PROBABILITY AROUND US

AUGUST 2014

PAVEL ETINGOF

Based on a previous version by Pavel Etingof and Alexander Kirillov Jr.

1. Lesson 1

1.1. **Basic probability.** The **probability of an event** A is a number $0 \le P(A) \le 1$ which characterizes the likelihood that A will occur. Here are some basic probability rules, which allow one to compute $P(A)$.

We have a collection of possible outcomes (cases) some of which are favorable (i.e., A occurs) while others are not (i.e., A does not occur).

Counting rule. If there are finitely many outcomes which are equally likely, then

$$P(A) = \frac{\text{number of outcomes giving } A}{\text{total number of possible outcomes}}$$

Example 1.1. The probability to pull out a face card (i.e., Jack, Queen, or King) out of a deck of cards is $12/52 = 3/13$.

Addition rule.

$$P(A \text{ or } B) = P(A) + P(B)$$

if A and B can't happen together. More generally,

$$P(A_1 \text{ or } A_2 \text{ or ... or } A_n) = P(A_1) + ... + P(A_n)$$

if no two of the events $A_1, ..., A_n$ can happen together.

Complement rule.

$$P(\text{not } A) = 1 - P(A).$$

Product rule. If we are doing two tests (e.g., tossing two coins), and the results of one do not depend on the results of the other, then the probability of getting event A in the first test and event B in the second one is

$$P(A \text{ and } B) = P(A) \cdot P(B).$$

This rule can be generalized to more than two tests: if we are doing n tests whose results are independent, then

$$P(A_1 \text{ and } A_2 \text{ and } \dots \text{and } A_n) = P(A_1)P(A_2)\dots P(A_n).$$

1.2. **Analyzing games of chance.** In this class, we will analyze and then play various games of chance. Each game will have two players, A and B, and each round of the game will have two possible outcomes: either A or B wins. After each round is played, the player who lost must pay the winner an amount (bet or stake), agreed upon by A and B before the beginning of the round. Note that the stakes of the two players could be different from each other, and change from round to round.

The goal of the analysis of the game is to determine whether on average it is profitable for A or for B, and to what extent. To this end, one computes **the expected gain** $E = E_A$ for this game for player A, i.e. the average amount of money A wins if he plays the game once.

The expected gain can be positive, negative, or zero.

If $E > 0$, the game is profitable for A; if $E < 0$, it is profitable for B; if $E = 0$, the game is fair: on average, neither A nor B wins.

The expected gain of a game is computed as follows. If in some game A gets X dollars if he wins, and pays Y dollars otherwise, then the expected gain is

$$E = P(A) \cdot X - P(\text{not } A) \cdot Y,$$

where $P(A)$ is the probability that A wins.

More generally, if there are n outcomes $A_1, ..., A_n$ and A wins X_i in outcome A_i (where X_i can be positive, negative, or zero), then the expected gain is

$$E = P(A_1)X_1 + ... + P(A_n)X_n.$$

Example 1.2. Suppose that the game is to roll three dice, and A wins \$ 1 if they give at least two equal numbers (otherwise B wins \$ 1). To compute the expected gain, we need to compute the probability of winning and losing for player A. In order for the player A to lose, the dice should yield three different numbers. There are 6 possible values for the first, 5 for the second (since it may not equal the first), and 4 for the third (since it may not be equal the first two). Thus, there are $6 \cdot 5 \cdot 4$ possible losing combinations, out of 6^3 possible combinations. Thus the probability of A losing is $\frac{6 \cdot 5 \cdot 4}{6 \cdot 6 \cdot 6} = 5/9$. Hence the probability of A winning is $1 - 5/9 = 4/9$. So the expected gain is $E = 4/9 - 5/9 = -1/9$. So this game is profitable for player B.

Example 1.3. Consider a lottery in which a lottery ticket costs \$ 1, and there are 10,000 tickets. Suppose there is one winning ticket with prize of $9,500$. In this case A is a buyer of a lottery ticket, and B is the lottery owner. We have $P(A) = 1/10,000$, so the expected gain is $0.95 - 0.9999 = -0.0499$. This means that this game is slightly unfair for A. In fact, any lottery is slightly unfair for the buyer of lottery tickets (otherwise it will not make sense for the lottery owner to organize a lottery). Yet, many people are willing to play since there is a possibility of winning a big prize, even though on average they lose.

1.3. **Rules of casino games.** In any real casino, any game is slightly unfair for the player and slightly profitable for the casino owner (otherwise it will not make sense for the owner to keep a casino). However, in our Casino "Pasha, Slava, and the Satan", we will let you choose, in each game, if you want to be A or B (we are happy to let you win, since in the end we will get your soul as payment in any case). Thus, for each game you should compute the expected gain, and choose to be A if it is positive and B if it is negative. If the expected gain is zero, the choice does not matter. Note that decisions based on intuition are often wrong (i.e., secretly dictated to you by the Satan!)

The exact rules of the Casino are as follows.

• To play in the Casino, you will use your points. Your points will be recorded in a special notebook under your name.

• At the beginning of each day, you'll be given 5 points, but you can also use the points remaining from the previous day, if any.

• In each class, you will be asked to analyze 4 games. You are required to compute the expected gain of all games, and say whether you want to be A or B. We will then look at your answers, pick two games and play them with you at least 20 times.

• You will be allowed to choose the size of your bet, as long as it does not exceed half the number of points you have, and as long as the Casino staff member playing with you agrees to it. Normally Casino staff won't agree to very large bets.

• At the end of the class we will compute the total numbers of points and pay out the winnings in special camp currency (gribnas).

So, if you compute correctly, you should have an advantage (since none of the games will be fair), but if you make an error (i.e. decide if it is better to be A or B incorrectly), we will choose the game which you computed incorrectly, and you will likely lose.

1.4. **Games-1.**

Game 1a. Roll 4 dice. If at least two are the same, A gets 1 point. Otherwise (if all are different), B gets 2 points.

Game 1b. Roll 3 dice. If there is a 6, A gets 1 point. Otherwise, B gets 1 point.

Game 1c. Roll 2 dice. If they are equal or 2 apart (e.g., 3 and 5), A gets 1 point. Otherwise, B gets 1 point.

Game 1d. Roll 3 dice. If exactly one number is bigger than 4, A wins 1 point. Otherwise B wins 1 point.

2. Lesson 2

2.1. **Combinatorics.** Here are the main combinatorics (counting) rules. In fact, we have applied them already in analyzing casino games last time.

- **Product rule**: if there are m ways to choose the first item, and independently n ways to choose the second, then there are $m \cdot n$ ways to choose the pair. More generally, if there are k items being chosen independently, and the j-th item can be chosen in n_j ways then the total number of choices is $n_1....n_k$. In particular, we have the **power rule**: If we need to choose k items independently, each of which can be selected from the same list of n, and **order matters, repetitions are allowed**, then there are n^k ways to do this.

