Probability and Computing

Chapter 2: Discrete Random Variables and Expectation

Zeng Yuxiang

The Hong Kong University of Science and Technology

July 2, 2018

Table of Contents

1. Random Variables and Expectation
2. The Bernolli and Binomial Random Variables
3. Conditional Expectation
4. The Geometric Distribution
5. Application: The Expected Run-Time of Quicksort

Table of Contents

1. Random Variables and Expectation
2. The Bernolli and Binomial Random Variables
3. Conditional Expectation
4. The Geometric Distribution
5. Application: The Expected Run-Time of Quicksort

Concept: Random Variable

random variable

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

Concept: Random Variable

random variable

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

Example

If X is the random variable representing the sum of the two dice, then the event $X = 4$ corresponds to the set of basic events $(1,3), (2,2), (3,1)$. Hence,

$$\Pr(X = 4) = \frac{4}{36} = \frac{1}{12}.$$

Independent Random Variables

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

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

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

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

Concept: Expectation

Expectation of Discrete Random Variable

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

$$E[X] = \sum_i i \Pr(X = i).$$

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.

Concept: Expectation

Expectation of Discrete Random Variable

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

$$E[X] = \sum_i i \Pr(X = i).$$

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.

Examples

1. The expectation of the random variable X representing the sum of two dice is $E[X] = \frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \frac{3}{36} \cdot 4 + \cdots + \frac{1}{36} \cdot 12 = 7$.

Concept: Expectation

Expectation of Discrete Random Variable

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

$$E[X] = \sum_i i\Pr(X = i).$$

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.

Examples

1. The expectation of the random variable X representing the sum of two dice is $E[X] = \frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \frac{3}{36} \cdot 4 + \cdots + \frac{1}{36} \cdot 12 = 7$.
2. Consider a random variable X that takes on the value 2^i with probability $1/2^i$ for $i = 1, 2, \cdots$. The expected value of X is $E[X] = \sum_{i=1}^{\infty} 2^i \cdot \frac{1}{2^i} = \sum_{i=1}^{\infty} 1 = \infty$.

Concept: Expectation

Theorem: Linearity of Expectations

For any finite collection of discrete random variables $X_1, X_2, ..., X_n$ with finite expectations,

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

Concept: Expectation

Theorem: Linearity of Expectations

For any finite collection of discrete random variables $X_1, X_2, ..., X_n$ with finite expectations,

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

Proof:

We prove for two random variables X and Y and induce to general case.

$$\begin{aligned} E[X+Y] &= \sum_i \sum_j (i+j) \cdot \Pr\big((X=i) \cap (Y=j)\big) \\ &= \sum_i i \sum_j \Pr\big((X=i) \cap (Y=j)\big) + \sum_j j \sum_i \Pr\big((X=i) \cap (Y=j)\big) \\ &= \sum_i i \cdot \Pr(X=i) + \sum_j j \cdot \Pr(Y=j) = E[X] + E[Y] \end{aligned}$$

Concept: Expectation

Theorem: Linearity of Expectations

For any finite collection of discrete random variables $X_1, X_2, ..., X_n$ with finite expectations,

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

Example

We compute the expected sum of two standard dice. Let $X = X_1 + X_2$, where X_i represents the outcome of i-th dice. Then

$$E[X_i] = \frac{1}{6}\sum_{j=1}^{6} j = \frac{7}{2}.$$

Applying Linearity of Expectations, we have

$$E[X] = E[X_1] + E[X_2] = 7.$$

Concept: Expectation

Theorem: Linearity of Expectations

1. It is worth emphasizing that linearity of expectations holds for any collection of random variables, even if they are not independent! For example, let the random variable $Y = X_1 + {X_1}^2$, $E[Y] = E[X_1] + E[{X_1}^2]$.

Concept: Expectation

Theorem: Linearity of Expectations

1. It is worth emphasizing that linearity of expectations holds for any collection of random variables, even if they are not independent! For example, let the random variable $Y = X_1 + {X_1}^2$, $E[Y] = E[X_1] + E[{X_1}^2]$.
2. Linearity of expectations also holds for countably infinite summations in certain cases. Specifically, it can be shown that

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

whenever $\sum_{i=1}^{\infty} E[|X_i|]$ converges.

