Probability

Applications

Analysis of Algorithms

Average Case Complexity

Monte Carlo Methods

Spam Filters

Basic Concepts

Probability Distribution

- $S = \{a_1, \ldots, a_n\}$ = finite set of outcomes = *sample space*
- $p : S \rightarrow [0,1]$

  -- $p(a_k) \geq 0$

  -- $\sum_{k=1}^{n} p(a_k) = 1$

Events

- An event E is a subset of the possible outcomes S

  -- $\Pr(E) = \sum_{a_i \in E} p(a_i)$

- For equally likely outcomes

  -- $\Pr(E) = \frac{|E|}{|S|}$

<u>**Example -- Dice**</u>

Dice

- Ordered Pairs: 36 possible outcomes (m, n)
- Sums: 11 possible outcomes $\{2, 3, \ldots, 12\}$
- Unordered Pairs: 21 possible outcomes, 6 doubles and 15 non-doubles
- *Moral: Must describe both the experiment and the possible outcomes.*

  -- Ordered Pairs: $\Pr(m, n) = 1/36$

  -- Sums: $\Pr(2) = 1/36,\ \Pr(4) = 3/36, \ldots, \Pr(7) = 6/36$

  -- Unordered Pairs: $\Pr(double) = 6/36,\ \Pr(each\ nondouble) = 2/36$

Example -- Poker

Poker Hands

- # poker hands = $C(52,5)$

- # 3 of a kind = $C(13,1)\times C(4,3)\times 48\times 44$

  -- $\Pr(3\ \ of\ \ a\ \ kind)=\dfrac{C(13,1)\times C(4,3)\times 48\times 44}{C(52,5)}\approx .01$

- # flushes = $C(4,1)C(13,5)$

  -- $\Pr(flush)=\dfrac{C(4,1)\,C(13,5)}{C(52,5)}\approx .00396$

Example -- Bridge

Splits in Bridge

- # opponent bridge hands = $C(26,13)$

- 4 Missing Trumps

  -- 2-2 split = $\dfrac{C(4,2)\,C(22,11)}{C(26,13)} \approx 40\%$

  -- 3-1 split = $2\dfrac{C(4,3)\,C(22,10)}{C(26,13)} \approx 50\%$

  -- 4-0 split = $2\dfrac{C(4,4)\,C(22,9)}{C(26,13)} \approx 10\%$

More Examples

Choose Up Sides

- 10 kids, 5 per team
- # possible teams with player x = $C(9,4)$
- # possible teams with players x and y = $C(8,3)$
- Pr(two friends on same team) = $\dfrac{C(8,3)}{C(9,4)} = \dfrac{4}{9} < \dfrac{1}{2}$

Hatcheck Problem

- # hat permutations: $P(n) = n!$
- # hat derangements: $D(n) = n!\left(1 - \dfrac{1}{1!} + \dfrac{1}{2!} - \cdots\right) \approx \dfrac{n!}{e}$
- $\Pr(\mathit{derangement}) \approx \dfrac{1}{e} = .3679 \qquad (n > 7)$

Complementary Events

Setup

- $S = \{a_1, \ldots, a_n\}$ = finite set of outcomes
- $S \supset E$ = set of possible outcomes
- $E^c = S - E$ = complementary set of outcomes

-- $P(E) = 1 - P(E^c)$

-- $P(E) = \sum_k p(e_k) = 1 - \sum_j p(e_j^c)$

Examples

Coin Tosses

- E = at least one head in n tosses
- E^c = no heads in n tosses

-- $p(E^c) = 1/2^n$

-- $p(E) = 1 - p(E^c) = 1 - 1/2^n = \dfrac{2^n - 1}{2^n}$

Birthday Problem

- E = at least 2 out of n people having same birthday (day and month)
- E^c = no 2 people have the same birthday

-- $p(E^c) = (364/365)(363/365)\cdots\big((365-n)/365\big)$

-- $p(E) = 1 - p(E^c) \Rightarrow p(E) > 1/2$ when $n > 26$

Binomial Distribution

Bernoulli Trials

- t = probability of event (success)
- $1-t$ = probability of nonevent (failure)
- $B_k^n(t)$ = probability of k events (successes) in n *identical, independent trials*

