Expectation

Expectation of Products

Example: R outcome of 6-sided die

- $E[R] = 7/2 = 3.5$
- $E[R]^2 = 49/4 = 12.25$
- $E[R^2]$?

$$\begin{aligned} E[R^2] &= \sum_{w \in S} R^2(w) \Pr[w] \\ &= \sum_{i=1}^{6} i^2 \Pr[R(w) = i] \\ &= \sum_{i=1}^{6} i^2/6 \\ &= 15.17 \end{aligned}$$

Note: $E[R_1 R_2] \neq E[R_1]E[R_2]$ for arbitrary r.v. R_1, R_2.

Claim: If R_1, R_2 *independent* then

$$E[R_1 R_2] = E[R_1]E[R_2].$$

Proof:

- Event $R_1 R_2 = r$ is $\{w | R_1(w) R_2(w) = r\}$.
- Equivalently, $\{w | R_1(w) = r_1, R_2(w) = r_1, r_1 r_2 = r\}$

$$\begin{aligned} E[R_1 R_2] &= \sum_{r_1} \sum_{r_2} r_1 r_2 \Pr[R_1 = r_1 \cap R_2 = r_2] \\ &= \sum_{r_1} \sum_{r_2} r_1 r_2 \Pr[R_1 = r_1] \Pr[R_2 = r_2] \\ &= \sum_{r_1} r_1 \Pr[R_1 = r_1] \sum_{r_2} r_2 \Pr[R_2 = r_2] \\ &= E[R_1]E[R_2] \end{aligned}$$

[[*Does converse hold? See end of lecture...*]]

Conditional Expectation

Def:

$$E[R|A] = \sum_{w \in S} R(w) \cdot \Pr[w|A]$$

[[*derived from definitions*]]

Example: 6-sided die

Question: Expected number given its even?

- R = number rolled
- A = event that number is even
- Then
  $E[R|A] = \sum_{i=1}^{6} i \Pr[i|A]$
  $= 2 \cdot (1/3) + 4 \cdot (1/3) + 6 \cdot (1/3) = 4$

Note: $E[R|A]$ just expectation w.r.t. prob. measure $\Pr_A[\cdot]$ (from pset).

Claim: (total expectation): If events A_i partition sample space, then

$$E[R] = \sum_i E[R|A_i] \Pr[A_i].$$

[[*For proof, see book.*]]

Applications of Probability

Birthday Paradox

Question: Probability two of us have same birthday?

Variables: m people, N days

Assumptions:

- for each person, all bdays equally likely

  [[*actually more likely to be born on a weekday; most common birthday Oct. 5th, least common May 22nd.*]]

- bdays mutually indep

  [[*not if there are twins, for example*]]

[[*Assumptions valid for CS applications, will see later.*]]

Four-step method:

1. sample space: map people i to bdays b_i

$$S = \{(b_1, \ldots, b_m) | b_i \in \{1, \ldots, N\}\}$$

2. events:

   A = event ≥ 2 people have same bday [[*hard to evaluate, use complement instead*]]

   A^c = event no two people have same bday

$$A^c = \{(b_1, \ldots, b_m) | \forall i \neq j, b_i \neq b_j\}$$

   Recall: $\Pr[A] = 1 - \Pr[A^c]$

3. outcome prob.:

   - $\Pr[b_i = k] = 1/N$ [[*by 1'st assumption*]]
   - $\Pr[(b_1, \ldots, b_m) = (k_1, \ldots, k_m)] = \prod_i \Pr[b_i = k_i] = (1/N)^m$ [[*by 2'nd assumption*]]

   so uniform

4. event prob.:

   - $|A^c| = N(N-1)\ldots(N-m+1) = N!/(N-m)!$
   - $|S| = N^m$

   so $\frac{N!}{N^m(N-m)!}$

Claim: Stirling approx: $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$

a lot of math... Pr[2 of 23 share bday] $> 1/2$.

