Twenty problems in probability

This section is a selection of famous probability puzzles, job interview questions (most high-tech companies ask their applicants math questions) and math competition problems. Some problems are easy, some are very hard, but each is interesting in some way. Almost all problems I have heard from other people or found elsewhere. I am acknowledging the source, or a partial source, in square brackets, but it is not necessarily the original source.

You should be reminded that all random choices (unless otherwise specified) are such that all possibilities are equally likely, and different choices within the same context are by default independent. Recall also that an *even bet* on the amount x on an event means a correct guess wins you x, while an incorrect guess means loss of the same amount.

1. [P. Winkler] One hundred people line up to board an airplane. Each has a boarding pass with assigned seat. However, the first person to board has lost his boarding pass and takes a random seat. After that, each person takes the assigned seat if it is unoccupied, and one of unoccupied seats at random otherwise. What is the probability that the last person to board gets to sit in his assigned seat?

2. [D. Knuth] Mr. Smith works on the 13th floor of a 15 floor building. The only elevator moves continuously through floors $1, 2, \dots, 15, 14, \dots, 2, 1, 2, \dots$, except that it stops on a floor on which the button has been pressed. Assume that time spent loading and unloading passengers is very small compared to the travelling time.

Mr. Smith complains that at 5pm, when he wants to go home, the elevator almost always goes up when it stops on his floor. What is the explanation?

Now assume that the building has n elevators, which move independently. Compute the proportion of time the first elevator on Mr. Smith's floor moves up.

3. [D. Barsky] *NCAA basketball pool.* There are 64 teams who play single elimination tournament, hence 6 rounds, and you have to predict all the winners in all 63 games. Your score is then computed as follows: 32 points for correctly predicting the final winner, 16 points for each correct finalist, and so on, down to 1 point for every correctly predicted winner for the first round. (The maximum number of points you can get is thus 192.) Knowing nothing about any team, you flip fair coins to decide every one of your 63 bets. Compute the expected number of points.

4. [E. Berlekamp] *Betting on the World Series.* You are a broker; your job is to accommodate your client's wishes without placing any of your personal capital at risk. Your client wishes to place an even \$1,000 bet on the outcome of the World Series, which is a baseball contest decided in favor of whichever of two teams first wins 4 games. That is, the client deposits his \$1,000 with you in advance of the series. At the end of the series he must receive from you either \$2,000 if his team wins, or nothing if his team loses. No market exists for bets on the entire world

series. However, you can place even bets, in any amount, on each game individually. What is your strategy for placing bets on the individual games in order to achieve the cumulative result demanded by your client?

5. From *New York Times*, Science Times, D5, April 10, 2001:

> "Three players enter a room and a red or blue hat is placed on each person's head. The color of each hat is determined by [an independent] coin toss. No communication of any sort is allowed, except for an initial strategy session before the game begins. Once they have had a chance to look at the other hats [but not their own], the players must simultaneously guess the color of their own hats or pass. The puzzle is to find a group strategy that maximizes the probability that at least one person guesses correctly and no-one guesses incorrectly."

The naive strategy would be for the group to agree that one person should guess and the others pass. This would have probability 1/2 of success. Find a strategy with a greater chance for success. (The solution is given in the article.)

For a different problem, allow every one of n people to place an even bet on the color of his hat. The bet can either be on red or on blue and the amount of each bet is arbitrary. The group wins if their combined wins are strictly greater than their losses. Find, with proof, a strategy with maximal winning probability.

6. [L. Snell] Somebody chooses two nonnegative integers X and Y and secretly writes them on two sheets of paper. The distrubution of (X, Y) is unknown to you, but you do know that X and Y are different with probability 1. You choose one of the sheets at random, and observe the number on it. Call this random number W and the other number, still unknown to you, Z. Your task is to guess whether W is bigger than Z or not. You have access to a random number generator, i.e., you can generate independent uniform (on $[0, 1]$) random variables at will, so your strategy could be random.

Exhibit a stategy for which the probability of being correct is $1/2 + \epsilon$, for some $\epsilon > 0$. This ϵ may depend on the distribution of (X, Y), but your strategy of course can not.

7. A person's birthday occurs on a day i with probability p_i, where $i = 1, \ldots, n$. (Of course, $p_1 + \cdots + p_n = 1$.) Assume independent assignment of birthdays among different people. In a room with k people, let $P_k = P_k(p_1, \ldots, p_n)$ be the probability that no two persons share a birthday. Show that this probability is maximized when all birthdays are equally likely: $p_i = 1/n$ for all i.

