Lecture 9: Probably ...

9.1 Vocabulary

A real-world experiment has various **outcomes**. The collection of all possible outcomes is called the **sample space**. This vocabulary discussion will be accompanied by two simple examples. Then more complicated examples will be considered.

Example 1 Flipping coins. The outcomes are heads (H) or tails (T). We simplify by ignoring coins which land on their edges, lost coins, etc. The sample space is $\{H, T\}$.

Example 2 Tossing a die (singular of dice). The outcomes can be labeled by the number of dots showing when the die stops moving. So there are 6 outcomes: $\{1, 2, 3, 4, 5, 6\}$.

We can collect the outcomes in various ways, and these collections are called **events**. Example 1 doesn't have too many events: $\{H\}$ and $\{T\}$ and $\{H, T\}$. The last event is the entire sample space. We also record the event with *no* outcomes, the empty event: $\emptyset$.

Although example 2 is still somewhat small, it has many events. Here are some.

The even throws: $\{2, 4, 6\}$; the squares: $\{1, 4\}$.

There are exactly 2^6 different events. Each event corresponds to a decision to include or exclude one of the 6 particular outcomes.

We could further complicate example 2 by, say, throwing the die twice. The outcomes of that "experiment" would be classified by ordered pairs (i, j) with $1 \leq i \leq 6$ and $1 \leq j \leq 6$. Here are some events of that experiment.

- $\mathcal{I}$, the increasing tosses: pairs (i, j) with $i < j$, so the outcome whose first toss is 3 and whose second toss is 5 is in $\mathcal{I}$. $\mathcal{I}$ contains $5+4+3+2+1 = 15$ distinct outcomes. The "4" in the sum corresponds to the outcomes $(2, 3)$, $(2, 4)$, $(2, 5)$, and $(2, 6)$.
- Sum_4, those tosses whose dots sum to 4. The toss 1 followed by the toss 3 is in Sum_4, which contains 3 distinct outcomes.

This experiment would have $2^{36} \approx 10^{10.8}$ different events.

Probability assigns numbers to events. The number represents the chance that the event happens. We should believe that if the experiment is repeated many times, the ratio of the times that the outcomes are in the event divided by the number of experimental runs should generally "approach" the probability of the event. The reason for the quotes on the word is that I can't be more precise, and "approach" (written with $\rightarrow$) is an indication of how well the abstractions considered here model reality. So if an experiment is run N times, and A is an event, the ratio (sometimes called the **relative frequency**) $\frac{\text{outcomes in } A}{N} \rightarrow P(A)$ as N gets large.

If A is an event, then the probability $P(A)$ is a number which is between 0 and 1: $0 \leq P(A) \leq 1$. This assignment of numbers to events should obey some rules. These rules in turn evolved from the origins of probability, which seems to have been discussions of gambling problems during the 1600's.

The rules of the game

- The probability of anything (or, perhaps, everything!) happening is 1: $P(\text{the whole sample space}) = 1$.

- The empty set, $\emptyset$ (which has no outcomes as members), has probability 0.
- Disjoint events, which are events which have no outcomes in common, have probabilities which add: if A and B are events sharing no outcomes, $P(A)+P(B) = P(\text{an outcome in either } A \text{ or } B \text{ occurs})$.

This can be rewritten using set notation*. $A \cap B$ means the intersection of events (the outcomes in both A and B) and $A \cup B$ means the union of events (the outcomes in at least one of A and B).

$$\text{If } A \cap B = \emptyset, \text{ then } P(A \cup B) = P(A) + P(B).$$

Of course this can be extended to more than two events and even to sequences of events.

This equation has several consequences. First, as events get "larger" (have more outcomes), their probabilities increase. My language here is imprecise. I really mean that their probabilities don't have to strictly increase, but they can't decrease. That is, if $A \subset C$ (A is a subset of C, so every outcome in the event A is also an outcome in the event C), then $P(A) \le P(C)$. This is because C is the disjoint union of the events A and $C \setminus A$ (outcomes in C and not in A) with $P(A \setminus C) \ge 0$.

We can take apart $A \cup B$ when A and B are *not* disjoint: $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$. The three sets on the right side of the equation are disjoint, so $P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B)$. Since $P(A \setminus B) \le P(A)$ and $P(B \setminus A) \le P(B)$, $P(A \cup B) \le P(A) + P(B)$.

The simplest example is a **fair coin** with $P(\{H\}) = P(\{T\}) = \frac{1}{2}$: not much to discuss. A fair die gives each event a probability equal to the number of outcomes in the event divided by 6.

9.2 Fair from unfair

But there is no reason our imaginary coin or die should be fair. We can already start to analyze an interesting problem.

> **Problem JvN** Suppose we're given a *biased* (or unfair) coin, so that $P(\{H\}) = p$ (with $0 < p < 1$) and $P(\{T\}) = q = 1 - p$. How can we use the tosses of this coin to generate a random sequence which simulates a sequence of fair coin tosses?

I call this problem "JvN" because the solution below is said to be due to John von Neumann. I hope you can see applications of this problem to cryptography and creating random bitstreams. Think about the problem before you read more.

Imagine the sequence of tosses of the biased coin occurring as pairs of tosses. So the sequence of coin tosses, which might be $HTHHTHHHTTTHTHTT\ldots$ could be rewritten as $HT\ HH\ TH\ HH\ TT\ TH\ TH\ TT\ldots$. What should the probability of these

---

* *Mathematicians are like a certain type of Frenchman: when you talk to them they translate it into their own language, and then it soon turns into something completely different.* This is Maxim #1278 by Goethe (1749-1832), appearing in his *Maxims and Reflections*. It is useful to remember when reading and learning mathematics. See `http://www.kirjasto.sci.fi/goethe.htm` for information about Goethe.

various pairs of coin tosses be? There are four possible outcomes: HT, TT, TH, and HH. We believe that the coin tosses are **independent**. This is an important technical word in probability. Here it means that the first coin toss shouldn't influence the second one. If the first toss results in heads some portion p of the time, then the next toss should result in a tail a q^{th} portion of *that* time. So $P(\{HT\})$ should be pq. And $P(\{TH\})$ should be qp (the same as pq, of course). Here's von Neumann's scheme: when a successive pair of biased coin flips is HT, signal HEADS for the imaginary fair coin. If the pair is TH, signal TAILS for the imaginary fair coin. Otherwise, flip another two times. Of course it is possible that the coin will *never* have successive pairs of flips HT or TH. The probability of TT is q^2 and the probability of HH is p^2. Since $p + q = 1$, we know that $p^2 + 2pq + q^2 = 1^2 = 1$, so $0 < p^2 + q^2 < 1$. The probability of a sequence composed only of HH's and TT's is certainly at most $(p^2 + q^2)^{\text{high power}}$ (again independence is being used here, so the probabilities multiply), and this $\to 0$ as the "high power" grows. The strategy of von Neumann will win "almost all" of the time. Later I'll try to analyze how expensive or efficient this strategy is. That is, we will discover how may coin flips of the biased coin are needed, on average, to get the report of one "unbiased" H or T using this method.