Concept: Expectation

Theorem: Linearity of Expectations

1. It is worth emphasizing that linearity of expectations holds for any collection of random variables, even if they are not independent! For example, let the random variable $Y = X_1 + {X_1}^2$, $E[Y] = E[X_1] + E[{X_1}^2]$.
2. Linearity of expectations also holds for countably infinite summations in certain cases. Specifically, it can be shown that

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

   whenever $\sum_{i=1}^{\infty} E[|X_i|]$ converges.
3. Lem: For any constant c and discrete random variable X, $E[cX] = c \cdot E[X]$.

Concept: Jensen's Inequality

Convex Function: Definition

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be convex if, for any x_1, x_2 and $0 \le \lambda \le 1$.

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

Concept: Jensen's Inequality

Convex Function: Definition

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be convex if, for any x_1, x_2 and $0 \leq \lambda \leq 1$.

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

Convex Function: Property

If f is twice differentiable function, then f is convex if and only if $f''(x) \geq 0$.

Concept: Jensen's Inequality

Convex Function: Definition

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be convex if, for any x_1, x_2 and $0 \leq \lambda \leq 1$.

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$

Convex Function: Property

If f is twice differentiable function, then f is convex if and only if $f''(x) \geq 0$.

Jensen's Inequality

If f is a convex function, then $E[f(X)] \geq f(E[X])$.

Concept: Jensen's Inequality

Jensen's Inequality

If f is a convex function, then $E[f(X)] \geq f(E[X])$.

Proof:

- Assume that f has a Taylor expansion. Let $\mu = E[X]$. By Taylor's theorem, there exists c such that

$$\begin{aligned} f(x) &= f(\mu) + f'(\mu)(x-\mu) + \frac{f''(c)(x-\mu)^2}{2} \\ &\geq f(\mu) + f'(\mu)(x-\mu) \end{aligned}$$

Concept: Jensen's Inequality

Jensen's Inequality

If f is a convex function, then $E[f(X)] \geq f(E[X])$.

Proof:

- Assume that f has a Taylor expansion. Let $\mu = E[X]$. By Taylor's theorem, there exists c such that
$$\begin{aligned} f(x) &= f(\mu) + f'(\mu)(x-\mu) + \frac{f''(c)(x-\mu)^2}{2} \\ &\geq f(\mu) + f'(\mu)(x-\mu) \end{aligned}$$
- Since $f''(c) > 0$ by convex property, taking expectation on both sides and applying linearity of expectation, we have
$$\begin{aligned} E[f(X)] &\geq E[f(\mu) + f'(\mu)(X-\mu)] = E[f(\mu)] + f'(\mu)E[X-\mu] \\ &= E[f(\mu)] + f'(\mu)(E[X]-\mu) = f(\mu) + f'(\mu)\cdot 0 \\ &= f(E[X]) \end{aligned}$$

Table of Contents

1. Random Variables and Expectation
2. The Bernolli and Binomial Random Variables
3. Conditional Expectation
4. The Geometric Distribution
5. Application: The Expected Run-Time of Quicksort

Concept: Bernoulli Random Variable

Definition: Bernoulli Random

Suppose that we run an experiment that succeeds with probability p and fails with probability $1 - p$. Let Y be a random variable such that

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

The variable Y is called a Bernoulli or an indicator random variable. Note that, for a Bernoulli random variable, $E[Y] = p = \Pr(Y = 1)$.

Concept: Bernoulli Random Variable

Definition: Bernoulli Random

Suppose that we run an experiment that succeeds with probability p and fails with probability $1 - p$. Let Y be a random variable such that

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

The variable Y is called a Bernoulli or an indicator random variable. Note that, for a Bernoulli random variable, $E[Y] = p = \Pr(Y = 1)$.

Example

If we flip a fair coin and consider the outcome "heads" a success, then the expected value of the corresponding indicator random variable is $1/2$.

Concept: 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 $j = 0, \cdots, n$:

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

Concept: 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 $j = 0, \cdots, n$:

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

Example

Network Router: We do not have the memory available to store all of the packets, so we choose to store a random subset or *sample* of the packets for later analysis. If each packet is stored with probability p and If n packets go through the router each day, then the number of sampled packets each day is a binomial random variable X with parameters n and p. If we want to know how much memory is necessary for such a sample, a natural starting point is to determine the expectation of the random variable X.

