Probability theory

H.W. Lenstra, Jr.

These notes contain material on probability for **Math 55**, Discrete mathematics. They were written to supplement sections 3.6 and 3.7 of C. L. Liu, *Elements of discrete mathematics* (McGraw-Hill, second edition, 1985), but they can be used in conjunction with other textbooks as well.   November 1988.

1. Sample spaces, events and conditional probabilities.

A *sample space* is a finite or countable set S together with a function

$$P\colon S \to [0,1] = \{y \in \mathbf{R} : 0 \le y \le 1\},$$

such that the following condition is satisfied:

$$\sum_{x \in S} P(x) = 1.$$

One thinks of S as the set of all possible outcomes of an experiment, with $P(x)$ equal to the probability that the outcome is x. We call P the *probability function*.

In many interesting cases S is finite and $P(x) = 1/\#S$ for all $x \in S$.

Let S be a sample space, with probability function P. An *event* is a subset A of S, and the probability of an event A is given by

$$P(A) = \sum_{x \in A} P(x).$$

If A, B are two events, and $P(B) \ne 0$, then the *conditional probability* of A given B is defined by

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

Examples of sample spaces, events and conditional probabilities, and exercises on them, can be found in many textbooks on discrete mathematics.

2. Independence.

Two events A, B are called *independent* if $P(A \cap B) = P(A) \cdot P(B)$. This means the same as $P(A|B) = P(A)$ (if $P(B) \neq 0$), so two events are independent if the occurrence of one of them does not make the other more likely or less likely to occur.

Example 1. Let $S = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$, the set of outcomes of rolling two dice, with each outcome having probability $\frac{1}{36}$. Consider the events

$$A = \text{the first die equals 3,}$$
$$B = \text{the second die equals 4,}$$
$$C = \text{the total equals 7,}$$
$$D = \text{the total equals 6,}$$

more precisely

$$A = \{3\} \times \{1, 2, 3, 4, 5, 6\},$$
$$B = \{1, 2, 3, 4, 5, 6\} \times \{4\},$$
$$C = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\},$$
$$D = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}.$$

Here A and B are independent, since $P(A \cap B) = \frac{1}{36} = P(A) \cdot P(B)$. Also A and C are independent, and B and C are independent. However, A and D are not independent, since $P(A \cap D) = \frac{1}{36}$, whereas $P(A) \cdot P(D) = \frac{1}{6} \cdot \frac{5}{36}$, which is smaller than $\frac{1}{36}$. Likewise B and D are not independent. Are C and D independent?

Warning. Any two of the three events A, B, C are independent, but nevertheless the three events A, B, C cannot be considered independent, since the occurrence of any two of them implies the occurrence of the third one. For three events A, B, C to be independent one requires not only that any two are independent, but also that $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$. For more than three events one has to look at even more combinations, see for example Exercise 25 (a).

3. Random variables.

It frequently happens that one is not interested in the actual outcome of the experiment, but in a certain function of the outcome. For example, if one throws two dice one is often just interested in the total score; or if one tosses a coin n times, it may be that one is just interested in the number of heads appearing, or in the longest consecutive sequence of tails. These are examples of *random variables*.

Let, generally, S be a sample space, with probability function P. A *random variable* is a function $f\colon S \to \mathbf{R}$. If r is a real number in the image of f, then the probability that f assumes the value r is defined by

$$P(f = r) = \sum_{x \in S,\, f(x) = r} P(x).$$

This is the same as the probability of the event $f(x) = r$. The *expectation* (or *expected value*, or *mean*) of a random variable f is defined by

$$E(f) = \sum_{x \in S} f(x) P(x).$$

This may be thought of as the "average" value of f if one repeats the experiment a great number of times. An alternative formula for $E(f)$ is

$$E(f) = \sum_{r \in f(S)} r \cdot P(f = r).$$

To prove this, notice that for each $r \in f(S)$ we have

$$\sum_{x \in S,\, f(x) = r} f(x) P(x) = r \cdot P(f = r),$$

by the definition of $P(f = r)$. Now sum this over $r \in f(S)$.