Can one reverse the von Neumann trick? In particular, describe a strategy to solve the following problem:

> **Problem NvJ** Suppose you have a fair coin. Describe how to simulate a sequence of H's and T's so that the probability of H is $\frac{2}{3}$ and the probability of T is $\frac{1}{3}$.

Once you've done this, try describe how to simulate a sequence of tosses of an "arbitrary" biased coin. That is, suppose p is between 0 and 1. Use a fair coin to simulate a string of tosses so the probability of a head is p and the probability of a tail is $1 - p$.

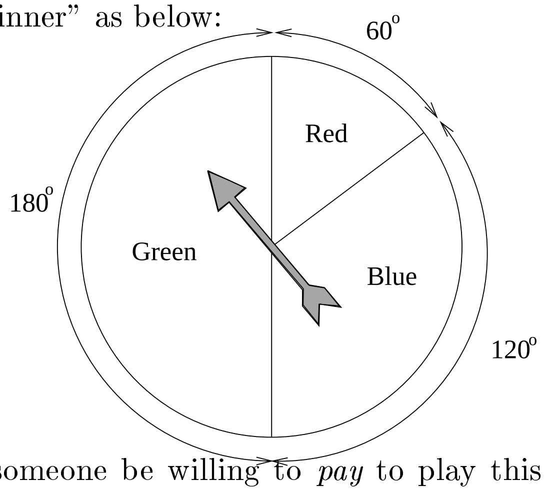
Before going on, here's an important formal definition:

Events A and B are **independent** if $P(E \cap F) = P(E) \cdot P(F)$.

Lecture 10: Gambling

10.1 Fair entry fees

It may be more interesting to use an extended gambling example to explain more about probability. So: suppose that we have a "spinner" as below:



What's the chance an "honest" or "fair" spin lands on **Green**? I think it is $\frac{180°}{360°}$, or $\frac{1}{2}$. Similarly, we can easily see that the chance of **Blue** is $\frac{120°}{360°}$, or $\frac{1}{3}$, while the chance of **Red** is $\frac{1}{6}$. So

$P(\textbf{Green}) = \frac{1}{2}$; $P(\textbf{Blue}) = \frac{1}{3}$; $P(\textbf{Red}) = \frac{1}{6}$.

So $0 \leq P(\text{something}) \leq 1$ and the sum of all the "somethings" that could happen is 1.

Now a new ingredient: what could one *win* in such a game? For example, suppose a **Green** spin pays \$30, a **Blue** spin pays \$15, and a **Red** spin pays \$75. What's an average spin worth?

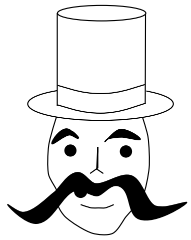
Phrased a bit differently, how much should someone be willing to *pay* to play this game? One naive answer might be: since there are three possible outcomes, and three possible rewards, the average reward of a spin is the average of the three outcomes, or \$ $\frac{30+15+75}{3}$ = \$40. Some consideration of extreme cases might show that's too simple. If **Green** were worth \$10,000 and the other two colors were worth nothing, then we'd expect about half of our spins to be **Green** in the long run, and about half the time to win \$10,000 per spin, so that the *average* winning per spin would be \$5,000. In our 30–15–75 payoff plan, we must compute a *weighted* average, and the weights are the chances, the probabilities, that each color will occur.

Outcome	Probability	Payoff per spin	Expected winnings
Green	$\frac{1}{2}$	\$30	\$15.00
Blue	$\frac{1}{3}$	\$15	\$5.00
Red	$\frac{1}{6}$	\$75	\$12.50

The total expected winnings (the **expectation**) will be \$32.50, the correct weighted average of the probabilities and the payoffs. An entry "fee" of less than \$32.50 would, in the long run, over many plays, yield a profit to the player. An entry fee greater than \$32.50 would, in the long run, over many plays, profit the "proprietor" of the spinner.

10.2 A game with many flips

Let's try some real gambling. A gentleman comes to you with a smile. He carries a fair coin, with two different sides, one side heads, H, and one side tails, T. He offers to play a game with you. He will toss the coin. If H appears, he will pay you \$1. If T appears, he will toss the coin again. If, on the second toss, H appears, he will pay you \$2. If T appears again, he will toss another time. And so on . . ..



Your friendly gambler

Let's specify this "game" more carefully. The set of outcomes is the collection of coin tosses, $TTT \ldots TH$. That is, for each non-negative integer, n, one possible outcome is $(n-1)$ T's followed by an H. Let's call this $\mathcal{S}_n$. What's the probability of $\mathcal{S}_n$? We're asking for n straight specified tosses of a fair coin, so $P(\mathcal{S}_n)$ must be $\left(\frac{1}{2}\right)^n$ (independence of the tosses). The gambler will pay n dollars if $\mathcal{S}_n$ occurs. The sample space here is *infinite* – very different from examples 1 and 2. Some questions we should consider follow.

- What if the coin *never* lands heads?

  The probability of n tosses of tails is $\left(\frac{1}{2}\right)^n$, and surely the event of the outcome "never heads" will be less than this for any positive integer n. The probability should be 0. This may be an example of a conceivable (?) event which never happens (??). In our probability modeling, it is more precisely a non-empty event with probability 0. The user of the model must decide if this makes sense.*

- Would you pay the gambler, an honest, genial individual, 50 cents to play this game?

  Of course. You've got to win at least a dollar.

- Would you pay the gambler, an honest, genial individual, one dollar to play this game?

  Surely, for the same reason as the previous question.

- Would you pay the gambler, an honest, genial individual, one million dollars to play this game?