Concept: Binomial Random Variable

Expectation of Binomial Random Variable

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

Concept: Binomial Random Variable

Expectation of Binomial Random Variable

Because

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

We have

$$\begin{aligned} E[X] &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} \\ &= np \cdot \big(p + (1-p)\big)^{n-1} \\ &= np \cdot 1 \\ &= np \end{aligned}$$

Concept: Binomial Random Variable

Expectation of Binomial Random Variable (Simpler Analysis)

The linearity of expectations allows for a significantly simpler argument.

- If X is a binomial random variable with parameters n and p, then X is the number of successes in n trials, where each trial is successful with probability p.

Concept: Binomial Random Variable

Expectation of Binomial Random Variable (Simpler Analysis)

The linearity of expectations allows for a significantly simpler argument.

- If X is a binomial random variable with parameters n and p, then X is the number of successes in n trials, where each trial is successful with probability p.
- Define a set of n indicator random variables $X_1, \cdots, X_n$, where $X_i = 1$ if the i-th trial is successful and 0 otherwise.

Concept: Binomial Random Variable

Expectation of Binomial Random Variable (Simpler Analysis)

The linearity of expectations allows for a significantly simpler argument.

- If X is a binomial random variable with parameters n and p, then X is the number of successes in n trials, where each trial is successful with probability p.
- Define a set of n indicator random variables $X_1, \cdots, X_n$, where $X_i = 1$ if the i-th trial is successful and 0 otherwise.
- Clearly, $E[X_i] = p$, $X = \sum_{i=1}^{n} X_i$ and so, by the linearity of expectations,

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

Table of Contents

1. Random Variables and Expectation
2. The Bernolli and Binomial Random Variables
3. Conditional Expectation
4. The Geometric Distribution
5. Application: The Expected Run-Time of Quicksort

Concept: Conditional Expectation

Definition

The conditional expectation of a random variable is defined as follows:

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

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

Concept: Conditional Expectation

Example

Suppose that we independently roll two standard six-sided dice. Let X_1 be the number that shows on the first dice, X_2 the number on the second dice, and X the sum of the numbers on the two dice. Then

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

Concept: Conditional Expectation

Example

Suppose that we independently roll two standard six-sided dice. Let X_1 be the number that shows on the first dice, X_2 the number on the second dice, and X the sum of the numbers on the two dice. Then

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

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

Concept: Conditional Expectation

Expectation and Conditional Expectation

For any random variables X and Y,

$$E[X] = \sum_y \Pr(Y = y) E[X \mid Y = y]$$

where the sum is over all values in Y and all of the expectations exist.

Concept: Conditional Expectation

Expectation and Conditional Expectation

For any random variables X and Y,

$$E[X] = \sum_y \Pr(Y = y) E[X \mid Y = y]$$

where the sum is over all values in Y and all of the expectations exist.

Proof:

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

Concept: Conditional Expectation

Linearity of Conditional Expectations

For any finite collection of discrete random variables $X_1, X_2, \cdots, X_n$ with finite expectations and for any random variable Y,

$$E\Big[\sum_{i=1}^{n} X_i \mid Y = y\Big] = \sum_{i=1}^{n} E[X_i \mid Y = y].$$

Concept: Conditional Expectation

Expectation as random variable

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

Concept: Conditional Expectation

Expectation as random variable

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

We emphasize that $E[Y \mid Z]$ is not a real value; it is actually a function of the random variable Z. Hence $E[Y \mid Z]$ is itself a function from the sample space to the real numbers and can therefore be thought of as a random variable.

Concept: Conditional Expectation

Expectation as random variable

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

Example

Suppose that we independently roll two standard six-sided dice. Let X_1 be the number that shows on the first dice, X_2 the number on the second dice, and X the sum of the numbers on the two dice. Then

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

If $E[X \mid X_1]$ is a random variable, then it makes sense to consider its expectation $E[E[X \mid X_1]]$.

