

Math 3070, Applied Statistics

Section 1

September 9, 2019

Section 3.3

- Independent Random Variables
- Binomial Probability Distribution
- Binomial Mean and Variance
- Examples

Independent Random Variables, Definition

Two discrete random variables X and Y are **independent** if for every pair of x and y ,

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

Multiple discrete random variables X_1, \dots, X_n are **independent** if for all n -tuples (x_1, \dots, x_n) ,

$$P(X_1 = x, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n).$$

Not the current topic, but needed to explain binomial random variables. Will not ask questions about the independence of random variables for now, only independence of events. Will go into more detail in later sections.

Binomial Distributions, Definition

Let Y_1, Y_2, \dots, Y_n be independent Bernoulli random variables with parameter p . Then their sum

$$X = Y_1 + Y_2 + \dots + Y_n$$

is a **binomial random variable** with parameters n and p . We write $X \sim \text{Bin}(n, p)$

In other words, given a sequence of n independent trials, each with probability p of success, the binomial random variable X counts the number of successes. The possible values of X are $0, 1, \dots, n$. For example, if we toss a fair coin 3 times, then the number of heads X is a binomial random variable with parameters $n = 3$ and $p = .5$.

Binomial Distributions, Derivation of PMF

Goal: Derive the PMF of a $X \sim \text{Bin}(n, p)$.

- 1 $X = Y_1 + \cdots + Y_n$ with $Y_i \sim \text{bern}(p)$.
- 2 Notice that the probability of observing a specific outcome with x successes can be calculated using the independence of the Y_i ,

$$\begin{aligned} &P(Y_1 = 1, \dots, Y_x = 1, Y_{x+1} = 0, \dots, Y_n = 0) \\ &= P(Y_1 = 1) \cdots P(Y_x = 1) P(Y_{x+1} = 0) \cdots P(Y_n = 0) \\ &= p^x (1 - p)^{n-x} \end{aligned}$$

Binomial Distributions, Derivation of PMF

Goal: Derive the PMF of a $X \sim \text{Bin}(n, p)$.

- 4 The Binomial Random Variable only counts the number of success, it does not care about their order. From a set of n Bernoulli trials, the number of outcomes with x successes is $\binom{n}{x}$ or the number of ways to pick x successes from n trials in any order.
- 5 Each trail with x successes is a distinct outcome. Each have the same probability. The event $X = x$ is the sum of the probabilities of the disjoint outcomes.

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

- 6 Account for all values of X .

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x}, & x = 1, \dots, n \\ 0, & \text{otherwise} \end{cases}$$

Binomial Distributions, Example

Suppose that a person's month of birth equally likely out of the twelve and independent from other birth months of others. What is probability that a three people out of 10 share the birth month of December?

Correspond $i = 1, \dots, 10$ with a distinct person.

Consider

$$Y_i = \begin{cases} 1, & \text{if person } i \text{ was born in December} \\ 0, & \text{otherwise} \end{cases}$$

For each Y_i , $Y_i \sim \text{bern}(1/12)$. Define $X = Y_1 + \dots + Y_{10}$.

Note that $X \sim \text{bin}(10, 1/12)$.

$P(\text{"three out of the ten are born in December"})$

$$= P(X = 3) = \binom{10}{3} (1/12)^3 (11/12)^7 \approx 0.0377674429$$

Binomial Distributions, Summary

- If Y_1, \dots, Y_n are independent $bern(p)$ random variables and $X = Y_1 + \dots + Y_n$, then $X \sim bin(m, p)$.
- If $X \sim bin(m, p)$ then the PMF is

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 1, \dots, n \\ 0, & \text{otherwise} \end{cases}$$

- Appears when counting successes of independent Bernoulli trials.

Binomial Mean and Variance, Mean Derivation

Goal: Compute $E(X)$ where $X \sim \text{bin}(n, p)$.

① $X = Y_1 + \cdots + Y_n$ where $Y_i \sim \text{bern}(p)$.

②

$$\begin{aligned} E(X) &= E(Y_1 + \cdots + Y_n) \\ &= E(Y_1) + \cdots + E(Y_n) && \text{(use linearity of } E) \\ &= p + \cdots + p \\ &= np && (E(Y_i) = p \text{ from last chapter}) \end{aligned}$$

③ $E(X) = np$

Note, we have not discussed linearity of expected value with multiple random variables, used in step 2. Will return to this later. Alternatively, you can directly apply the definition of $E(X)$.

Binomial Mean and Variance, Variance

We should discuss much more independent random variables before discussing this computation. Without more on independent RVs, this calculation is annoying, using the definition of $\text{Var}(X)$. Will address this calculation later in the course.