8. [Putnam Exam] Two real numbers X and Y are chosen at random in the interval $(0, 1)$. Compute the probability that the closest integer to X/Y is even. Express the answer in the form $r + s\pi$, where r and s are rational numbers.

9. [L. Snell] Start with n strings, which of course have $2n$ ends. Then randomly pair the ends and tie together each pair. (Therefore you join each of the n randomly chosen pairs.) Let L be the number of resulting loops. Compute $E(L)$.

10. [Putnam Exam] Assume C and D are chosen at random from $\{1, \ldots, n\}$. Let p_n be the probability that $C + D$ is a perfect square. Compute $\lim_{n\to\infty}(\sqrt{n} \cdot p_n)$. Express the result in the form $(a\sqrt{b} + c)/d$, where a, b, c, d are integers.

11. [D. Griffeath] Let $\alpha \in [0, 1]$ be an arbitrary number, rational or irrational. The only randomizing device is an unfair coin, with probability $p \in (0, 1)$ of heads. Design a game between Alice and Bob so that Alice's winning probability is exactly α. The game of course has to end in a finite number of tosses with probability 1.

12. [Putnam Exam] Let $(X_1, \ldots, X_n)$ be a random vector from the set $\{(x_1, \ldots, x_n) : 0 < x_1 < \cdots < x_n < 1\}$. Also let f be a continuous function on $[0, 1]$. Set $X_0 = 0$. Let R be the Riemann sum

$$R = \sum_{i=0}^{n-1} f(X_{i+1})(X_{i+1} - X_i).$$

Show that $ER = \int_0^1 f(t)P(t)\,dt$, where $P(t)$ is a polynomial of degree n, independent of f, with $0 \le P(t) \le 1$ for $t \in [0, 1]$.

13. [R. Stanley] You have $n > 1$ numbers $0, 1, \ldots, n-1$ arranged on a circle. A random walker starts at 0 and at each step moves at random to one of its two nearest neighbors. For each i, compute the probability p_i that, when the walker is at i for the first time, all other points have been previously visited, i.e., that i is the last new point. For example, $p_0 = 0$.

14. [R. Stanley] Choose $X_1, \ldots, X_n$ from $[0, 1]$. Let p_n be the probability that $X_i + X_{i+1} \le 1$ for all $i = 1, \ldots, n-1$. Prove that $\lim_{n\to\infty} p_n^{1/n}$ exists and compute it.

15. [Putnam Exam] Each of the triples (r_i, s_i, t_i), $i = 1, \ldots, n$, is a randomly chosen permutation of (1, 2, 3). Compute the three sums $\sum_{i=1}^n r_i$, $\sum_{i=1}^n s_i$, and $\sum_{i=1}^n t_i$, and label them (not necessarily in order) A, B, C so that $A \le B \le C$. Let a_n be the probability that $A = B = C$ and let b_n be the probability that $B = A + 1$ and $C = B + 1$. Show that for every $n \ge 1$, either $4a_n \le b_n$ or $4a_{n+1} \le b_{n+1}$.

16. [Putnam Exam] Four points are chosen on the unit sphere. What is the probability that the origin lies inside the tetrahedron determined by the four points?

17. [Putnam Exam] An $m \times n$ checkerboard is colored randomly: each square is randomly painted white or black. We say that two squares, p and q, are in the same *connected monochromatic*

component (or *component*, in short) if there is a sequence of squares, all of the same color, starting at p and ending at q, in which successive squares in the sequence share a common side. Show that the expected number of components is greater than $mn/8$ and smaller than $(m+2)(n+2)/6$.

18. Choose, at random, three points on the circle $x^2 + y^2 = 1$. Interpret them as cuts that divide the circle into three arcs. Compute the expected length of the arc that contains the point $(1, 0)$.

Remark. Here is a "solution." Let L_1, L_2, L_3 be the lengths of the three arcs. Then $L_1 + L_2 + L_2 = 2\pi$ and by symmetry $E(L_1) = E(L_2) = E(L_3)$, so the answer is $E(L_1) = 2\pi/3$. Explain why this is wrong.

19. You are in possession of n pairs of socks (hence a total of $2n$ socks) ranging in shades of grey, labeled from 1 (white) to n (black). Take the socks blindly from a drawer and pair them at random. What is the probability that they are paired so that the colors of any pair differ by at most 1? You have to give an explicit formula, which may include factorials.

20. [P. Winkler] Choose two random numbers from $[0, 1]$ and let them be the endpoints of a random interval. Repeat this n times. What is the porbability that there is an interval which intersects all others.