Concept: Conditional Expectation

Expectation as random variable

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

Example

Suppose that we independently roll two standard six-sided dice. Let X_1 be the number that shows on the first dice, X_2 the number on the second dice, and X the sum of the numbers on the two dice. Then

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

Concept: Conditional Expectation

More generally, we have

Expected Conditional Expectation

$$E[Y] = E[E[Y \mid Z]].$$

More generally, we have

Expected Conditional Expectation

$$E[Y] = E[E[Y \mid Z]].$$

Proof:

We have $E[Y \mid Z] = f(Z)$, where $f(Z)$ takes on the value $E[Y \mid Z = z]$ when $Z = z$. Hence,

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

RHS equals to $E[Y]$ according to the Lemma of Expectation and Conditional Expectation.

Application: Conditional Expectation

Example

Consider a program that includes one call to a process $\mathcal{S}$. Assume that 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. We assume that 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?

Application: Conditional Expectation

Example

Consider a program that includes one call to a process $\mathcal{S}$. Assume that 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. We assume that 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?

To analyze this process, we introduce the idea of generation.

Definition: Generation

The initial process $\mathcal{S}$ is in 0-th generation. Otherwise, we say that a process $\mathcal{S}$ is in i-th generation if it was spawned by another process $\mathcal{S}$ in $(i-1)$-th generation. Let Y_i denote the number of $\mathcal{S}$ processes in i-th generation. Since we know that $Y_0 = 1$, the number of processes in the first generation has a binomial distribution. Thus, $E[Y_1] = np$.

Application: Conditional Expectation

Analysis: Recursive Spawning

Suppose we knew that the number of processes in $(i-1)$-th generation was y_{i-1}, so $Y_{i-1} = y_{i-1}$ Let Z_k be the number of copies spawned by the k-th process spawned in the $(i-l)$-th generation for $1 \leq k \leq y_{i-1}$. Each Z_k is a binomial random variable with parameters n and p. Then

Application: Conditional Expectation

Analysis: Recursive Spawning

Suppose we knew that the number of processes in $(i-1)$-th generation was y_{i-1}, so $Y_{i-1} = y_{i-1}$ Let Z_k be the number of copies spawned by the k-th process spawned in the $(i-l)$-th generation for $1 \leq k \leq y_{i-1}$. Each Z_k is a binomial random variable with parameters n and p. Then

$$
\begin{aligned}
E[Y_i \mid Y_{i-1} = y_{i-1}] &= E\Big[\sum_{k=1}^{y_{i-1}} Z_k \mid Y_{i-1} = y_{i-1}\Big] \\
&= \sum_{j \geq 0} j \Pr\Big(\sum_{k=1}^{y_{i-1}} Z_k = j \mid Y_{i-1} = y_{i-1}\Big) \\
&= \sum_{j \geq 0} j \Pr\Big(\sum_{k=1}^{y_{i-1}} Z_k = j\Big) \ (Z_k \text{ are indepedent})
\end{aligned}
$$

Application: Conditional Expectation

Analysis: Recursive Spawning (Cont.)

Suppose we knew that the number of processes in $(i-1)$-th generation was y_{i-1}, so $Y_{i-1} = y_{i-1}$ Let Z_k be the number of copies spawned by the k-th process spawned in the $(i-l)$-th generation for $1 \leq k \leq y_{i-1}$. Each Z_k is a binomial random variable with parameters n and p. Then

$$
\begin{aligned}
E[Y_i \mid Y_{i-1} = y_{i-1}] &= \sum_{j \geq 0} j \Pr\Big(\sum_{k=1}^{y_{i-1}} Z_k = j\Big) \text{ (Z_k are indepedent)} \\
&= E\Big[\sum_{k=1}^{y_{i-1}} Z_k\Big] \\
&= \sum_{k=1}^{y_{i-1}} E[Z_k] \text{ (Linearity of expectations)} \\
&= y_{i-1} np
\end{aligned}
$$

Application: Conditional Expectation

Analysis: Recursive Spawning (Cont.)

Applying Theorem of Expected Conditional Expectation, we can compute the expected size of the *i*-th generation inductively.

$$E[Y_i] = E[E[Y_i \mid Y_{i-1}]] = E[Y_{i-1}np] = npE[Y_{i-1}].$$

Application: Conditional Expectation

Analysis: Recursive Spawning (Cont.)

