STEVENS
INSTITUTE OF TECHNOLOGY
1870

MA 540: Introduction to Probability Theory

Lecture 4: Discrete random variables

Graph coloring – Probabilistic reasoning without finding an explicit solution

- The complete graph having n vertices is defined to be a set of n points (called vertices) in the plane and the $\binom{n}{2}$ lines (called edges) connecting each pair of vertices. Suppose now that each edge in a complete graph having n vertices is to be colored either red or blue. For a fixed integer k, a question of interest is, Is there a way of coloring the edges so that no set of k vertices has all of its $\binom{k}{2}$ connecting edges the same color?

Beginning of proof:

- Let every edge be i.i.d. selected to be R or B with probability ½
- Number each set of *k* vertices 1, 2, ... , $\binom{n}{k}$ - there are this many unique sets of *k* vertices
- Let A_i be the event that all edges connecting vertices in the i^{th} set of *k* vertices are the same color
- Rest shown in class last week

Question from previous midterm

Assume at the time you took a Covid test, 20% of the population has Covid, and 10% of the population has the flu, and the rest of the population is healthy. Nobody has both Covid and the flu.

1. When you have Covid, the test is 100% accurate. When you are healthy, the test is 70% accurate. When you have the flu, you will test positive for Covid 20% of the time.

If you test positive, what is the probability you have Covid? (You can leave this value as a fraction)

2. Assume you take a second test that uses ***a different detection mechanism*** from the first test. When you have Covid, the second test is 95% accurate. When you are healthy, the test is accurate 2/3 of the time. When you have the flu, you will test positive for Covid 50% of the time.

If you also test positive on this second test, what is the probability you are healthy? (You can leave this value as a fraction)

Random Variable Definition

- A function from possible outcomes in a sample space to a measurable space, that is, to one where we have probability measure on the set of possible values of the random variables.

- We try to distinguish between the random variable, which is denoted as a capital (e.g. X) and represents the possible outcomes of the experiment, from the value of particular outcome, which is denoted as a small letter (e.g. x)
  - The term $P(X = x)$ translates to " The probability that random variable X takes on the value x"
  - If X is a dice roll, $P(X = 5) = 1/6$
  - Can think of X as the daily high temperature in July in Hoboken, $P(X > 100)$ means the probability the temperature exceeds 100

- As an RV is a function from sample space to a measurable space, we must define the probabilities of all possible values in the image of the function

Example

A life insurance agent has 2 elderly clients, each of whom has a life insurance policy that pays \$100,000 upon death. Let Y be the event that the younger one dies in the following year, and let O be the event that the older one dies in the following year.

Assume that Y and O are independent, with respective probabilities $P(Y) = .05$ and $P(O) = .10$. If X denotes the total amount of money (in units of \$100, 000) that will be paid out this year to any of these clients' beneficiaries, then X is a random variable that takes on one of the possible values 0, 1, 2 with respective probabilities:

$P\{X = 0\} = P(Y^c O^c) = P(Y^c)P(O^c) = (.95)(.9) = .855$

$P\{X = 1\} = P(Y O^c) + P(Y^c O) = (.05)(.9) + (.95)(.1) = .140$

$P\{X = 2\} = P(YO) = (.05)(.1) = .005$

Example (truncated geometric distribution)

Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total of n flips is made. If we let X denote the number of times the coin is flipped, then X is a random variable taking on one of the values 1, 2, 3, … , n with respective probabilities

$P\{X = 1\} = P\{H\} = p$

$P\{X = 2\} = P\{(T, H)\} = (1 - p)p$

$P\{X = 3\} = P\{(T,T, H)\} = (1 - p)^2 p$

…

$P\{X = n - 1\} = (1 - p)^{n-2}\, p$

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

- (the result of the last flip doesn't matter

Note these outcomes are disjoint so

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n}\{X = i\}\right) &= \sum_{i=1}^{n} P\{X = i\} \\
&= \sum_{i=1}^{n-1} p(1-p)^{i-1} + (1-p)^{n-1} \\
&= p\left[\frac{1-(1-p)^{n-1}}{1-(1-p)}\right] + (1-p)^{n-1} \\
&= 1 - (1-p)^{n-1} + (1-p)^{n-1} \\
&= 1
\end{aligned}
$$

