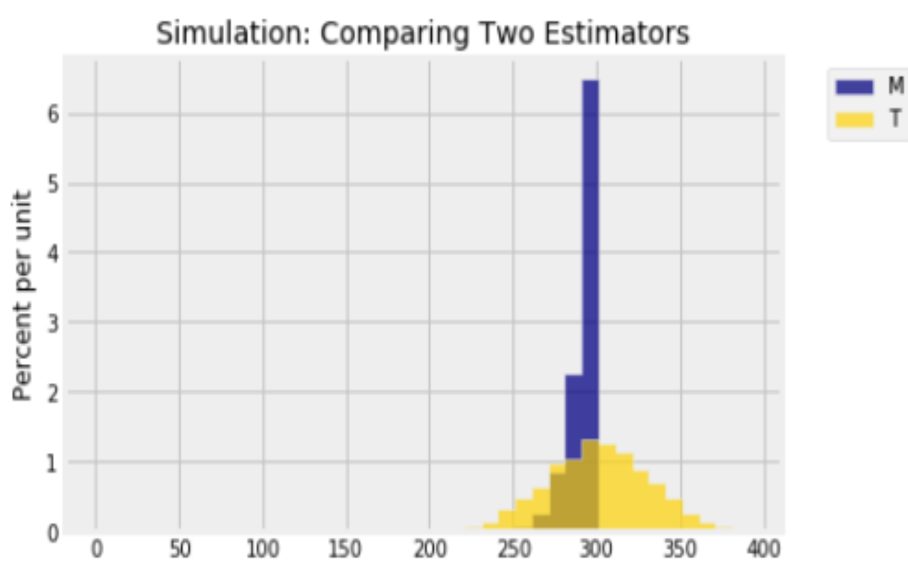
Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim}$ Uniform$\{1, 2, \ldots, N\}$ for some fixed but unknown N. To estimate N, there are two possible estimators we can come up with:

1. $M = \max\{X_1, \ldots, X_n\}$. Note that this is a biased estimator.
2. We know that the population mean is $\mu = (N+1)/2$ and thus $E(\bar{X}) = (N+1)/2$ since it is unbiased. Then what is an estimator T such that

$$E(T) = N?$$

Lets look at sampling distribution of (1) $M = \max(X_1, \ldots, X_n)$ and (2) $T = 2\bar{X} - 1$.

```
N = 300
n = 30
5000 repetitions
```



The histograms show that both estimators have pros and cons.

M - Pros: small spread of values; Cons: biased.

T - Pros: unbiased; Cons: big spread of values.

Unbiasedness is a good property, but so is low variability. Bias-variance tradeoff

5.5. Conditional Expectation

Let's first review how to find expectation of a joint distribution. A joint distribution for two random variables X and S is given below:

	$X = 1$	$X = 2$	$X = 3$
$S = 2$	0.0625	0	0
$S = 3$	0.125	0.125	0
$S = 4$	0.0625	0.25	0.0625
$S = 5$	0	0.125	0.125
$S = 6$	0	0	0.0625

The marginal distribution of S is given by summing along the rows:

s	2	3	4	5	6
$P(S = s)$	0.0625	0.25	0.375	0.25	0.0625

Conditional Distribution Suppose someone runs the experiment and tells you that $S = 3$. Given this information, what is the distribution of X?

$$P(X = 1|S = 3) = \frac{P(X = 1, S = 3)}{P(S = 3)} = \frac{0.125}{0.25} = 0.5.$$

Similarly we can get $P(X = 2|S = 3) = 0.5$ and $P(X = 3|S = 3) = 0$.

If X and S are two random variables on the same outcome space, then for a fixed value s of S, the conditional distribution of X given $S = s$ is

- the set of all possible values of X under the condition that $S = s$, and
- all the corresponding conditional probabilities $P(X = x|S = s)$.

The distribution of X changes depending on the given value of S:

	X=1	X=2	X=3
Conditional Dist'n of X given S=3	0.5	0.5	0

	X=1	X=2	X=3
Conditional Dist'n of X given S=4	0.1667	0.6667	0.1667

Conditional Expectation The expectation of X, also called the unconditional expectation of X, is easy to see from the distribution table:

x	1	2	3
$P(X = x)$	0.25	0.5	0.25

$E(X) =$

Given that S has the value s, the conditional distribution of X is just an ordinary distribution and thus has an expectation. This is called the conditional expectation of X given $S = s$ and is denoted $E(X|S = s)$.

$E(X|S = 3) =$

Generally, $E(X|S=s) = \sum_{\text{all } x} x P(X|S=s)$

Unconditional Expectation

$E(X) = \sum_{\text{all } x} x P(X=x)$

What is relationship between expectation and conditional expectation?

$$E(X) = \sum_{\text{all } x} xP(X = x) = \sum_{\text{all } x} \sum_{\text{all } s} xP(X = x, S = s).$$

By multiplication rule,

$$P(X = x, S = s) = P(X = x | S = s)P(S = s).$$

So

$$E(X) = \sum_{\text{all } s} \sum_{\text{all } x} xP(X = x, S = s) = \underbrace{\sum_{\text{all } s} \sum_{\text{all } x} xP(X = x | S = s)}_{=E(X|S=s)} P(S = s).$$

Therefore

$$E(X) = \sum_{\text{all } s} E(X|S = s)P(S = s).$$

Important: $E(X|S = s)$ is a function of s. For example,

s	2	3	4	5	6
$E(X \mid S = s)$	1	1.5	2	2.5	3

5.6. Expectation by Conditioning

To find expectation of one random variable, it sometimes helps to condition on another random variable.

Time to Reach Campus A student has two routes to campus. Each route has a random duration. The student prefers Route A because its expected duration is 15 minutes compared to 20 minutes for Route B. However, on 10% of her trips she is forced to take Route B because of road work on Route A. On the remaining 90% of the days she takes Route A.

What is the expected duration of her trip on a random day?

Example: (Exercise 5.7.13) A survey organization in a town is studying families with children. Among the families that have children, the distribution of the number of children is as follows.

Number of Children n	1	2	3	4	5
Proportion with n **Children**	0.2	0.4	0.2	0.15	0.05

Suppose each child has chance 0.51 of being male, independently of all other children. What is the expected number of male children in a family picked at random from those that have children?

Example: You flip a fair coin N times where N is a random variable $N \sim \text{Poisson}(5)$. What is the expected number of heads you will get?