  The gambler asserts that there are many, many positive integers. Only finitely many of these integers are less than one million, and infinitely many of them are more than one million. Therefore (according to the gambler) there has infinitely *more* chances of paying more than one million than you have of losing one million. (!!!) So pay the million and play the game. (The gambler is, of course, ignoring the fact that different outcomes occur with different probabilities.)

We can try to compute this game's average payoff (its "expectation") in a way that's similar to the spinner game. So:

Outcome	**Probability**	**Payoff per game**	**Expected winnings**
H	$\frac{1}{2}$	\$1	\$0.50
TH	$\frac{1}{4}$	\$2	\$0.50
TTH	$\frac{1}{8}$	\$3	\$0.375
. . . . .	...	...	...
$\mathcal{S}_n$	$\frac{1}{2^n}$	$\$n$	$\$\frac{n}{2^n}$
. . . . .	...	...	...

* **A literary note** Such coin-flipping is described and analyzed with much imagination in the first few pages of the play *Rosencrantz & Guildenstern Are Dead* by Tom Stoppard. More than 30 years after it was written, *R & G* is in the top 6500 at `Amazon.com`.

There's several observations we can make about this table. The sum of the probabilities of the various outcomes (an infinite list of outcomes which is abbreviated!) is

$$\sum_{n=1}^{\infty} P(\mathcal{S}_n) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1$$

Although we've got an "infinite series", we have the friendliest of such series, a geometric series. A geometric series is one whose terms are created by taking a first term, a, and multiplying that term repeatedly by one fixed number, r, the ratio, to create $a + ar + ar^2 + ar^3 + \ldots$. If S is the sum of this series, then

$$S = a + ar + ar^2 + ar^3 + \ldots = a + r(a + ar + ar^2 + \ldots) = a + rS$$

and this linear equation in S can be solved to get the formula $S = \frac{a}{1-r}$. Our specific geometric series has both a and r equal to $\frac{1}{2}$, so its sum is 1, as we should hope, since we've made a list of all (non-negligible) possible outcomes.

What about the average payoff? If the entry fee is *less* than the average payoff, in the long run, over repeated plays, we'd expect to profit, and the honest gambler tossing the coin for us would expect to lose money. If the entry fee is *more* than the average payoff, in the long run, over repeated plays, we'd expect to loose, and our honest, genial friend would expect to profit.

We need to look at the sum of the expected winnings:

$$\frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \frac{4}{2^4} + \ldots = \sum_{n=1}^{\infty} \frac{n}{2^n}$$

It is perfectly possible that the expectation *could* be infinite. To emphasize this, modify the game to one in which the gambler offers to pay 2^n dollars if the first head occurs on the n^{th} toss: that is, outcome $\mathcal{S}_n$. Then every play of the game offers us a chance to win (on average) $\sum_{n=1}^{\infty} \frac{2^n}{2^n} = \infty$ dollars – quite a lot! But what does the result ∞ mean? Certainly no one play or coin toss sequence will win infinitely many dollars. It really means that there's no upper bound on the average amount of winnings in the changed game. There's no fair entry fee that the gambler could charge – the gambler would always lose money in the long run.

Return to the original game, where the payoff is $\sum_{n=1}^{\infty} \frac{n}{2^n}$. This series is more complicated than the first. It is not a geometric series since the ratio between successive terms changes. For example, the ratio between the first and second terms is 1, and the ratio between the second and third terms is $\frac{3}{4}$. Let's see how to compute the average payoff.

10.3 Computing the payoff with algebra & magic

We can use the geometric series formula to write the following:

$$\begin{array}{rcl}
1 & = & \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \ldots \\
\frac{1}{2} & = & \qquad \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \ldots \\
\frac{1}{2^2} & = & \qquad\qquad \frac{1}{2^3} + \frac{1}{2^4} + \ldots \\
\frac{1}{2^3} & = & \qquad\qquad\qquad \frac{1}{2^4} + \ldots \\
& \vdots &
\end{array}$$

If we add this array *vertically* we'll get an equation whose left-hand side is 2 (because it is just the sum of another geometric series with $a = 1$ and $r = \frac{1}{2}$), and whose right-hand side magically is $\frac{1}{2} + 2\left(\frac{1}{2}\right)^2 + 3\left(\frac{1}{2}\right)^3 + 4\left(\frac{1}{2}\right)^4 + \ldots$ This sum can also be evaluated using calculus, but calculus isn't needed for everything!

The average payoff from the coin-flipping game is therefore 2. So an entry fee of less than \$2 will result, over the long term, in profit for the player, while a fee of more than \$2 will give the gambler an edge.

10.4 JvN efficiency

We can now evaluate the likely efficiency of von Neumann's solution, converting a biased bitstream to an unbiased bitstream. I'll be more precise. We have a biased or unfair coin which when flipped is heads with probability p with $0 < p < 1$ and is tails with probability $q = 1 - p$. We begin flipping this coin, and consider the outcome of pairs of flips. If the coin lands HH or TT we try again. If the ordered pair of flips is HT we conclude that our imaginary fair coin has landed H. If the ordered pair of flips is TH, then we imagine it has landed T. How many flips are necessary, on average, for us to make one imaginary toss of our fair coin? That is, what is the expectation of the number of coin tosses for one decision?

For example, with probability $2pq$ the first pair is either HT or TH. So $2 \cdot 2pq = 4pq$ is part of the sum for the expectation. The first "2" comes from the 2 unbiased flips and the $2pq$ is really $pq + qp$, the probabilities of HT and TH, respectively.

How can we wind up with 4 tosses? The first two could be HH or TT. The probability of these exclusive events is $p^2 + q^2$. Then we would need to multiply by $2pq$ for the third and fourth tosses of the biased coin. And, finally, we'd need to multiply by 4 for the number of flips needed. So another piece of the sum for the expectation is $8pq(p^2 + q^2)$.

Now to go for 6 tosses. The first 4 tosses must be $HH\,HH$ or $HH\,TT$ or $TT\,HH$ or $TT\,TT$. These exclusive events have probability $p^4 + p^2q^2 + q^2p^2 + q^4$. This is $\left(p^2 + q^2\right)^2$. Then we multiply by $2pq$ and then by 6 to get $12pq\left(p^2 + q^2\right)^2$.

