Random Variable

Definition

A random variable X on a sample space Ω is a real-valued probability function on Ω; that is, $X : \Omega \rightarrow \mathcal{R}$. A discrete random variable is a random variable that takes on only a finite or countably infinite number of values.

Examples:

1. In rolling a dice, the number that comes up is a random variable.
2. Consider a gambling game in which a player flips two coins, if he gets head in both coins we wins \$3, else he losses \$1. The payoff of the game is a random variable.

Independence

Definition

Two random variables X and Y are independent if and only if

$$\Pr((X = x) \cap (Y = y)) \quad = \quad \Pr(X = x) \cdot \Pr(Y = y)$$

for all values x and y. Similarly, random variables $X_1, X_2, \ldots X_k$ are mutually independent if and only if for any subset $I \subseteq [1, k]$ and any values x_i,$i \in I$,

$$\Pr\left(\bigcap_{i \in I} X_i = x_i\right) \quad = \quad \prod_{i \in I} \Pr(X_i = x_i).$$

Expectation

Definition

The expectation of a discrete random variable X, denoted by $\mathbf{E}[X]$, is given by

$$\mathbf{E}[X] = \sum_i i \Pr(X = i),$$

where the summation is over all values in the range of X. The expectation is finite if $\sum_i |i| \Pr(X = i)$ converges; otherwise, the expectation is unbounded.

The expectation (or mean or average) is a weighted sum over all possible values of the random variable.

Examples:

- The expected value of one dice roll is:

$$E[X] \;=\; \sum_{i=1}^{6} iPr(X=i) \;=\; \sum_{i=1}^{6} \frac{i}{6} \;=\; 3\frac{1}{2}.$$

- The expectation of the random variable X representing the sum of two dice is

$$\mathbf{E}[X] = \frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \frac{3}{36} \cdot 4 + \ldots \frac{1}{36} \cdot 12 = 7.$$

- Let X take on the value 2^i with probability $1/2^i$ for $i = 1, 2, \ldots$.

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} \frac{1}{2^i} 2^i = \sum_{i=1}^{\infty} 1 = \infty.$$

Median

Definition

The **median** of a random variable X is a value m such

$$Pr(X < m) \leq 1/2 \quad \text{and} \quad Pr(X > m) < 1/2.$$

Consider a game in which a player chooses a number in $[1, ..., 6]$ and then rolls 3 dice.
The player wins \$1 for each dice the matches the number, he losses \$1 if no dice matches the number.
What is the expected outcome of that game:

$$-1(\frac{5}{6})^3 + 1 \cdot 3(\frac{1}{6})(\frac{5}{6})^2 + 2 \cdot 3(\frac{1}{6})^2(\frac{5}{6}) + 3(\frac{1}{6})^3 = -\frac{17}{216}.$$

Linearity of Expectation

Theorem

For any two random variables X and Y

$$E[X + Y] = E[X] + E[Y].$$

Proof.

$$E[X+Y] =$$

$$\sum_{i \in range(X)} \sum_{j \in range(Y)} (i+j) Pr((X=i) \cap (Y=j)) =$$

$$\sum_i \sum_j i Pr((X=i) \cap (Y=j)) +$$

$$\sum_j \sum_i j Pr((X=i) \cap (Y=j)) =$$

$$\sum_i i Pr(X=i) + \sum_j j Pr(Y=j).$$

$\square$

(Since we sum over all possible choices of i (j).)

Lemma

For any constant c and discrete random variable X,

$$\mathbf{E}[cX] = c\mathbf{E}[X].$$

Proof.

The lemma is obvious for $c = 0$. For $c \neq 0$,

$$\begin{aligned}
\mathbf{E}[cX] &= \sum_j j \Pr(cX = j) \\
&= c \sum_j (j/c) \Pr(X = j/c) \\
&= c \sum_k k \Pr(X = k) \\
&= c\mathbf{E}[X].
\end{aligned}$$

$\square$

Examples:

- The expectation of the sum of two dice is 7, even if they are not independent.
- The expectation of the outcome of one dice plus twice the outcome of a second dice is $10\frac{1}{2}$.
- Assume that we flip N coins, what is the expected number of heads?
  Using linearity of expectation we get $N \cdot \frac{1}{2}$.
  By direct summation we get $\sum_{i=0}^{N} i\binom{N}{i}2^{-N}$.
  Thus we prove
  $$\sum_{i=0}^{N} i\binom{N}{i}2^{-N} = \frac{N}{2}.$$

Assume that N people checked coats in a restaurants. The coats are mixed and each person gets a random coat.
How many people got their own coats?
It's hard to compute $E[X] = \sum_{k=0}^{N} kPr(X = k)$. Instead we define N 0-1 random variables X_i, where $X_i = 1$ iff i got his coat.