  Example: If we roll a die 10 times, there are 6^{10} possible combinations.
- **Arrangements**: If we need to choose k items, each of which can be selected from the same list of n, and **order matters, repetitions are not allowed**, then there are

$$A_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

  ways to do this. In other words, A_n^k is the number of **arrangements** of k **distinguishable** items into n slots; e.g., it is the number of ways to seat k people into n chairs.

  Example: There are $A_{20}^3 = 20 \cdot 19 \cdot 18$ ways to select the first, second, and third place winner from 20 competitors. There are $A_6^3 = 6 \cdot 5 \cdot 4$ ways to roll a die three times in such a way that all three outcomes are different. There are A_{10}^5 ways to put 5 pigeons into 10 cages (at most one per cage).
- **Permutations (orderings):** Applying the previous rule when $n = k$, we see that there are $k! = 1 \cdot 2 \cdots k$ ways to order k distinct items.

  Example: there are 52! ways to shuffle a deck of cards.
- **Combinations**: If we need to choose k items, each of which can be selected from the same list of n, and **order does not matter, repetitions are not allowed**, then there are

$$C_n^k = \frac{A_n^k}{k!} = \frac{n(n-1)...(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

  ways to do this. In other words, C_n^k, also denoted $\binom{n}{k}$ (pronounced "n choose k") is the number of arrangements of k **indistinguishable** items into n slots (i.e., of marking some k out of n slots as "full"); e.g. it is the number of ways to reserve k out of n chairs by putting white sheets of paper on them. More mathematically speaking, it is the number of selections (choices) of a k-element subset from a set of n elements. Such subsets are called **combinations**.

  Note that without interpreting the right hand side of the last equality as the number of combinations, it is not even obvious that this is an integer! Also, note that $\binom{n}{k} = \binom{n}{n-k}$.

  Examples: There are $\binom{52}{6}$ ways to choose 6 cards out of a deck of 52; if we toss a coin 10 times, there are $\binom{10}{4}$ combinations in which we have 4 heads and 6 tails.

The number of combinations $\binom{n}{k}$ is also called the **binomial coefficient**, since it arises in the expansion of the Newton binomial:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Bulgakov's Korovyev may have thought it complicated, but it is actually quite simple. To prove this formula, it suffices to write a product of n brackets $(x+y)$:

$$(x+y)^n = (x+y)(x+y)...(x+y);$$

then, when you open brackets, to get $x^k y^{n-k}$, you have to choose k brackets out of n where you will pick x, which implies the formula.

Also, binomial coefficients can be arranged in Pascal's triangle:

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
......

in which every entry is the sum of the one immediately above it and one to the left of the first one, i.e.

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

(where we agree that if $k < 0$ or $k > n$, then $\binom{n}{k} = 0$). This identity can be proved directly from the explicit formula, or by marking one of the elements of the n-element set from which we have to choose k elements. Then when we choose k elements, we either do choose the marked one or we don't which gives the desired formula.

Note that the Newton binomial formula with $x = y = 1$ implies that the sum of the n-th row of Pascal's triangle is 2^n. Also, columns of Pascal triangle represent "triangular numbers" in various dimensions. E.g. the third column is the triangular numbers, the fourth is pyramidal numbers, etc.

Binomial coefficients also count the number of unordered partitions of a number into a fixed number of parts. Namely, let us compute the number of unordered partitions of n into k positive parts. To do so, draw n dots in a row. A partition would be obtained if we draw $k-1$ lines in spaces between dots:

$$\bullet\ \ \bullet\ \ \bullet|\bullet\ \ \bullet|\bullet$$

There are $n-1$ positions where we can put these lines, so the number of partitions is $\binom{n-1}{k-1}$.

If we need parts to be nonnegative rather than positive (i.e., allow 0), then we can make them positive by adding 1 to each of them. The sum will then become $n+k$. So the number of partitions of n into k nonnegative parts is $\binom{n+k-1}{k-1}$. These numbers stand on diagonals in the Pascal triangle.

This method also allows us to find the number of k-tuples of numbers (nonnegative or positive) with sum not exceeding n. It suffices to consider the case of nonnegative numbers (the case of positive numbers reduces to it by subtracting 1 from each summand). In this case, such a partition is just a partition of n into $k+1$ nonnegative parts (the remaining

$k+1$-th part is how much the sum falls short of n). So the number is $\binom{n+k}{k}$. Thus for strictly positive summands it is $\binom{n}{k}$.

The number $\binom{n+k}{k}$ also counts ways to choose k items from the same list of $n+1$ if **order does not matter** and **repetitions are allowed**. Indeed, creating a partition into nonnegative parts is nothing but picking spots for k dividing lines, but now the dividing lines can be at the extreme left and extreme right spot (so there are $n+1$ spots) and two lines can be at the same spot:

$$|\bullet \ \bullet \ \bullet||\bullet \ \bullet|\bullet|$$

Example. Suppose you play the following game. You roll three dice, and if the sum is ≤ 7, A wins 5 points. Otherwise B wins 1 point. What is the expected gain of this game?

Using the above rules, we see that the number of combinations with sum ≤ 7 is $\binom{7}{3} = 35$, so the probability is $35/216$. Thus the expected gain is

$$\frac{175}{216} - \frac{181}{216} = -\frac{1}{36}$$

So this game is slightly more profitable for player B.

2.2. **Games-2.**

Game 2a. Roll 4 dice. If there are exactly two odd numbers, A gets 1 point. Otherwise, B gets 1 point.

Game 2b. Roll 4 dice. If exactly one of the numbers is divisible by 3 (i.e., is 3 or 6), A wins 1 point. Otherwise B wins 1 point.

Game 2c. Roll 3 dice. If the sum of the numbers is at most 8, A gets 2 points. Otherwise B gets 1 point.

Game 2d. Roll 2 dice. If the sum is at most 7, A gets 1 point. Otherwise B gets 1 point.

3. Lesson 3

3.1. **Conditional probability.** There is a famous joke about a man who was arrested for taking a bomb onboard of a plane. When the police asked him why he did it, he responded: "For personal safety. The chance of having a bomb on the plane is 1 in a million, so the chance of having two bombs is 1 in a million squared!"

