

Class 4 in-class problems, 18.05, Spring 2022

Concept questions

Concept question 1.

Suppose X a random variable with the CDF shown.

values of X :	1	3	5	7
cdf $F(a)$:	0.5	0.75	0.9	1

What is $P(X \leq 3)$?

- (a) 0.15 (b) 0.25 (c) 0.5 (d) 0.75

Solution: (d) 0.75. $P(X \leq 3) = F(3) = 0.75$.

What is $P(X = 3)$? Same distribution as above

- (a) 0.15 (b) 0.25 (c) 0.5 (d) 0.75

Solution: (b) $P(X = 3) = F(3) - F(1) = 0.75 - 0.5 = 0.25$.

Board questions

Problem 1. Computing expectation

Suppose X is a random variable with the following pmf.

X :	1	2	3
pmf:	1/4	1/2	1/4

Find $E[X]$ and $E[1/X]$.

Solution:

$$E[X] = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} = 2$$
$$E[1/X] = 1 \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{4} = \frac{7}{12}.$$

Problem 2. Interpreting expectation

(a) Would you accept a gamble that offers a 10% chance to win \$95 and a 90% chance of losing \$5?

(b) Would you pay \$5 to participate in a lottery that offers a 10% percent chance to win \$100 and a 90% chance to win nothing?

Hint: find the expected value of your winnings in each case.

Solution: (a) Here is the table for this random variable.

Values:	95	-5
PMF:	0.1	0.9

So, the expected value is $\$95 \cdot 0.1 - \$5 \cdot 0.9 = \$5$

(b) Has the same table, e.g. if you pay \$5 and win \$100, your net gain is \$95. So, it has the same expected value. If you play enough times, you will average winning \$5 for every game played.

Discussion

Framing bias / cost versus loss. The two situations are identical, with an expected value of gaining \$5. In a study, 132 undergrads were given these questions (in different orders) separated by a short filler problem. 55 gave different preferences to the two events. Of these, 42 rejected (a) but accepted (b). One interpretation is that we are far more willing to pay a cost up front than risk a loss. (See *Judgment under uncertainty: heuristics and biases* by Tversky and Kahneman.)

Loss aversion and cost versus loss sustain the insurance industry: people pay more in premiums than they get back in claims on average (otherwise the industry wouldn't be sustainable), but they buy insurance anyway to protect themselves against substantial losses. Think of it as paying \$1 each year to protect yourself against a 1 in 1000 chance of losing \$100 that year. By buying insurance, the expected value of the change in your assets in one year (ignoring other income and spending) goes from negative 10 cents to negative 1 dollar. But whereas without insurance you might lose \$100, with insurance you always lose exactly \$1.

Problem 3. Musical chairs or linearity of expectation

Suppose that there are n people at your table and everyone got up, ran around the room, and sat back down randomly (i.e., all seating arrangements are equally likely).

What is the expected value of the number of people sitting in their original seat?

Solution: Number the people from 1 to n . Let X_i be the Bernoulli random variable with value 1 if person i returns to their original seat and value 0 otherwise. Since person i is equally likely to sit back down in any of the n seats, the probability that person i returns to their original seat is $1/n$. Therefore $X_i \sim \text{Bernoulli}(1/n)$ and $E[X_i] = 1/n$.

Now, let X be the number of people sitting in their original seat following the rearrangement. Then

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

By linearity of expected values, we have

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

Extra discussion of derangements (Not required – for fun only.)

- It's neat that the expected value is 1 for any n .
- If $n = 2$, then both people either retain their seats or exchange seats. So $P(X = 0) = 1/2$ and $P(X = 2) = 1/2$. In this case, X never equals $E[X]$.
- The X_i are not independent (e.g. for $n = 2$, $X_1 = 1$ implies $X_2 = 1$).
- Expectation behaves linearly even when the variables are dependent.

Neat fact: A permutation of n people in which nobody returns to their original seat is called a derangement. The number of derangements turns out to be the nearest integer to

$n!/e$. Since there are $n!$ total permutations, we have:

$$P(\text{everyone in a different seat}) \approx \frac{n!/e}{n!} = 1/e \approx 0.3679.$$

It's surprising that the probability is about 37% regardless of n , and that it converges to $1/e$ as n goes to infinity.

<https://en.wikipedia.org/wiki/Derangement>

Problem 4. Bernoulli

(a) Suppose $X \sim \text{Bernoulli}(p)$. Find $E[X]$.