Formulas

- $B_k^n(t) = \binom{n}{k} t^k (1-t)^{n-k}$

  -- t^k = probability of k successes

  -- $(1-t)^{n-k}$ = probability of $n-k$ failures

  -- $\binom{n}{k}$ = number of ways *exactly* k successes can occur in n trials

Examples

Coin Tossing

- t = probability of heads
- $B_k^n(t) = \binom{n}{k} t^k (1-t)^{n-k}$ = probability of exactly k heads in n tosses

Urn Models -- Sampling with Replacement

- w white balls, b black balls
- $t = w/(w+b)$ = probability of selecting a white ball
- $B_k^n(t) = \binom{n}{k} t^k (1-t)^{n-k}$ = probability of selecting exactly k white balls in n trials

Random Walk in Pascal's Triangle

- t = probability of turning right
- $B_k^n(t) = \binom{n}{k} t^k (1-t)^{n-k}$ = probability of landing in *kth* bin at the bottom

<u>**Monte Carlo Methods**</u>

Computation of π

Simulation of Random Walks

Conditional Probability

Conditional Probability

Formula

- $$\Pr(E \mid F) = \frac{\Pr(E \cap F)}{\Pr(F)}$$ provided that $\Pr(F) > 0$

Proof (for equally likely event)

- $$\Pr(E \mid F) = \frac{|E \cap F|}{|F|} = \frac{|E \cap F| / |S|}{|F| / |S|} = \frac{\Pr(E \cap F)}{\Pr(F)}$$

Proof (for arbitrary probability distributions)

- F is the new Sample Space

Observation

- Conditional probability is often tricky -- see below.

Example -- Cards

Playing Cards

- Draw one card out of 52 -- 52 possible outcomes
- $\Pr(red) = \frac{26}{52} = \frac{1}{2}$
- $\Pr(diamond) = \frac{13}{52} = \frac{1}{4}$
  - -- $\Pr(diamond \mid red) = \frac{1/4}{1/2} = \frac{1}{2}$
  - -- $\Pr(red \mid diamond) = \frac{1/4}{1/4} = \frac{1}{1}$
  - -- $\Pr(diamond \mid ace) = \frac{1/52}{4/52} = \frac{1}{4}$

Example -- Coins

Coins

- Flip two pennies (distinguished by dates) -- 4 possible outcomes

  -- $\Pr(2\ heads \mid first\ \ penny\ \ is\ \ a\ \ head) = \dfrac{1/4}{1/2} = 1/2$

  -- $\Pr(2\ heads \mid 1\ head) = ?$

Example -- Coins

Probabilities

- Flip two pennies (distinguished by dates) -- 4 possible outcomes

  -- $\Pr(2\ heads \mid first\ \ penny\ \ is\ \ a\ \ head) = \dfrac{1/4}{1/2} = \dfrac{1}{2}$

  -- $\Pr(2\ heads \mid 1\ head) = \dfrac{1/4}{3/4} = 1/3$ (??? see below)

Possible Events

- Possible Events = $\{HH, HT, TH, TT\}$
- Possible Events with First Penny Head = $\{HH, HT\}$
- Possible Events with at Least One Head = $\{HH, HT, TH\}$

<u>Example -- Children</u>

Boy-Girl

- Boys and Girls are equally likely.
- You ask: *Do you have any boys?* Man responds: *Yes*.
- Man volunteers: *I have two children; one is a boy*.
- Man says: *I have two children: the firstborn is a boy*.

Question

- $\Pr(2 \; boys \mid 1 \; boy) = ?$

Example -- Children

Analysis

- You ask: *Do you have any boys?* Man responds: *Yes*.

  -- $\Pr(2\ boys \mid 1\ boy) = \frac{1}{3}$

  -- Possible Events = $\{BB, BG, GB\}$

- Man volunteers: *I have two children; one is a boy*.

  -- $\Pr(2\ boys \mid 1\ boy) = \frac{1/4}{2/4} = \frac{1}{2}$

  -- Possible Events

  -- BB -- $(1/4)1$

  -- BG -- $(1/4)(1/2)$

  -- GB -- $(1/4)(1/2)$

Analysis (continued)

- Man says: *I have two children: the firstborn is a boy*.