Now assemble the beginning of the sum for the expectation.

$$4pq + 8pq(p^2 + q^2) + 12pq\left(p^2 + q^2\right)^2 + \ldots$$

I hope the pattern is clear. The expectation for the average number of biased coin flips needed to simulate one fair coin flip is $E = \sum_{n=1}^{\infty} 4npq\left(p^2+q^2\right)^{n-1}$. The $\left(p^2+q^2\right)^{\text{a power}}$ comes from a collection of double heads and double tails (in any order) preceding either HT or TH. All the ways of multiplying out the p^2 and q^2 each represent one path of HH's and TT's. Then the same ideas we've already seen can be used to find the sum of this series, which is $\dfrac{4pq}{\left(1-(p^2+q^2)\right)^2}$.

Here's a different way to do this problem. Suppose E is the expectation we want to compute. We can get a simple equation for E by "pure thought". Either we succeed with the first pair of coin tosses and the number of tosses needed is 2, or we do not succeed with the first pair. If we do not, then the number of coin tosses increases by 2. So: the chance of success with the first pair of tosses is $2pq$. The chance of failure with the first pair of tosses is $1-2pq$. Therefore $E = 2pq\cdot 2 + (1-2pq)\cdot(2+E)$ and we can solve this equation for E. We can multiply and get $E = 4pq + 2 + E - 4pq - 2pqE$ so that $E = \frac{1}{pq}$. Is this the same as the answer above? Since $p+q=1$, $p^2+2pq+q^2=1$ and $2pq = 1-(p^2+q^2)$. Then $\left(1-(p^2+q^2)\right)^2 = (2pq)^2$ and $\dfrac{4pq}{\left(1-(p^2+q^2)\right)^2} = \dfrac{4pq}{(2pq)^2} = \dfrac{1}{pq}$. The two answers are the same. The second method ("conditioning on the first pair of flips") is very clever*.

If the biased coin is only slightly defective (so p and q are both about $\frac{1}{2}$), then the sum is about 4. If, though, $p = .99$ (a *very* biased coin!) on average about 101 tosses of the biased coin will be needed to get one "fair" toss simulated. This seems wasteful. Is there a more efficient solution?

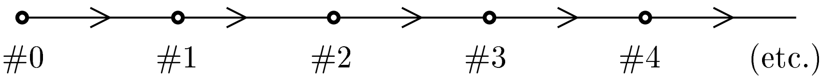
* A comment about logic: the second method computes E after first assuming (without proof) that it exists and is finite. Some people find this uncomfortable.

Lecture 11: The transmission network

11.1 Two models

We'll apply the logic developed to study some models of computer networks, and to evaluate their chance of failure or inaccuracy. We'll begin by examining a very simple network. We can imagine a computer at #0 transmitting some information, say, for simplicity, just a bit (0 or 1). Ideally the information is received by computer #1 which retransmits it to #2 etc.: a sequence of computers or switches relaying information.

Transmission network



11.2 Model #1: things break

Suppose each link has a probability $q = 1 - p$, a number between 0 and 1, of breaking. What's the long-term effect on the network? The network is unbroken up to computer n if all of the first n links are unbroken. We assume that broken links between the computers occur independently. Therefore the probability that the first n links are unbroken is p^n: the probabilities multiply because of the independence assumption. As n increases, this probability goes to 0. For example, if $p = .999$ (so one would assume the network is rather reliable), then there's 1 chance in 20 that the network is broken after about 50 links since $(.999)^{50} \approx .95121$. It is highly unlikely that the network stays unbroken link after link.

11.3 Model #2: bit flipping

Here's a different model of the transmission network. The first computer transmits a bit (0 or 1). Assume that there is some probability $q = 1 - p$ between 0 and 1 that the bit gets flipped: if 0 is received, then 1 gets transmitted; if 1 is received, then 0 gets transmitted. Now the network never "fails" in the sense of not transmitting *something*, but it may not transmit correct information. We also may get lucky: if the bit gets flipped twice (or, in fact, an even number of times) the true message will be transmitted.

Suppose T_n is the probability that the **T**rue value of the bit has been transmitted through the n^{th} link, and F_n is the probability that the opposite bit (a **F**alse bit) is transmitted. T_n and F_n are both between 0 and 1, and $T_n + F_n = 1$. Let's compute a few T_n's and F_n's to gain some familiarity.

$T_1 = p$ and $F_1 = q$. What about T_2? Two disjoint events could occur giving us the correct output. First, the truth could be transmitted twice. This has probability p^2. Or we could be lucky and both links could flip the bit. This has probability q^2. So $T_2 = p^2 + q^2$. To get an error after the second link, there had to be just one error (either in the first computer or the second, but not both). So $F_2 = qp + pq = 2pq$. For T_3 there can be exactly 0 or 2 errors, so $T_3 = p^3 + 3pq^2$. The 3 appears because there are 3 distinct outcomes with 2 errors, so there must be one "true" transmission in each of 3 positions. F_3 occurs when there are exactly 1 or 3 errors, so $F_3 = 3p^2q + q^3$. Here the 3 occurs because the 1 error or bit flip can occur in one of 3 links. This should help you understand the general answer. T_n happens when we sprinkle an even number of errors in the network, and F_n, when there's an odd number of errors. The binomial coefficients describe the total number of ways of distributing the errors.

Here's a more systematic way to get the formulas. We know that

$$\begin{aligned} T_n =& pT_{n-1} + qF_{n-1} \\ F_n =& qT_{n-1} + pF_{n-1} \end{aligned}$$

because the event whose probability is T_n occurs when the truthful bit obtained with probability T_{n-1} is transmitted truthfully and when the false bit previously obtained is flipped. There's parallel logic for the equation giving F_n. So $T_n - F_n = (p-q)T_{n-1} + (q-p)F_{n-1} = (p-q)(T_{n-1} - F_{n-1})$. We know that $T_1 - F_1 = p - q$, so $T_n - F_n = (p-q)^n$. The name for the official proof technique needed here is "mathematical induction" but I hope convincing evidence has been given. Since $T_n + F_n$ must be 1, we see that $T_n = \dfrac{1+(p-q)^n}{2}$. If $p = .999$, then 50 links give about 95.2% reliability (slightly more than the first model, which gives 95.1% reliability).

Both transmission models get unreliable when the number of links increases. The first one is quite likely to be broken, and, under the conditions of the second model, $T_n \to \frac{1}{2}$ as n gets large because $|p-q| < 1$ so powers of $p-q$ go to 0. This means that the chance of detecting the original value of the transmitted bit is heading towards 50%, a random guess.

