CS109: Probability for Computer Scientists

Lecture 16 Worksheet — Algorithmic Analysis

Feb 13, 2026

---

1 Quick Review

If X_i is an indicator RV where it is a 1 if event E_i occurs, show that $E[X_i] = P(E_i)$.

Solution: $E[X_i] = 1 \cdot P(E_i) + 0 \cdot (1 - P(E_i)) = P(E_i)$.

2 Differential Privacy

$X_1, \ldots, X_{100}$ are i.i.d. Bernoulli(p). For each i and Y_i is the output of the program in the slides.

(a) Compute $E[Y_i]$.

Solution: With probability 1/2, output is a fair random bit (mean 1/2). With probability 1/2, output is X_i (mean p). So

$$E[Y_i] = \tfrac{1}{2} \cdot \tfrac{1}{2} + \tfrac{1}{2} \cdot p = \tfrac{1}{4} + \tfrac{p}{2}.$$

(b) Let $Z = \sum_{i=1}^{100} Y_i$. Compute $E[Z]$.

Solution: By linearity,

$$E[Z] = \sum_{i=1}^{100} E[Y_i] = 100\left(\tfrac{1}{4} + \tfrac{p}{2}\right) = 25 + 50p.$$

(c) How could you estimate p? Leave your answer in terms of Z.

Solution: Set $Z \approx E[Z] = 25 + 50p$, so

$$\hat{p} = \frac{Z - 25}{50}.$$

(d) **Challenge**: what is the probability that your estimate is good?

Solution: Use a concentration bound (like Hoeffding/Chernoff) on Z around $E[Z]$, then convert that bound through $\hat{p} = (Z - 25)/50$.

3 Computer Cluster Utilization

A cluster has k servers. Each request independently goes to server i with probability p_i.

Let A_i be the event that server i receives no requests in the first n requests. Let

$$X = \#\{\text{events } A_1, \ldots, A_k \text{ that occur}\}.$$

(a) Can you express X in terms of the sum of another RV? Recall that A_i is an event and not a random variable.

Solution: Let $B_i = \mathbf{1}\{A_i\}$. Then

$$X = \sum_{i=1}^{k} B_i.$$

(b) Compute $E[X]$.

Solution: By linearity,

$$E[X] = \sum_{i=1}^{k} E[B_i] = \sum_{i=1}^{k} P(A_i).$$

For server i, each request misses it with probability $(1 - p_i)$, so $P(A_i) = (1 - p_i)^n$. Hence

$$E[X] = \sum_{i=1}^{k} (1 - p_i)^n.$$

4 Toy Collection

You are trying to collect n distinct toys. On each purchase, each toy is equally likely.

Let X be the number of purchases until you have at least one of each toy.

(a) Can you define X as the sum of some other RVs?

Solution: Define X_i = number of purchases needed to get a new toy after already having i distinct toys, for $i = 0, 1, \ldots, n-1$. Then

$$X = X_0 + X_1 + \cdots + X_{n-1}.$$

(b) What is the other RV? And what is its expectation?

Solution: After i distinct toys, success probability is $p_i = \frac{n-i}{n}$ (next purchase is unseen). So $X_i \sim \text{Geo}(p_i)$ and

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

(c) What is $E[X]$?

Solution: By linearity,

$$E[X] = \sum_{i=0}^{n-1} E[X_i] = \sum_{i=0}^{n-1} \frac{n}{n-i} = n \sum_{j=1}^{n} \frac{1}{j}.$$

5 Expected Runtime (Netflix)

Let L be movie file location with:

Location ℓ	$E[\text{Runtime} \mid L = \ell]$	$P(L = \ell)$
Palo Alto	0.3s	0.10
SoCal	1.6s	0.50
Japan	300s	0.37
Space	7200s	0.03

Compute the expected runtime of `database.get_movie(movie_name)`.

Solution: By law of total expectation,

$$E[\text{Runtime}] = 0.3(0.10) + 1.6(0.50) + 300(0.37) + 7200(0.03) = 327.83 \text{ s}$$

(about 5.46 minutes).

6 Analyzing Recursive Code

```
int Recurse() {
   int x = randomInt(1, 3);  // equally likely
   if (x == 1) return 3;
   else if (x == 2) return 5 + Recurse();
   else return 7 + Recurse();
}
```

Let Y be the value returned by `Recurse()`. Compute $E[Y]$.

Solution: Condition on the first random choice:

$$E[Y] = \frac{1}{3}(3) + \frac{1}{3}(5 + E[Y]) + \frac{1}{3}(7 + E[Y]).$$

So

$$E[Y] = 5 + \frac{2}{3}E[Y] \quad \Rightarrow \quad \frac{1}{3}E[Y] = 5 \quad \Rightarrow \quad E[Y] = 15.$$