  -- $\Pr(2\ boys \mid 1\ boy) = \dfrac{1}{2}$

  -- Possible Events

    -- BB -- $1/2$

    -- BG -- $1/2$

Moral

- Protocol Matters
- See Teasers Paper -- Probability Depends on the Protocol

Example -- Children

Boy-Girl

- Boys and Girls are equally likely
- Woman says: *I have a girl*.
- Woman says: *I have a girl named Alice*.

Questions

- $\Pr(2\ \ girls \mid 1\ \ girl) = ?$
- $\Pr(2\ \ girls \mid 1\ \ girl\ \ named\ \ Alice) = ?$

Example -- Children

Analysis

- *I have a girl.*
  - -- $\Pr(2\ girls \mid 1\ girl) = \frac{1/4}{3/4} = \frac{1}{3}$
  - -- Possible Events = $\{GG, BG, GB\}$

- *I have a girl named Alice.*
  - -- $\Pr(2\ girls \mid 1\ girl\ named\ Alice) = \frac{1}{2}$
  - -- Possible Events = $\{AB,\ BA,\ AG,\ GA\}$

Conclusion

- Protocol Matters

Information Leaks

Logical Puzzles

- Island of Perfect Logicians
- A Daughter named Alic*e*

Cryptography

- Frequency of Letters in Alphabet
- Timing Channel
- Power Channel
- Subliminal Channels -- Message in the Noise

Example -- Cards

Colored Cards

- Given 2 Cards -- red/red and red/white

- Pick a card at random and select a side at random

Question

- Probability of (2 red | 1 red) = ?

Example -- Cards

Colored Cards

- Given 2 Cards -- red/red and red/white
- Pick a card at random and select a side at random
- Probability of (2 red | 1 red) = $\dfrac{1/2}{3/4} = \dfrac{2}{3}$

  -- RR -- $(1/2)1 = 1/2$

  -- RW -- $(1/2)(1/2) = 1/4$

- More likely to be red on back, since red/red has two chances to land on red.

<u>Example -- Monte Hall Problem</u>

Protocol

- Three Doors
- One Fabulous Prize
- You Pick a Door at Random

Host (Monte Hall)

- Opens a Different Door -- No Prize
- Offers to Trade Your Door for His Remaining Door

Question

- Should you Make the Deal?
- Does it Matter?

Example -- Monte Hall Problem

Analysis

- Pr(Prize Behind Your One Door) = $\frac{1}{3}$
- Pr(Prize Behind His Two Doors) = $\frac{2}{3}$
- Pr(Prize Behind Door He Opens) = 0
- Pr(Prize Behind Door He Does Not Open) = $\frac{2}{3}$
- Solution: Make the Deal!

Monte Hall Problem -- Variations

Variation 1

- You choose your door AFTER Monte Hall opens his door.
- Pr(Prize Behind Your One Door) = $\frac{1}{2}$.
- Switching Doors does NOT Change Your Odd of Winning.

Variation 2

- Monte Hall opens one of his doors at RANDOM.

  -- If Monte finds the prize, you lose.

- -- If Monte does not find the prize,

  Pr(Prize Behind Your One Door) = $\frac{1/3}{2/3} = \frac{1}{2}$.

- Switching Doors does NOT Change Your Odd of Winning.

Example -- Prisoner Problem

Protocol

- Judge sentences Tom or Dick or Harry to hang
  - -- pr(Tom will hang) = 1 / 3
- Tom asks jailer to tell him a name of one of the other two who will NOT be hanged.
- Jailer says: *Dick will not be hanged*.

Questions

- Have the Odds Changed for Tom?
- Does it matter to Tom that Dick will NOT be hanged?
- What is the probability that Tom will be hanged?

Example -- Prisoner Problem

Protocol

- Judge sentences Tom or Dick or Harry to hang
  - -- pr(Tom hanged) = $1/3$
- Tom asks jailer to tell him a name of one of the other two who will NOT be hanged.
- Jailer says: *Dick will not be hanged*.

