Lecture 10 Bins and Balls and Coupon Collectors, Michael Monagan

February 3, 2021 12:07 PM

Lecture 10: Applications of Discrete Random Variables

The bins and balls problem

Suppose we throw m balls into n bins randomly.
On average, how many bins will be empty?
On average, how many bins will have one ball in them?

Arises in "hashing" CMPT 307.

The coupon collectors problem

Arises in "probabilistic algorithms".

Suppose we have a bin containing n types of coupons and we draw coupons one at a time from the bin at random. Assume the probability of drawing each type of coupon is $1/n$ and the bin has a very large number of coupons. On average, how many draws do we need to make till we get all n coupons?

Definition

Let S be a sample space and X a random variable on S. Let x be a value from the range of X. The probability of x, denoted by $Pr(X = x)$ is the sum of the probabilities of all outcomes s of S such that $X(s) = x$.

Example 1 Let S be the set of all binary sequences of size $n = 3$ bits.
Let $X(s)$ be the number of 1 bits in a binary string $s \in S$.
Here the range of X denoted $r(X)$ is $\{0, 1, 2, 3\}$.

$S = \{000, 001, \ldots, 111\}$ $|S| = 8$.

$X(011) = 2$

$\Pr(X = 0) =$ $1/8$ $S = 000$

$\Pr(X = 1) =$ $3/8$ $S = 100$ 010 001

$\Pr(X = 2) =$ $3/8$ $S = 11\underline{0}$ $1\underline{0}1$ $\underline{0}11$ $\binom{3}{2} = 3.$

$\Pr(X = 3) =$ $1/8$ $S = 111$

$\Pr(X = k) =$ $\binom{3}{k}/8$

Definition

The **expected value** of a random variable X on a sample space S is defined by

$$E(X) = \sum_{x \in r(X)} x Pr(X = x) = \sum_{s \in S} X(s) Pr(s).$$

Example 1 (cont.)

x	0	1	2	3
$Pr(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$E[X] = \sum_{x \in r(X)} x Pr(X = x) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = \frac{3}{2}$$

$$E(X) = \sum_{s \in S = \{000, 001, \ldots\}} X(s) \cdot \underbrace{Pr(s)}_{= 1/8} = \frac{1}{8}\Big[\underbrace{X(000)}_{0} + \underbrace{X(001)}_{1} + \underbrace{X(010)}_{1} + \underbrace{X(100)}_{1} + \underbrace{X(110)}_{2} + \underbrace{X(011)}_{2} + \underbrace{X(101)}_{2} + \underbrace{X(111)}_{3}\Big] = \frac{12}{8} = \frac{3}{2}$$

Theorem (Linearity of Expection)

Let X and Y be two random variables on the same sample space S and $a \in \mathbb{R}$. Then

(1) $E(aX) = aE(X)$ *and*

(2) $E(X + Y) = E(X) + E(Y)$. $\Rightarrow E(X + (Y + Z)) \overset{(2)}{=} E(X) + E(Y + Z) \overset{(2)}{=} E(X) + E(Y) + E(Z)$

Proof (1) Exercise

(2) $E(X+Y) = \sum_{s \in S} (X+Y)(s) \cdot Pr(s).$

$(\sin + \cos)(x) = \sin x + \cos x.$

$$= \sum_{s \in S} (X(s) + Y(s)) \cdot Pr(s).$$

$$= \sum [X(s) \cdot Pr(s) + Y(s) Pr(s)]$$

$$= \Big(\underbrace{\sum X(s) \cdot Pr(s)}_{E(X)}\Big) + \Big(\underbrace{\sum Y(s) \cdot Pr(s)}_{E(Y)}\Big).$$

The bins and balls problem

$m=4$

$n=3.$

bin 1 bin 2 bin 3.

Suppose we throw m balls into n bins randomly.

Question 1: What is the probability that bin i has k balls?

$Pr(\text{a ball is in bin } i) = 1/n \qquad Pr(\text{ball is not in bin } i) = \frac{n-1}{n} = 1-\frac{1}{n}$

$Pr(\text{bin } i \text{ has 0 balls}) = \underset{\text{ball 1}}{(1-\tfrac{1}{n})} \cdot \underset{\text{ball}}{(1-\tfrac{1}{n})} \cdot \cdots \cdot \underset{\text{ball } m}{(1-\tfrac{1}{n})} = (1-\tfrac{1}{n})^m$

$m=3$ balls.

$Pr(\text{bin } i \text{ has 1 ball}) = \binom{m}{1}(\tfrac{1}{n}) \cdot (1-\tfrac{1}{n})^{m-1}$

any one of m tosses one goes in $m-1$ don't go in bin i

$Pr(\text{bin } i \text{ has 2 balls}) = \binom{m}{2} \cdot (\tfrac{1}{n})^2 (1-\tfrac{1}{n})^{m-2}$

i

$\binom{m}{2}$ ways for 2 balls to go in two in $m-2$ don't.