$$E[X_i] \ = \ 1 \cdot Pr(X_i = 1) \ + \ 0 \cdot Pr(X_i = 0) =$$

$$Pr(X_i = 1) = \frac{1}{N}.$$

$$E[X] \ = \ \sum_{i=1}^{N} E[X_i] \ = \ 1.$$

Bernoulli Random Variable

A Bernoulli or an indicator random variable:

$$Y = \begin{cases} 1 & \text{if the experiment succeeds,} \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbf{E}[Y] = p \cdot 1 + (1-p) \cdot 0 = p = \Pr(Y = 1).$$

Binomial Random Variable

Definition

A binomial random variable X with parameters n and p, denoted by $B(n, p)$, is defined by the following probability distribution on $j = 0, 1, 2, \ldots, n$:

$$\Pr(X = j) = \binom{n}{j} p^j (1 - p)^{n-j}.$$

Expectation of a Binomial Random Variable

$$
\begin{aligned}
\mathbf{E}[X] &= \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j} \\
&= \sum_{j=0}^{n} j \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \\
&= \sum_{j=1}^{n} \frac{n!}{(j-1)!(n-j)!} p^j (1-p)^{n-j} \\
&= np \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!((n-1)-(j-1))!} p^{j-1} (1-p)^{(n-1)-(j-1)} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} = np.
\end{aligned}
$$

Using linearity of expectations

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

Quicksort

Procedure Q_S(S);
Input: A set S.
Output: The set S in sorted order.

1. Choose a random element y uniformly from S.
2. Compare all elements of S to y. Let

$$S_1 = \{x \in S - \{y\} \mid x \leq y\}, \qquad S_2 = \{x \in S - \{y\} \mid x > y\}.$$

3. Return the list:

$$Q\_S(S_1), y, Q\_S(S_2).$$

Let T = number of comparisons in a run of QuickSort.

Theorem

$$E[T] = O(n \log n).$$

Let $s_1, ..., s_n$ be the elements of S is sorted order.
For $i = 1, ..., n$, and $j > i$, define 0-1 random variable $X_{i,j}$, s.t. $X_{i,j} = 1$ iff s_i is compared to s_j in the run of the algorithm, else $X_{i,j} = 0$.
The number of comparisons in running the algorithm is

$$T = \sum_{i=1}^{n} \sum_{j>i} X_{i,j}.$$

We are interested in $E[T]$.

What is the probability that $X_{i,j} = 1$?
s_i is compared to s_j iff either s_i or s_j is chosen as a "split item" before any of the $j - i - 1$ elements between s_i and s_j are chosen.
Elements are chosen uniformly at random $\rightarrow$ elements in the set $[s_i, s_{i+1}, ..., s_j]$ are chosen uniformly at random.

$$Pr(X_{i,j} = 1) = \frac{2}{j - i + 1}.$$

$$E[X_{i,j}] = \frac{2}{j - i + 1}.$$

$$E[T] \;=\; E[\sum_{i=1}^{n} \sum_{j>i} X_{i,j}] =$$

$$\sum_{i=1}^{n} \sum_{j>i} E[X_{i,j}] = \sum_{i=1}^{n} \sum_{j>i} \frac{2}{j-i+1} \leq$$

$$n \sum_{k=1}^{n} \frac{2}{k} \leq 2nH_n = n \log n + O(n).$$

A Deterministic QuickSort

Procedure DQ_S(S);
Input: A set S.
Output: The set S in sorted order.

1. Let y be the first element in S.
2. Compare all elements of S to y. Let

   $$S_1 = \{x \in S - \{y\} \mid x \leq y\}, \qquad S_2 = \{x \in S - \{y\} \mid x > y\}.$$

   (Elements is S_1 and S_2 are in th same order as in S.)
3. Return the list:

   $$DQ\_S(S_1), y, DQ\_S(S_2).$$

Probabilistic Analysis of QuickSort

Theorem

The expected run time of DQ_S on a random input, uniformly chosen from all possible permutation of S is $O(n \log n)$.

Proof.

Set $X_{i,j}$ as before.
If all permutations have equal probability, all permutations of $S_i, ..., S_j$ have equal probability, thus

$$Pr(X_{i,j}) = \frac{2}{j - i + 1}.$$

$$E[\sum_{i=1}^{n} \sum_{j>i} X_{i,j}] = O(n \log n).$$

□

Randomized Algorithms:

- Analysis is true for **any** input.
- The sample space is the space of random choices made by the algorithm.
- Repeated runs are independent.

Probabilistic Analysis;

- The sample space is the space of all possible inputs.
- If the algorithm is **deterministic** repeated runs give the same output.