- 0.4 for $n = 20$
- 0.7 for $n = 30$
- 0.999998876 for $n = 100$

[[*Poll class to see how many people have same bday.*]]

Alternatively:

1. sample space: $S = \{(b_1, \ldots, b_m)\}$

2. events:

   B_i = event $b_i \notin \{b_1, \ldots, b_{i-1}\}$

$$\begin{aligned}\Pr[A^c] &= \Pr[B_1 \cap \ldots \cap B_m] \\ &= \Pr[B_1]\ldots\Pr[B_n | B_1, \ldots, B_{n-1}]\end{aligned}$$

3. outcome prob.: uniform

4. event prob.:

   $\Pr[B_i | \cap_{j<i} B_j] = 1 - \Pr[B_i^c | \cap_{j<i} B_j] = 1 - (i-1)/N$

Claim: $(1-x) \leq e^{-x}$ for all x (good approx. if x close to 0)

$$\begin{aligned} \Pr[A^c] &= \Pr[B_1] \ldots \Pr[B_m | B_1, \ldots, B_{m-1}] \\ &= (1-0/d)...(1-(n-1)/d) \\ &\leq e^{-0/d} e^{-1/d} ... e^{-(m-1)/N} \\ &= e^{-m^2/2N} \end{aligned}$$

Note: Constant prob. of collision for $m \geq \sqrt{2N}$.

Hashing (Balls and Bins)

- want to store m records
- using N keys
- function h maps record to key
- *if* h maps randomly, need $N = m^2$-sized array to avoid collisions
- resolve collisions with linked list (want to bound max bin size)

Question: suppose choose $N = m$, then how long are the linked lists?

- let n be number records/keys ($n = N = m$)
- let X be rand. var. equal to max size.
- find smallest k s.t. $P[X \geq k] \leq 1/2$
- Let X_i = size of list i.
- Then

$$(X \geq k) =$$

$$(X_i \geq k) \cup \ldots \cup (X_n \geq k).$$

- By *union bound*

$$P[X \geq k] \leq \sum_{i=1}^{n} P[X_i \geq k]$$

so find k s.t. $P[X_i \geq k] \leq \frac{1}{2n}$.

$$\begin{aligned} P[X_i \geq k] &= P[(X_i = k) \cup (X_i = k+1) \cup \\ &\quad \ldots \cup (X_i = n)] \\ &= \sum_{j=k}^{n} P[X_i = j] \\ &= \sum_{j=k}^{n} \binom{n}{j} \left(\frac{1}{n}\right)^j \left(1 - \frac{1}{n}\right)^{n-j} \\ &\leq \sum_{j=k}^{n} \left(\frac{1}{j}\right)^j \\ &\leq \sum_{j=k}^{n} \left(\frac{1}{k}\right)^j \\ &= \left(\frac{1}{k}\right)^k \sum_{j=0}^{\infty} \left(\frac{1}{k}\right)^j \\ &\leq 2\left(\frac{1}{k}\right)^k \end{aligned}$$

Conclusion:

$$2\left(\frac{1}{k}\right)^k \leq \frac{1}{2n}$$

if

$$k > 2 \ln n / \ln \ln n$$

[[*Plug* $k = 2\ln n/\ln\ln n$ *into* $k \ln k$ *to check it's sufficient (i.e.,* $\geq \ln n$*).*]]

E.g., $n = 10^6$, w.p.r. 1/2, just 10 collisions max.

Probability Practice

Question: Roll a 6-sided die twice.

- A = event that first roll is odd
- B = event that sum of rolls is odd

Are A and B independent?

Question: Roll a 6-sided die twice. Let X be the sum of the rolls and Y be the difference.

1. Are X and Y independent?
2. What is $E[X]$?
3. What is $E[Y]$?
4. What is $E[XY]$?

[[*See MIT OpenCourseWare Recitation Notes*]]

Question: Pick a number $n \in \{1, \ldots, 6\}$

- Roll 2 die
- If n doesn't come up, lose \$1
- If n comes up once, win \$1
- If n comes up twice, win \$2
- If n comes up three times, win \$4

What is expected payoff?

Question: Roll two die. Advance in board game as follows:

- go forward # squares equal to sum of rolls
- if a double, roll again and advance additional # squares equal to sum of rolls
- if a double, roll again and advance additional # squares equal to sum of rolls
- third time doubles, go back to start position

How many squares do you advance in expectation?

Question: Given urn with r red and b black balls:

- remove balls one at a time, at random without replacement
- let X be # red balls removed before first black ball

What is $E[X]$?

Hint: look for nice random variables X_i s.t. $X = \sum_i X_i$ and use linearity of expectation.