$Pr(\text{bin } i \text{ has } k \text{ balls}) = \binom{m}{k} (\tfrac{1}{n})^k (1-\tfrac{1}{n})^{m-k}.$

This is binomially distributed with $p=\frac{1}{n}$ and $n=m=\#$balls.

Question 2: On average, how many bins are empty?

Let X be the number of empty bins $\qquad X(\sqcup \sqcup \sqcup)=1.$

Let $X_i = \begin{cases} 1 & \text{if bin is empty} \\ 0 & \text{otherwise.} \end{cases}$ $\qquad X_i$ is called an indicator random variable.

$X = X_1 + X_2 + X_3 + \cdots + X_n$

$E(X) = E(X_1 + X_2 + \cdots + X_n) \overset{\text{Thm(2)}}{=} E(X_1) + E(X_2) + \cdots + E(X_n).$

$E(X_i) = 1 \cdot Pr(X_i=1) + 0 \cdot Pr(X_i=0) = 1 \cdot (1-\tfrac{1}{n})^m + 0.$ $\qquad \lim_{n\to\infty} (1-\tfrac{1}{n})^n = \frac{1}{e}$

$E(X) = n \cdot E(X_i) = n \cdot (1-\tfrac{1}{n})^m = n(\underbrace{(1-\tfrac{1}{n})^n}_{\to 1/e})^{m/n}$

$\approx n \cdot e^{-m/n}$

Case $m=n$: $\quad E(X) \approx n \cdot e^{-1} = .368n$

So for $m=n$, on average 37% of the bins will be empty.

Exercise: On average, how many bins have one ball? $\qquad$ Let $X_i = \begin{cases} 1 & \text{if bin } i \text{ has 1 ball} \\ 0 & \text{otherwise.} \end{cases}$

The coupon collectors problem

$n=3$ 1 2 3

$C_1\ C_2\ C_1\ C_2\ C_3$

$t_1=1$ $t_2=1$ $t_3=3.$

$T=5$

Suppose a large bin contains many copies of $n = \cancel{10}$ 3 coupons. Assuming there are an equal number of each coupon, if we draw coupons at random from the bin, on average, how many draws will it take to get all n coupons?

Let T be the # coupons we draw to one of each.

Let t_i be the # coupons we draw to get the i'th coupon assuming we already have $i-1$ coupons.

The $T = t_1 + t_2 + t_3 + \cdots + t_n$.

Let $p_i = Pr(t_i) = Pr(\text{next draw is new}) = \frac{(n-(i-1))}{n}$

$p_1 = \frac{n-(1-1)}{n} = 1 \qquad P_n = \frac{(n-(n-1))}{n} = \frac{1}{n}.$

$E(T) = E(t_1 + t_2 + \cdots + t_n) = \underline{E(t_1)} + E(t_2) + \cdots + E(t_n).$

Each t_i is geometrically distributed with parameter $p = p_i$ so $E(t_i) = 1/p_i$.

The coupon collectors problem continued.

$$E(T) = 1/p_1 + 1/p_2 + \cdots + 1/p_n = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1}$$

$$= n\Big[\underbrace{\tfrac{1}{n} + \tfrac{1}{n-1} + \cdots + \tfrac{1}{2} + 1}_{H_n \text{ the } n\text{'th Harmonic number.}}\Big]$$

$$= n \cdot H_n.$$

For $n=3$: $E(T) = 3\cdot\left[\frac{1}{3}+\frac{1}{2}+1\right] = 3\cdot\left(\frac{11}{6}\right) = \frac{11}{2} = \underline{5.5}$

Exercise: On average, how many times must you toss a fair coin before you get a head and a tail?

Midterm: 45 mins + 10 mins for upload to crowdmark.

Connect to zoom. $\le$ 12:25 pm

Crowdmark $\xrightarrow[\text{Midterm}]{\text{Email}}$ 12:28pm, Due 1:25pm.

−10% for each minute late.

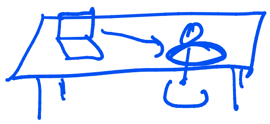
Written by hand on paper. No tablets.

⟶ Breakouts with a TA.

Written by hand on paper. No tablets.

→ Breakouts with a TA.

→ Must show us your <u>hands</u> + <u>desktop</u>. + face.



Material. Letures 1-8 Assignments 1 & 2 only.

Study. Cheat Sheet. (write on both sides).

Go through notes & write formulas + examples. (2 hours)

Review problems. (practice midterm). <u>90 mins max.</u>

Exam Resources