Solutions

1. Look at the situation when the k'th passenger enters. Neither of the previous passengers showed any preference for the k'th seat vs. the seat of the first passenger. This in particular is true when $k = n$. But the n'th passenger can only occupy his seat or the first passenger's seat. Therefore the probability is $1/2$.

2. In the one-elevator case, we can reasonably assume that the elevator is equally likely to be at any point between floor 1 and floor 15 at any point in time. We can also assume that the probability that the elevator is exactly on the 13th floor when Smith arrives is negligible. This gives the probability $2/14 = 1/7 \approx 0.1429$ that it is above floor 13 (which is when it will go down when it goes by this floor) when Smith wants to go home.

Let's have n elevators now. Call the *unbiased portion* the part of the elevators route up from floor 9 to the top and then down to floor 13. Any elevator at a random spot of the unbiased portion is equally likely to go up or down when it goes by the 13th floor. Moreover, if there is at least one elevator in the unbiased portion, all elevators out of it do not matter. However, if no elevator is in the unbiased portion, then the first one to reach the 13th floor goes up. Therefore the probability that the first elevator to stop at 13th floor goes down equals $\frac{1}{2}(1 - (10/14)^n)$. (For $n = 2$ it equals approximately 0.2449.)

3. If you have n round and 2^n teams, the answer is $\frac{1}{2}(2^n - 1)$, so 31.5 when $n = 6$.

This is another evidence of how useful linearity of expectation is. Fix a game g and let I_g be the indicator of the event that you collect points on this game, i.e., correctly predict its winner. If $s = s(g)$ is this game's round, then your winnings on this game are $2^{s-1} \cdot I_g$. However, EI_g is the probability that you have correctly predicted the winner of this game in this, and all previous rounds, that is 2^{-s}. So your expected winnings on this game are $2^{s-1} \cdot 2^{-s} = \frac{1}{2}$. This is independent of g, so your answer is one half of the total number of games.

4. Let's assume that the money unit is \$1,000, call your team A and your client's team B. Call each sequence of games *terminal* if the series may end with it. To each terminal sequence at which A wins, say AAABA, attach value 2, and to each terminal sequence at which B wins, say BBAAABB, attach 0. These are of course the payoffs we need to make. Each non-terminal sequence, say AABA, will have a value which is the average of the two sequences to which it may be extended by the next game, AABAA and AABAB in this case. This recursively defines the values of all possible sequences. It is important to note that the value of the empty sequence (before games start) is 1, as the average on the sequences of length 7, and then at each shorter level, is 1. Now simply bet, on A, your value minus the lower value of your two successors at each sequence.

Note that you can extend, with 2's or respectively 0's, to length 7 all sequences in which A or respectively B wins. The value is the amount you have provided you use the above betting strategy. Also note that you do not need to split a penny because the values of sequences of length 1 have at most 2^5 in the denominator (and we know that the value is an integer for the sequence of length 0).

5. For the first question, here is the strategy. If you see same colors, guess the color you do not see. If you see different colors, pass. The probability of a win is then 3/4. (Maximizing the probability in this context is a difficult problem for a large number of people.)

For the second question, call the two colors + and $-$, label people $1, \dots, n$ and put them in this order on a line, from left to right. Every possible strategy can be described as n functions F_i, $i = 1, \dots, n$, $F_i : \{+, -\}^{n-1} \to \mathbb{R}$, which could be interpreted as i's bet on + provided i sees the configuration of signs in given order. (The negative values of F are of course bets on $-$.) For example, the payoff at configuration $+--$ (for $n = 3$) then is $F_1(--) - F_2(+-) - F_3(+-)$. There are 2^n configuration, hence these many payoffs. We need to make as many of these positive as possible. On the other hand, to specify a strategy we need to specify $n \cdot 2^{n-1}$ numbers. This makes it look like all payoffs can be made positive, but this is not so. Denote by x a generic n–configuration and x^i the $(n-1)$–configuration obtained by removing the i'th coordinate from x. Then the expected payoff is

$$\frac{1}{2^n} \sum_x F_i(x^i) = 0,$$

as every $F_i(y)$ appears in the sum twice, with different signs. As a consequence, at most $2^n - 1$ payoffs can be made positive. To show that this is indeed possible, we will give an explicit strategy. The i'th person looks only to his left. If he sees no + hats, he bets 2^{i-1} (on his hat being a +). Otherwise, he places no bet, i.e., bets 0. Under this strategy, the first person always places a bet of 1, and if there is a single +, the group wins. Indeed, if the leftmost + is at position $i+1$, $0 \le i \le n-1$, the group wins

$$-1-2-\cdots-2^{i-1}+2^i=1.$$

Needless to say, this is balanced by the huge negative balance of bets $-(2^n-1)$ in the case there are only $-$ hats. The probability of success therefore is $1-1/2^n$.