Comment The Binomial Theorem for positive integer exponents states that

$$(A+B)^n = \sum_{k=0}^{n} \binom{n}{k} A^{n-k}B^k \text{ with } \binom{n}{k} = \frac{n!}{(n-k)!k!}$$

and the **binomial coefficients** $\binom{n}{k}$ have some interesting interpretations such as "the number of ways of choosing k distinct objects from n distinct objects".

For example, suppose we have five animals: **L**ion, **T**iger, **E**lephant, **H**orse, and **G**oat. How many different collections of three animals could we get from these five? There are five ways of selecting the first animal, then four ways of selecting the second, and finally three ways of selecting the third. One selection is $\{\mathbf{T}, \mathbf{H}, \mathbf{L}\}$, in that order. There are $5 \cdot 4 \cdot 3 = 60$ such ordered selections. But we could also have chosen (in order) $\{\mathbf{L}, \mathbf{H}, \mathbf{T}\}$. There are several different ways of ordering each of the collections of three animals. How many? The first animal could be chosen in one of three different ways, and then the second animal could be chosen in one of two different ways, and the third animal is determined by the other choices. So for each collection of three animals, there are $3 \cdot 2 \cdot 1 = 6$ choices of order. Therefore to count the number of distinct and different collections of three animals from the five animals named, we should take 60 and divide by 6. Let's see: $\frac{60}{6} = \frac{5\cdot4\cdot3}{3\cdot2\cdot1} = \frac{5\cdot4\cdot3\cdot2\cdot1}{3\cdot2\cdot1\cdot2\cdot1} = \frac{5!}{3!\cdot2!} = \binom{5}{3} = 10$. You could check this calculation by listing all 10 collections of these animals.

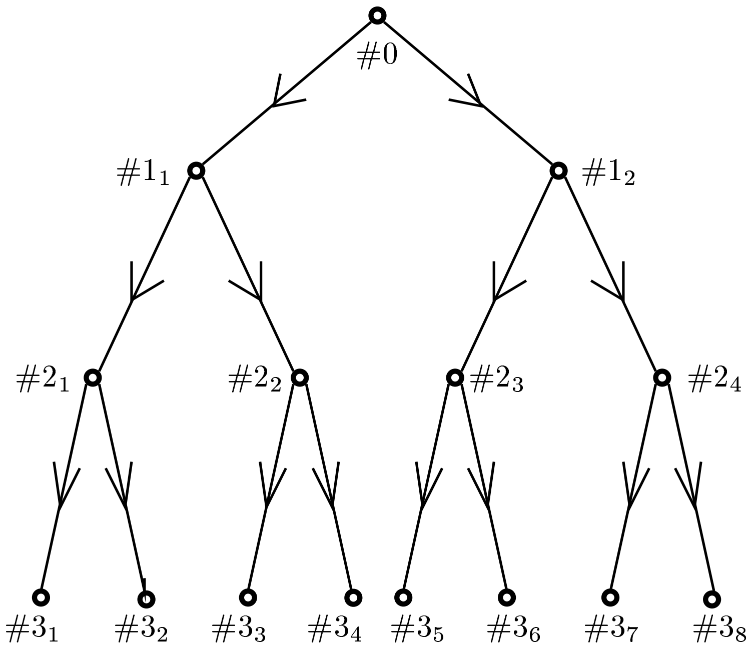
We did not prove the Binomial Theorem. A direct proof can be given using mathematical induction.

What's above suggests that $\dfrac{(A+B)^n + (A-B)^n}{2}$ is equal to that part of the sum written about which involves only even powers of B and that's true also. In the case we analyzed above, $A = p$ and $B = q$ so $A + B = p + q = 1$ and $(p+q)^n = 1$.

Lecture 12: Broadcasting – statement & heuristics*

There are other networks where, probabilistically, the truth will prevail. Here is one of them. First, we give a description of the ideal behavior of this network. The computer at #0 broadcasts a bit. The bit is received by two computers, labeled here as $\#1_1$ and $\#1_2$. Each of these then broadcasts the bit to two distinct computers, and each of these in turn to two others, etc.

Binary broadcasting network



(etc.)

This is a more complicated model. We will analyze only the "links breaking" behavior. Bit flipping is definitely harder to understand (it is discussed in [5]). The structure presented here is officially called a rooted binary tree of depth n. Each path from the root (labeled #0) down to a leaf has n edges, and there are 2^n leaves on the n^{th} level.

So we suppose that a link breaks with probability $q = 1 - p$, with q and p strictly between 0 and 1. Therefore the link is unbroken with probability p. We also require that breaking of links occurs independently.

> **Question** What is the probability that there is an unbroken path from the root to some computer at the n^{th} level?

Here the probabilities of various paths are not independent, so we must be careful. For example, if the link from $\#1_1$ to $\#2_2$ breaks, both $\#3_3$ and $\#3_4$ can't get information from the root.

I used `Maple` to create some data about this complicated model. The model is described by several variables. One variable is p, the probability that a link does *not* break.

---

* The dictionary says "heuristic" means "allowing or assisting to discover" (as an adjective). The word is frequently extended to a noun. Here I mean exploring a complicated situation with computer simulation, trying to guess what the correct, precise answer is.

Another variable is L, the number of levels in the "tree" of computers which we are trying to understand.

I had a `Maple` program use the random number generator to consider all the links, and cut each link, independently, with probability $1 - p$. The program checked if there was then at least one unbroken path to level L. I repeated this experiment N times and divided the number of times that there was an path from the root to level L by N to get the relative frequency. This could be an approximation to the probability that there is a path from the root to level L.

For example, I asked `Maple` to look at a tree with $L = 8$ levels and with $p = .3$ one hundred times. I got .01 and when I requested this again I got .02 and then .04 and .01 again. The results are different because `Maple`'s "random" number generator cut the tree in many different places. When I changed to $p = .8$ with the same number of levels again with one hundred trials, .91 and .89 were the first two reported relative frequencies. When I changed to $L = 4$ levels with $p = .3$, the first two reported relative frequencies were .09 and .11. With $p = .8$ and $L = 4$, they were .94 and .97. There are too many numbers. What's going on and how can we understand these results and use them to help us rather than confuse us?*

Suppose $\mathcal{A}(p, L, N)$ is the following "function":

> Consider the binary tree from the root to level L. Cut links independently with probability $1-p$ (and maintain them with probability p) using a pseudo-random number generator. Do this N times. $\mathcal{A}(p, L, N)$ is the relative frequency that there at least one path from the root to level L: divide by N the total number of times that there is such a path.