This is of course a joke - the actions of this man are indeed ridiculous. Nevertheless, many people do very similar things in daily life without even noticing. I know a person who believed that in a casino, one should sit down at a table where people have been losing to the casino for a while - then your chances are better (while it is quite clear that the chances are exactly the same; maybe at that particular table they are even worse - what if the dealer at that table is a crook?). Many people don't understand that if a (fair) coin yielded tails 10 times in a row, then the probability of it yielding a head at the 11th throw is just the old good 1/2, no more and no less. There is even a proverb "Two shells don't fall into the same crater". Apart from the fact that when the first shell falls, there is no crater yet (which we will forget), this is essentially correct (as is the statement about two bombs on the plane): indeed, if the probability of one shell falling into a crater is p, then the probability of two shells falling into the same crater is p^2, and since p is small, p^2 is almost zero. What is wrong is the way this proverb is usually interpreted: if one shell already fell into this crater, then I can safely hide in it since the probability of another shell falling into this crater is much less than the probability of it falling somewhere else. Which is completely false, since both probabilities are the same and equal to p.

What goes on here is confusion between the **absolute** probability that two shells will fall into the crater (which is p^2), and the **conditional** probability that a shell will fall into the crater given that another shell has already fallen into it (which is p). The conditional probability is the one relevant in this situation, but we are "hard-wired" to confuse the two probabilities.

If A and B are two events, then the **conditional probability** $P(A|B)$ is the probability of A occuring **given that** B occurs. This conditional probability is computed by the formula

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}.$$

Indeed, let N be the number of all outcomes, $N(B)$ the number of outcomes in which B occurs, and $N(AB)$ is the number of outcomes in which both A and B occur. Then $P(B) = N(B)/N$, $P(A \text{ and } B) = N(AB)/N$, and $P(A|B) = N(AB)/N(B)$, which gives the formula.

Example 3.1. You roll 3 dice. If not all numbers are different, the dice are rolled again. If all numbers are different, and there is a 3, A wins 1 point. Otherwise B wins 1 point. What is the expected gain?

To compute it, we need to compute the conditional probability that there will be a 3 given that all the numbers are different. The number of combinations with all numbers different is, as we know, $6 \cdot 5 \cdot 4 = 120$. But what is the probability that there will be a 3 and all the numbers are different? You can get 3 on the 1st, 2nd, and 3d die (3 options), and there are $5 \cdot 4 = 20$ choices for the other two dice. So the total number of combinations is 60, and the conditional probability is 1/2. So this is a fair game, $E = 0$.

Example 3.2. There are three cards: one is red on both sides, the other blue on both sides, and the third one is blue on one side and red on the other. The dealer draws one of these cards without looking and puts it on the table; it is red. He offers you a bet: if the other side of this card is also red, you pay him 2 dollars; if it is blue, he pays you 3 dollars. Will you accept the bet? What is your expected gain?

Solution: It may seem that it is a good deal, since the probability of seeing red or blue is 1/2. But this is absolute probability, and the conditional probability of red is actually 2/3 (there are 3 possible other sides of the card, and two are red). So the expected gain is $3 \cdot (1/3) - 2 \cdot (2/3) = -1/3$, and the game is profitable to the dealer.

3.2. **Covariance and correlation.** The **covariance** of two events A and B is the number

$$\operatorname{cov}(A, B) = P(A \text{ and } B) - P(A)P(B).$$

If this number is zero, the events are statistically independent, and the conditional probability of either of them is the same as the absolute one; such events are called **uncorrelated**. If $\operatorname{cov}(A, B)$ is positive, one says that the events are **positively correlated** (i.e. if one occurs, the other is more likely to occur; the conditional probability is bigger than the absolute one). Similarly, if $\operatorname{cov}(A, B)$ is negative, the events are **negatively correlated** (if one occurs, the other is less likely to occur; the conditional probability is smaller than the absolute one).

For instance, on a summer day, the event "it is sunny" is positivcely correlated with the event "The high temperature is above 85 degrees". That is, if the high is above 85, it is more likely that it was sunny than cloudy, and vice versa. On the other hand, the event "It has not rained for 2 weeks" is negatively correlated with "Pasha found more than 20 kg of mushrooms".

Positive and negative correlation is used in all areas of human activities for decision making in situations where there are unknown or random factors. E.g. risk factors for heart attack (cholesterol levels) etc. Also genetic diseases (if someone in your family had the disease, are you more likely to get it?)

Example 3.3. In one psychology book it is written that a person having ADD is more likely to have dyslexia than an average person, but a person having dyslexia is not more likely to have ADD than an average person. This can't be the case since $\operatorname{cov}(A, B) = \operatorname{cov}(B, A)$. But there could, of course, have been two studies suggesting two different results.

Example 3.4. Correlation can be used to explain why marriage between close relatives (such as sister and brother) leads to a much larger risk of birth defects in children. Here is a very rough version of this explanation. Birth defects are caused by "bad" genes. The child gets two versions of each type of gene - from father and from mother. The good thing is that bad genes are rare (small probability p) and recessive (you need both of the versions to be bad for it to take effect). So if two parents are not connected in any way, the probability of a birth defect is p^2, which is very small. But if they are brother and sister, the events that they carry a bad gene are strongly positively correlated, so the probability is close to p. Now, if there are about $1/p$ such genes, we get $(1 - p^2)^{1/p}$ (small number) versus $(1 - p)^{1/p}$ (about 0.37, i.e. pretty significant).

3.3. **Monty Hall problem.** Suppose there is a car dealer, let us call him Monty Hall. He runs a lottery selling tickets for \$ 10 each. Out of each 1,000 tickets, just one is a winning ticket that takes the winner to the second stage of the lottery. Let us call him Lucky Guy.

At the second stage, Lucky Guy is given the choice of three doors (and will win what is behind the chosen door). Behind one door is a car costing \$ 15,000; behind the others, there is nothing. The car is placed randomly behind one of the doors (with probability 1/3) beforehand. The rules are as follows: After the Lucky Guy has chosen a door, the door remains closed for the time being. Monty Hall, who knows what is behind the doors, now has to open one of the two remaining doors, and the door he opens must have nothing behind it. If both remaining doors have nothing behind them, he chooses one at random (with probability 1/2). After Monty Hall opens a door with nothing behind it, he will ask you to decide whether you want to stay with your first choice or to switch to the last remaining door. In fact, he will offer you \$ 1,000 for not switching (ahead of the game). Would you accept his offer?

It seems that he is just clearly giving you a gift of \$ 1,000, since switching couldn't possibly matter, if the car was placed behind a door randomly. Yet, it turns out that taking \$ 1,000 puts the Lucky Guy at a disadvantage. Our intuition fails us here since it makes us confuse absolute and conditional probabilities.

Let us say that Lucky Guy chose door 1, and Monty Hall opened door 3. Clearly, the absolute probability that the car is behind door 1 is 1/3, regardless of what Monty Hall does. So someone who decides that he will never switch will win 1/3 of the time. Hence someone who decides to switch will win 2/3 of the time.

On the other hand, the conditional probability that the car is behind door number 1 given that it is not behind door number 3 is equal to 1/2. But this is not relevant, although our false intuition thinks it is.