6. Let G be an exponential random variable with expectation 1 (or any other random variable with density which is positive everywhere on nonnegative x-axis), which you can obtain as $-\log U$, where U is uniform on $[0,1]$. The strategy is to guess that $W>Z$ if $W>G$ and that $W<Z$ if $W<G$. In this case

$$\begin{aligned}
&P(\text{correct guess})\\
&= P(W>Z, W>G) + P(W<Z, W<G)\\
&= (1/2)[P(X>Y, X>G) + P(Y>X, Y>G) + P(X<Y, X<G) + P(Y<X, Y<G)]\\
&= (1/2)[P(X>Y) + P(X>Y, Y<G<X) + P(X<Y) + P(X<Y, X<G<Y)]\\
&= 1/2 + (1/2)P(G \text{ between } X \text{ and } Y) > 1/2.
\end{aligned}$$

7. We have

$$P_k = k! \sum_{1\le i_1<\cdots<i_k\le k} p_{i_1}\dots p_{i_k}.$$

(The sum is known as the k'th *symmetric polynomial* in $p_1,\dots,p_n$.) This is obtained selecting some k different birthdays and then deciding which of them belongs to which person.

For $i<j$, write

$$P_n = Ap_ip_j + B(p_i+p_j) + C,$$

where A, B , and C do not depend on either p_i or p_j. Let $p_i'=p_j'=(p_i+p_j)/2$. Then (as it easy to verify by squaring out), $p_i'p_j' \ge p_ip_j$, with strict inequality unless $p_i=p_j$. Of course, $p_i'+p_j'=p_i+p_j$. Now if you replace p_i and p_j by p_i' and p_j', then $p_i'p_j' \ge p_ip_j$.

Now assume that P_n is maximized while not all p_i are equal, say $p_i \ne p_j$. We can then also assume that P_n is nonzero (when it is zero it is obviously not maximal) and therefore that some n of p's are nonzero. Then $A \ne 0$ (even though p_i or p_j might be zero). Now replace p_i and p_j by p_i' and p_j'; The sum of p's is still 1, while P_n has strictly increased. This contradiction shows that $p_i=p_j$ for all i and j.

7. Let N be the closest integer to X/Y. Then $N=0$ if $X<2Y$. Moreover, for $n>0$, $N=2n$ if $2X/(4n+1)<Y<2X/(4n-1)$. Hence the required probability is $1/4+(1/3-1/5)+(1/7-1/9)+\cdots=5/4-\pi/4$, as $\pi/4=1-1/3+1/5-1/7+\dots$

8. Let $e_n=EL$. Then $e_1=1$. Furthermore, solve the problem with $n-1$ strings. When you add another string, take one of its ends. You either tie this end to the new string's other end

(with probability $1/(2n-1)$) which isolates it from the other strings, or you tie it to some other end (with probability $(2n-2)/(2n-1)$), in which case you reduce the number of strings by 1. Therefore

$$e_n = \frac{1}{2n-1}(e_{n-1}+1) + \frac{2n-2}{2n-1}\, e_{n-1} = \frac{1}{2n-1} + e_{n-1}.$$

So

$$e_n = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-1}.$$

9. The number of pairs (c, d) with sum m^2 is m^2-1 for $m^2 \le n$, and $2n-m^2+1$ for $n < m^2 \le 2n$. Therefore

$$p_n = \frac{1}{n^2}\left(\sum_{1}^{\lfloor\sqrt{n}\rfloor} (m^2-1) + \sum_{\lfloor\sqrt{n}\rfloor+1}^{\lfloor\sqrt{2n}\rfloor} (2n - m^2 + 1) \right).$$

Now we use $\sum_1^k m^2 = k^3/3 + \mathcal{O}(k^2)$. Also removing the integer parts gives an error of $\mathcal{O}(n)$ in the sums. So we get:

$$p_n = \frac{1}{3} n^{-1/2} + 2(\sqrt{2}-1)n^{-1/2} - \frac{1}{3}(2\sqrt{2}\, n^{-1/2} - n^{-1/2}) + \mathcal{O}(n^{-1}).$$

Therefore

$$\lim_{n\to\infty} p_n \sqrt{n} = \frac{4(\sqrt{2}-1)}{3}.$$

10. Assume first that $\alpha = 0.5$. Toss the coin twice: if the two tosses are the same repeat, otherwise stop. Whne you finally stop, the Heads-Tails and Tails-Heads have the same probability. We can thus assume, from now on, that $p = 1/2$. Toss the fair coin until it comes up heads, and let N be the number of tosses required.