Analysis

- pr(Tom and ~Dick | ~Dick) = $\dfrac{(1/3)\,(2/3)}{(2/3)} = 1/3$
- pr(Tom will hang) = $1/3 \neq 1/2$

Reason

- Tom hangs $\Rightarrow$ pr(Dick selected) = $1/2$
- Tom does NOT hang $\Rightarrow$ pr(Dick selected) = 1

Bayes' Theorem

Motivation

Problem

- Find $p(F)$ given that E has occurred.
- Find $p(F \mid E)$ if we know $p(E \mid F)$.

<u>**Basic Relations**</u>

Lemma 1: $p(E \cap F) = p(E \mid F)\, p(F)$

Proof: $p(E \mid F) = \dfrac{p(E \cap F)}{p(F)} \Rightarrow p(E \cap F) = p(E \mid F)\, p(F)$

Lemma 2: $p(F \mid E)\, p(E) = P(E \mid F)\, p(F)$

Proof:

- $p(E \cap F) = p(E \mid F)\, p(F)$
- $p(F \cap E) = p(F \mid E)\, p(E)$

$\therefore\ p(F \mid E)\, p(E) = P(E \mid F)\, p(F)$

Bayes' Theorem

Bayes' Formula

- $$p(F \mid E) = \frac{p(E \mid F)\, p(F)}{p(E \mid F)\, p(F) + p(E \mid F^c)\, p(F^c)}$$

Proof: By Lemma 1

- $$p(F \mid E) = \frac{P(E \cap F)}{p(E)} = \frac{p(E \mid F)\, p(F)}{p(E)}$$

But $E = (E \cap F) \cup (E \cap F^c)$ is a disjoint union, so again by Lemma 1

- $$p(E) = p(E \cap F) + p(E \cap F^c) = p(E \mid F)\, p(F) + p(E \mid F^c)\, p(F^c)$$

Example: Tests for Rare Diseases

Notation

- F = the event that a person is sick with a very rare disease
- E = the event that a person tests positive for this rare disease

Protocol

- $p(F) = 1/100{,}000$ very rare disease
- $p(E \mid F) = 99/100$ test correct 99% of the time for sick people
- $p(E^c \mid F^c) = 995/1000$ test correct 99.5% of time for healthy people

Problems

- $p(F \mid E)$ = probability that a person that tests positive is actually ill = ?
- $p(F^c \mid E^c)$ = probability that a person that tests negative is actually healthy = ?

Example: Tests for Rare Diseases (continued)

Probabilities

- $p(F) = 1/100{,}000 = .00001$ — Probability of Illness
- $p(F^c) = 1 - 1/100{,}000 = .99999$ — Probability of Health
- $p(E \mid F) = 99/100 = .99$ — Probability Sick Person Tests Positive
- $p(E^c \mid F) = 1 - 99/100 = .01$ — Probability Sick Person Tests Negative
- $p(E^c \mid F^c) = 995/100 = .995$ — Probability Healthy Person Tests Negative
- $p(E \mid F^c) = 1 - 995/100 = .005$ — Probability Healthy Person Tests Positive

$$p(F \mid E) = \frac{p(E \mid F)\,p(F)}{p(E \mid F)\,p(F) + p(E \mid F^c)\,p(F^c)} = \frac{(.99)(.00001)}{(.99)(.00001) + (.005)(.99999)} \approx .002$$

$$p(F^c \mid E^c) = \frac{p(E^c \mid F^c)\,p(F^c)}{p(E^c \mid F^c)\,p(F^c) + p(E^c \mid F)\,p(F)} = \frac{(.995)(.99999)}{(.995)(.99999) + (.01)(.00001)} \approx .9999999$$

Tests for Rare Diseases (continued)

Rare Diseases -- $p(F) = 1/100{,}000$

- Test Positive -- Do Not Worry

  -- $p(F \mid E) \approx .002$

- Test Negative $\Rightarrow$ Healthy

  -- $p(F^c \mid E^c) \approx .9999999$

Common Diseases -- $p(F) = 1/10$

- Test Positive $\Rightarrow$ Sick

  -- $$p(F \mid E) = \frac{p(E \mid F)\,p(F)}{p(E \mid F)\,p(F) + p(E \mid F^c)\,p(F^c)} = \frac{(.99)(.1)}{(.99)(.1) + (.005)(.9)} \approx .9565$$

