Balls & Bins

Anil Maheshwari

School of Computer Science
Carleton University
Canada

Outline

Basics

Basic Definition

Definitions

Sample Space S = Set of Outcomes.
Events $\mathcal{E}$ = Subsets of S.
Probability is a function from subsets $A \subseteq S$ to positive real numbers between $[0, 1]$ such that:

1. $Pr(S) = 1$
2. For all $A, B \subseteq S$ if $A \cap B = \emptyset$, $Pr(A \cup B) = Pr(A) + Pr(B)$.
3. If $A \subset B \subseteq S$, $Pr(A) \leq Pr(B)$.
4. Probability of complement of A, $Pr(\bar{A}) = 1 - Pr(A)$.

Examples

1. Flipping a fair coin:
   $S = \{H, T\}$;
   $\mathcal{E} = \{\emptyset, \{H\}, \{T\}, S = \{H, T\}\}$
2. Flipping fair coin twice:
   $S = \{HH, HT, TH, TT\}$;
   $\mathcal{E} = \{\emptyset, \{HH\}, \{HT\}, \{TH\}, \{TT\},$
   $\{HH, TT\}, \{HH, TH\}, \{HH, HT\},$
   $\{HT, TH\}, \{HT, TT\}, \{TH, TT\},$
   $\{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\},$
   $\{HT, TH, TT\}, S = \{HH, HT, TH, TT\}\}$
3. Rolling fair die twice:
   $S = \{(i, j) : 1 \leq i, j \leq 6\}$;
   $\mathcal{E} = \{\emptyset, \{1, 1\}, \{1, 2\}, \ldots, S\}$

Random Variable

Expectation

Definition

A random variable X is a function from sample space S to real numbers, $X : S \to \Re$.

Expected value of a discrete random variable X:

$E[X] = \sum_{s \in S} X(s) * Pr(X = X(s))$.

Example: Flip a fair coin. Let r.v. $X : \{H, T\} \to \Re$ be

$$X = \begin{cases} 1 & \text{Outcome is Heads} \\ 0 & \text{Outcome is Tails} \end{cases}$$

$$E[X] = \sum_{s \in \{H,T\}} X(s) * Pr(X = X(s)) = 1 * \frac{1}{2} + 0 * \frac{1}{2} = \frac{1}{2}$$

Linearity of Expectation

Consider two random variables $X, Y : S \to \Re$, then
$E[X+Y] = E[X] + E[Y]$.
In general, consider n random variables $X_1, X_2, \ldots, X_n$ such that
$X_i : S \to \Re$, then $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$.

Example: Flip a fair coin n times and define n random variable $X_1, \ldots, X_n$ as

$$X_i = \begin{cases} 1 & \text{Outcome is Heads} \\ 0 & \text{Outcome is Tails} \end{cases}$$

$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = \frac{1}{2} + \cdots + \frac{1}{2} = \frac{n}{2}$
(Expected # of Heads in n tosses)

Geometric Distribution

Geometric Distribuition

Definition

Perform a sequence of independent trials till the first success. Each trial succeeds with probability p (and fails with probability $1-p$).
A geometric r.v. X with parameter p is defined to be equal to $n \in N$ if the first $n-1$ trials are failures and the n-th trial is success. Probability distribution function of X is $Pr(X=n) = (1-p)^{n-1}p$.

Let Z to be the $r.v.$ that equals the # failures before the first success, i.e. $Z = X - 1$.

Problem: Evaluate $E[X]$ and $E[Z]$.

Computation of $E[Z]$

$Z =$ # failures before the first success.
To show: $E[Z] = \frac{1-p}{p}$ and $E[X] = 1 + E[Z] = \frac{1}{p}$

Examples

Examples:

1. Flipping a fair coin till we get a Head:
   $p = \frac{1}{2}$ and $E[X] = \frac{1}{p} = 2$
2. Roll a die till we see a 6:
   $p = \frac{1}{6}$ and $E[X] = \frac{1}{p} = 6$
3. Keep buying LottoMax tickets till we win (assuming we have 1 in 33294800 chance).
   $p = \frac{1}{33294800}$ and $E[X] = \frac{1}{p} = 33,294,800$.