Next, write $\alpha = 0.\alpha_1\alpha_2, \ldots$ in the binary form. Declare that Alice wins if $\alpha_N = 1$. This happens with probability

$$P(N=1)\cdot\alpha_1 + P(N=2)\cdot\alpha_2 + P(N=3)\cdot\alpha_3 + \cdots = \frac{1}{2}\cdot\alpha_1 + \frac{1}{2^2}\cdot\alpha_2 + \frac{1}{2^3}\cdot\alpha_3 + \cdots = \alpha,$$

by definition of binary expansion.

11. The set of outcomes has volume $1/n!$. This is obtained by iterated integration. Then we

need to compute

$$
\begin{aligned}
&E[(X_{i+1} - X_i) f(X_i)] \\
&= \int_0^1 dx_{i+1} \int_0^{x_{i+1}} dx_i \, (x_{i+1} - x_i) f(x_{i+1}) \\
&\qquad \cdot \left(\int_0^{x_i} dx_{i-1} \ldots \int_0^{x_2} dx_1 \right) \cdot \left(\int_{x_{i+1}}^1 dx_{i+2} \ldots \int_{x_{n-1}}^1 dx_n \right) \\
&= \int_0^1 dx_{i+1} \int_0^{x_{i+1}} dx_i \, (x_{i+1} - x_i) f(x_{i+1}) \cdot \frac{x_i^{i-1}}{(i-1)!} \cdot \frac{(1 - x_{i+1})^{n-i-1}}{(n-i-1)!} \\
&= \int_0^1 f(x_{i+1}) \cdot \frac{x_{i+1}^{i+1}}{(i+1)!} \cdot \frac{(1 - x_{i+1})^{n-i-1}}{(n-i-1)!} \, dx_{i+1} = \int_0^1 f(t) P_i(t) \, dt,
\end{aligned}
$$

where P_i is a polynomial of degree n. It now follows that

$$
ER = \int_0^1 f(t) P(t) \, dt,
$$

where

$$
\begin{aligned}
P(t) = n! \sum_{i=0}^{n} P_i(t) &= nt(1-t)^{n-1} + \binom{n}{2} t^2 (1-t)^{n-2} + \cdots + t^n \\
&= (t + 1 - t)^n - (1-t)^n = 1 - (1-t)^n.
\end{aligned}
$$

12. For this to happen, the random walker must visit both neighbors of i before he visits i. Therefore, when he is first adjacent to i, he must hit the other neighbor before hitting i. This is a necessary and sufficient condition. Therefore, p_i are the same for all $i \neq 0$. It follows that $p_i = 1/(n-1)$ for all $i \neq 0$.

13. The question gives a hint — n'th root test for convergence of power series. So we consider

$$
f(x) = \sum_{n=0}^{\infty} p_n x^n.
$$

Hopefully we can compute f explicitly. Now,

$$
p_n = \int_0^1 dx_1 \int_0^{1-x_1} dx_2 \ldots \int_0^{1-x_{n-1}} dx_n.
$$

The form of iterated integrals suggests a slight generalization $p_n(y)$, which differs from p_n in that the upper bound in the first integral is $1 - y$ instead of 1. Then $p_0(y) = 1$, and

$$
p_n(y) = \int_0^{1-y} p_{n-1}(t) \, dt \quad .
$$

Moreover, $p_n = p_n(0)$. This way, we can compute $p_0 = 1$, $p_1 = 1$, $p_2 = 1/2$, $p_3 = 1/6$. Furthermore, let

$$f(x,y) = \sum_{n=0}^{\infty} p_n(y) x^n.$$

Then,

$$\begin{aligned} f(x,y) &= 1 + \sum_{n=1}^{\infty} x^n \int_0^{1-y} p_{n-1}(t)\,dt \\ &= 1 + x \int_0^{1-y} \sum_{n=1}^{\infty} x^{n-1} p_{n-1}(t)\,dt \\ &= 1 + x \int_0^{1-y} f(x,t)\,dt. \end{aligned}$$

Then $f(x,1) = 1$, $f_y(x,y) = -xf(x,1-y)$, $f_y(x,0) = -x$, and

$$f_{yy} = x f_y(x, 1-y) = -x^2 f(x,y).$$

Therefore

$$f(x,y) = A(x)\sin(xy) + B(x)\cos(xy).$$

From $f_y(x,0) = -x$ we get $A(x) = -1$. Then, from $f(x,1) = 1$, we get $B(x) = (1+\sin x)/\cos x$. This gives

$$f(x,y) = \frac{\sin(x(1-y)) + \cos(xy)}{\cos x}$$

and finally

$$f(x) = f(x,0) = \frac{1+\sin x}{\cos x}.$$

This is finite and an analytic function of complex $x \in \mathbb{C}$ for $|x| < \pi/2$ and therefore the lim sup in question is $2/\pi$.