There are quotes around the word "function" because the number $\mathcal{A}(p, L, N)$ will likely be different each time the relative frequency is computed. Sometimes we may be lucky and there may be a large number of successes and sometimes we may be unlucky. But **if** there is some acceptable probability, we expect that as N increases, unless we are fantastically unlucky, the number $\mathcal{A}(p, L, N)$ would get close to that probability.

Experiment #1 For example, here is one computation for $p = .4$ and $L = 5$.

$N = 10$: .3; $N = 100$: .2; $N = 1{,}000$: .179; $N = 10{,}000$: .1882

And here is another computation with the same p and L.

$N = 10$: .2; $N = 100$: .18; $N = 1{,}000$: .193; $N = 10{,}000$: .1819

The results are different since the random number generator cuts different links.

Experiment #2 Change p to .6 but leave L at 5, unchanged.

$N = 10$: .4; $N = 100$: .68; $N = 1{,}000$: .634; $N = 10{,}000$: .6363

When I ran the program again I got these numbers.

$N = 10$: .7; $N = 100$: .67; $N = 1{,}000$: .635; $N = 10{,}000$: .6354

Theory We'll see shortly that theory predicts the true value of the first probability to be .18516, and the true value of the second, .63785. We would hope, as I wrote before, that as N increases, the numbers $\mathcal{A}(p, L, N)$ should "stabilize" near the true probability.

---

* A great applied mathematicians of the last century, Richard Hamming, wrote:

The purpose of computing is insight, not numbers.

Always remember this when doing computations.

General behavior of $\mathcal{A}(p, L, N)$

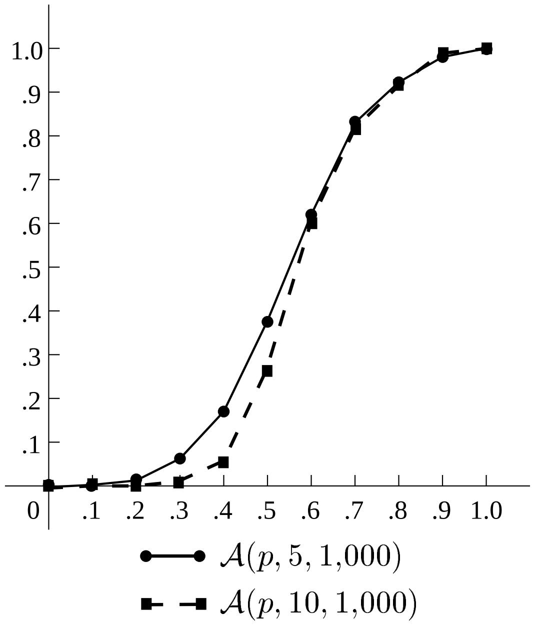
As p increases with L unchanged we should expect $\mathcal{A}(p, L, N)$ to increase. This is because more links are likely to be left on the tree. As L increases with p unchanged we should expect $\mathcal{A}(p, L, N)$ to decrease. Each additional level down the tree increases the risk that a path could be cut. All of this is the same qualitative behavior as the transmission models.

I computed more numbers. They are reported below. The computations were done to 10 decimal places. The trailing zeros are omitted.

Some values of $\mathcal{A}(p, L, N)$ for $L = 5$ and $L = 10$ ($N = 1{,}000$ trials)

p	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$L = 5$	0	0	.009	.053	.175	.385	.621	.84	.936	.986	1
$L = 10$	0	0	0	.003	.051	.269	.603	.825	.923	.993	1

The reported values for $\mathcal{A}(.9, 5, 1{,}000)$, .986, and $\mathcal{A}(.9, 10, 1{,}000)$, .993, are not in the "correct" order: even with 1,000 repetitions, the simulation was just unlucky. Here's a picture of this data which might be more instructive. The horizontal axis is p.



The most interesting part of these experimental results is that some of the probabilities *don't* seem to be getting close to 0 as the number of levels increases. This is strikingly different from the transmission models, where it seems that inevitably information is lost as the number of links increases. Something different is happening.

Lecture 13: Broadcasting – problem analysis

Certainly the probability that there is *some* path from the root to a computer at the n^{th} level is $\leq$ the sum over all paths of the probabilities that any specific path is unbroken (this uses the rule $P(A \cup B) \leq P(A) + P(B)$). But each specific path has n links, so each specific path is unbroken with probability p^n. There are 2^n distinct paths, so the probability at least one path is unbroken is *overestimated* by $2^n p^n$. If $p < \frac{1}{2}$ then $2p < 1$ and the overestimate is powers of $2p$, a number between 0 and 1. These powers, as in the transmission models, decay to 0 rapidly.

If $p \geq \frac{1}{2}$, we can't draw any conclusions based on what what's done so far, because the overestimates don't go to 0, so they can't "force" the probabilities to have any asymptotic behavior when n gets large. We need to work a bit more.

We define P_n to be the probability that there is some unbroken path from #0 to an n^{th} level leaf. We will analyze P_n in detail, primarily by relating P_n and P_{n+1}. There can be a path to a leaf at the $(n+1)^{\text{st}}$ level in two disjoint ways. Either both links from #0 are unbroken, or exactly one link from #0 is unbroken.

- Both links are unbroken with probability p^2 (multiplication again, since the breaking events are independent). We now need to look at the subtrees branching from 1_1 and 1_2. At least one of them should have a path to a leaf n levels below. The probability that one subtree will not have such a path is $1 - P_n$, so the event both will not have a path has probability $(1 - P_n)^2$. Therefore at least one has a path to a terminal leaf with probability $1 - (1 - P_n)^2$. The cumulative chance that this happens is $p^2\left(1 - (1 - P_n)^2\right)$.
- One link from #0 is broken and one is unbroken with probability $p(1-p)$. There are two disjoint ways for this to happen, so the total probability is $2p(1-p)$. The subtree from the unbroken link has a path to the bottom with probability P_n, so there is a path to a terminal leaf in this case with probability $2p(1-p)P_n$.