- Test Negative $\Rightarrow$ Healthy

  -- $$p(F^c \mid E^c) = \frac{p(E^c \mid F^c)\,p(F^c)}{p(E^c \mid F^c)\,p(F^c) + p(E^c \mid F)\,p(F)} = \frac{(.995)(.9)}{(.995)(.9) + (.01)(.1)} \approx .99888$$

Example: Spam Filters

Notation

- S = the event that the message is spam
- E = the event that the message contains the word w

Protocol

- $p(S) = 9/10$ most of my messages are spam
- $p(E \mid S) = p(w)$ probability that w appears in spam
- $p(E \mid S^c) = q(w)$ probability that w appears in a real message

Problems

- $p(S \mid E)$ = probability that the message is spam if w appears = ?
- $p(S^c \mid E^c)$ = probability that the message is not spam if w does not appear = ?

Example: Spam Filters (continued)

Probabilities

- $p(S) = 9/10 = .9$ — Probability of Spam
- $p(S^c) = 1 - 9/10 = .1$ — Probability of Real Message
- $p(E \mid S) = p(w)$ — Probability w appears in Spam
- $p(E^c \mid S) = 1 - p(w)$ — Probability w does NOT appears in Spam
- $p(E \mid S^c) = q(w)$ — Probability w appears in Message
- $p(E^c \mid S^c) = 1 - q(w)$ — Probability w does NOT appear in Message
- $$p(S \mid E) = \frac{p(E \mid S)\,p(S)}{p(E \mid S)\,p(S) + p(E \mid S^c)\,p(S^c)} = \frac{9\,p(w)}{9\,p(w) + q(w)}$$
- $$p(S^c \mid E^c) = \frac{p(E^c \mid S^c)\,p(S^c)}{p(E^c \mid S^c)\,p(S^c) + p(E^c \mid S)\,p(S)} = \frac{1 - q(w)}{\big(1 - q(w)\big) + 9\big(1 - p(w)\big)}$$

Monte Hall Problem

Notation

- F = event that a person selects the door with the prize
- E = event that the door opened by Monte Hall does not contain the prize

Protocol

- $p(F) = 1/3$ You have a 1/3 chance of selecting the prize
- $p(E \mid F) = 1$ Monte CANNOT have prize if you do
- $p(E \mid F^c) = 1$ Monte NEVER opens the door with the prize

Problem

- $p(F \mid E)$ = probability that your door has the prize if Monte Hall's does not = ?

Monte Hall Problem (continued)

Probabilities

- $p(F) = 1/3$ Probability your door has the prize
- $p(F^c) = 2/3$ Probability your door does not have prize
- $p(E \mid F) = 1$ Monte CANNOT have prize if you do
- $p(E \mid F^c) = 1$ Monte NEVER opens the door with the prize

Conditional Probability

- $$p(F \mid E) = \frac{p(E \mid F)\,p(F)}{p(E \mid F)p(F) + p(E \mid F^c)p(F^c)} = \frac{(1)(1/3)}{(1)(1/3) + (1)(2/3)} = 1/3$$

Conclusion

- Switch doors with Monte Hall

Monte Hall Problem -- Variation

Probabilities

- $p(F) = 1/3$ Probability your door has the prize
- $p(F^c) = 2/3$ Probability your door does not have prize
- $p(E \mid F) = 1$ Monte CANNOT have prize if you do
- $p(E \mid F^c) = 1/2$ Monte selects a door at random

Conditional Probability

- $$p(F \mid E) = \frac{p(E \mid F)\,p(F)}{p(E \mid F)p(F) + p(E \mid F^c)p(F^c)} = \frac{(1)(1/3)}{(1)(1/3) + (1/2)(2/3)} = 1/2$$

Conclusion

- Switching does NOT change the odds.

Daughter Problem

Notation

- F = event that a person with two children has two daughters
- E = event that a person with two children has at least one daughter

Protocol

- $p(F) = 1/4$ one of four possible case: BB, BG, GB. <u>GG</u>
- $p(F^c) = 3/4$ three of four possible case: <u>BB</u>, <u>BG</u>, <u>GB</u>. GG
- $p(E \mid F) = 1$ if you have two daughters, you have at least one
- $p(E \mid F^c) = 2/3$ two out of three cases: BB, BG, GB

Problem

- $p(F \mid E)$ = probability person has two daughters if they have one daughter = ?