The expected gain in this game if the player switches is computed as follows. He wins the car with probability $(2/3) \cdot 1/1000$ and loses \$10 otherwise. So the expected gain is approximately 0. But if he takes $1,000$ and does not switch, then he has the car with probability $(1/3) \cdot 1/1000$, the \$ 1,000 with probability $1/1000$, and loses \$ 10 otherwise. So the expected gain is about minus \$ 4!

3.4. **Games-3.**

Game 3a. You roll 3 dice until you get at least one 6. If the sum of the other two numbers is 4, 5 or 6, A wins 1 point. Otherwise B wins 1 point.

Game 3b. You roll 2 dice until you get two different numbers. If their sum is odd, A wins 1 point. Otherwise B wins 1 point.

Game 3c. You draw 7 cards. If two of them have the same denomination, A wins 1 point. Otherwise B wins 3 points.

Game 3d. You draw 2 cards. If they are at most 2 apart in denomination, A wins 3 points. Otherwise B wins 2 points (we agree that $J = 11, Q = 12, K = 13, A = 1$).

4. Lesson 4

4.1. **The doubling technique.** The doubling technique is a "sure win" technique in the coin tossing game. The strategy is that if you lose a round of this game, you double the bet in the next round. In this case you will surely be in a winning position at some point.

Indeed, suppose you start with a bet of \$ 1. If you win the first round, you have won \$ 1. If you lose it, you bet \$ 2 in the second round. If you win, your total win is $-1+2=1$. If you lose again, you bet \$ 4 on the third round, and if you win, you gain $-1-2+4=1$, and so on. Now, after you win, you switch back to the bet of \$ 1 and repeat the procedure.

Things look even better if you triple the bet. Suppose you start with a bet of \$ 1. If you win the first round, you have won \$ 1. If you lose it, you bet \$ 3 in the second round. If you win, your total win is $-1+3=2$. If you lose again, you bet \$ 9 on the third round, and if you win, you gain $-1-3+9=5$, and so on. Now, after you win, you switch back to the bet of \$ 1 and repeat the procedure. Thus, you always win, and the more times you lose in a row at the start, the more you win!

So it looks like you are able to consistently win money in a fair game. Where is the catch?

Well, the catch is that this assumes that you are always able to double (or even triple) the bet, which is based on a tacit assumption that you have an infinite amount of money. In practice, if you have \$ 1,000, and you lose 10 (or even 9) times in a row, you are no longer able to double the bet.

If you have a \$ 1,000,000 on your bank account, then, of course, it is very unlikely that you will run out of money, since for this, if you use the doubling strategy, you need a streak of 20 losses, and the probability of this is about $1/1,000,000$. So even if you play until the first win every hour during daytime for 100 years, this will most probably never happen. But you'll be making \$ 12 per day, i.e., about \$ 4,400 per year, which is 0.44 % of the amount you have in the bank. Any bank would give you a better interest!

The moral of this story is that if you have a fair game (with $E=0$), then no matter what your strategy is, on average you can win nothing!

4.2. **The two boxes paradox.** Suppose you are given two identical boxes with random sums of money, but it is known that one of them contains twice as much money as the other. You are allowed to take the money only from one of them. When you open one of the boxes, you are asked if you'd like to keep it or to switch. It seems that switching makes no sense whatsoever (since your probability to pick the box with the larger amount on the first try is clearly 1/2).

Yet there is a very convincing argument in favor of switching. Namely, assume that the amount of money in the box is X dollars. Then the amount in the other box is either $X/2$ or $2X$, with equal probabilities, so on average $5X/4$, which is more than X.

Where is the mistake?

Actually, one problem is that we are making a tacit assumption that all amounts of money are equally likely ("uniform distribution"). More precisely, let us assume for simplicity that the amount of money can only take values 2^k, where k is an integer (positive or negative). Then our assumption is that all the amounts 2^k are equally likely. But this assumption does not make sense since then the probability of every such outcome is 0, and hence the sum of probabilities is 0 and not 1, which is a contradiction. So the problem is not correctly formulated.

However, this problem can be fixed in a way that the paradox still remains. Imagine that the probability of the smaller amount being 2^k is p_k. Then, if you take a box and it has 2^m dollars, then the other box has 2^{m+1} dollars with probability $p := \frac{p_m}{p_m + p_{m+1}}$, and 2^{m-1} dollars with probability $1-p$, so the expected gain in the case of switching is

$$p \cdot 2^{m+1} + (1-p) \cdot 2^{m-1} = 2^{m-1}(1+3p).$$

To make switching profitable, this needs to be more than 2^m, so $1+3p > 2$, $p > 1/3$. Thus, we get $p_{m+1} > p_m/2$ as the condition for switching. There are many nonnegative sequences satisfying this condition with $\sum_k p_k = 1$. For example, $p_k = 0$ for $k < 0$, and $p_k = \frac{1}{3}\left(\frac{2}{3}\right)^k$ for $k \geq 0$.

So what is the real problem?

Let us try to understand why we are puzzled. In fact, our intition on why this can't be is based on the following argument: if it is always better to switch, then by switching, regardless of the amount we see, we can always increase our average gain in this game, which makes no sense, since blind switching clearly cannot increase our gain. But it happens that a probability distribution which satisfies the switching inequality never has an average, i.e. $\sum_k p_k \cdot 2^k$ is infinite! Indeed, the inequality says that the sequence $p_k \cdot 2^k$ has to increase, and then the sum will be clearly infinite.

So the problem is pretty subtle. We are dealing here with a quantity that can take infinitely many values and does not have an average, which causes the paradox.

4.3. **Benford's law.** Benford's law is based on the following simple observation. Consider powers of 2. Then the first digit is 1 more than 5 times more often than it is 9. In fact, among the first 100 powers of 2, 30 have first digit 1 and only 5 have first digit 9. If you take a very large number of powers of 2, the percentages of the first digits will be as follows:

1: 30.1%
2: 17.6%
3: 12.5%
4: 9.7%
5: 7.9%
6: 6.7%
7: 5.8%
8: 5.1%
9: 4.6%

This seems surprising since it would seem that all digits should be equally likely. But if one thinks a little, one can come up with a heuristic explanation (which can actually be made into a proof) of why 1 should be about 6 times more frequent than 9. Indeed, consider the intervals

$$A_1 = [1], A_2 = [10, 19], A_3 = [100, 199], ...$$

and

$$B_1 = [9], B_2 = [90, 99], B_3 = [900, 999], ...$$

To start with 1, a number should be in one of the intervals $A_1, A_2, ...$ and to start with 9 it should be in $B_1, B_2, ...$ Of course, A_1 has the same size as B_1, A_2 as B_2, and so on. This makes us think that 1 and 9 should be equally likely. But in fact powers of 2 get sparser as they get larger. More precisely, their sparseness is proportional of their size. Since numbers in B_N are on average about 6 times larger than numbers in A_N (staring already from $N = 2$), the sparseness is also about 6 times larger, and it is about 6 times less likely to see 9 than 1.