In fact, the limit exists, as one can get from asymptotics for Taylor coefficients of tan and sec. An even better method is the standard trick of separating the nearest singularity,

$$f(x) = \frac{-2}{x - \pi/2} + g(x).$$

(Note that f is analytic at $-\pi/2$.) Here $g(x)$ is analytic for $|x| < 3\pi/2$ and the rational function has explicitly, and easily, computable Taylor coefficients. So

$$p_n = q_n + r_n,$$

where

$$q_n = \frac{4}{\pi}\left(\frac{2}{\pi}\right)^n, \quad \limsup |r_n|^{1/n} < \frac{2}{\pi}.$$

To finish the proof, write

$$p_n^{1/n} \geq q_n^{1/n}\left(1 - \frac{|r_n|}{q_n}\right)^{1/n} = q_n^{1/n}(1 + o(1)),$$

as $|r_n|/q_n$ goes to 0 exponentially fast.

14. Label $B' = B - A$, $B' = C - A$. Then $a_n = P((B', C') = (0, 0))$, $b_n = P((B', C') = (1, 1))$, and let also $c_n = P((B', C') = (2, 2))$, $d_n = P((B', C') = (0, 3))$.

Start with n permutations, and add another one. One can check that the $(0, 0)$ state can be only obtained from $(1, 1)$, while the $(1, 1)$ state can be only obtained from $(0, 0)$, $(2, 2)$, and $(0, 3)$. In fact,

$$\begin{aligned} a_{n+1} &= \frac{1}{6} b_n, \\ b_{n+1} &= a_n + \frac{1}{3} b_n + \frac{1}{6} c_n + \frac{1}{3} d_n, \\ c_{n+1} &\geq \frac{1}{6} b_n, \\ d_{n+1} &\geq \frac{1}{3} b_n. \end{aligned}$$

Note that $c_n \geq b_{n-1}/6 = a_n$ and $d_n \geq b_{n-1}/3 = 2a_n$ so that $b_{n+1} \geq 11a_n/6 + b_n/3$. Assume also that $b_n < 4a_n$. Then

$$b_{n+1} > \frac{11}{24} b_n + \frac{1}{3} b_n = \frac{19}{24} b_n = \frac{19}{4} a_{n+1},$$

which proves the inequality with room to spare. (In fact, 4 can be replaced by $1 + 2\sqrt{3}$.)

15. Let X_1, X_2, X_3, X_4 be the positions of the 4 points. Note that you can almost surely uniquely write $\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 = 0$, with $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$, and the question becomes to compute $P(\alpha_1, \alpha_2, \alpha_3, \alpha_4 > 0)$. The trick is to express, equivalently $-X_1 = \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$, and compute $P(\beta_2, \beta_3, \beta_4 > 0)$. Define the eight events $S_1 = \{\beta_2, \beta_3, \beta_4 > 0\}$, $S_1 = \{\beta_2 < 0, \beta_3, \beta_4 > 0\}$, and so all, for all combination of signs of β_i. Then S_i are pairwise disjoint. (E.g., if $S_1 \cap S_2 \neq 0$, then $\beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 = -\beta_2' X_2 + \beta_3' X_3 + \beta_4' X_4$, for positive β_i, β_i'. But this is impossible, by linear independence.) Also $P(\cup_i S_i) = 1$ and all $P(S_i)$ are the same (since X_i equals, in distribution, to $-X_i$). So $P(S_1) = 1/8$.

16. Let's start with the upper bound. One should start with the observation that *many components are very small.* So we see if counting the smallest ones suffices. For example, the probability that a site x is itself a component of size 1 is at least $1/16$. Therefore, the expected number of components consisting of single sites is at least $mn/16$. Unfortunately, it's cumbersome to improve $1/16$ to $1/8$ by counting other small components.