Example 2. Let $S = \{0,\, 1\}$, with $P(0) = P(1) = \frac{1}{2}$, and let $f\colon S \to \mathbf{R}$ be the inclusion map (so $f(0) = 0$, $f(1) = 1$). Then $E(f) = \frac{1}{2}$.

Example 3. Let $S = \{1,\, 2,\, 3,\, 4,\, 5,\, 6\}$, with $P(x) = \frac{1}{6}$ for each $x \in S$ (throwing one die), and let $f\colon S \to \mathbf{R}$ again be the inclusion map. Then $E(f) = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3\frac{1}{2}$.

If f, g are two random variables, then $f + g$ is also a random variable; namely, $(f + g)(x)$ is defined to be $f(x) + g(x)$, for all $x \in S$. One has

$$E(f + g) = E(f) + E(g).$$

This follows directly from the definition.

Example 4. Let S, P be as in Example 1 (rolling two dice). If f is the value of the first die (so $f\big((a,b)\big) = a$), and g is the value of the second die $\big(g\big((a,b)\big) = b\big)$, then $E(f) = E(g) = 3\frac{1}{2}$. The total is $f + g$, so the expectation of the total score equals $E(f + g) = E(f) + E(g) = 7$.

4. Events as random variables.

Let S be a sample space, with probability function P. Let A be an event; so this is a subset of S. It gives rise to a random variable f that is defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

This random variable assumes only the values 0 and 1; and any random variable that assumes only the values 0 and 1 arises in this way from some event. The probability of the event A is the same as the expectation $E(f)$ of the corresponding random variable f.

Example 5. Let $p \in \mathbf{R}$, $0 < p < 1$, and write $q = 1 - p$. Let S be the sample space $\{0, 1\}$, with probability function given by $P(1) = p$, $P(0) = q$. This sample space can be thought of as the set of outcomes of tossing a coin that is not fair (unless $p = \frac{1}{2}$). The probability of the event $A = \{1\}$ is the same as the expectation of the inclusion map (see Example 2), namely p.

Example 6. Repeat the experiment from the previous example n times; so S must then be replaced by $\{0, 1\}^n$, with $x \in S$ having probability $p^k q^{n-k}$ if x contains exactly k 1's (any $x \in S$ is a sequence of n 0's and 1's). For each $i = 1, 2, \ldots, n$, let A_i be the event that the i-th toss of the coin gives a 1, and let f_i be the corresponding random variable; so if $x = (x_1, x_2, \ldots, x_n)$, then $f_i(x) = x_i$. Each A_i has probability p, so each f_i has expectation p.

Now let $f = f_1 + f_2 + \cdots + f_n$. Then $f(x)$ is the total number of 1's in x, and

$$E(f) = E(f_1 + f_2 + \cdots + f_n) = E(f_1) + E(f_2) + \cdots + E(f_n) = np.$$

So the expected number of 1's is np. We can also verify this by a direct calculation. First one notes that

$$(*) \qquad\qquad P(f = k) = \binom{n}{k} p^k q^{n-k},$$

because there are $\binom{n}{k}$ elements $x \in S$ that have exactly k 1's, and each such x has probability $p^k q^{n-k}$. Now

$$E(f) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k} = np(p+q)^{n-1} = np.$$

A random variable $f: S \to \mathbf{R}$ is said to be a *binomial random variable*, or to have a *binomial distribution*, if there exists an integer $n \geq 0$ and a real number p, $0 < p < 1$, such that ($*$) holds for all k, $0 \leq k \leq n$. If we take $p = \frac{1}{2}$, then we see (not surprisingly) that the average number of elements of a subset of a set of n elements equals $\frac{1}{2}n$.