The same law, for the same reason, applies to powers of any number which is not a (rational) power of 10.

It turns out that the same distribution of first digits arises in many collections of numbers that occur in real life, such as amounts of money in bank documents, numbers of bacteria in a jar, populations of cities, countries, data of physical experiments, etc. This statement is usually called "Benford's law"; it was discovered by physicist F. Benford in 1938. The reason it appears in many contexts is that we have exponentially growing quantities, i.e. ones that multiply by a certain number every unit of time. Populations and amounts of money grow exponentially. Also, this is the only distribution which is independent on the units of measurement, which is essentially the reason it appears in physics data. But there are many kinds of numerical data where Benford's law does not appear, such as a phone book or list of heights of people. Also there are many settings where it appears but reasons for it are not completely understood. In general, if a quantity is restricted to one or two orders of magnitude, Benford's law may fail, but if there are several orders of magnitude, its effects often start to show up.

Benford's law is admissible in US courts as evidence for proof of bank fraud. Also it was used to prove that the results of Iranian elections of 2009 were rigged. So if someone wants to create fake bank documents or election results, one should remember about Benford's law. To create data satisfying Benford's law, one should simply take a uniformly distributed list of numbers x and then exponentiate it, i.e., take, say, 2^x. Luckily, people who know such things usually have better things to do.

4.4. **Games-4.**

Game 4a. Roll 4 dice. If there is a 1 or 2, A gets 1 point. Otherwise B gets 3 points.

Game 4b. Roll 3 dice. If the difference between the smallest and the largest numbers is at most 2, A wins 1 point. Otherwise B wins 1 point.

Game 4c. You roll 4 dice until you have two equals. If there is a 3, A wins 1 point. Otherwise B wins 1 point.

Game 4d. Pick a random two-digit number N using cards or 10-faced dice. If 2^N starts with $1, 2, 3$, A gets 1 point. Otherwsie B gets 1 point.

5. Lesson 5

5.1. **Geometric probabilities.** Suppose you have a region R which is contained in the unit square, and suppose we randomly spit or throw darts at this unit square. Then the probability of hitting R is the area of R. This principle is the basis of a computational method for areas called the **Monte-Carlo method.** Namely, you generate a pair of independent random numbers x, y between 0 and 1 a large number of times, and look in what percentage of cases the point (x, y) belongs to R. This percentage is an approximation to the area of R.

This method can be used, for instance, to compute an approximation of π. Let us, for instance, use a 10-faced die to produce a random 2-digit numbers x, y, and put a 1 if we have $x^2 + y^2 \leq 10^4$ and a 0 otherwise. Then the percentage of ones for a large number of trials is close to $\pi/4$. Indeed, the above condition is the condition that $(x, y) \in R$, where R is a quarter-disk contained in the unit square.

5.2. **Buffon's needle.** Buffon's needle problem is a question first posed in the 18th century by Georges-Louis Leclerc, Comte de Buffon: "Suppose we have a floor made of parallel strips of wood, each the same width, and we drop a needle onto the floor. What is the probability that the needle will lie across a line between two strips?"

Assume that the length of the needle is L and the interval between strips is 1. Assume $L \leq 1$, then it is (almost) certain that the needle will intersect at most one of the lines. So the probability p_L of intersecting the line is the same as the average number of lines intersected by the needle, $E(L)$. Also, if we combine two needles (not necessarily in a straight way), the numbers $E(L)$ add, so $E(L + M) = E(L) + E(M)$. Thus, for any closed curve C made out of wire, the average number of intersections $E(C)$ is given by the formula

$$E(C) = K \cdot \mathrm{length}(C)$$

where K is some constant independent on the curve (since any curve approximately consists of many short straight needles). Now to find K, take C to be the unit circle of diameter 1. Its length is π and it always has two intersections with the lines, so we get $K = 2/\pi$. Thus $E_L = 2L/\pi$.

This can be used to compute π by using the Monte-Carlo method.

Also this implies Barbier's theorem: the length of any curve of constant width 1 is π (indeed, such a curve, when tossed on the floor, will always have exactly two intersections with the lines).

5.3. **Computing π using relatively prime numbers.** It turns out that the probability that two randomly chosen positive integers x, y (in some large interval $[1, N]$) are relatively prime is about $6/\pi^2$ (i.e., tends to this value as $N \to \infty$), which is approximately 0.6. Indeed, for every prime p the probability that x, y are both divisible by p (in the limit $N \to \infty$) is $\frac{1}{p} \cdot \frac{1}{p} = \frac{1}{p^2}$. Thus the probability that one of them is not divisible by p is $1 - \frac{1}{p^2}$. Moreover, for different primes these events are independent, hence the probability that this happens for all primes is $P = \prod_{p \text{ prime}} (1 - \frac{1}{p^2})$. Thus

$$\tfrac{1}{P} = \prod_{p \text{ prime}} \frac{1}{1 - \frac{1}{p^2}} = \prod_{p \text{ prime}} (1 + \tfrac{1}{p^2} + \tfrac{1}{p^4} + \ldots) = \sum_{n=1}^{\infty} \tfrac{1}{n^2},$$

which is known to equal $\pi^2/6$. Thus $P = 6/\pi^2$.

This can also be used to find π using the Monte-Carlo method.

5.4. **Games-5.**

Game 5a. You draw 3 cards from the standard deck. If at least one of them is a King or a Queen, A wins 1 point. Otherwise B wins 1 point.

Game 5b. You draw 3 cards from the standard deck. If there are two of the same suit, A wins 1 point. Otherwise B wins 1 point.

Game 5c. Roll 3 dice. If the sum is at most 9, A gets 1 point. Otherwise B gets 1 point.

Game 5d. You deal triples of cards until they are not of the same color (if the same color, put the cards back and shuffle). If there is no spade, A wins 1 point. Otherwise B wins 1 point.

6. Lesson 6

6.1. **The law of large numbers.**

Problem 1. Would you bet 10 against 1 that you will get more than 30% of heads if you toss a coin
1) 10 times?
2) 20 times?

A direct computation of probabilities gives that in the first case the probability of getting no more than 30% of heads is about 0.17, and in the second case about 0.07. Thus the bet is unsafe for 10 tosses but safe (on average) for 20 tosses. So we may suspect that this bet gets better and better as the number of tosses increases. This turns out to be indeed the case. In fact, if you make a very large number of tosses (say, 400), then this bet essentially becomes a sure bet, in the sense that the probability of it losing is less than 0.001.

A similar phenomenon would occur if the figure 30% is replaced by any other one $< 50\%$. This is a manifestation of the so called **Law of Large Numbers**, which for coin tossing is formulated as follows.