A very slick solution instead proceeds by putting down edges between neighboring squares with the same color. For any pair of neighboring squares, the probability that there is an edge between them is $1/2$. There are $(m - 1)n + (n - 1)m$ pairs of neighboring squares and let A be the random number of edges. Note that N, the number of components, is at least $mn - A$ (every new edge can decrease the number of components by at most one), but this is clearly not enough, as the $E(A)$ is about mn for large m and n. So we need another insight, which is that any edge which creates a new cycle does not increase the number of components.

This will happen for at least one edge in every monochromatic 2×2 square. Why? Imagine ordering such 2×2 squares so that every next square in the ordering is not included in the union of the previous ones. (For example, we can order their upper left corners alphabetically.) Then every new 2×2 square has at least one new site, hence at least two new edges, one of which will create a new cycle.

Let B be the number of monochromatic squares. Then $E(B) = (m-1)(n-1)/8$ and $N \geq mn - A + B$, so that $E(N) \geq m + n + (m-1)(n-1)/8 > mn/8$.

Now for the upper bound. Attach to every component the unique member that is the leftmost of its top sites. Call such points *component sites*. Then we need to estimate the probability that a point $x = (a, b)$ is a component site. Here we imagine the rectangle as a matrix and a, b as indices, so that $(1,1)$ is the top left corner. Clearly when $a = b = 1$ the probability is 1. Still easy is the case when $a = 1$, and $b \geq 2$, when the probability is at most $1/2$, as the necessary requirement is that the left neighbor be of opposite color.

Otherwise, a necessary condition for x being a component site is as follows. Let $i \geq 0$ be the number of contiguous squares of the same color as x to the right of x. Then the left neighbor of x (if any), the top neighbors of x and the said i sites, and the right neighbor of the i'th site (if any), all have to be of the opposite color. For $a, b \geq 2$, this gives

$$P(x \text{ is a component site}) \leq \frac{1}{2}\left(\sum_{i=0}^{n-b-1} \frac{1}{2^{2(i+1)}} + \frac{1}{2^{2(n-b)+1}}\right) = \frac{1}{6} + \frac{1}{12}\cdot\frac{1}{2^{2(n-b)}},$$

while if $b = 1$, $a \geq 2$, the upper bound is twice as large. This implies

$$\begin{aligned} E(N) &\leq 1 + \sum_{b=2}^{n}\frac{1}{2} + \sum_{a=2}^{m}\left(\frac{1}{3} + \frac{1}{6}\cdot\frac{1}{2^{2(n-1)}}\right) + \sum_{a=2}^{m}\sum_{b=2}^{n}\left(\frac{1}{6} + \frac{1}{12}\cdot\frac{1}{2^{2(n-b)}}\right) \\ &= \frac{1}{6}(m-1)(n-1) + \frac{1}{2}(n-1) + \frac{4}{9}(m-1) + 1 + \frac{1}{18}(m-1)2^{-2(n-1)} \end{aligned}$$

Then

$$6E(N) \leq (m+2)(n+2) - 3 - \frac{1}{3}(m-1)\left(1 - 2^{-2(n-1)}\right),$$

which ends the proof.

13. The arc containing $(1,0)$ is likely to be larger than the others. Imagine choosing a point at random instead of $(1,0)$. This point falls into any part proportionally to its length. So the answer via the suggested route actually is $3E(L_1^2)/(2\pi)$. But is is better to proceed directly than to compute $E(L_1^2)$.

The arc containing $(1,0)$ consists of two pieces, the clockwise one and the counterclockwise one. Their lengths are equal in expectation. So let L be the length of counterclockwise piece. Then

$$P(L \geq x) = P(\text{no pt. with angle in } [0,x)) = (1 - x/(2\pi))^3$$

for $x \in [0, 2\pi]$, and so

$$E(L) = \int_0^{2\pi} (1 - x/(2\pi))^3\,dx = \frac{\pi}{2}$$

and the answer is $2E(L) = \pi$.

19. The number of possible pairings is $1 \cdot 3 \cdot \ldots (2n-1)$. Let a_n be the number of pairings that satisfy the condition. Write the recursive equation:

$$a_n = a_{n-1} + 2a_{n-2}.$$

You get this by considering the two socks $2n$, call them $(2n)_1$ and $(2n)_2$. Decide upon the pair for $(2n)_1$ first. If it is $(2n)_2$, (with probability $1/(2n-1)$), then you are left with the same problem with $n-1$ pairs. If not, it must be paired with one of the socks $(2n-1)$ (two possibilities) and then $(2n)_2$ needs to be paired with the other $(2n-1)$ sock.

