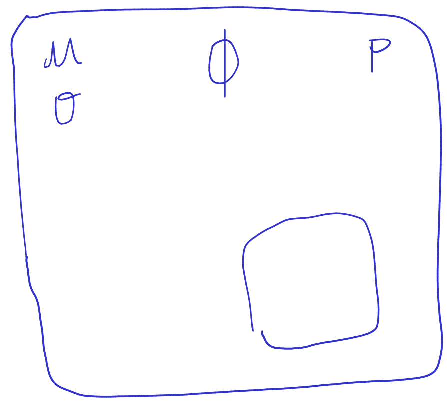
Discrete Distributions - Bernoulli and Binomial

David Armstrong

UCI

Distributions - General Form

- Statisticians have been studying distributions for a long time.
- We are trying to find the probability of seeing a specific observation from a sample, given some population parameters.
- We say $X \sim Distribution(\Phi)$ $\longrightarrow$ x follows distribution
- We define Φ as representing the population parameters.



Our first look at a distribution.

- Set $X = \{0, 1\}$.
- Let $P(X = 1) = p$ and $P(X = 0) = 1 - p$.
- Calculate the expectation of X.

$$E(X) = \sum_{x \in S_X} x f(x) = 0(1-p) + 1(p) = p$$

x	0	1	T
$f(x)$	$1-p$	p	1

- Calculate the variance of X.

$$E(X^2) = \sum_{x \in S_X} x^2 f(x) = 0^2(1-p) + 1^2(p) = p$$

$$VAR(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p)$$

The Bernoulli Distribution

- Assume we perform one event.
- If we set $X = \{0, 1\}$.
- Let p be the probability of success.
- Let $P(X = 1) = p$ and $P(X = 0) = \boxed{1 - p}$. q
- Then we say X follows a Bernoulli distribution with parameter p
- Denoted $X \sim Bernoulli(p)$
- $f(x) = P(X = x) = p^x(1 - p)^{1-x}$
  - $\mathrm{E}(X) = p$.
  - $\mathrm{Var}(X) = p(1 - p)$. $p\,q$

$P(x=0) = p^0(1-p)^{1-0} = 1-p$

$P(x=1) = p^1(1-p)^{1-1} = p$

Example: Assume we perform an experiment with one trial where probability of success is 42%. Write out a relative frequency table for the event.

x	0	1	T
$f(x)$	0.58	0.42	1

$x \sim Bernoulli\ (p = 0.42)$

$x \sim Bernoulli\ (0.42)$

Binomial Distribution Motivation

- **Example:** Assume we perform an experiment with 4 repeated trials where the probability of success for each trial is p.

- What is the support of X?

$$S_X = \{0, 1, 2, 3, 4\}$$

- How many total outcomes are there?

$$n = 2^4 = 16$$

- Draw out a table of all possible outcomes for each element in the support.

$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$
qqqq	pqqq	ppqq	pppq	pppp
	qpqq	pqpq	ppqp	
	qqpq	pqqp	pqpp	
	qqqp	qppq	qppp	
		qpqp		
		qqpp		
$\binom{4}{0} p^0 q^4$	$\binom{4}{1} p^1 q^3$	$\binom{4}{2} p^2 q^2$	$\binom{4}{3} p^3 q^1$	$\binom{4}{4} p^4 q^0$

- Below each column, write probability of each element in the **Support**

The Binomial Distribution

- The binomial random variables with n trials and p parameter can be characterized as the number of success' in n independent trials.
- We define the probability of success on any given trial to be p, and the probability of no success on a given trial to be $1-p$.
- The probability of each independent event does not change. p is constant
- Then we say X follows a Binomial distribution with parameters n and p.
- Denoted $X \sim Binomial(n, p)$
  - $\mathrm{E}(X) = np$.
  - $\mathrm{Var}(X) = np(1-p)$. npq

pmf
- $f(x) = P(X = x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} \qquad \text{for } x = 0, 1, 2, ..., n.$

R Code

- The pmf of the binomial is dbinom(x, n, p) = P(X=x).
- The cdf of the binomial is pbinom(x, n, p) = P(X≤x).

Examples of Binomial random variables.

- The number of heads to show up when flipping a fair coin 1000 times. ($n = 1000$ and $p = 0.5$).
  n
  independent
- A company has 123 employees. All employees are independent of one another, and the probability that a single employee has certification is p. Then the number of employees that have certificates (out of the 123) is a binomial random variable.
- You roll a fair die 21650 times. The number of times a 3 shows is a binomial random variable. Here, $p = \frac{1}{6}$ is the probability of seeing a 3 on a given roll. The probability of not seeing a 3 (1,2,4,5, or 6) is $1 - p = \dfrac{5}{6}$.

$X \sim B(22, \frac{1}{4})$

Binomial Distribution - Set Values Example

Let X be a random variable that follows a binomial distribution with $n = 22$ and $p = \frac{1}{4}$

- What is the distribution of X

  $X \sim B(22, \frac{1}{4})$

- Calculate the expectation of X.

  $E(X) = np = 22(0.25) = 5.5$

- Calculate the variance of X.

  $VAR(X) = npq = 22(0.25)(0.75) = 4.125$

- Calculate the standard deviation of X.

  $\sigma = \sqrt{VAR(X)} = \sqrt{4.125} = 2.0310$

- What is the probability of seeing 5 successes?

  $P(X=5) = \text{dbinom}(5, 22, 0.25)$ ← R code

  $= 0.1933$

- What is the probability that we see less than 5 successes?

  $P(X<5) = P(X \leq 4) = \text{pbinom}(4, 22, 0.25)$

  $= 0.3235$

- What is the probability that we see at most 3 successes?

  $P(X \leq 3) = \text{pbinom}(3, 22, 0.25)$

  $= 0.1624$

- What is the probability that we see at least 4 successes?

  $P(X \geq 4) = 1 - P(X<4) = 1 - P(X \leq 3)$

  $= 1 - 0.1624 = 0.8376$

Binomial Distribution - Wireless Households

Example According to CTIA, 32% of all U.S. housholds are wireless-only households, meaning they have no landline. In a random sample of 20 households, what is the probability that:

$n = 20 \quad p = 0.32$

- Exactly 5 are wireless-only?

$P(X=5) = \text{dbinom}(5, 20, 0.32)$

$= 0.1600$

- Fewer that 3 are wireless-only?

$P(X<3) = P(X \leq 2) = \text{pbinom}(2, 20, 0.32)$

$= 0.0235$

- How many households in your sample would you expect to be wireless-only?

$E(X) = np = 20(0.32)$

$= 6.4$ households

- What is the standard deviation of homes in your sample that would be wireless-only?

$\sigma = \sqrt{VAR(X)} = \sqrt{npq} = \sqrt{20(0.32)(0.68)}$

$= 2.0861$ households