Applying Theorem of Expected Conditional Expectation, we can compute the expected size of the i-th generation inductively.

$$E[Y_i] = E[E[Y_i \mid Y_{i-1}]] = E[Y_{i-1}np] = npE[Y_{i-1}].$$

By induction on i, and using the fact that $Y_0 = 1$, we then obtain

$$E[Y_i] = (np)^i.$$

Application: Conditional Expectation

Analysis: Recursive Spawning (Cont.)

Applying Theorem of Expected Conditional Expectation, we can compute the expected size of the i-th generation inductively.

$$E[Y_i] = E[E[Y_i \mid Y_{i-1}]] = E[Y_{i-1}np] = npE[Y_{i-1}].$$

By induction on i, and using the fact that $Y_0 = 1$, we then obtain

$$E[Y_i] = (np)^i.$$

The expected total number of copies of process $\mathcal{S}$ generated by the program is given by

$$E\Big[\sum_{i \geq 0} [Y_i]\Big] = \sum_{i \geq 0} E[Y_i] = \sum_{i \geq 0} (np)^i.$$

Application: Conditional Expectation

Analysis: Recursive Spawning (Cont.)

Applying Theorem of Expected Conditional Expectation, we can compute the expected size of the i-th generation inductively.

$$E[Y_i] = E[E[Y_i \mid Y_{i-1}]] = E[Y_{i-1}np] = npE[Y_{i-1}].$$

By induction on i, and using the fact that $Y_0 = 1$, we then obtain

$$E[Y_i] = (np)^i.$$

The expected total number of copies of process $\mathcal{S}$ generated by the program is given by

$$E\Big[\sum_{i\geq 0}[Y_i]\Big] = \sum_{i\geq 0} E[Y_i] = \sum_{i\geq 0}(np)^i.$$

If $np \geq 1$, then the expectation is unbounded;
If $np < 1$, then the expectation is $1/(np-1)$.

Table of Contents

1. Random Variables and Expectation
2. The Bernolli and Binomial Random Variables
3. Conditional Expectation
4. The Geometric Distribution
5. Application: The Expected Run-Time of Quicksort

Geometric Distribution

Definition

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

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

Geometric Distribution

Definition

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

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

Example

Suppose that we flip a coin until it lands on heads. What is the distribution of the number of flips? This is an example of a geometric distribution, which arises in the following situation: we perform a sequence of independent trials until the first success, where each trial succeeds with probability p.

Geometric Distribution

Property

For a geometric random variable X with parameter p and for $n > 0$,
$$\sum_{n \geq 1} \Pr(X = n) = 1.$$

Geometric random variables are said to be *memoryless* because the probability that you will reach your first success n trials from now is independent of the number of failures you have experienced. Informally, one can ignore past failures because they do not change the distribution of the number of future trials until first success. Formally, we have the following statement.

Geometric Distribution

Property

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

$$\sum_{n \geq 1} \Pr(X = n) = 1.$$

Geometric random variables are said to be *memoryless* because the probability that you will reach your first success n trials from now is independent of the number of failures you have experienced. Informally, one can ignore past failures because they do not change the distribution of the number of future trials until first success. Formally, we have the following statement.

Conditioned Probability

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

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

Geometric Distribution

Conditioned Probability

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

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

Proof:

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

Geometric Distribution

Expectation of Geometric Random Variable

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

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

Geometric Distribution

Expectation of Geometric Random Variable

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

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

Proof:

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

Expectation of Geometric Random Variable

Since

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

Geometric Distribution

Expectation of Geometric Random Variable

Since

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

we have

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

Geometric Distribution

Expectation of Geometric Random Variable (Simpler Version)

Recall that X corresponds to the number of flips until the first heads given that each flip is heads with probability p. Let $Y = 1$ if the first flip is tails and $Y = 1$ if the first flip is heads.

Expectation of Geometric Random Variable (Simpler Version)

Recall that X corresponds to the number of flips until the first heads given that each flip is heads with probability p. Let $Y = 1$ if the first flip is tails and $Y = 1$ if the first flip is heads. We have

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

Geometric Distribution

Expectation of Geometric Random Variable (Simpler Version)

Recall that X corresponds to the number of flips until the first heads given that each flip is heads with probability p. Let $Y = 1$ if the first flip is tails and $Y = 1$ if the first flip is heads. We have

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

If $Y = 1$ then $X = 1$, so $E[X \mid Y = 1] = 1$. If $Y = 0$, then $X > 1$. In this case, let the number of remaining flips (after the first flip until the first heads) be Z. Then, by the linearity of expectations,