Daughter Problem (continued)

Probabilities

- $p(F) = 1/4$ — Probability of two daughters
- $p(F^c) = 3/4$ — Probability of at most one daughter
- $p(E \mid F) = 1$ — If you have two daughter, you have at least one
- $p(E \mid F^c) = 2/3$ — Two out of three cases: BB, BG, GB

Conditional Probability

- $$p(F \mid E) = \frac{p(E \mid F)\,p(F)}{p(E \mid F)\,p(F) + p(E \mid F^c)\,p(F^c)} = \frac{(1)(1/4)}{(1)(1/4) + (2/3)(3/4)} = 1/3$$

<u>A Daughter Named *Alice*</u>

Notation

- F = event that a person with two children has two daughters
- E = event that a person with two children has one daughter named *Alice*
- p = percentage of girls named *Alice*

Protocol

- $p(F) = 1/4$ one of four possible case: BB, BG, GB. <u>GG</u>
- $p(F^C) = 3/4$ three of four possible case: <u>BB</u>, <u>BG</u>, <u>GB</u>. GG
- $p(E \mid F) = 2p$ *Alice* is a girl's name and there are two girls
- $p(E \mid F^C) = (2/3)p$ two of three possible cases: BB, <u>GB</u>, <u>BG</u>

Problem

- $p(F \mid E)$ = probability person has two daughters if they daughter named *Alice* = ?

A Daughter Named *Alice* (continued)

Probabilities

- $p(F) = 1/4$ — Probability of two daughters
- $p(F^c) = 3/4$ — Probability of at most one daughter
- $p(E \mid F) = 2p$ — *Alice* is a girl's name and there are two girls
- $p(E \mid F^c) = 2p/3$ — two of three possible cases: BB, $\underline{\text{GB}}$, $\underline{\text{BG}}$

Conditional Probability

- $$p(F \mid E) = \frac{p(E \mid F)\,p(F)}{p(E \mid F)p(F) + p(E \mid F^c)p(F^c)} = \frac{(2p)(1/4)}{(2p)(1/4) + (2p/3)(3/4)} = 1/2$$

Expectation and

Average Case Complexity

Random Variables

Definition

- A *Random Variable* X is a Function from the sample space S of an experiment to the real numbers.

- $X : S \rightarrow R$

Examples

- Coin Tossing

    -- $S = \{HH, HT, TH, TT\}$

    -- X = Number of heads

- Dice

    -- $S = \{(1,1), (1, 2), \ldots, (6,6)\}$

    -- X = Sum of Dots

Expectation

Setup

- $S = \{a_1, \ldots, a_n\}$ = sample space
- $pr : S \to [0,1]$ = probability distribution
- $X : S \to R$ = random variable

Expectation

- $E(X) = \sum_{i=1}^{n} \Pr(a_i)\, X(a_i) = \sum_{i=1}^{n} \Pr(X = r_i)\; r_i$
- Weighted Average

Additivity

- $E(X+Y) = E(X) + E(Y)$

Additivity

Formula

- $E(X+Y)=E(X)+E(Y)$

Proof

- $E(X+Y)=\sum_{i=1}^{n}\Pr(a_i)\big(X(a_i)+Y(a_i)\big)$

$$=\sum_{l=1}^{n}\Pr(a_i)X(a_i)+\sum_{i=1}^{n}\Pr(a_i)Y(a_i)=E(X)+E(Y)$$

Advice

- Try to use Additivity, NOT the Definition of Expectation

Dice

Direct Method

- $S = \{(1,1), (1,2), \ldots, (6,6)\}$
- X = Sum of Dots
- $E(X) = \sum_{k=0}^{n} \Pr(s_i) X(s_i) = \frac{12\times 1 + 11\times 2 + 10\times 3 + \cdots}{36} = \frac{252}{36} = 7$

Summation Method

- X_1 = Number of Dots of First Die
- X_2 = Number of Dots of Second Die
- $X = X_1 + X_2$
- $E(X_k) = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$
- $E(X) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7$