Theorem 6.1. As you keep tossing a coin, the percentage of heads will tend to 50% with probability 1. In particular, for any $M < 50\%$, the probability that the percentage of heads will be at most M out of N tosses approaches zero as N goes to infinity.

The Law of Large Numbers is not only one of the most important theorems in mathematics, but it is also one of the most fundamental laws of nature. In the next subsection, we will explain how to prove it.

Remark 6.2. It is important to note that if there are infinitely many outcomes, then to say that an event occurs with probability 1 is not quite the same thing as to say that it occurs with certainty (although it may be the same for all practical purposes). Similarly, to say that the probability of an event is zero is not quite the same as to say that it is impossible. For example, imagine that you toss a coin infinitely many times (not that you can actually do this, but still) and record a 0 when you get a heads and 1 when you get a tails. Then you get an infinite sequence of numbers. Assuming we don't make a recording error, it is **impossible** (in the most direct sense of the word) that this sequence will contain the number 2; it **certainly** will not. But it is **almost impossible** (although not completely impossible) that this sequence will consist of only zeros (in the sense that the probability of this event is 0). In other words, it is **almost certain** (but not completely certain) that this sequence will contain a 1 (the probability of this event is 1). This is why when a probability of some event is 1, one says that this event will occur **almost surely**.

This, however, generates the following curious paradox. Suppose we toss a coin infinitely many times and write down the corresponding sequence of zeros and ones. The probability to get this particular sequence is zero, so it is almost impossible, and yet we did get it!

6.2. **Random variables.**

A **random variable** is any numerical function of a random outcome of a test.

Example 6.3. 1. The number x you get by rolling a die is a random variable.

2. If you roll several dice, the sum of all numbers and the biggest of all numbers are random variables.

3. Another random variable y may be defined as follows: you toss a coin, and $y = 1$ if you get a tails and -1 if you get a heads.

4. If the test is tossing a coin infinitely many times, then the number N of tosses until you get a heads is a random variable (with can take any nonnegative integer value).

5. The gain in a casino game is a random variable.

Every (discrete[1]) random variable z has a **distribution density** $p_z(a)$, which assigns to every possible value a of z the probability $p_z(a)$ of z taking this value. Thus we have the identity

$$\sum_a p_z(a) = 1. \tag{1}$$

Example 6.4. In Example 6.3:

1. $p_x(a) = 1/6$ for each $a = 1, ..., 6$.

2. $p_y(a) = 1/2$ for $a = 1, -1$.

3. $p_N(a) = 1/2^{a+1}$ for $a = 0, 1, 2, ...$ On the other hand, if the coin is biased (i.e., the probability of a tails is p, which is not necessarily equal to $1/2$), then $p_N(a) = (1-p)p^a$.

The expectation value of a random variable z, denoted $E(z)$ or Ez, is the sum

$$Ez = \sum_a ap(a).$$

In other words, it is the mean (or average) value of a. For instance, the expected gain in a casino game is nothing but the expectation value of the gain.

For example, if z takes values $a_1, ..., a_n$ with equal probability $1/n$, like in the case with a die ("uniform distribution"), then Ez is the arithmetic mean

$$Ez = \frac{a_1 + ... + a_n}{n}.$$

Examples.

1. $Ex = 3.5$.

2. $Ey = 0$.

Problem 2. Show that EN, for a biased coin, equals

$$EN = \sum_{a \geq 0} a(1-p)p^a = \frac{p}{1-p}.$$

Problem 3. Show that for any two random variables z, w one has

$$E(z + w) = Ez + Ew.$$

Problem 4. Show that if z, w are independent random variables (i.e., they are obtained from two independent tests) then

$$E(zw) = Ez \cdot Ew.$$

[1] "Discrete" means that every value ever taken by this variable is taken with positive probability, i.e., there are no values that are almost impossible but not impossible. All the random variables we'll consider will be discrete.

Problem 5. Suppose that you toss a biased coin with probability of a tails equal to p. Let $M(n)$ be the number of tails out of n tosses.

(a) Find $E(M(n))$.

(b) Find $E(M(n)^2)$.

6.3. **Proof of the law of large numbers.** Let $y_1, ..., y_n$ be the variables $y = \pm 1$ associated to n independent coin tosses. Let $Y = y_1 + ... + y_n$. Then it is easy to see that $Y = 2m - n$, where m is the number of tails you get out of the n tosses. So

$$\frac{Y}{2n} = \frac{m}{n} - \frac{1}{2}$$

is what we are interested in – the deviation of the share of tails from its mean value 1/2. More precisely, we are interested in the absolute value deviation, $|Y/2n|$.

So it seems that we should calculate $E|Y/2n|$. But it turns out that it is more convenient to calculate the expectation value of the square of this quantity.

So let us calculate the expectation value EY^2. We have

$$E(Y^2) = \sum_{i=1}^{n} E(y_i^2) + 2\sum_{i<j} E(y_i y_j).$$

Since y_i^2 is always equal to 1, we have $E(y_i^2) = 1$. Also, since the i-th and the j-th tosses are independent, the random variable $y_i y_j$ takes values 1 and -1 with probabilities 1/2, so $E(y_i y_j) = 0$. Thus we find

$$E(Y^2) = n,$$

and hence

$$E((\tfrac{Y}{2n})^2) = \tfrac{1}{4n}. \tag{2}$$

But if z is any random variable with nonnegative values, then it is clear from the definition of the expectation value that

$$Ez \geq aP(z \geq a)$$

for any $a \geq 0$ (where $P(z \geq a)$ is the probability that $z \geq a$). Hence for $a > 0$ we get

$$P(z \geq a) \leq Ez/a.$$

In particular, for $z = (Y/2n)^2$ we get

$$P(z \geq a) \leq \frac{1}{4na}.$$

Let us now take $a = 1/4n\varepsilon$, where ε is an arbitrarily small positive number. We then get

$$P\left(|\tfrac{m}{n} - \tfrac{1}{2}| \geq \sqrt{\tfrac{1}{4n\varepsilon}}\right) \leq \varepsilon.$$

This implies the law of large numbers.

Equation (2) also shows that the deviation of the number of tails m from its mean value $n/2$ for large n is, on average, about $\sigma = \sqrt{n}/2$ in absolute value. This number is called **the standard deviation**. So we see that a typical deviation is of the order of $\sqrt{n}$.

In statistics there is a "rule of three σ", which says that deviations of more than 3σ almost never occur (the probability of this is less than 1/1000). Deviations of more than 2σ have the likelihood of $< 5\%$. For instance, if you toss a coin a 100 times, the number of tails is practically guaranteed to be between 35 and 65, and there is a $> 95\%$ chance that it'll be between 40 and 60.

Problem 6. Let z be a random variable (taking finitely many values) which depends on the result of a test T. Suppose the test T is repeated infinitely many times, and denote the corresponding independent random variables z by $z_1, ..., z_n, ...$ Prove the following general version of the law of large numbers: the mean value $\frac{z_1+...+z_n}{n}$ almost surely approaches Ez.