Same logic as in gambler's ruin problem

Discrete RV

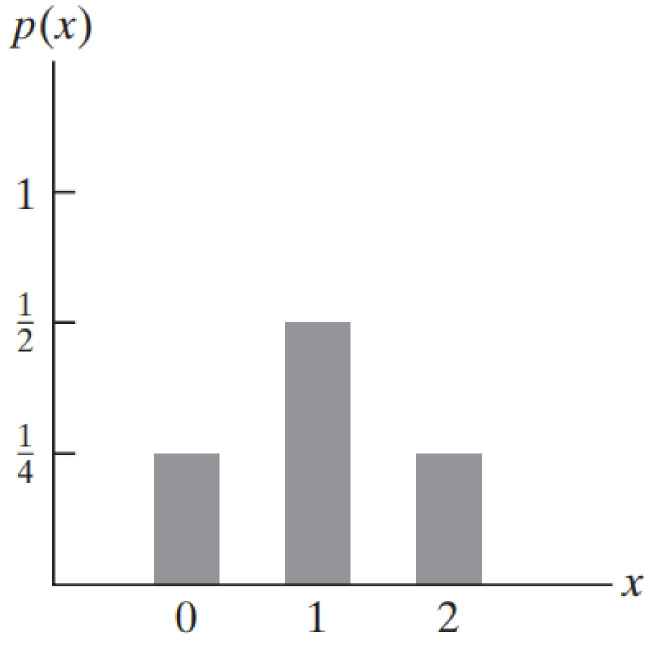
- A random variable that can take on **at most a countable number of possible values** is said to be discrete. For a discrete random variable X, we define the probability mass function p(a) of X by p(a) = P{X = a}

- The probability mass function p(a) is positive for at most a countable number of values of a. That is, if X must assume one of the values x_1, x_2, … , then
- $p(x_i) > 0$ for i = 1, 2, …
- p(x) = 0 for all other values of x

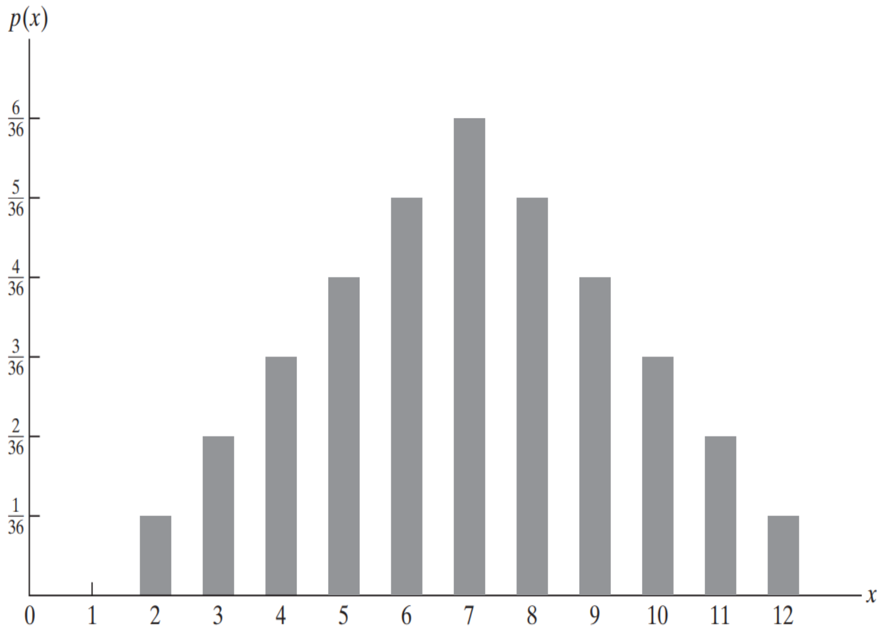
- Since X must take on one of the values x_i, we have $\sum_{i=1}^{\infty} p(x_i) = 1$

Graphs of pmfs

$$p(0) = \frac{1}{4} \qquad p(1) = \frac{1}{2} \qquad p(2) = \frac{1}{4}$$



Sum of 2 dice



Example

- The probability mass function of a random variable X is given by

p(i) = $\frac{c\lambda^i}{i!}$ for i = 0, 1, 2, … where λ is some positive value.

- (a) Find P{X = 0}
- (b) Find P{X > 2}

Hint 1: Figure out the value of c

Hint 2: the pmf for any RV has the property $\sum_{i=1}^{\infty} p(x_i)$ = 1

Solution

- The probability mass function of a random variable X is given by

p(i) = $\frac{c\lambda^i}{i!}$ for i = 0, 1, 2, ... where λ is some positive value.

- (a) Find P{X = 0}
- (b) Find P{X > 2}
- We know $\sum_{i=0}^{\infty} \frac{c\lambda^i}{i!}$ = 1 and that $\sum_{i=0}^{\infty} \frac{x^i}{i?} = e^x$ so c = $e^{-\lambda}$

Hence P(X =0) = $\frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda}$

- P(X > 2) = 1- P(X ≤ 2) = 1- P(X = 0) - P(X = 1) - P(X = 2)

= 1 - $e^{-\lambda}$ - $e^{-\lambda}\lambda - \frac{e^{-\lambda}\lambda^2}{2}$

Cumulative Distribution Function for discrete RV

- The cumulative distribution function (CDF) F of an RV can be expressed in terms of p(a) by

