Probability

Sam Polgar

August 4, 2024

1. week 1

Definition 1.1. Expected Value

Outcomes of an experiment or random process are real numbers $a_1, \ldots a_n$ with probabilities $p_1, \ldots, p_n$

$$\sum_{k=1}^{n} a_k p_k = a_1 p_1 + \cdots + a_n p_n$$

Example 1.2. Lottery example: 500,000 people pay \$5 with winners 1 x \$1,000,000, 10 x \$1,000, 1000 x \$500, 10,000 x \$10. What's the expected value of the ticket?
$p_k = \frac{1}{500,000}$ for each $k = 1, \ldots, 500000$ p_k is the probability of each outcome occuring. a_i is the net gain for a ticket a_i where $a_1 = 999,995$ the net gain for winning minus the \$5 cost. 2nd prize $a_2, \ldots, a_{11} = 995$, 3rd = $a_{12}, \ldots, a_{1011} = 495$ 4th = $a_{1011}, \ldots, a_{1012} = 10$, 5th = $a_{1012}, \ldots, a_{11011} = 10$, remainder $a_{11011}, \ldots, a_{500000} = -5$ Expected value of a ticket is

$$\sum_{k=1}^{500000} a_k p_k. \, Given \; p_k = \frac{1}{500000}, then \frac{1}{500000} \cdot \sum_{k=1}^{500000} a_k$$

$$\frac{1}{500000}(999,995 + 10 \cdot 995 + 100 \cdot 495 + 10000 \cdot 5 + (-5) \cdot 488989) = -1.78$$

A person who plays this lottery will on average lose 1.78 per ticket!

Example 1.3. How many consecutive pairs of the same suit are expected in a deck of cards?

1. Define Indicator variables: Deck of cards = 52, 13 each suit. Let $X_i = 1$ if i and $i+1$ are same suit, 0 otherwise.
2. Calculate $P(X_i = 1)$. First $P = \frac{1}{52}$, Second card matches first = $\frac{12}{51}$ for each i.
3. Linearity of expectation: Total no. pairs

Experiment Shuffle a deck of cards, go through in order. How many times do 2 consecutive cards have the same suit?

1.1. Linearity of expectation

...The sum of each little thing

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

No assumption of independence or anything. Surprisingly useful.
Prove Expectation of selecting a card of type 1 1/13

sum of all expecations Xi depends on 2 cards. What's Pr

Probability of 52 C 13 E[Xi] = Pr[Xi = 1]

Note: difference between Expectation and Probability. Probability = the likelihood of the event e.g. selecting 2 consecutive suit cards from a deck of 52 Expectation = the average outcome. Multiply each outcome by it's probability.

need to do th

discuss differe

Monte Carlo, Las Vegas Running time Output quality

Question I have an array with n = 100 index of an even number what is time complexity of getting even numbers Theta n

- why isn't this constant? because you can create an algorithm that only selects

"on expectation", the las Vegas

exepcted time For all expectation [Ta] = sum from infiintiy i = 1, i Pr [takes i attempts to find an even number]

$$\mathbb{E}[T_a] = \sum_{i=1}^{\infty} i \cdot \Pr[\text{takes } i \text{ attempts to find an even number}]$$

$$= \sum_{i=1}^{\infty} \frac{i}{2^i} = O(1)$$

Why Randomization?

Faster, Simpler Algo'S - miller rabin, it's a monte carlo algo. Runs in $Otilden^2$

Algos Quicksort Expected running time Is it Las Vegas or Monte Carlo? It's always going to return the sorted array, so it's Las Vegas. Proof: T(n) = Expectations[runtime on array size n] T(n) = E[T(|A1|)] + E[T(|A2|)] + O(n) We know |A1| + |A2| = n-1 |Why n-1?

Expected time analysis vs worst case. expected time analysis: for randomized algorithms average time analysis: using input from a known probability distribution amortized analysis: reusing algorithm on a sequence of inputs, and look at the worst-case sequence of input for the algorithm divided by the length of the sequence

Tutorial 1

Expectation, Discrete Random Variable

- $E[X]$ = expectation of random variable X
- X is a discrete random variable having probability mass function $p(x)$, then $E[X]$ is defined by

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

- Probability Mass Function (PMF) $p(x)$ gives Pr that a discrete random variable X is equal to some value x
- PMF: $p(x) \geqq 0$ for all x, $\sum p(x) = 1$ the sum of probabilities over all possible values

Expectation, Continuous Random Variable

- X is a continuous random variable having probability mass function $f(x)$, then $E[X]$ is defined by

$$E[X] = \int_{x:p(x)>0} xf(x)d(x)$$

- PDF = probability density function

Variance

- Def: average of the squared differences from the mean

$$X : Var(X) = \mathbb{E}[(X - \mathbb{E}[X]^2)] \text{ By Linearity } = \mathbb{E}[X^2] - (E[X])^2$$

- $Var(aX + b) = a^2 Var(X)$ for constants a and b
- For independent random variables: $Var(X + Y) = Var(X) + Var(Y)$
- Indicates how much a data point deviates from the mean, low variance = close to mean, high variance = far from mean, wider range
- Example: population $\sigma^2 = \sum(X - \mu)^2/N$. Sample: $s^2 = \sum \frac{(X-X)^2}{n-1}$.
- σ^2/s^2 is the variance, X is each value in a data set, $\mu or X$ is the mean, $N or n$ is number of data points
- Expressed as squared units of the original data, sometimes confusing
- Standard deviation is the square root of variance, in the same units of the original data
-

Problem 1 Consider a deck of $4n$ cards with 'S', 'H', 'D', 'C', after shuffled randomly, what's the expected number of consecutive pairs of the same suit.

1. Define Indicator Variable X_i for each iteration required. We have $4n - 1$ because $4n$ cards, minus 1 match because the last card can't be matched with the null pointer next door. Let S = shuffled card

$$X_i = \begin{cases} 1 & \text{if } S_i = S_{i+1} \\ 0 & \text{else} \end{cases}$$

2. Sum the number of consecutive pairs found, i.e., sum all 1 cases

$$X = \sum_{i=1}^{4n-1} X_i \text{ we need to find } E[X] = E[\sum_{i=1}^{4n-1} X_i]$$

By linearity

$$= \sum_{i=1}^{4n-1} E[X_i] \;=\; \sum_{i=1}^{4n-1} \frac{\text{no. cards in a suit}}{\text{no. cards in a deck}}$$

We need to find $E[X_i]$ for each i = Pr that $S_i = S_{i+1}$:

$$\Pr(X_i = 1) = \frac{n-1}{4n-1} \;= E[X_i] = \frac{n-1}{4n-1}$$

Problem 2 Similar
Problem 3 Similar Variance part explained here https://claude.ai/chat/9210ca3a-e032-4137-80b1-455acfe9835a
Problem 4 explained here https://claude.ai/chat/fc539a53-9b45-4e59-9638-fcd3e8532612
Problem 5

Quiz 0

Q1: Handshaking Lemma states if $G = (V, E)$ is an undirected graph, $\sum_{v \in V} degv$ is equal to: $2|E|$.
Recall E0–v–0E

Q2: Linearity of expectation means if X, Y are 2 arbitrary random variables and a, b are 2 arbitrary random numbers then $\mathbb{E}[aX + bY] = \mathbb{E}[aX] + \mathbb{E}[bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ if:

1. as long as both expectations are defined
2. only if X, Y are independent
3. only if a, b are positive
4. only if X, Y are uncorrelated

If you want the math to appear in its own line, the standard way is to use:

```
\[
    \sqrt{x+y}
\]
```