(This is important! Remember it!)

(b) Suppose $Y = X_1 + X_2 + \dots + X_{12}$, where each $X_i \sim \text{Bernoulli}(0.25)$. Find $E[Y]$.

Solution: (a) Here's the probability table for a Bernoulli random variable.

$X:$	0	1
pmf:	$1 - p$	p

So,

$$E[X] = (1 - p) \cdot 0 + p \cdot 1 = p$$

(b) By the linearity of expectation

$$E[X] = E[X_1] + E[X_2] + \dots + E[X_{12}] = 12 \cdot (0.25) = 3$$

(Remember, if X_1, \dots, X_{12} are independent, we say $X \sim \text{Binomial}(12, 0.25)$.)

In general, if $X \sim \text{Binomial}(n, p)$ then $E[X] = np$.

Problem 5. Don't let one failure stop you

Let $X = \#$ of successes before the second failure of a sequence of independent Bernoulli(p) trials. Find the pmf of X .

Hint: this requires a small amount of counting.

Solution: X takes values 0, 1, 2, We will show the pmf is $p(n) = (n + 1)p^n(1 - p)^2$.

For concreteness, we'll derive this formula for $n = 3$. Let's list the outcomes with three successes before the second failure. If $n = 3$, then we had 3 successes and 2 failures and the fifth trial was a failure. So each such outcome must have the form

_ _ _ _ F

with three S and one F in the first four slots. So we just have to choose which of these four slots contains the F . We can list all the possibilities

$$\{FSSSF, SFSSF, SSFSF, SSSFF\}$$

In other words, there are $\binom{4}{1} = 4$ such outcomes. Each of these outcomes has three S and two F , so each outcome has probability $p^3(1 - p)^2$. The probability that $n = 3$ is therefore

$$p(3) = P(X = 3) = 4p^3(1 - p)^2.$$

The same reasoning works for general n :

$$p(n) = \binom{n + 1}{1} p^n (1 - p)^2 = (n + 1)p^n (1 - p)^2.$$

In class examples and discussion

Gambler's fallacy. [roulette]

Fallacy: *If black comes up several times in a row then the next spin is more likely to be red.*

Truth: *$P(\text{red})$ remains the same. The roulette wheel spins are independent.*

Discussion: “On August 18, 1913, at the casino in Monte Carlo, black came up a record twenty-six times in succession [in roulette]. [There] was a near-panicky rush to bet on red, beginning about the time black had come up a phenomenal fifteen times. In application of the maturity [of the chances] doctrine, players doubled and tripled their stakes, this doctrine leading them to believe after black came up the twentieth time that there was not a chance in a million of another repeat. In the end the unusual run enriched the Casino by some millions of francs.” (I've lost the source for this. Wikipedia has a good article on the Gambler's Fallacy.)

Hot hand

Theory: *NBA players get 'hot'.*

Data: *The data show that player who has made 5 shots in a row is no more likely than usual to make the next shot.*

Discussion: See *The Hot Hand in Basketball: On the Misperception of Random Sequences* by Gilovich, Vallone and Tversky. (A link that worked when these slides were written is [https://doi.org/10.1016/0010-0285\(85\)90010-6](https://doi.org/10.1016/0010-0285(85)90010-6))

There is still some controversy about this. Some statisticians feel that the authors of the above paper erred in their analysis of the data and the data do support the notion of a hot hand in basketball.

Discussion: This reverses the Gambler's Fallacy, but is less obviously false, since we cannot model player psychology as well as we can model the Roulette wheel. The paper by Gilovich et al uses NBA and college game data to conclude that the hot hand is hot air.

Extra problems

Extra 1. *Suppose $X \sim \text{Binomial}(n, p)$, i.e. the number of successes in n , independent Bernoulli(p) trials. Explain why X is the sum of n Bernoulli(p) random variables.*

Solution: Let X_1, X_2, \dots, X_n be the n Bernoulli trials. X_i is 1 if it's a success and 0 if it's a failure. If there are k successes, then k of the X_i are 1 and $n - k$ are 0. If you sum, k ones and $n - k$ zeros you get k . That is, the sum of the n Bernoulli variables is the number of successes. In symbols

$$\sum_{i=1}^n X_i = \text{number of successes} = X.$$

MIT OpenCourseWare

<https://ocw.mit.edu>

18.05 Introduction to Probability and Statistics

Spring 2022

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.