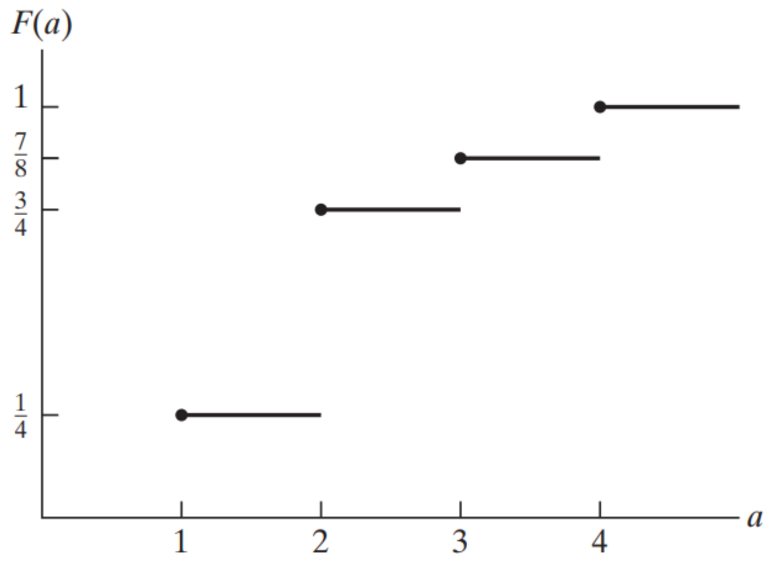
$$F(a) = \sum_{\text{all } x \leq a} p(x)$$

- For example, a random variable X with the following probabilities:

$p(1) = 1/4$ $p(2) = 1/2$ $p(3) = 1/8$ $p(4) = 1/8$

Will have the CDF

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \leq a < 2 \\ \frac{3}{4} & 2 \leq a < 3 \\ \frac{7}{8} & 3 \leq a < 4 \\ 1 & 4 \leq a \end{cases}$$



Using Cumulative Distribution Functions

- The cumulative distribution function F can be expressed in terms of p(a) by $F(a) = \sum_{\text{all } x \leq a} p(x)$
  - Note this is for a discrete RV

- It's easy to calculate the probability of a random variable taking a value over an interval (a, b] as

  $P(a < X \leq b) = F(b) - F(a)$

  - **Note the inclusion of b but not a in the interval**. For intuition, think of a random variable that always takes on single value a with probability 1. Then the formula $F(b) - F(a)$ would give $1 - 1 = 0$, even though the probability that $P(a \leq X \leq b) \geq P(X = a) = 1$

Properties of CDF

1. F is a nondecreasing function; that is, if $a < b$, then $F(a) \leq F(b)$.
2. $\lim_{b \to \infty} F(b) = 1$.
3. $\lim_{b \to -\infty} F(b) = 0$.
4. F is right continuous. That is, for any b and any decreasing sequence $b_n, n \geq 1$, that converges to b, $\lim_{n \to \infty} F(b_n) = F(b)$.

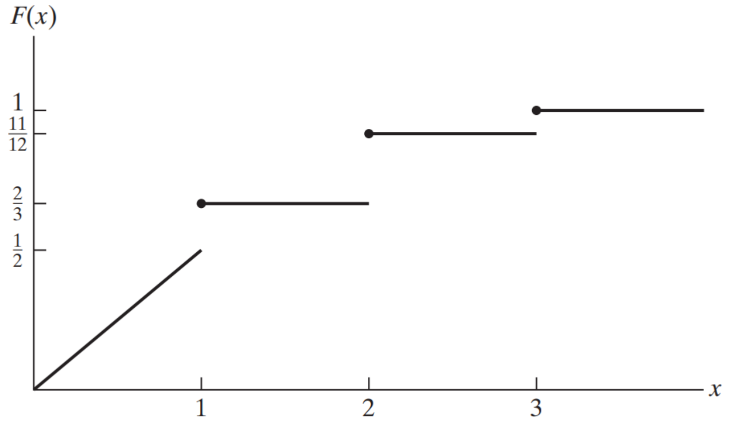
Note: we use rules about the limit of an increasing/ decreasing sequence of events based on what we proved in lecture 2

- 1 comes from the event $\{X \le a\} \subseteq \{X \le b\}$ therefore $P(\{X \le a\}) \le P(\{X \le b\})$
- 2: let b_n be an increasing sequence going to ∞, then the events $\{X \le b_n\}$ are increasing events whose union is $\{X \le \infty\}$, so $\lim_{n\to\infty} P\{X \le b_n\}$ = $P\{X \le \infty\} = 1$
- 3: similar to property 2 using a decreasing sequence
- 4: let b_n be a decreasing sequence going to b, then then the events $\{X \le b_n\}$ are decreasing events whose intersection is $\{X \le b\}$, so $\lim_{n\to\infty} P\{X \le b_n\}$ = $P\{X \le b\}$

Example

- Take a random variable with the following CDF

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \le x < 1 \\ \frac{2}{3} & 1 \le x < 2 \\ \frac{11}{12} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

