STAT 88: Lecture 13

Contents

Warm up:

(a) (X_1, X_2) has joint distribution:

	$X_2 = 0$	$X_2 = 1$
$X_1 = 0$	$\frac{5}{36}$	$\frac{25}{36}$
$X_1 = 1$	$\frac{1}{36}$	$\frac{5}{36}$

Are X_1 and X_2 independent?

(b) A die is rolled 10 times. Find the expectation of the number of times an odd number of spots appears.

Last time

The expectation of a random variable X, denoted $E(X)$, is the average of the possible values of X weighted by their probabilities:

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

Recall (Bernoulli (indicator) random variable)

$$X = \begin{cases} 1 & \text{with probability } p \quad \leftarrow \text{success} \\ 0 & \text{with probability } 1-p \quad \leftarrow \text{failure} \end{cases}$$

Then

$$E(X) = 1 \cdot p + 0 \cdot (1-p) = p.$$

Additivity of Expectation $E(X_1 + X_2) = E(X_1) + E(X_2)$.

Method of indicator to find $E(X)$

Counting the number of successful trials is the same as adding zeros and ones.

Example: A success is blue and failure non blue.

B R R G B R B B
1 0 0 0 1 0 1 1

$$\#\text{blue} = 1 + 0 + 0 + 0 + 1 + 0 + 1 + 1 = 4.$$

Suppose a trial is blue with probability p. Find the expected # blue in n trials.

Step 1: Write down what X is.

$X = \#$ trials out of n that are blue. $\sim \text{Binomial}(n, p)$

Step 2: Find I_j (jth trial).

$$I_j = \begin{cases} 1 & \text{if } j\text{th trial is blue} \quad \leftarrow p \\ 0 & \text{otherwise} \end{cases}$$

Step 3: Find p.

Step 4: Write X as a sum of indicators:

$$X = I_1 + I_2 + \cdots + I_n.$$

Step 5: Find $E(X)$.

$$E(X) = E(I_1 + I_2 + \cdots + I_n) = E(I_1) + E(I_2) + \cdots + E(I_n) = nE(I_1) = np.$$

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

5.3. Method of Indicators (Continued)

Example: Let X be the number of spades in 7 cards dealt with replacement from a well shuffled deck of 52 cards containing 13 spades. Find $E(X)$.

Step 1: Write down what X is.

Step 2: Find I_j (jth trial).

Step 3: Find p.

Step 4: Write X as a sum of indicators.

Step 5: $E(X) =$

Example: Let X be the number of spades in 7 cards dealt without replacement from a well shuffled deck of 52 cards containing 13 spades. Find $E(X)$.

If X is not binomial or hypergeometric be thoughtful how define your indicator. You want each indicator to have same p.

Example: (Exercise 5.7.6) A die is rolled 12 times. Find the expectation of

(a) the number of times the face with five spots appears.

(c) the number of faces that don't appear.

Example: 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 n.

(b) The expected number of matches is 1

(c) The number of matches is hypergeometric.

(d) More than one of the above.

Example: A building has 10 floors above the basement. If 12 people get into an elevator at the basement, and each chooses a floor at random to get out, independently of the others, at how many floors do you expect the elevator to make a stop to let out one or more of these 12 people?

5.4. Unbiased Estimators

Preliminary: Linear Function Rule Let X be a random variable and let $Y = aX+b$. Then Y is a linear function of X. Then

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

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

In the context of estimation, a **parameter** is a fixed number associated with the population.

A **statistic** is a number computed based on the data in your sample.

If a statistic is being used to estimate a parameter, the statistic is called an **estimator** of the parameter.

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. The sample average is the statistic $\bar{X}$ defined as the function

$$\bar{X} = g(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Important:

$\bar{X}$ is unbiased if $E(\bar{X}) = \mu$. In fact,

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