This discussion was an effort to convince you that the following equation is correct.

$$P_{n+1} = p^2\left(1 - (1 - P_n)^2\right) + 2p(1-p)P_n$$

Now I'll do some algebraic massaging:

$$\begin{aligned} p^2(1 - (1-P_n)^2) + 2p(1-p)P_n &= p^2\left(1 - 1 + 2P_n - (P_n)^2\right) + 2pP_n - 2p^2P_n \\ &= 2p^2P_n - p^2(P_n)^2 + 2pP_n - 2p^2P_n = 2pP_n - p^2(P_n)^2 = 1 - (1 - pP_n)^2 \end{aligned}.$$

so that

$$P_{n+1} = 1 - (1 - pP_n)^2.$$

Suppose that P_n has some nice behavior as n gets large: that is, P_n gets close to $\mathcal{P}$ for n large. The equation above implies that $\mathcal{P} = 1 - (1 - p\mathcal{P})^2$ or $\mathcal{P} = 2p\mathcal{P} - p^2\mathcal{P}^2$. Either $\mathcal{P}$ is 0 or $\mathcal{P} = \frac{2p-1}{p^2}$. Remember that $\frac{1}{2} \leq p \leq 1$ so $2p - 1$ is between 0 and 1. If $p < \frac{1}{2}$ then $2p - 1 < 0$. $\mathcal{P}$, a limit of non-negative probabilities, can't be $\frac{2p-1}{p^2}$.

13.1 An example: p=.75 and n from 1 to 10

This is complicated. An particular example may give some insight and inspiration. If $p = .75$ (so three-quarters of the time a link is *not* cut), then $P_{n+1} = 1 - (1 - .75P_n)^2$ and the number $\frac{2p-1}{p^2}$ is just $\frac{8}{9}$ or about .88889. Here is a table of values of the first 10 P_n's gotten using my silicon friend, `Maple`.

n	1	2	3	4	5	6	7	8	9	10
P_n	.93750	.91186	.90080	.89441	.89163	.89026	.88957	.88923	.88906	.88897

These P_n's seem to steadily decrease, and they seem to decrease towards $\mathcal{P}$. Please note that the numbers here are logically different from the data presented before. These numbers are computed from a theoretical "deterministic" model of the situation. The data shown earlier was obtained by simulating probabilistic or random actions on the tree.

13.2 Detailed analysis of $\{P_n\}$

For any p between $\frac{1}{2}$ and 1, some intricate algebra actually verifies what is suggested by the table above. The basic tool is again mathematical induction.

• **What happens when $n = 0$**

$P_0 = 1$ since we can always reach #0 from #0 (!) and $P_1 = p^2 + 2p(1-p) = 2p - p^2 = p(2-p)$, a quadratic whose maximum value is 1 (achieved when $p = 1$, halfway between the roots). So we know $P_1 \le P_0$. Also, $\frac{2p-1}{p^2}$ must be $\le P_0$ since $2p - 1 \le p^2$ (because $p^2 - 2p + 1 = (p-1)^2 \ge 0$). So

$$P_1 \le P_0 \qquad \text{and} \qquad \frac{2p-1}{p^2} \le P_0.$$

We combine these observations for $n = 0$ with the following facts.

• **What happens with bigger n's**

Fact 1 If $P_{n+1} \le P_n$ then $P_{n+2} \le P_{n+1}$.

Proof Take $P_{n+1} \le P_n$ and multiply by the positive number p to get $pP_{n+1} \le pP_n$ then multiply by -1 (reversing the inequality) and add 1: $1 - pP_{n+1} \ge 1 - pP_n$. These quantities are all between 0 and 1, so squaring doesn't change the inequality: $(1 - pP_{n+1})^2 \ge (1 - pP_n)^2$. Multiply by -1 again (changing the inequality) and add 1: $1 - (1 - pP_{n+1})^2 \le 1 - (1 - pP_n)^2$. This is exactly $P_{n+2} \le P_{n+1}$.

Fact 2 If $\frac{2p-1}{p^2} \le P_n$ then $\frac{2p-1}{p^2} \le P_{n+1}$.

Proof Take $\frac{2p-1}{p^2} \le P_n$ and multiply by the positive number p to get $\frac{2p-1}{p} \le pP_n$ then multiply by -1 (reversing the inequality) and add 1: $1 - \left(\frac{2p-1}{p}\right) \ge 1 - pP_n$. Of course $1 - \left(\frac{2p-1}{p}\right) = \frac{1-p}{p}$ These quantities are all between 0 and 1, so squaring doesn't change the inequality: $\left(\frac{1-p}{p}\right)^2 \ge (1 - pP_n)^2$. Multiply by -1 again (changing the inequality) and add 1: $1 - \left(\frac{1-p}{p}\right)^2 \le 1 - (1 - pP_n)^2$. The right-hand side is $\frac{p^2-(1-p)^2}{p^2} = \frac{-1+2p}{p^2}$, exactly what we wanted to prove.

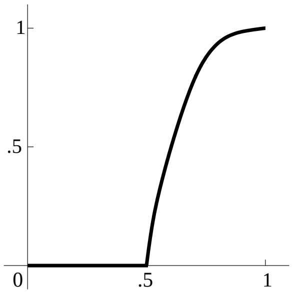
The sequence $\{P_n\}$ decreases as the integer n increases. Each of the terms is bounded below by $\frac{2p-1}{p^2}$. Such a sequence *must* converge to some $\mathcal{P}$ satisfying the equation $\mathcal{P} = 2p\mathcal{P} - p^2\mathcal{P}^2$. We saw there were two choices, and certainly we must have $\mathcal{P} = \frac{2p-1}{p^2}$. What can one say in (more-or-less) plain English about this model? The binary tree broadcasting model each of whose links independently breaks less than half the time *can* successfully transmit an initial bit to some computer at **any level** of the tree with *positive* probability.

Lecture 14: Broadcasting – solution and discussion

If $p > \frac{1}{2}$ is the probability that an individual link does not break, then the computer at the root of the tree can transmit to a computer at every level of the tree with probability at least $\mathcal{P}$, where

$$\mathcal{P} = \begin{cases} 0 & \text{if } 0 \le p \le \frac{1}{2} \\ \frac{2p-1}{p^2} & \text{if } \frac{1}{2} < p \le 1 \end{cases}$$

and a graph of $\mathcal{P}$ is shown to the right.

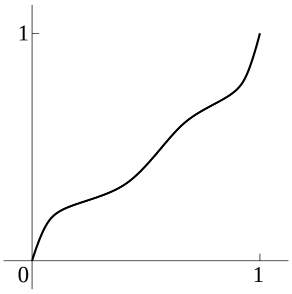


$\mathcal{P}$ as a function of p

This is a subtle result. Again, we've shown that when $p \le \frac{1}{2}$ the probability of a path from the root to at least one n^{th} leaf goes to 0 as n gets large. If $p > \frac{1}{2}$, the probability remains strictly above 0 (indeed, above the corresponding value of $\mathcal{P}$) as n grows. Here is the answer to the question with which we began the discussion of the tree network.

