Stat 88 lec 14

Warmup 2:00 – 2:10

A drawer contains s black socks and s white socks (s> 0). I pull two socks out at random without replacement and call that my first pair. Then I pull out two socks at random without replacement and call that my second pair. I proceed in this way until I have s pairs and the drawer is empty. Find the expected number of pairs in which two socks are different colors.

X = number of pairs (out of s) of mismatching

$I_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ pair is mismatch} \\ 0 & \text{else} \end{cases}$ P

$P = \frac{\binom{s}{1}\binom{s}{1}}{\binom{2s}{2}}$

$X = I_1 + \cdots + I_s$

$E(X) = sP$

alternatively $P = 2\frac{s}{2s}\frac{s}{2s-1}$ if think of chance get WB or BW.

midterm: Chapters 1-5

review materials coming soon

Last Time

Sec 5.3 method of indicators

Step 1 Describe X (a count of what?)

Step 2 $I_2 =$

Step 3 $p =$ (same for all indicators)

Step 4 $X =$

Step 5 $E(X) =$

If $X \sim \text{Binomial}(n, p)$, $E(X) = np$.

If $X \sim HG(N, G, n)$, $E(X) = n\frac{G}{N}$

Properties of expectation :

① $E(X_1+X_2) = E(X_1)+E(X_2)$

additivity

② $E(aX+b) = aE(X)+b$,

linear function rule

Preliminary: Linear Function Rule

Let X be a random variable and let $Y = aX + b$. Then Y is a linear function of X. By our method for finding the expectation of a function of a random variable,

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

$$\begin{aligned} E(Y) &= E(aX+b) = \sum_{\text{all } x}(ax+b)P(X=x) \\ &= a\sum_{\text{all } x} xP(X=x) + b\sum_{\text{all } x} P(X=x) \\ &= aE(X)+b \end{aligned}$$

<u>Today</u> ① go over concept test from last time.

② sec 5.4 <u>unbiased estimator</u>

1

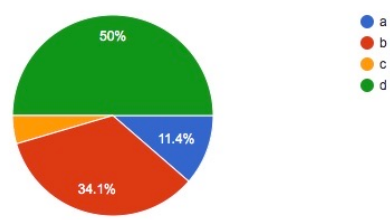
n people with hats have had a bit too much to drink at a party. As they leave the party, each person randomly grabs a hat. A match occurs if a person gets his or her own hat.

a The expected number of matches depends on n

b The expected number of matches is 1

c The number of matches is hypergeometric

d more than one of the above



d	The larger n is, the smaller the probability of having a match is. It's hypergeometric because each trial is dependent on the outcome of the other trials. So A and C are correct
b	We chose this because the nature of the problem, given n people that each grab a hat, means it is NOT hypergeometric. In addition, if X is the # out of n, then it stands that when counting matches, the likelihood that someone gets their hat is 1/n and n indicators => 1

X = # people (out of n) matching

say $n=3$

$$\left.\begin{aligned} P(X=0) &= 1/3 \\ P(X=1) &= 1/2 \\ P(X=2) &= 0 \\ P(X=3) &= \tfrac{1}{6} \end{aligned}\right\}$$

no way this can be zero if $X \sim HG$

a b c — 3 matches

$\left.\begin{matrix} a\ c\ b \\ c\ b\ a \\ b\ a\ c \end{matrix}\right\}$ 1 match

$\left.\begin{matrix} b\ c\ a \\ c\ a\ b \end{matrix}\right\}$ 0 matches

$$I_2 = \begin{cases} 1 & \text{if } 2^{\text{nd}} \text{ person matches} \\ 0 & \text{else} \end{cases}$$

$p = 1/n$

$X = I_1 + \cdots + I_n$

$E(X) = n\left(\frac{1}{n}\right) = \boxed{1}$

(2) Unbiased Estimators

Data scientists often want to estimate a parameter of a population.

A statistic is a number based on your sample such as the sample mean.

Statistic = a RV.

Parameter = a fixed unknown constant

An estimator, is a statistic used to approximate a parameter θ.

(An unbiased estimator of a parameter is an estimator whose expected value is equal to the parameter

Sample Mean as estimator of population mean

ex Estimate the average annual income in California, μ.

Suppose you draw a random sample size n.

$X_1, \ldots, X_n$ are sample incomes

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the <u>sample mean</u>

$\bar{X}$ is <u>unbiased</u> if

$E(\bar{X}) = \mu$ Population mean

$\mu = E(X_i)$

$E(aX) = aE(X)$

<u>Check</u>

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

$$= \frac{1}{n}\sum_{i=1}^{n} \underbrace{E(X_i)}_{=\mu} = \frac{1}{n} \cdot n\mu = \mu \checkmark$$

Which of these estimators of μ is unbiased?