Algorithm classification

A **Monte Carlo Algorithm** is a randomized algorithm that may produce an incorrect solution.
For decision problems: A **one-side error** Monte Carlo algorithm errs only one one possible output, otherwise it is a **two-side error** algorithm.
A **Las Vegas** algorithm is a randomized algorithm that **always** produces the correct output.
In both types of algorithms the run-time is a random variable.

Compund events:

- A program that has one call to a process $\mathcal{S}$.
- Each call to process $\mathcal{S}$ recursively spawns new copies of the process $\mathcal{S}$, where the number of new copies is a binomial random variable with parameters n and p.
- These random variables are independent for each call to $\mathcal{S}$.
- What is the expected number of copies of the process $\mathcal{S}$ generated by the program?

Conditional Expectation

Definition

$$\mathbf{E}[Y \mid Z = z] = \sum_{y} y \Pr(Y = y \mid Z = z),$$

where the summation is over all y in the range of Y.

Example

We role two dice. X_1 be the number that shows on the first die, X_2 be the number on the second die, and X be the sum of the numbers on the two dice.

$$\mathbf{E}[X \mid X_1 = 2] = \sum_x x \Pr(X = x \mid X_1 = 2) = \sum_{x=3}^{8} x \cdot \frac{1}{6} = \frac{11}{2}.$$

As another example, consider $\mathbf{E}[X_1 \mid X = 5]$.

$$\begin{aligned}
\mathbf{E}[X_1 \mid X = 5] &= \sum_{x=1}^{4} x \Pr(X_1 = x \mid X = 5) \\
&= \sum_{x=1}^{4} x \frac{\Pr(X_1 = x \cap X = 5)}{\Pr(X = 5)} \\
&= \sum_{x=1}^{4} x \frac{1/36}{4/36} \\
&= 5/2.
\end{aligned}$$

Lemma

For any random variables X and Y,

$$\mathbf{E}[X] = \sum_{y} \Pr(Y = y) E[X \mid Y = y],$$

where the sum is over all values in the range of Y.

Proof.

$$
\begin{aligned}
& \sum_y \Pr(Y = y) E[X \mid Y = y] \\
= \; & \sum_y \Pr(Y = y) \sum_x x \Pr(X = x \mid Y = y) \\
= \; & \sum_x \sum_y x \Pr(X = x \mid Y = y) \Pr(Y = y) \\
= \; & \sum_x \sum_y x \Pr(X = x \cap Y = y) \\
= \; & \sum_x x \Pr(X = x) = \mathbf{E}[X].
\end{aligned}
$$

$\square$

Conditional Expectation as a Random variable

Definition

The expression $\mathbf{E}[Y \mid Z]$ is a random variable $f(Z)$ that takes on the value $\mathbf{E}[Y \mid Z = z]$ when $Z = z$.

Consider the outcome of rolling two dice $X_1, X_2, X = X_1 + X_2$.

$$\mathbf{E}[X \mid X_1] = \sum_x x \Pr(X = x \mid X_1) = \sum_{x=X_1+1}^{X_1+6} x \cdot \frac{1}{6} = X_1 + \frac{7}{2}.$$

If $\mathbf{E}[Y \mid Z]$ is a random variable, it has an expectation.

Theorem

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y \mid Z]].$$

$$\mathbf{E}[X \mid X_1] = X_1 + \frac{7}{2}.$$

Thus

$$\mathbf{E}[\mathbf{E}[X \mid X_1]] = \mathbf{E}\left[X_1 + \frac{7}{2}\right] = \frac{7}{2} + \frac{7}{2} = 7.$$

Proof.

$\mathbf{E}[Y \mid Z] = f(Z)$, where $f(Z)$ takes on the value $\mathbf{E}[Y \mid Z = z]$ when $Z = z$.

$$
\begin{aligned}
\mathbf{E}[\mathbf{E}[Y \mid Z]] &= \sum_z \mathbf{E}[Y \mid Z = z] \Pr(Z = z) \\
&= \sum_z \left(\sum_y y \Pr(Y = y \mid Z = z) \right) \Pr(Z = z) \\
&= \sum_z \sum_y y \Pr(Y = y \mid Z = z) \Pr(Z = z) \\
&= \sum_z \sum_y y \Pr(Y = y \cap Z = z) \\
&= \sum_y y \sum_z \Pr(Y = y \cap Z = z) \\
&= \sum_y y \Pr(Y = y) = \mathbf{E}[Y].
\end{aligned}
$$

$\square$

Back to the Spawning Process

- The initial process $\mathcal{S}$ is in generation 0.
- A process $\mathcal{S}$ is in generation i if it was spawned by another process $\mathcal{S}$ in generation $i-1$.
- Let Y_i denote the number of $\mathcal{S}$ processes in generation i.
- $Y_0 = 1$, and Y_1 has a binomial distribution.
- $$\mathbf{E}[Y_1] = np.$$