If $X \sim \text{bin}(n, p)$,

$$\text{Var}(X) = np(1 - p).$$

IOU proof.

Binomial Mean and Variance, Example

You divide your farm plots into groups of five. Each group is independent and they all have the same probability of failing to yield. Your family knows that expected number of failed fields from any group is 0.4. What is the variance?

$X = \text{"number of failed fields in a group"} , \quad X \sim \text{bin}(5, p)$

$$0.4 = E(X) = np = 5p \rightarrow p = 0.4/5 = 0.08$$

$$\text{Var}(X) = np(1 - p) = 5 \cdot 0.08(1 - 0.08) = 0.368$$

Binomial Mean and Variance, Summary

- If $X \sim \text{bin}(n, p)$ then

$$E(X) = np \text{ and } \text{Var}(X) = np(1 - p).$$

- Can translate statistical information into a model. Of course, we need to discuss estimation.
- Missing proofs. Will return to them after discussing independence.

Examples, Missing Information

Return to the potato example, $X \sim \text{bin}(5, p)$. Assume $E(X)$ is not known and $\text{Var}(X) = 1$. Solve for $E(X)$.

$$\text{Var}(X) = np(1 - p) = 5p(1 - p) = 1 \rightarrow p(1 - p) = 1/5$$

$$-p^2 + p - 1/5 = 0 \rightarrow p = \frac{1}{2} \pm \frac{1}{2\sqrt{5}} \approx 0.276, 0.723$$

Two solutions with different behaviors!

Examples, Complement Example

Suppose that **one in twenty five Tweets** contain false information. For a survey, a researcher collects **1000 independent tweets**. What is the probability that 3 or more of them contain false information?

Tweets with false information out of 1000 = $X \sim \text{bin}(1000, 1/25)$

$$P(\text{3 or more with false information}) = P(X \geq 3) = 1 - P(X < 3)$$

$$\begin{aligned} P(X < 3) &= P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{1000}{0} \left(\frac{1}{25}\right)^0 \left(\frac{24}{25}\right)^{1000} + \binom{1000}{1} \frac{1}{25} \left(\frac{24}{25}\right)^{999} \\ &\quad + \binom{1000}{2} \left(\frac{1}{25}\right)^2 \left(\frac{24}{25}\right)^{998} \approx 1.697 \times 10^{-15} \end{aligned}$$

Examples, Complement Example

Suppose that one in twenty five Tweets contain false information. For a survey, a researcher collects 1000 independent tweets. What is the probability that 3 or more of them contain false information?

$$\begin{aligned}P(X \geq 3) &= 1 - P(X \leq 2) \\ &\approx 1 - 1.697 \times 10^{-15}\end{aligned}$$

Examples, Conditioning Example

There is a 0.5% chance that a machine breaks on any given day. When the machine breaks in runs 10 processes with a 95% chance of failure. What is the probability that the machine breaks and completes exactly 1 process without failure?

event A = "machine breaks", X = "number of successful processes"

$$X|A \sim \text{bin}(10, 1 - 0.95)$$

Note, this notation means X conditioned on the right event.

$$P(X = 1, A) = P(X = 1|A)P(A) = \binom{10}{1} 0.05^1 0.95^9 0.005$$

$$P(X = 1, A) \approx 0.0015756235243115234375$$

Note, " $X = 1, A$ " means both events occur, the intersection.

Examples, Law of Total Prob. Example

Suppose that with 25% probability four fair coins are flipped otherwise twelve fair coins are flipped. What is probability that exactly two heads are observed?

event A = "four coins flipped", event B = "twelve coins flipped"

X = "number of heads"

$$X|A \sim \text{bin}(4, 0.5), \quad X|B \sim \text{bin}(12, 0.5)$$

Note, this notation means X conditioned on the right event.

Examples, Law of Total Prob. Example

Suppose that with 25% probability four fair coins are flipped otherwise twelve fair coins are flipped. What is probability that exactly two heads are observed?

$$\begin{aligned}P(X = 2) &= P(X = 2|A)P(A) + P(X = 2|B)P(B) \\&= \left[\binom{4}{2} 0.5^2 (0.5^2) \right] 0.25 + \left[\binom{12}{2} 0.5^2 (0.5^{10}) \right] 0.75 \\&\approx 0.1058349609375\end{aligned}$$

Midterm September 18, Information

- There is a midterm on September 18th in class. Calculator and notes allowed.
- Study the quizzes, summary slides, and homework. See the Canvas 'Files' tab for information.
- Review on September 16th. Come with questions.
- Reschedule by Wednesday, September 11, if needed. No makeup for late exams.