Coupon Collector Problem

Coupon's Collector Problem

Problem Definition

A cereal manufacturer has ensured that each cereal box contains a coupon among a possible n coupon types. Probability that a box contains any particular type of coupon is $\frac{1}{n}$. Show that the expected number of boxes that we need to buy to collect all the n coupons is $n \ln n$.

Is $E[N] = nH_n = n \ln n$ a good estimate?

Balls & Bins

Balls & Bins

Model

We have m Balls and n Bins. We throw each ball in a bin uniformly at random.

What is the probability of following events:

1. Balls i and j are in the same bin.
2. Bin $\#i$ receives (a) 0 balls, (b) k balls, and (c) $\geq k$ balls.
3. All bins have $\leq \frac{c \ln n}{\ln \ln n}$ balls.

Applications: Birthday Paradox, Load Balancing, Perfect Hashing

Probability[Balls i and j in the same bin]

Number of Balls = m
Number of Bins = n.

$Pr[\text{Balls } i \text{ and } j \text{ in same bin}] = \frac{1}{n}$

Expected number of collisions

Number of Balls = m

Number of Bins = n

Show that Expected number of collisions is $\frac{1}{n}\binom{m}{2}$

Birthday Paradox

Number of Balls = m = Number of Students
Number of Bins = n = Number of days in a Year.

For two students to have same Birthday:
What value of m will result in $E[X] = \frac{1}{n}\binom{m}{2} \geq 1$

Answer: $m = 28$, since $E[X] = \frac{1}{365}\binom{28}{2} = 1.04 > 1$

Birthday Paradox Contd.

What is minimum value of m so that the probability that two students share the same birthday is $\geq \frac{1}{2}$?

Number of Balls in Bin i

Number of Balls = m; Number of Bins = n.

Problem I

What is the probability that Bin i receives no balls?

$$\left(1 - \frac{1}{n}\right)^m \leq e^{-\frac{m}{n}}$$

If $n = m$, $(1 - \frac{1}{n})^n \leq e^{-1} = 0.37$.

Problem II

What is the probability that Bin i receives exactly k balls?

$$\binom{m}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{m-k}$$

Number of Balls in Bin i contd.

Number of Balls = m; Number of Bins = n.

Problem III

What is the probability that Bin i receives $\geq k$ balls?

$$\leq \binom{m}{k}\left(\frac{1}{n}\right)^k$$

If $n = m$ and using Stirling's approximation ($\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$), we have $\binom{n}{k}\left(\frac{1}{n}\right)^k \leq \left(\frac{e}{k}\right)^k$

Expected Number of Balls in a Bin

Number of Balls = m; Number of Bins = n.

Problem IV

Show that the Expected # of Balls in a Bin is $\frac{m}{n}$

Expected Number of Empty Bins

Number of Balls = m; Number of Bins = n.

Problem V

What is Expected # of Empty Bins?

Define a r.v. X_i such that

$$X_i = \begin{cases} 1 & \text{if Bin } i \text{ is empty} \\ 0 & \text{Otherwise} \end{cases}$$

From Problem I, $Pr(X_i = 1) \leq e^{-\frac{m}{n}}$ and $E[X_i] \leq e^{-\frac{m}{n}}$

Thus, $E[\text{\# of Empty Bins}] = \sum_{i=1}^{n} E[X_i] \leq ne^{-\frac{m}{n}}$

When $n = m$, $E[\text{\# of Empty Bins}] \leq \frac{n}{e}$

Max # Balls in Bins

Number of Balls = Number of Bins = n.

Max # of Balls in Bins

With probability $\geq 1 - \frac{1}{n}$ all bins receive fewer than $3\frac{\ln n}{\ln \ln n}$ balls.

References

1. Probability and Computing by Mitzenmacher and Upfal, Cambridge Univ. Press 2005.
2. Introduction to Probability by Blitzstein and Hwang, CRC Press 2015.
3. Courses Notes of COMP 2804 by Michiel Smid.
4. My Notes on Algorithm Design.
5. Introduction to Probability by Blitzstein and Hwang, CRC Press 2015.