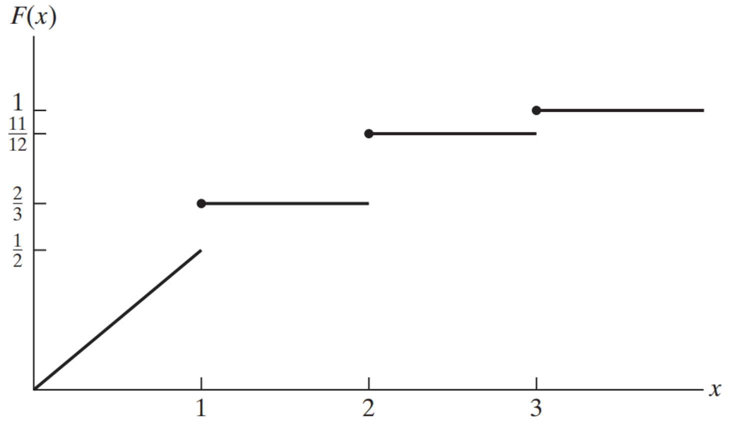


- Compute **(a)** P{X < 3} **(b)** P{X = 1}, **(c)** P{X > 1.2 }, and **(d)** P{2 < X ≤ 4}

Example

- Take a random variable with the following CDF

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \le x < 1 \\ \frac{2}{3} & 1 \le x < 2 \\ \frac{11}{12} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$



- Compute **(a)** P{X < 3} **(b)** P{X = 1}, **(c)** P{X > 1.2 }, and **(d)** P{2 < X ≤ 4}
- A) P{X < 3} = $\lim_n P\{X \le 3 - \frac{1}{n}\} = \lim_n F(3 - \frac{1}{n})$ = 11/12
- B) P{X ≤ 1} – P{X < 1} = F(1) - $\lim_n F(1 - \frac{1}{n})$ = 2/3 – 1/2 = 1/6
- C) 1 - P{X ≤ 1/2} = 1 – F(1/2) = ¾
- D) F(4) – F(2) = 1/12

Expected value of discrete RV

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

- Can be thought of as a weighted average of the possible values of X, weighted by the probability that X assumes that value

- Find E[X], where X is the outcome when we roll a fair die

- We say that I is an indicator variable for the event A if $I = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A^c \text{ occurs} \end{cases}$
  - Find E[I]

Expected value of discrete RV

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

- Can be thought of as a weighted average of the possible values of X, weighted by the probability that X assumes that value

- Find E[X], where X is the outcome when we roll a fair die.
  - (1)(1/6) + (2)(1/6) + (3)(1/6) + (4)(1/6) + (5)(1/6) + (6)(1/6) = 3.5

- 3B We say that I is an indicator variable for the event A if $I = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A^c \text{ occurs} \end{cases}$
  - Find E[I]
  - 1(P(A)) + (0)(1-P(A)) = P(A)

Example

- A school class of 120 students is driven in 3 buses to a symphonic performance. There are 36 students in one of the buses, 40 in another, and 44 in the third bus. When the buses arrive, one of the 120 students is randomly chosen. Let X denote the number of students on the bus of that randomly chosen student, and find $E[X]$.

Example

- A school class of 120 students is driven in 3 buses to a symphonic performance. There are 36 students in one of the buses, 40 in another, and 44 in the third bus. When the buses arrive, one of the 120 students is randomly chosen. Let X denote the number of students on the bus of that randomly chosen student, and find E[X].

$$P\{X = 36\} = \frac{36}{120} \quad P\{X = 40\} = \frac{40}{120} \quad P\{X = 44\} = \frac{44}{120}$$

$$E[X] = 36\left(\frac{3}{10}\right) + 40\left(\frac{1}{3}\right) + 44\left(\frac{11}{30}\right) = \frac{1208}{30} = 40.2667$$

- **friendship paradox:** On average, your friends have more friends than you!

St. Petersburg paradox

A casino offers the following bet:

A fair coin is flipped until it comes up heads. At that point the player wins a prize worth 2^n dollars, where n is the number of times the coin was flipped.

To be more explicit

If H – get $2

If HT – get $4

If HHT – get $8

If HHHT – get $16

…

How much would you pay for the bet above?

Proposition and proof about expectation of a function of a discrete RV

- If X is a discrete random variable that takes on one of the values x_i, $i \geq 1$, with respective probabilities $p(x_i)$, then for any real valued function *g:*

$$E[g(X)] = \sum_i g(x_i)p(x_i)$$

- First: group together all terms in $\sum_i g(x_i)p(x_i)$ having the same value of $g(x_i)$,

where the values are represented by y_j

PROOF

$$\begin{aligned}
\sum_i g(x_i)p(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \\
&= \sum_j \sum_{i:g(x_i)=y_j} y_j p(x_i) \\
&= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) \\
&= \sum_j y_j P\{g(X) = y_j\} \\
&= E[g(X)]
\end{aligned}$$

Corollaries to previous proposition

- If a and b are constants, then E[aX + b] = aE[X] + b

$$\begin{aligned} E[aX + b] &= \sum_{x:p(x)>0} (ax + b)p(x) \\ &= a \sum_{x:p(x)>0} xp(x) + b \sum_{x:p(x)>0} p(x) \\ &= aE[X] + b \end{aligned}$$