Geometric Distribution

Expectation of Geometric Random Variable (Simpler Version)

Recall that X corresponds to the number of flips until the first heads given that each flip is heads with probability p. Let $Y = 1$ if the first flip is tails and $Y = 1$ if the first flip is heads. We have

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

If $Y = 1$ then $X = 1$, so $E[X \mid Y = 1] = 1$. If $Y = 0$, then $X > 1$. In this case, let the number of remaining flips (after the first flip until the first heads) be Z. Then, by the linearity of expectations,

$$E[X] = (1 - p)E[Z + 1] + p \cdot 1 = (1 - p)E[Z] + 1$$

Geometric Distribution

Expectation of Geometric Random Variable (Simpler Analysis)

Recall that X corresponds to the number of flips until the first heads given that each flip is heads with probability p. Let $Y = 1$ if the first flip is tails and $Y = 1$ if the first flip is heads.

If $Y = 0$, then $X > 1$. In this case, let the number of remaining flips (after the first flip until the first heads) be Z. We have

$$E[X] = (1 - p)E[Z] + 1$$

Expectation of Geometric Random Variable (Simpler Analysis)

Recall that X corresponds to the number of flips until the first heads given that each flip is heads with probability p. Let $Y = 1$ if the first flip is tails and $Y = 1$ if the first flip is heads.

If $Y = 0$, then $X > 1$. In this case, let the number of remaining flips (after the first flip until the first heads) be Z. We have

$$E[X] = (1 - p)E[Z] + 1$$

By the memoryless property of geometric random variables, Z is also a geometric random variable with parameter p. Hence $E[Z] = E[X]$, since they both have the same distribution. We therefore have

Geometric Distribution

Expectation of Geometric Random Variable (Simpler Analysis)

Recall that X corresponds to the number of flips until the first heads given that each flip is heads with probability p. Let $Y = 1$ if the first flip is tails and $Y = 1$ if the first flip is heads.

If $Y = 0$, then $X > 1$. In this case, let the number of remaining flips (after the first flip until the first heads) be Z. We have

$$E[X] = (1-p)E[Z] + 1$$

By the memoryless property of geometric random variables, Z is also a geometric random variable with parameter p. Hence $E[Z] = E[X]$, since they both have the same distribution. We therefore have

$$E[X] = (1-p)E[X] + 1$$

which yields $E[X] = 1/p$.

Example: Coupon Collector's Problem

Example

Suppose that each box of cereal contains one of n different coupons. Once you obtain one of every type of coupon, you can send in for a prize. Assuming that the coupon in each box is chosen independently and uniformly at random from the n possibilities and that you do not collaborate with others to collect coupons, how many boxes of cereal must you buy before you obtain at least one of every type of coupon?

This simple problem arises in many different scenarios and will reappear in several places in the book.

Example: Coupon Collector's Problem

Example

Suppose that each box of cereal contains one of n different coupons. Once you obtain one of every type of coupon, you can send in for a prize. Assuming that the coupon in each box is chosen independently and uniformly at random from the n possibilities and that you do not collaborate with others to collect coupons, how many boxes of cereal must you buy before you obtain at least one of every type of coupon?
This simple problem arises in many different scenarios and will reappear in several places in the book.

Preliminary

Let X be the number of boxes bought until at least one of every type of coupon is obtained. We only need to compute $E[X]$. If X_i is the number of boxes bought while you had exactly $i-1$ different coupons, then clearly $X = \sum_{i=1}^{n} X_i$.

Example: Coupon Collector's Problem

Analysis

Obviously, each X_i is a geometric random variable. When exactly $i-1$ coupons have been found, the probability of obtaining a new coupon is

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

Example: Coupon Collector's Problem

Analysis

Obviously, each X_i is a geometric random variable. When exactly $i-1$ coupons have been found, the probability of obtaining a new coupon is

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

Hence, X_i is a geometric random variable with parameter p_i, and

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

Example: Coupon Collector's Problem

Analysis

Obviously, each X_i is a geometric random variable. When exactly $i-1$ coupons have been found, the probability of obtaining a new coupon is

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

Hence, X_i is a geometric random variable with parameter p_i, and

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

Applying the the linearity of expectations, we have that

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

$$= \sum_{i=1}^{n} \frac{n}{n-i+1} = n\sum_{i=1}^{n} \frac{1}{i}$$

Example: Coupon Collector's Problem

Harmonic Number

The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.