> **Answer** Suppose p is between 0 and 1. If P_n is the probability that the root can access the n^{th} level, then $P_n \to 0$ for $0 \le p < \frac{1}{2}$ as n increases, and $P_n \ge \frac{2p-1}{p^2} = \mathcal{P}$ for $p \ge \frac{1}{2}$ as n increases. In fact P_n steadily decreases to $\mathcal{P}$ as n gets large in the latter case.

We wrote P_1 explicitly earlier: it was $2p - p^2$. The equation $P_{n+1} = 1 - (1 - pP_n)^2$ allows computation of any P_n. So the P_n's are all *polynomials* in p, and students who know and love polynomials may find their relationship to the "strange" graph above, a function whose graph has a corner, somewhat strange. What should the graph of a nice "tame" polynomial look like? Certainly it should be smooth. A polynomial occurring in this problem should increase from left to right because the probability of having a path increases as p increases, and it should go through $(0,0)$ and $(1,1)$.

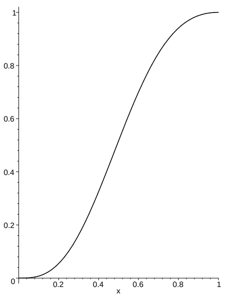


A tame polynomial

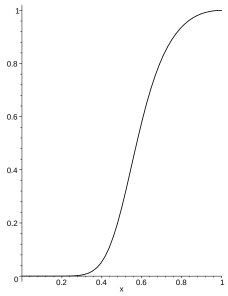
Most of our experience is with degree 1 and 2 polynomials (lines and parabolas). Maybe you've seen a few cubics and other higher degree polynomials. P_3 is a polynomial of degree 14:

$$-p^{14} + 8p^{13} - 24p^{12} + 28p^{11} + 8p^{10} - 48p^9$$
$$+28p^8 + 14p^7 - 8p^6 - 8p^5 - 4p^4 + 8p^3$$

which seems l-o-n-g and complicated. `Maple`'s picture of its graph looks simple: a smooth curve, increasing from $(0,0)$ to $(1,1)$.

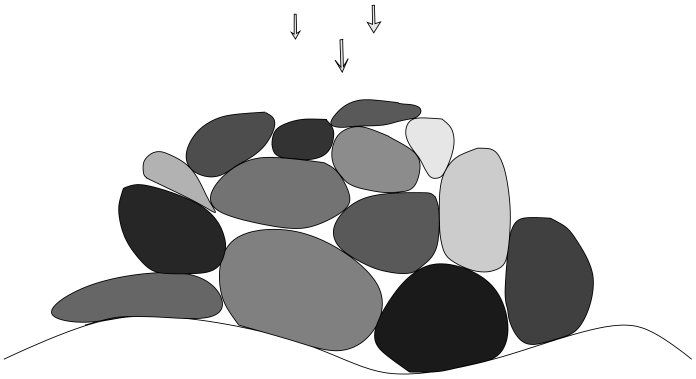


P_{10} is more complicated. It has degree is 2,046 and it has 2,037 non-zero terms. I will *not* write it out! In general, P_n has degree $2^n - 2$. I don't know how many non-zero terms it has. Here is a picture of P_{10}. I hope you can see the resemblance to the graph of $\mathcal{P}$ above. As n increases, the polynomial graph begins to resemble more and more the strange broken graph above.



Much of this mathematics was invented by physicists, who called the subject, "percolation". The picture below shows a physical example, with drips between the stones. The binary tree is supposed to be a model of the water trickling between the rocks.

{RAIN ON ROCKS}



The corner in the graph of $\mathcal{P}$ represents something MYSTERIOUS called a **phase change**. Similar models are used to discuss changes of state from solid to liquid or from liquid to gas: an abrupt change in the way a complex system behaves. The change can't be modeled by some nice function. Phase changes appear in many studies of complex phenomena.

14.1 Thanks and bibliography

Many conversations helped me prepare this material. Several faculty members (Professor Eugene Speer, whose specialty is mathematical physics, and Professor Charles Sims, whose specialty is computer algebra) spent time talking with me so I could better understand what's going on. Several graduate students (David Galvin, Vincent Vatter, and Jason Tedor) helped me with some of the computations. I wanted to show an "elementary" change-of-phase. Mr. Galvin and Professor Speer each suggested the specific example analyzed here.

Now some references.

[1] Almost *any* book on probability will be useful, and will have easy exercises on the basic vocabulary. The examples shown here aren't likely to appear in a basic book. The text used at Rutgers for the introductory probability course is Sheldon Ross, *A First Course in Probability*, 6$^{\text{th}}$ edition, 2002, $95. There are less expensive books covering this material.*

[2] Geoffry Grimmett, *Percolation*, Springer-Verlag, 1989, 296 pages, $99. This is an advanced text. Section 8.1 discusses percolation on a tree.

[3] Russell Lyons and Yuval Peres, *Probability on Trees and Networks*. The authors write: "This book is still being written. Most parts that are available are in close-to-finished form, but some are definitely in progress. . . . The final product will be published by Cambridge University Press. We hope that will be in the year 2003." A version is available on the web: `http://www.math.gatech.edu/~rdlyons/prbtree/prbtree.html`

---

* Hint: See Dover Publications.

[4] B. P. Watson and P. L. Leath, *Conductivity in the two-dimensional-site percolation problem*, Phys. Rev. B 9, 4893-4896 (1974). This can be seen at `http://prola.aps.org/thumbnail/PRB/v9/i11/p4893_1?start=0`. This classic physics paper reports on a phase change problem similar to what is discussed here using a simple (and witty!) experimental setup. Take a rectangular mesh window screen, and establish a potential difference across diagonal corners. Start clipping (breaking) some of the wire connections "at random", and observe how the resistance changes. A "change of state" occurs after a portion of the wires are cut. The experiment reflects phenomena very similar to what was analyzed here.

[5] William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman, *Broadcasting on Trees and the Ising Model*, Ann. Appl. Prob. 10, (2000), 410–433. This is #23 on `http://stat-www.berkeley.edu/~peres/recent.html`. The paper discusses bit flipping and information transmission on some networks, including binary trees.