- The expected value of a random variable X, E[X], is also referred to as the mean or the first moment of X. The quantity $E[X^n]$, n ≥ 1, is called the nth moment of X. By the earlier Proposition

$$E[X^n] = \sum_{x:p(x)>0} x^n p(x)$$

Lemma before sums of random variables

- Let X(s) denote the value of X when s ∈ *S* is the outcome of the experiment, where *S* is at most countably infinite

Lemma: $$E[X] = \sum_{s \in S} X(s)\, p(s)$$

Denote: $$S_i = \{s : X(s) = x_i\}$$

Proof:

$$\begin{aligned} E[X] &= \sum_i x_i P\{X = x_i\} \\ &= \sum_i x_i P(S_i) \\ &= \sum_i x_i \sum_{s \in S_i} p(s) \\ &= \sum_i \sum_{s \in S_i} x_i p(s) \\ &= \sum_i \sum_{s \in S_i} X(s) p(s) \\ &= \sum_{s \in S} X(s) p(s) \end{aligned}$$

Linearity of Expectation for Sums of RVs

If X is a random variable and Y is a random variable, their sum, Z = X + Y, is also a random variable

For random variables $X_1, X_2, \ldots, X_n$,

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Proof Let $Z = \sum_{i=1}^{n} X_i$. Then, by Proposition 9.1,

On previous slide

$$\begin{aligned}
E[Z] &= \sum_{s \in S} Z(s)p(s) \\
&= \sum_{s \in S} \big(X_1(s) + X_2(s) + \ldots + X_n(s)\big)\, p(s) \\
&= \sum_{s \in S} X_1(s)p(s) + \sum_{s \in S} X_2(s)p(s) + \ldots + \sum_{s \in S} X_n(s)p(s) \\
&= E[X_1] + E[X_2] + \ldots + E[X_n]
\end{aligned}$$

Variance

- Consider the 3 following random variables:

$$W = 0 \quad \text{with probability } 1$$

$$Y = \begin{cases} -1 & \text{with probability } \frac{1}{2} \\ +1 & \text{with probability } \frac{1}{2} \end{cases}$$

$$Z = \begin{cases} -100 & \text{with probability } \frac{1}{2} \\ +100 & \text{with probability } \frac{1}{2} \end{cases}$$

- These all have the same expected value, but their possible values are spread out very differently

- If X is a random variable with mean μ, then the variance of X, denoted by Var(X) is defined as

$$\mathbf{Var(X) = E[(X - \mu)^2]}$$

Equivalent definition of variance

$$
\begin{aligned}
\mathrm{Var}(X) &= E[(X - \mu)^2] \\
&= \sum_x (x - \mu)^2 p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) p(x) \\
&= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - \mu^2
\end{aligned}
$$

$$
\boxed{\mathrm{Var}(X) = E[X^2] - (E[X])^2}
$$

Example

- Calculate Var(X) if X represents the outcome when a fair die is rolled

Example

- Calculate Var(X) if X represents the outcome when a fair die is rolled

- Use $E[X^2] - (E[X])^2$
- $E[X^2] = (1^2)(1/6) + (2^2)(1/6) + (3^2)(1/6) + (4^2)(1/6) + (5^2)(1/6) + (6^2)(1/6) = 91/6 = 182/12$
- $(E[X])^2 = (7/2)^2 = 49/4 = 147/12$

- $Var(X) = 35/12$

Variance when multiplying and adding RVs, Standard deviation

- $\text{Var}(aX + b) = a^2\,\text{Var}(X)$

$$\begin{aligned}\text{Var}(aX + b) &= E[(aX + b - a\mu - b)^2] \\ &= E[a^2(X - \mu)^2] \\ &= a^2 E[(X - \mu)^2] \\ &= a^2 \text{Var}(X)\end{aligned}$$

- However, if we take *a* different i.i.d. random variables $X_1, X_2, \ldots X_a$, $\text{Var}(\sum_{i=1}^{a} X_i) = a\text{Var}(X)$
  - Not a^2 Var(X)
  - Think 100 fair \$1 bets vs 1 fair \$100 bet.
  - In the latter case, you either win or lose \$100.
  - The former case looks like a binomial(n=100, p= .5) distribution minus \$50. You are much more likely to have results near 0 than near -100 or 100.
- **The square root of Var(X) is called the standard deviation of X, and we denote it by SD(X).**

Bernoulli Random variable

- An experiment where there are only two possible outcomes: success or failure
  - Success: $X = 1$
  - Failure: $X = 0$

- The probability mass function of X is:
  - $P(1) = p$
  - $P(0) = 1-p = q$

- Different experiments may have different rates of success (p), for example people get an easier question right more frequently than a harder question

Binomial random variable

- Now imagine we have n independent trials, all with probability p
  - The whole class tries to answer the same question, without collaborating

- If X represents the number of successes that occur in the n trials, then X is said to be a binomial random variable with parameters (n, p)
  - A Bernoulli RV is a Binomial(1, p)

- The pmf of a Binomial(n, p) is: $p(i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad i = 0, 1, \ldots, n$
  - The probability of any sequence of *i* successes and *n-i* failures is $p^i (1-p)^{n-i}$

- From the binomial theorem: $$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = [p + (1-p)]^n = 1$$

Example

- Five fair coins are flipped. If the outcomes are assumed independent, find the probability mass function of the number of heads obtained.