Remark 6.5. This is what makes experimental science possible. Every experiment designed to measure some quantity z has an error, but since the different experiments are independent, the average of the results of many identical experiments gives a much better value of z than an individual experiment.

6.4. **Random walk ("drunk man's walk").** Coin tossing can be interpreted in terms of a random (or drunk man's) walk. Namely, imagine a drunk man who is walking along a street and makes steps of length exactly 1 meter, but because he is drunk, after each step he chooses the direction of the next step randomly (forward or backward with probability $1/2$). Our arguments show that after n steps he will be at a distance of about $\sqrt{n}/2$ from his initial position.

Molecules of gas are in fact quite similar to this drunk man, because they collide with each other and change the direction of their motion. This gives rise to the following application of the theory of the drunk man's walk.

Problem 7. It is known that average velocity of air molecules at room temperature is about 500 m/s, and that each molecule runs about $6.6 \cdot 10^{-8}$ meters between collisions. Based on this, can you estimate how far an air molecule moves over one second? If we open a bottle of perfume, how long will it take for the smell to reach the far end of a 5m long room (assuming that the air is not moving)? a 50m long ballroom?

Hint: you may assume that the air molecule moves along a line and changes direction randomly after each collision.

Solution: Each molecule has $500/6.6 \cdot 10^{-8} = 7.57 \cdot 10^9$ collisions per second. So in time t the molecule will make $7.57 \cdot 10^9 t$ steps, so will on average travel the distance about

$$8.7 \cdot 10^4 \sqrt{t} \cdot 6.6 \cdot 10^{-8} \approx 5.74 \cdot 10^{-3} \sqrt{t} \text{ meters.}$$

So for the perfume to reach the end of a 5m long room, one needs about $7.57 \cdot 10^5$ seconds, i.e., about 8 days. Of course, a considerable percentage of molecules will travel twice as far, and will arrive at the end of the room in just 2 days (as the time depends on the distance quadratically).

Of course, in real life the perfume smell may spread even faster if the air is moving.

6.5. **How much should you bet?** If you are playing a game favorable for you, then to maximize your expected gain, you should, of course, bet the largest amount allowed. But this may not be the best strategy in practical terms.

E.g., suppose the casino gets liberal and lets you bet any amount you have. Then betting all your money at each round is best for maximizing expected gain, for any number of rounds. If you play N rounds starting with 1 dollar, you'll end up with $2^{N+1} - 1$ with probability p^N (where p is your probability of winning a round), and 0 with probability $1 - p^N$. So the expected gain is

$$E_N = (2^{N+1} - 1)p^N.$$

As $p > 1/2$, this number grows exponentially with N.

But is it truly a good strategy? Not really! Indeed, unless the game is a sure win (i.e., if $p < 1$) then eventually you will lose a round and all your money is gone. So if you play N times, your expected gain is large, but the probability of ending up with a nonzero amount of money is only p^N, a pretty small number for large N. So you are basically in the same position as someone participating in a lottery, but since on average the game is favorable for you (unlike a lottery), you can actually do better than that.

So how much should you really bet? For simplicity assume that payout of both players is the same, so $E = 2p - 1$. Suppose each time your bet a percentage x of your capital M. So if you win, you end up with $(1 + x)M$ and if you lose, you end up with $(1 - x)M$. Thus $\log M$ changes by $\log(1 + x)$ with probability p and by $\log(1 - x)$ with probability $1 - p$. So by the Law of Large Numbers, after a large number N of rounds $\log M$ will change by about

$$N(p \log(1 + x) + (1 - p) \log(1 - x))$$

(with error of the order $\sqrt{N}$). Thus the best value of x (for maximizing the likely amount you win in a large number of rounds) is the one maximizing the function

$$f(x) = p \log(1 + x) + (1 - p) \log(1 - x)$$

Since $f(x) \to -\infty$ when $x \to 1$ and $f(0) = 0$ the maximum is attained in $(0, 1)$. Using differential calculus, we have

$$f'(x) = \frac{p}{1 + x} - \frac{1 - p}{1 - x} = \frac{p(1 - x) - (1 - p)(1 + x)}{1 - x^2} = \frac{2p - 1 - x}{1 - x^2}.$$

Thus the best value of x is the one where $f'(x) = 0$, i.e., $x = 2p - 1 = E$. In this case the average increment of $\log M$ in a round is $p \log(2p) + (1 - p) \log 2(1 - p) = \log 2 - \mathcal{E}(p)$, where

$$\mathcal{E}(p) := -p \log p - (1 - p) \log(1 - p).$$

is the **entropy** of the game. Thus after N rounds you are likely to have the amount of money about

$$M_N = M_0(2p^p(1 - p)^{1-p})^N = M_0 2^N e^{-\mathcal{E}(p)N}.$$

Note that $\mathcal{E}(p) \leq \log 2$, achieving this value only for $p = 1/2$ (which one can see again by differential calculus). Thus M_N grows exponentially with N.

Example 6.6. Suppose you play a game with expected gain $1/4$, i.e., $p = 5/8$. Then $2p^p(1 - p)^{1-p} \approx 1.032$. So each time you should bet 25% of your money, and then after 50 rounds you will likely have about

$$M = M_0 \cdot 1.032^{50} \approx 5M_0.$$

Note that if you start with \$ 5, you'll end up with about \$ 25, and this is better than betting \$ 1 each time, which, even if you are allowed to go into the red (which you are not) would produce an average final amount $5 + 50/4 = 17.5$.

On the other hand, if each time you bet 50% (the maximum the casino allows) then the increment will be $2 \cdot 1.5^{5/8} \cdot 0.5^{3/8} \approx 0.994 < 1$, so most likely after a large number of rounds you will end up with less money than you started with.

6.6. **Games-6.**

Game 6a. You deal hands of 5 cards until they are all of different denomination. If there is a King, A wins 1 point. Otherwise B wins 1 point.

Game 6b. You deal triples of cards until there is at least one face card. If there is no card below 5, A gets 1 point. Otherwise, B gets 1 point. (We agree that an ace is not a face card and has denomination 1).

Game 6c. You roll 5 dice. If exactly 2 or 3 are odd, A wins 1 point. Otherwise B wins 1 point.

Game 6d. You roll 6 dice. If exactly 2 or 3 are odd, A wins 1 point. Otherwise B wins 1 point.

7. Solutions of games

Game 1a. The chance of all different is $6 \cdot 5 \cdot 4 \cdot 3/6^4 = 5/18$. So $E = (13 - 2 \cdot 5)/18 = 1/6$.
Answer: $E = 1/6$, A.