Example: Coupon Collector's Problem

Harmonic Number

The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.

Proof:

Since $1/x$ is monotonically decreasing, we can write

$$\ln n = \int_{x=1}^{n} \frac{1}{x}\, dx \le \sum_{k=1}^{n} \frac{1}{k}$$

Example: Coupon Collector's Problem

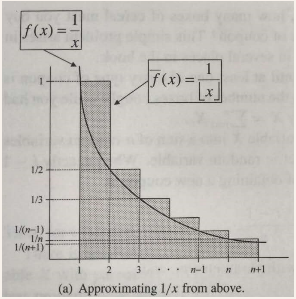
Harmonic Number

The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.

Proof:

Since $1/x$ is monotonically decreasing, we can write

$$\ln n = \int_{x=1}^{n} \frac{1}{x}\, dx \le \sum_{k=1}^{n} \frac{1}{k}$$



Example: Coupon Collector's Problem

Harmonic Number

The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.

Example: Coupon Collector's Problem

Harmonic Number

The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.

Proof:

Since $1/x$ is monotonically decreasing, we also have

$$\sum_{k=2}^{n} \frac{1}{k} \le \int_{x=1}^{n} \frac{1}{x}\, dx = \ln n$$

Example: Coupon Collector's Problem

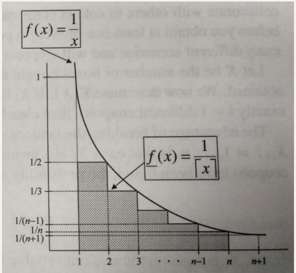
Harmonic Number

The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.

Proof:

Since $1/x$ is monotonically decreasing, we also have

$$\sum_{k=2}^{n} \frac{1}{k} \leq \int_{x=1}^{n} \frac{1}{x}\, dx = \ln n$$



(b) Approximating 1/x from below.

Example: Coupon Collector's Problem

Harmonic Number

The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.

Proof:

Since $1/x$ is monotonically decreasing, we have

$$\ln n = \int_{x=1}^{n} \frac{1}{x}\, dx \leq \sum_{k=1}^{n} \frac{1}{k}$$

$$\sum_{k=2}^{n} \frac{1}{k} \leq \int_{x=1}^{n} \frac{1}{x}\, dx = \ln n$$

Example: Coupon Collector's Problem

Harmonic Number

The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.

Proof:

Since $1/x$ is monotonically decreasing, we have

$$\ln n = \int_{x=1}^{n} \frac{1}{x}\, dx \leq \sum_{k=1}^{n} \frac{1}{k}$$

$$\sum_{k=2}^{n} \frac{1}{k} \leq \int_{x=1}^{n} \frac{1}{x}\, dx = \ln n$$

Example: Coupon Collector's Problem

Harmonic Number

The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.

Proof:

Since $1/x$ is monotonically decreasing, we have

$$\ln n = \int_{x=1}^{n} \frac{1}{x}\, dx \leq \sum_{k=1}^{n} \frac{1}{k}$$

$$\sum_{k=2}^{n} \frac{1}{k} \leq \int_{x=1}^{n} \frac{1}{x}\, dx = \ln n$$

Hence,

$$\ln n \leq H(n) \leq \ln n + 1$$

Table of Contents

1. Random Variables and Expectation
2. The Bernolli and Binomial Random Variables
3. Conditional Expectation
4. The Geometric Distribution
5. Application: The Expected Run-Time of Quicksort

The Expected Run-Time of Quicksort

Details Refer to the Book

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} \\
&= \sum_{k=2}^{n} \sum_{i=1}^{n-k+1} \frac{2}{k} = \sum_{k=2}^{n} (n-k+1) \cdot \frac{2}{k} \\
&= (n+1) \sum_{k=2}^{n} \frac{2}{k} - 2(n-1) \\
&= (2n+2) \sum_{k=1}^{n} \frac{1}{k} - 4n \\
&= 2n \ln n + \Theta(n).
\end{aligned}
$$

Q & A

Q & A

Thank you.