Binomial RV Example

- Five fair coins are flipped. If the outcomes are assumed independent, find the probability mass function of the number of heads obtained.

$$P\{X=0\} = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

$$P\{X=1\} = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = \frac{5}{32}$$

$$P\{X=2\} = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{32}$$

$$P\{X=3\} = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{10}{32}$$

$$P\{X=4\} = \binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 = \frac{5}{32}$$

$$P\{X=5\} = \binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 = \frac{1}{32}$$

Binomial RV Example 2

- Suppose that a particular trait (such as eye color or left-handedness) of a person is classified on the basis of one pair of genes, and suppose also that d represents a dominant gene and r a recessive gene. Thus, a person with dd genes is purely dominant, one with rr is purely recessive, and one with rd is hybrid. The purely dominant and the hybrid individuals are alike in appearance. Children receive 1 gene from each parent. If, with respect to a particular trait, 2 hybrid parents have a total of 4 children, what is the probability that 3 of the 4 children have the outward appearance of the dominant gene?

Example

- Suppose that a particular trait (such as eye color or left-handedness) of a person is classified on the basis of one pair of genes, and suppose also that d represents a dominant gene and r a recessive gene. Thus, a person with dd genes is purely dominant, one with rr is purely recessive, and one with rd is hybrid. The purely dominant and the hybrid individuals are alike in appearance. Children receive 1 gene from each parent. If, with respect to a particular trait, 2 hybrid parents have a total of 4 children, what is the probability that 3 of the 4 children have the outward appearance of the dominant gene?

Assume each child is equally likely to inherit each gene from either parent, so the probability that the child of 2 hybrid parents will have dd = 1/4, rr =1/4, and dr=1/2. That means each child will have probability of outward appearance being dominant of 3/4. Hence, we have a binomial(4, 3/4), so the probability of 3/4 children having dominant appearance is

$$\binom{4}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1 = \frac{27}{64}$$

Expected Value and Variance of Binomial

- If X ~ Binomial(n, p), then E(X) = $E[\sum_{i=1}^{n} X_i]$ = $\sum_{i=1}^{n} E[X_i]$ = $\sum_{i=1}^{n} p$ = np

- If X~ Binomial(n, p), then Var(X) = $Var[\sum_{i=1}^{n} X_i]$ = $\sum_{i=1}^{n} Var[X_i]$
- Question is: what's the variance of a single Bernoulli(p)?
  - $E(X^2) – (E[X])^2$ = $p – p^2$ = p(1-p)
  - Alternatively:
    - E[X] is p.
    - With probability p, $(X-E[X])^2 = (1-p)^2$
    - With probability (1-p), $(X-E[X])^2 = p^2$
    - In expectation: $(X-E[X])^2 = p(1-p)^2 + (1-p)p^2 = p(1-p)$
- Hence: Var(X) = np(1-p)

- Textbook shows an alternative derivation for both of these quantities

Proposition

- If X is a binomial random variable with parameters (n, p), where $0 < p < 1$, then as k goes from 0 to n, $P\{X = k\}$ first increases monotonically and then decreases monotonically, reaching its largest value when k is the largest integer less than or equal to $(n + 1)p$.

$$\frac{P\{X = k\}}{P\{X = k - 1\}} = \frac{\dfrac{n!}{(n-k)!k!}p^k(1-p)^{n-k}}{\dfrac{n!}{(n-k+1)!(k-1)!}p^{k-1}(1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)}$$

- $(n-k+1)p \geq k(1-p) \leftrightarrow$ if $k \leq (n+1)p \leftrightarrow P(X=k) \geq P(X=k-1)$

Poisson RV

- p(i) = P{X =i} = $e^{-\lambda}\frac{\lambda^i}{i!}$

We've already shown this will sum to 1 for any positive λ

- History
  - number of wrongful convictions in a given country
  - number of soldiers in the Prussian army killed accidentally by horse kicks

- Uses: Models the number of events that will occur if each event is independent with the same rate
  - Number of people entering a queue over a certain period of time
  - The number of people in a community who survive to age 100

Examples using Poisson: p(i) = $e^{-\lambda}\frac{\lambda^i}{i!}$

- Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = 1/2$. Calculate the probability that there is at least one error on this page.