Game 1b. The probability of no 6 is $p = 5^3/6^3 = 125/216$, so the chance of having a 6 is $1 - p = 91/216$, thus $E = \frac{91-125}{216} = -\frac{17}{108}$.
Answer: $E = -17/108$, B.

Game 1c. Options with equal: 6; options with two apart: 8, so probability of A winning is $p = (6+8)/36 = 7/18$. Thus $E = (7 - 11)/18 = -2/9$.
Answer: $E = -2/9$, B.

Game 1d. First bigger than 4, others not: $2 \cdot 4 \cdot 4 = 32$ options. Total: 96 options. Thus $E = (96 - 120)/216 = -1/9$.
Answer: $E = -1/9$, B.

Game 2a. Two odd, two even: $6 \cdot 3^2 \cdot 3^2$ options, probability $3/8$. So $E = (3 - 5)/8 = -1/4$.
Answer: $E = -1/4$, B.

Game 2b. Exactly 1 divisible by 3: $4 \cdot 2 \cdot 4^3$ options, probability $32/81$. So $E = (32 - 49)/81 = -17/81$.
Answer: $E = -17/81$, B.

Game 2c. This is the number of unordered partitions of 9 into 4 positive parts. So we get $\binom{8}{3} = 56$ options. So the expected gain is $E = (2 \cdot 56 - 160)/216 = -48/216 = -2/9$.
Answer: $E = -2/9$, B

Game 2d. The number of options for A is $\binom{7}{2} = 21$. So $E = (21 - 15)/36 = 1/6$.
Answer: $E = 1/6$, A.

Game 3a. The number of options with at least one 6 is 91. The options for non-6 digits are 31,13,22,41,14,23,32,51,15,24,42,33, so 12 options, each 3 times, so 36. Thus we have $E = (36 - 55)/91 = -19/91$.
Answer: $E = -19/91$, B.

Game 3b. 30 options with 2 different numbers, variants with odd sum: 65,63,61,54,52,43,41,32,21, and flips, so 18. Thus $E = (18 - 12)/30 = 1/5$.
Answer: $E = 1/5$, A.

Game 3c. All different denominations: $52 \cdot 48 \cdot 44 \cdot 40 \cdot 36 \cdot 32 \cdot 28$. Probability is this divided by $52 \cdot 51 \cdot ... \cdot 46$, which is $p \approx 0.21$. So $E = 1 - p - 3p \approx 0.16$.
Answer: $E \approx 0.16$, A.

Game 3d. Good options for A:
same denomination: $12 \cdot 13 = 156$;
differ by 1: $32 \cdot 12 = 384$;
differ by 2: $32 \cdot 11 = 352$.
So the total is 892. Probability $892/(52 \cdot 51) = 0.336$. $E = 3 \cdot 0.336 - 2 \cdot 0.664$, which is about -0.33.
Answer: $E \approx -0.33$, B.

Game 4a. Good options for B: 4^4 out of 6^4, so $E = 1 - 4 \cdot 4^4/6^4 = 1 - 64/81 = 17/81$.
Answer: $E = 17/81$, A.

Game 4b. Difference 0: 6 options. Difference 1: $5 \cdot 2 \cdot 3 = 30$ options. Difference 2: $4 \cdot 6 + 4 \cdot 6 = 48$ options. Total 84 options. $E = (84 - 132)/216 = -48/216 = -2/9$.

Answer: $E = -2/9$, B.

Game 4c. The number of options with two equals is $6(6^3 - 5 \cdot 4 \cdot 3) = 936$. Those that have no 3 are $5(5^3 - 4 \cdot 3 \cdot 2) = 505$. So $E = (431 - 505)/936 = -37/418$.

Answer: $E = -37/418$, B.

Game 4d. By Benford's law, about 0.6 of instances is $1, 2, 3$, so the expected gain is about $E = 0.6 - 0.4 = 0.2$.

Answer: $E \approx 0.2$, A.

Game 5a. Probability of no King or Queen: $44 \cdot 43 \cdot 42/52 \cdot 51 \cdot 50 \approx 0.6$. So $E \approx 1 - 2 \cdot 0.6 = -0.2$.

Answer: $E \approx -0.2$, B

Game 5b. All three different suits: $p = 52 \cdot 39 \cdot 26/(52 \cdot 51 \cdot 50) \approx 0.4$. $E \approx 1 - 2 \cdot 0.4 = 0.2$.

Answer: $E \approx 0.2$, A.

Game 5c. Number of options for A is $\binom{9}{3} - 3$ (since 711 and permutations are not allowed). This is 81. So get $E = (81 - 135)/216 = -1/4$.

Answer: $E = -1/4$, B.

Game 5d. Combinations of the same color: $2 \cdot 26 \cdot 25 \cdot 24$. So not of the same color: $52(51 \cdot 50 - 25 \cdot 24)$. Not of the same color, no spade: must have a club. So either one club or two clubs. One club: $3 \cdot 13 \cdot 26 \cdot 25$. Two clubs: $3 \cdot 13 \cdot 12 \cdot 26$. So probability $p \approx 0.37$. So $E \approx 0.74 - 1 = -0.26$.

Answer: $E \approx -0.26$, B.

Game 6a. The number of combinations with different denominations: $52 \cdot 48 \cdot 44 \cdot 40 \cdot 36$. Among them, those without a king: $48 \cdot 44 \cdot 40 \cdot 36 \cdot 32$. Probability for B to win: $p = 32/52 = 8/13$. $E = (5 - 8)/13 = -3/13$.

Answer: $E = -3/13$, B.

Game 6b. Number of triples of cards: $52 \cdot 51 \cdot 50$. Without a face card: $40 \cdot 39 \cdot 38$. So with at least one face card: $52 \cdot 51 \cdot 50 - 40 \cdot 39 \cdot 38 = 73320$. Number of combinations with at least one face card and no card below 5: $36 \cdot 35 \cdot 34 - 24 \cdot 23 \cdot 22 = 30696$. Thus the probability for A to win is $p = 30696/73320 \approx 0.42$. So $E \approx 0.42 - 0.58 = -0.16$.

Answer: $E \approx -0.16$, B.

Game 6c. All odd or all even: $2 \cdot 3^5$ options. 1 or 4 odd: $2 \cdot 5 \cdot 3^5$ options. So 2 or 3 odd: $6^5 - 12 \cdot 3^5 = 20 \cdot 3^5$ options. Thus the probability of A winning is $20/32 = 5/8$. So $E = (5 - 3)/8 = 1/4$.

Answer: $E = 1/4$, A.

Game 6d. 2 odd: $\binom{6}{2} \cdot 3^6$. 3 odd: $\binom{6}{3} \cdot 3^6$. Total: $35 \cdot 3^6$ options. Probability of A winning: $35/64$. So $E = (35 - 29)/64 = 3/32$.

Answer: $E = 3/32$, A.