Bernoulli Trials

Binomial Distribution

- $X(Experiment) = Number\ \ of\ \ Successes$

- $E(X) = \sum_{k=0}^{n} k B_k^n(t) = nt$

- Direct Method -- By Brute Force Algebra (see next page)

Summation Method

- $X_k = 1$ for success on the *kth* trial

  $= 0$ for failure on the *kth* trial

- $X = \sum_{k=1}^{n} X_k$

- $E(X) = \sum_{k=1}^{n} E(X_k) = \sum_{k=1}^{n} t = nt$

Expectation: Binomial Distribution

Algebra

$$E(X) = \sum_{k=0}^{n} k B_k^n(t)$$

$$= \sum_{k=1}^{n} k \binom{n}{k} t^k (1-t)^{n-k}$$

$$= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} t^k (1-t)^{n-k}$$

$$= nt \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\big((n-1)-(k-1)\big)!} t^{k-1} (1-t)^{(n-1)-(k-1)}$$

$$= nt \sum_{k=1}^{n} \binom{n-1}{k-1} t^{k-1} (1-t)^{(n-1)-(k-1)}$$

$$= nt \sum_{k=1}^{n} B_{k-1}^{n-1}(t)$$

$$= nt$$

Another Example

Birthday Problem

- Find number of people, n, in a room, so that expectation is at least one that two people have the same birthday.

Setup

- $X(n)$ = # people with the same birthday (day and month)
- $$X_{ij}(a_i, a_j) = 1 \quad \text{if } a_i \text{ and } a_j \text{ born on same day and month}$$
  $$= 0 \quad \text{otherwise}$$
- $$X(n) = \sum_{i<j} X_{ij}(a_i, a_j)$$

Solution

- $X(n)$ = # people with the same birthday

- $X(n) = \sum_{i<j} X_{ij}(a_i, a_j)$

  -- $p(X_{ij} = 1) = 1/365$

  -- $E(X_{ij}) = 1/365$

  -- $\# X_{ij} = C(n, 2)$

  -- $E(X) = (1/365)\, C(n,2)$

  -- $n \geq 28 \Rightarrow C(n, 2) \geq 378 \Rightarrow X(n) \geq 1$

- NOT quite the same as the number needed to make the probability 1/2.

Average Case Complexity

Setup

- $S = \{a_1, \ldots, a_n\}$ = possible inputs to an algorithm
- $X(a_i)$ = number of operations used by the algorithm for input a_i

Formula

- $E(X) = \sum_{i=1}^{n} \Pr(a_i) X(a_i)$ = Average Case Complexity

Example -- Linear Search

Assumptions

- n elements in UNORDERED list: $a_1, \ldots, a_n$
- $\Pr(x = a_i) = \frac{1}{n}$ (equal likelihood)
- $X(a_i)$ = number of comparisons used to locate a_i is i

Average Case Complexity

- $$E(x) = \sum_{i=1}^{n} \frac{i}{n} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n}\left(\frac{n(n+1)}{2} \right) = \frac{n+1}{2}$$

  -- Result makes good sense as an average -- half more, half less

Worst Case Complexity = n

Linear Search -- Revisited

Assumptions

- n elements in UNORDERED list: $a_1, \ldots, a_n$
- $\Pr(x = a_i) = \frac{1}{n}$ (equal likelihood)
- $\Pr(x \ in \ list) = p$
- $X(a_i)$ = number of comparisons used to locate a_i is $2i+1$
  - -- compare to current element
  - -- check for end of list -- inside and outside the loop
- $X(a \ not \ in \ list)$ = number of comparisons to determine a not in list is $2n+2$
  - -- one additional comparison on $(n+1)^{st}$ time through the loop

Linear Search -- Revisited (continued)

Average Case Complexity

- $$E(x) = (2n+2)(1-p) + \sum_{i=1}^{n} \frac{(2i+1)p}{n}$$

$$= (2n+2)(1-p) + \frac{p}{n}\sum_{i=1}^{n}(2i+1)$$

$$= (2n+2)(1-p) + \frac{\left((n+1)^2 - 1\right)p}{n}$$

$$= (2n+2)(1-p) + (n+2)p$$

$$= (2n+2) - np$$

Variance and

Standard Deviation

Variance and Standard Deviation

Setup

- $S = \{a_1, \ldots, a_n\}$ = sample space
- $p : S \to [0,1]$ = probability distribution
- X = random variable