  - $P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1/2} = .393$

- Consider an experiment that consists of counting the number of α particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that on the average, 3.2 such α particles are given off, what is a good approximation to the probability that no more than 2 α particles will appear?
  - $P\{X \leq 2\} = P\{X = 0\} + P\{X = 1\} + P\{X = 2\} = e^{-3.2} + 3.2e^{-3.2} + ((3.2)^2/2)\, e^{-3.2}$

Expectation and Variance of Poisson RV

$$
\begin{aligned}
E[X] &= \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} \\
&= \lambda \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \quad \text{by letting } j = i - 1 \\
&= \lambda \qquad \text{since } \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}
\end{aligned}
$$

$$
\begin{aligned}
E[X^2] &= \sum_{i=0}^{\infty} \frac{i^2 e^{-\lambda}\lambda^i}{i!} \\
&= \lambda \sum_{i=1}^{\infty} \frac{ie^{-\lambda}\lambda^{i-1}}{(i-1)!} \\
&= \lambda \sum_{j=0}^{\infty} \frac{(j+1)e^{-\lambda}\lambda^j}{j!} \quad \text{by letting } j = i - 1 \\
&= \lambda \left[\sum_{j=0}^{\infty} \frac{je^{-\lambda}\lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \right] \\
&= \lambda(\lambda + 1)
\end{aligned}
$$

- Taking both results together implies E(X) = Var(X) = λ

Poisson as approximation for Binomial

- For n large and moderate p, letting $\lambda = np$, the binomial distribution is the following

$$
\begin{aligned}
P\{X = i\} &= \frac{n!}{(n-i)!\,i!} p^i (1-p)^{n-i} \\
&= \frac{n!}{(n-i)!\,i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}
\end{aligned}
$$

Now, for n large and λ moderate,

$$
\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda} \quad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1 \quad \left(1 - \frac{\lambda}{n}\right)^i \approx 1
$$

Hence, for n large and λ moderate,

$$
P\{X = i\} \approx e^{-\lambda} \frac{\lambda^i}{i!}
$$

Application to birthday problem

- Think of the birthday problem as independent trials where trial i,j is a success when persons i and j have the same birthday
- In a room with n people, there are $\binom{n}{2}$ different trials
- In any given room with n people, we can think that the number of shared birthdays is a Poisson with rate $\binom{n}{2}$/365 = n(n-1)/730
- Using this formula, we can estimate the probability no 2 people have the same birthday as P(X=0) $=\exp\left\{\frac{-n(n-1)}{730}\right\}$
- We can solve for when this is less than 1/2
- $\exp\left\{\frac{-n(n-1)}{730}\right\} < ½$ → $\exp\left\{\frac{n(n-1)}{730}\right\} > 2$ → n(n-1) >730 * log 2 = 505.997
- Solving for n gives 23, which is what we've shown before

Poisson processes

- If the number of events occurring in any fixed interval of length t is a Poisson random variable with mean λt, and we say that the events occur in accordance with a Poisson process having rate λ.
  - There are certain conditions where this is the case, but will not be shown here (see textbook pages 144-145)
- The value λ can be shown to equal the rate per unit time at which events occur, but is treated as a constant that must be empirically determined.

- In other words, if the number of events occurring in an interval t is a Poisson RV with parameter λt, for some nonnegative constant a, the number of events occurring in the interval *at* is a Poisson RV with parameter λat

Example

- Suppose that earthquakes occur in the western portion of the United States in accordance with assumptions to make it a Poisson process, with $\lambda = 2$ and with 1 week as the unit of time. (That is, earthquakes occur at a rate of 2 per week.)
  - (a) Find the probability that at least 3 earthquakes occur during the next 2 weeks.
  - (b) Find the probability distribution of the time, starting from now, until the next earthquake.

Example

- Suppose that earthquakes occur in the western portion of the United States in accordance with assumptions to make it a Poisson process, with λ = 2 and with 1 week as the unit of time. (That is, earthquakes occur at a rate of 2 per week.)
  - (a) Find the probability that at least 3 earthquakes occur during the next 2 weeks.
  - (b) Find the probability distribution of the time, starting from now, until the next earthquake.

A. p(i) = $e^{-\lambda}\frac{\lambda^i}{i!}$

$$P\{N(2) \geq 3\} = 1 - P\{N(2) = 0\} - P\{N(2) = 1\} - P\{N(2) = 2\}$$
$$= 1 - e^{-4} - 4e^{-4} - \frac{4^2}{2}e^{-4}$$
$$= 1 - 13e^{-4}$$

B. $P\{X > t\} = P\{N(t) = 0\} = e^{-\lambda t}$ → $F(t) = P\{X \leq t\} = 1 - P\{X > t\} = 1 - e^{-\lambda t}$
$$= 1 - e^{-2t}$$

Geometric distribution

- Suppose that independent trials, each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs.
  - Can think about limited supplies, lottery tickets, bitcoin hashing

This RV measures the number of trials it takes until there is a success

- $P\{X = n\} = (1 - p)^{n-1} p, \quad n= 1, 2, \ldots$

$$\sum_{n=1}^{\infty} P\{X = n\} = p \sum_{n=1}^{\infty} (1 - p)^{n-1} = \frac{p}{1 - (1 - p)} = 1$$

So this is a valid probability distribution

Geometric Example

- An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each ball selected is replaced before the next one is drawn, what is the probability that
  - (a) exactly n draws are needed?
  - (b) at least k draws are needed?