Thus, if $a_n = 1 \cdot 3 \cdot \ldots (2n-1)p_n$, then $a_n = a_{n-1} + 2a_{n-2}$, $a_1 = 1$, $a_2 = 3$. This is a linear recursion, with solution

$$a_n = \frac{2^{n+1} + (-1)^n}{3}$$

and therefore

$$p_n = \frac{(2^{n+1} + (-1)^n)2^n n!}{3(2n)!}.$$

20. The answer is $\frac{2}{3}$, independently of n. Here is the proof, from the original paper by J. Justicz, E. R. Scheinerman, and P. M. Winkler, American Mathematical Monthly 97 (1990), 881–889.

The question will first be reformulated: take $2n$ points $1, 2, \ldots, 2n$ are pair them at random, so that all $\frac{(2n)!}{2^n\, n!} = (2n-1) \cdot (2n-3) \cdot \ldots \cdot 3 \cdot 1$ pairings are equally likely. The event that an interval intersects all others has the same probability as in the original formulation.

We have a lot of freedom to organize the random pairing, and this freedom will be exploited in the proof. Namely, we can choose pairs one by one — then, if we know the pairings of $2j$ points, we may choose the $(2j+1)$'st point, which we call A_{j+1}, in any way we wish from among $2n - 2j$ remaining points, and then pair it with the random one of the remaining $2n - (2j+1)$ points. It is advantageous to choose A_j close to the center, and so as to make the paired points as balanced as possible among the two sides: the *left* side $\{1, \ldots, n\}$, and the *right* side $\{n+1, \ldots, 2n\}$.

Here is how we do it. Start with $A_1 = n$, and let B_1 be its mate. Assuming we know $A_1, \ldots, A_j$ and there respective mates $B_1, \ldots, B_j$, we choose A_{j+1} as follows: if B_j is on the left side let A_{j+1} be the leftmost unpaired point on the right side, and if B_j is on the right side let A_j be the rightmost unpaired point on the left side. Note that $A_j < B_j$ exactly when B_j is on the right side. The second step of this recursive procedure is to choose the random mate B_{j+1} of A_{j+1} among the other $2n - (2j+1)$ unpaired points.

We call the j'th pair *AB-type* if $A_j < B_j$, and *BA-type* otherwise. Now the left and the right side will have equal number of paired points while the pairs are AB-types. After the first BA-type, the left side has two more paired points. This imbalance by two points is maintained while BA-types continue, until the next AB-type, when balance is restored. Observe also that there are never any unpaired points between the leftmost and the rightmost point among $A_1, \ldots, A_j$.

After the $(n-2)$'nd pair is chosen, there are four unpaired points $a < b < c < d$. We have two possibilities

1. Balanced case: a and b are on the left side, c and d on the right side.

2. Unbalanced case: a on the left side, b, c and d on the right side.

There is a single choice, among three equally likely ones, still left to be made, that is, to choose the mate of a.

Now all A_j, $j \leq n-2$ must be between a and c, as a and c are unpaired, and a is on the left and c on the right. Therefore, if a is paired with c, the interval $[a, c]$ will intersect all others. Clearly, the interval $[a, d]$ will intersect all others if a is paired with d. We will assume that a is paired with b from now on, and prove that then *no* interval intersects all others. This will end the proof, as an interval that intersects all others occurs in two out of three choices for the pairing of a, hence with probability $\frac{2}{3}$.

Certainly neither $[a, b]$ nor $[c, d]$ can intersect all others, as they are disjoint. Suppose that some interval defined by a pair A_{j_0}, B_{j_0}, for some $j_0 \leq n-2$, intersects all others. In the balanced case defined above, all A_j, $j \leq n-2$ must all be between b and c and thus the interval defined by A_{j_0}, B_{j_0} cannot intersect *both* $[a, b]$ and $[c, d]$.

The final possibility is thus the unbalanced case, when the $(n-2)$'nd pair must be a BA-type. On the other hand, the interval defined by A_{j_0}, B_{j_0} intersects $[c, d]$ and so the j_0'th pair must be an AB-type. Thus there is a $k \in [j_0+1, n-2]$ so that the $(k-1)$'st pair is an AB-type while the k'th pair is a BA-type. In particular, $A_k < n$ (which it always is when $k > 1$ and the previous pair is AB-type) and $A_k < A_{j_0}$ (for the same reason, except that now $k > j_0$). This means that $[B_k, A_k] \cap [A_{j_0}, B_{j_0}] = \emptyset$. This is a contradiction, so no interval intersects all others in this case.