- Z_k^{i-1} = number of copies spawned by the kth process spawned in the $(i-1)$-st generation.
- Z_k^{i-1} is a binomial random variable with parameters n and p.
- $$\begin{aligned} \mathbf{E}[Y_i \mid Y_{i-1} = y_{i-1}] &= \mathbf{E}\left[\sum_{k=1}^{y_{i-1}} Z_k\right] \\ &= \sum_{k=1}^{y_{i-1}} \mathbf{E}[Z_k] \\ &= y_{i-1} np. \end{aligned}$$

- $$E[Y_i] = E[E[Y_i \mid Y_{i-1}]] = E[Y_{i-1}np] = npE[Y_{i-1}].$$
- By induction on i, and using $Y_0 = 1$,
$$E[Y_i] = (np)^i.$$
- $$E[\sum_{i \geq 0} Y_i] = \sum_{i \geq 0} E[Y_i] = \sum_{i \geq 0} (np)^i.$$
- If $np \geq 1$, the expectation is unbounded, and if $np < 1$, the expectation is $1/(1 - np)$.

The Geometric Distribution

Definition

A geometric random variable X with parameter p is given by the following probability distribution on $n = 1, 2, \ldots$.

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

memoryless property

Lemma

For a geometric random variable with parameter p and $n > 0$,

$$\Pr(X = n + k \mid X > k) \;=\; \Pr(X = n).$$

Proof.

$$\begin{aligned}
\Pr(X = n + k \mid X > k) &= \frac{\Pr((X = n + k) \cap (X > k))}{\Pr(X > k)} \\
= \frac{\Pr(X = n + k)}{\Pr(X > k)} &= \frac{(1-p)^{n+k-1}p}{\sum_{i=k}^{\infty}(1-p)^i p} \\
= \frac{(1-p)^{n+k-1}p}{(1-p)^k} &= (1-p)^{n-1}p \\
&= \Pr(X = n).
\end{aligned}$$

Lemma

Let X be a discrete random variable that takes on only non-negative integer values. Then

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} \Pr(X \geq i).$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{\infty} \Pr(X \geq i) &= \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \Pr(X = j) \\
&= \sum_{j=1}^{\infty} \sum_{i=1}^{j} \Pr(X = j) \\
&= \sum_{j=1}^{\infty} j \Pr(X = j) = \mathbf{E}[X].
\end{aligned}$$

For a geometric random variable X with parameter p,

$$\Pr(X \geq i) = \sum_{n=i}^{\infty} (1-p)^{n-1} p = (1-p)^{i-1}.$$

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{i=1}^{\infty} \Pr(X \geq i) \\
&= \sum_{i=1}^{\infty} (1-p)^{i-1} \\
&= \frac{1}{1-(1-p)} \\
&= \frac{1}{p}
\end{aligned}$$

Alternative Proof

$Y = 1$ if $X = 1$, else $Y = 0$.

$$\begin{aligned} \mathbf{E}[X] &= \Pr(Y = 0)\mathbf{E}[X \mid Y = 0] + \Pr(Y = 1)\mathbf{E}[X \mid Y = 1] \\ &= (1 - p)\mathbf{E}[X \mid Y = 0] + p\mathbf{E}[X \mid Y = 1]. \end{aligned}$$

When $X > 1$, let $Z = X - 1$.

$$\mathbf{E}[X] = (1 - p)\mathbf{E}[Z + 1] + p \cdot 1 = (1 - p)\mathbf{E}[Z] + 1,$$

By the memoryless property Z is also a geometric random variable with parameter p. Hence $\mathbf{E}[Z] = \mathbf{E}[X]$.

$$\mathbf{E}[X] = (1 - p)\mathbf{E}[Z] + 1 = (1 - p)\mathbf{E}[X] + 1,$$

which yields $\mathbf{E}[X] = 1/p$.

Example: Coupon Collector's Problem

Suppose that each box of cereal contains a random coupon from a set of n different coupons.

How many boxes of cereal do you need to buy before you obtain at least one of every type of coupon?

Let X be the number of boxes bought until at least one of every type of coupon is obtained.

Let X_i be the number of boxes bought while you had exactly $i-1$ different coupon.

$$X = \sum_{i=1}^{n} X_i$$

.

X_i is a geometric random variable with parameter

$$p_i = 1 - \frac{i-1}{n}.$$

$$\mathbf{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}.$$

$$\begin{aligned}
\mathbf{E}[X] &= E\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathbf{E}[X_i] \\
&= \sum_{i=1}^{n} \frac{n}{n-i+1} \\
&= n\sum_{i=1}^{n} \frac{1}{i} = n \ln n + \Theta(n).
\end{aligned}$$