Example

- An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each ball selected is replaced before the next one is drawn, what is the probability that
  - (a) exactly n draws are needed?
  - (b) at least k draws are needed?

A. $\left(\frac{N}{N+M}\right)^{n-1}\left(\frac{M}{N+M}\right)$

B. Easy to see that this is just the probability that the first k-1 draws are white balls so

$$\left(\frac{N}{N+M}\right)^{k-1}$$

Expected value of geometric RV

- Let q = 1-p

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{\infty} i q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i - 1 + 1) q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i - 1) q^{i-1} p + \sum_{i=1}^{\infty} q^{i-1} p \\
&= \sum_{j=0}^{\infty} j q^{j} p + 1 \\
&= q \sum_{j=1}^{\infty} j q^{j-1} p + 1 \\
&= qE[X] + 1
\end{aligned}
$$

$pE[X] = 1 \rightarrow E[X] = 1/p$

Variance of geometric RV

$$
\begin{aligned}
E[X^2] &= \sum_{i=1}^{\infty} i^2 q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i - 1 + 1)^2 q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i - 1)^2 q^{i-1} p + \sum_{i=1}^{\infty} 2(i - 1) q^{i-1} p + \sum_{i=1}^{\infty} q^{i-1} p \\
&= \sum_{j=0}^{\infty} j^2 q^j p + 2 \sum_{j=1}^{\infty} j q^j p + 1 \\
&= qE[X^2] + 2qE[X] + 1
\end{aligned}
$$

$$pE[X^2] = \frac{2q}{p} + 1$$

$$E[X^2] = \frac{2q + p}{p^2} = \frac{q + 1}{p^2}$$

$$\mathrm{Var}(X) = \frac{q + 1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = \frac{1 - p}{p^2}$$

Negative Binomial RV

- Suppose that independent trials, each having probability p, $0 < p < 1$, of being a success are performed until a total of r successes is accumulated.
  - Can think about picking up produce at the grocery store, where you need several of a given item, but some of the produce has gone bad

- P{X = n} = $\binom{n-1}{r-1} p^r (1-p)^{n-r}$ for n=r, r+1, ...

  - There must have been r-1 successes in the first n-1 trials, and the nth trial had to be a success

- The number of trials to each of the r successes is an independent geometric RV, so since each geometric is finite with probability 1, so is their sum, that is the negative binomial RV

Negative Binomial Example

- If independent trials, each resulting in a success with probability p, are performed, what is the probability of r successes occurring before m failures?

Negative Binomial Example

- If independent trials, each resulting in a success with probability p, are performed, what is the probability of r successes occurring before m failures?
  - Sum of probabilities of achieving r successes in the first r trials to the first r+m-1 trials, otherwise m successes must have occurred first

$$\sum_{n=r}^{r+m-1} \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

Expected value and Variance for Negative Binomial

- We can use the sum of independent geometric random variables
- If X~ Negative Binomial(r, p), then E(X) = E[$\sum_{i=1}^{r} X_i$] = $\sum_{i=1}^{r} E[X_i]$ = $\sum_{i=1}^{r} 1/p$ = r/p
- If X~ Negative Binomial(r, p), then Var(X) = Var[$\sum_{i=1}^{r} X_i$] = $\sum_{i=1}^{r} Var[X_i]$ = $\sum_{i=1}^{r}(1-p)/p^2$ = $\frac{r(1-p)}{p^2}$
- Find the expected value and the variance of the number of times one must throw a die until the outcome 1 has occurred 4 times.

Hypergeometric RV

- Suppose that a sample of size n is to be chosen randomly (without replacement) from an urn containing N balls, of which m are white and N – m are black. If we let X denote the number of white balls selected, then

$$P\{X = i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \quad i = 0, 1, \ldots, n$$

- E(X) = nm/N
- Var(X) = $np(1-p)\left(1-\frac{n-1}{N-1}\right)$ where p = m/n

Zipf distribution

$$P\{X = k\} = \frac{C}{k^{\alpha+1}} \quad k = 1, 2, \ldots$$

$$C = \left[\sum_{k=1}^{\infty} \left(\frac{1}{k} \right)^{\alpha+1} \right]^{-1}$$

- Uses (heavy tailed distributions):
  - Distribution of family incomes/wealth
  - Frequency of word use in a language
  - Number of times a website is visited
  - Degree distribution in real world networks

Homework 4

- Due Wednesday 10/11
- Problems from textbook Chapter 4:
  - 4.2
  - 4.13
  - 4.22
  - 4.45
  - 4.52
  - 4.55
  - 4.78
- Theoretical problems from textbook Chapter 4:
  - 4.11