- (iii) X_{15} $\quad E(X_{15}) \stackrel{?}{=} \mu$ unbiased.
- (iv) $(X_1 + X_{15})/15$ biased.
- (v) $(X_1 + 2X_{100})/3$ unbiased

$$E\left(\frac{X_1 + 2X_{100}}{3}\right) = \frac{1}{3}\left(\underbrace{E(X_1)}_{\mu} + 2\underbrace{E(X_{100})}_{\mu}\right) = \frac{1}{3}(3\mu) = \mu$$

If we have a biased estimator, how can we make it unbiased?

Lets make $\frac{X_1 + X_{15}}{3}$ unbiased

$$E\left(\frac{X_1 + X_{15}}{3}\right) = \frac{1}{3}\left(E(X_1) + E(X_{15})\right)$$
$$= \frac{1}{3}(2\mu)$$
$$= \frac{2}{3}\mu$$

$$\frac{3}{2} E\left(\frac{X_1+X_{15}}{3}\right) = \mu.$$

$$E\left(\frac{3}{2}\left(\frac{X_1+X_{15}}{3}\right)\right) = \mu$$

$$E\left(\frac{X_1+X_{15}}{2}\right) = \mu$$

$\boxed{\frac{X_1+X_{15}}{2}}$ is unbiased estimator of μ.

Sample Proportion as estimator of Population proportion

When the population consists of zeros and ones, the population mean is the population proportion of ones.

ex population $\boxed{0\ 0\ 1\ 1\ 1}$ — population mean $= p = 3/5$

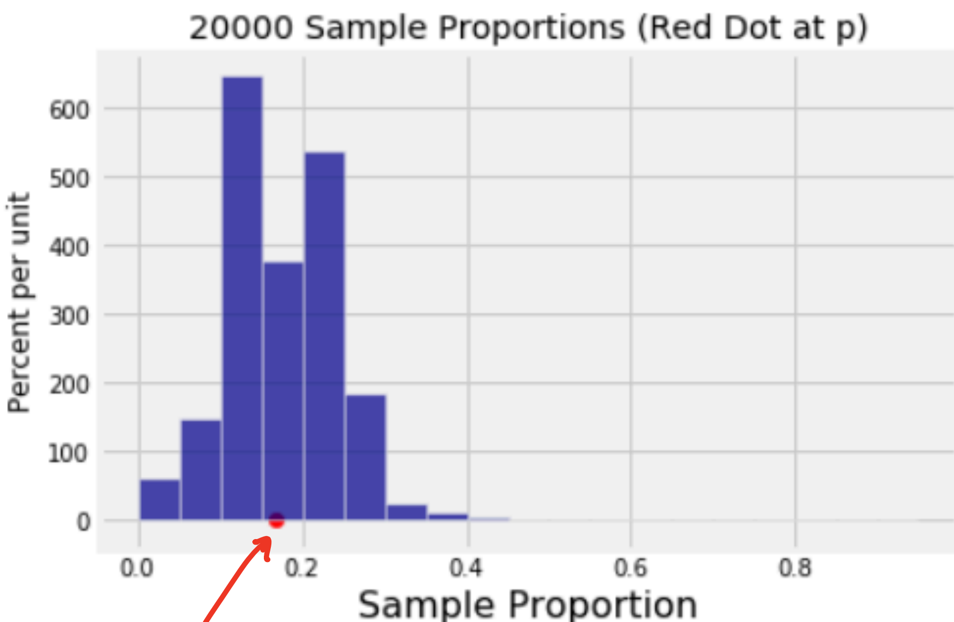
ex You roll a die 30 times and find the sample proportion of sixes.

The population consists of $\boxed{0\ 0\ 0\ 0\ 0\ 1}$ (1 2 3 4 5 6)

Repeat experiment 20,000 times and plot distribution of sample proportions.

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

sampling distribution



p since E(sample proportion) = p

Estimating the largest possible value

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Uniform}\{1, 2, \dots, N\}$ for some fixed but unknown N.

To estimate N you may think $M = \max(X_1, \dots, X_n)$ and this is an estimator but lets find a different one:

The pop mean is $\frac{1+N}{2}$ and $E(\bar{X}) = \frac{1+N}{2}$ since it is unbiased.

what is estimator s.t

$$E(\text{estimator}) = N$$

$$2E(\bar{x}) = 1+N$$

$$\underbrace{2E(\bar{x}) - 1}_{E(2\bar{x}-1)} = N$$

so $\boxed{2\bar{x}-1}$ is an unbiased estimator of N.