Variance

- $$V(X) = \sum_{i=1}^{n} \left(X(s_i) - E(X) \right)^2 p(s_i)$$

Standard Deviation

- $$\sigma(X) = \sqrt{V(X)} = \sqrt{\sum_{i=1}^{n} \left(X(s_i) - E(X) \right)^2 p(s_i)}$$

Objective

- To measure deviation of a random variable X from its average value $E(X)$ (expectation)

Variance and Expectation

Theorem: $V(X) = E(X^2) - \left(E(X)\right)^2$

Proof:

$$V(X) = \sum_{i=1}^{n} \left(X(s_i) - E(X)\right)^2 p(s_i)$$

$$= \sum_{i=1}^{n} X(s_i)^2 p(s_i) - 2E(X) \sum_{i=1}^{n} X(s_i) p(s_i) + E(X)^2 \sum_{i=1}^{n} p(s_i)$$

$$= E(X^2) - 2E(X)E(X) + E(X)^2$$

$$= E(X^2) - \left(E(X)\right)^2$$

Example

On Die

- $E(X)=\frac{1+2+3+4+5+6}{6}=\frac{21}{6}=\frac{7}{2}$

- $(E(X))^2=\left(\frac{1+2+3+4+5+6}{6}\right)^2=\left(\frac{7}{2}\right)^2=\frac{49}{4}$

- $E(X^2)=\frac{1^2+2^2+3^2+4^2+5^2+6^2}{6}=\frac{91}{6}$

- $V(X)=E(X^2)-(E(X))^2=\frac{91}{6}-\frac{49}{4}=\frac{35}{12}\approx 3$

Expectation and Independent Variables

Definition

- X,Y are called *independent random variables* if

  $p(X = r \text{ and } Y = s) = P(X = r)P(X = s)$ for all r,s

Theorem: X,Y independent $\Rightarrow E(XY) = E(X)E(Y)$

Proof:

$$E(XY) = \sum_{i=1}^{n} \Pr(a_i)X(a_i)Y(a_i)$$

$$= \sum_{r_1,\, r_2} r_1 r_2 \left(\Pr(X = r_1) \text{ and } \Pr(Y = r_2)\right)$$

$$= \left(\sum_{r_1} r_1 \Pr(X = r_1)\right)\left(\sum_{r_2} r_2 \Pr(Y = r_2)\right)$$

$$= E(X)E(Y)$$

Variance and Independent Variables

Theorem: $X,Y \textit{ independent} \Rightarrow V(X+Y) = V(X) + V(Y)$

Proof: $V(X+Y) = E\left((X+Y)^2\right) - \left(E(X+Y)\right)^2$

$$= E(X^2 + 2XY + Y^2) - \left(E(X)\right)^2 - 2E(X)E(Y) - \left(E(Y)\right)^2$$

$$= E(X^2) + 2E(XY) + E(Y^2) - \left(E(X)\right)^2 - 2E(X)E(Y) - \left(E(Y)\right)^2$$

$$= E(X^2) - \left(E(X)\right)^2 + E(Y^2) - \left(E(Y)\right)^2 \qquad \{E(XY) = E(X)E(Y)\}$$

$$= V(X) + V(Y)$$

Two Dice

Notation

- X_1 = number of dots on first die
- X_2 = number of dots on second die
- $X = X_1 + X_2$ = sum of dots on both dice

Variance

- $V(X_1) = V(X_2) = \frac{35}{12}$
- $V(X) = V(X_1 + X_2) = \frac{35}{12} + \frac{35}{12} = \frac{35}{6} \approx 6$

Bernoulli Trials

1 *Bernoulli Trial*

- $X(t) = 1 \quad success$
  $\quad\; = 0 \quad failure$

- $E(X) = t$

- $V(X) = E(X^2) - \left(E(X)\right)^2 = t - t^2 = t(1-t)$

n Bernoulli Trials

- $V(X) = V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n)$

- $V(X) = n\,t(1-t)$