Example 7. Ten men went to a party and checked their hats when they arrived. The hats were randomly returned to them when they departed. We want to know the expected number of men that get their own hats back. In this example, S is the set of bijective functions of a set M of 10 men to itself, all these functions having the same probability, which is $1/\#S = 1/10! = 1/3628800 \doteq 0.0000002755732$. The bijective function that is the outcome of the experiment maps i to j if i receives j's hat. The number of men that receive their own hats back is given by the random variable $f: S \to \mathbf{R}$ that sends each bijective function $b: M \to M$ to the number of elements that it fixes: $f(b) = \#\{i \in M : b(i) = i\}$. We want to calculate the expectation of f. As in the previous example, this expectation can be written as the sum of the probabilities of 10 events A_i, where A_i is the event that the i-th man gets his own hat back. It is not hard to see that $\#A_i = 9!$ for each i, so $P(A_i) = \frac{1}{10}$ and $E(f) = 10 \cdot \frac{1}{10} = 1$: on the average, *one* man gets his own hat back! This remains true for any (positive) number of men instead of 10.

Example 8. Consider the experiment of shooting at a target until there is a hit. Let p be the probability of hitting if one shoots once, with $0 < p < 1$, and write $q = 1 - p$ for the probability of missing. Assume that all shots are independent. What is the expected number of shots? Here $S = \{h, mh, mmh, mmmh, \ldots\}$, where the n-th element has probability $q^{n-1}p$ (one factor q for each miss, and p for the hit); you may wish to check that all these numbers add up to 1. Let $f: S \to \mathbf{R}$ count the number of shots; so f maps the n-th element of S to n. The expectation of f is

$$E(f) = \sum_{n=1}^{\infty} n q^{n-1} p.$$

Using the formula

$$\sum_{n=1}^{\infty} n x^{n-1} = \frac{1}{(1-x)^2}$$

(for $|x| < 1$) one finds that

$$E(f) = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

This answer is intuitively clear: if one needs k shots for a hit, on the average, then a hit occurs once every k times, so $p = 1/k$ and $k = 1/p$.

One can also obtain this result by using Exercise 19 to write $E(f) = \sum_{k=1}^{\infty} P(A_k)$, where A_k is the event that one shoots at least k times, which happens with probability q^{k-1}.

With $p = \frac{1}{2}$ we see that a fair coin has to be tossed twice, on the average, before the first head comes up.

5. Independence of random variables.

Let f, $g: S \to \mathbf{R}$ be two random variables. We call f and g *independent* if for each $r_1 \in f(S)$ and each $r_2 \in g(S)$ one has

$$P(\{x \in S : f(x) = r_1 \text{ and } g(x) = r_2\}) = P(f = r_1)P(g = r_2).$$

In other words, the events $f(x) = r_1$ and $g(x) = r_2$ are independent, for each $r_1 \in f(S)$ and each $r_2 \in g(S)$. It means that the value of the first random variable gives no information about the value of the second random variable.

This notion generalizes the notion of independence of events that we saw earlier: two events are independent if and only if the two corresponding random variables are independent.

Example 9. Let S be as in Example 1 (rolling two dice). The first die (more precisely: the random variable that gives the value of the first die) and the second die are independent. The first die and the total score are not independent.

If f, g are two random variables, then $f \cdot g$ is also a random variable; namely, $(f \cdot g)(x)$ is defined to be $f(x)g(x)$, for all $x \in S$. If in addition f and g are independent, then one has

$$E(f \cdot g) = E(f)E(g).$$

Proof:
$$E(f \cdot g) = \sum_{x \in S} f(x)g(x)P(x) = \sum_{r_1, r_2} r_1 r_2 P(f = r_1)P(g = r_2)$$
$$= \left(\sum_{r_1} r_1 P(f = r_1)\right) \cdot \left(\sum_{r_2} r_2 P(g = r_2)\right) = E(f)E(g).$$

6. Standard deviation and variance.

Example 10. Let $S = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$, with each $(i, j) \in S$ having probability $\frac{1}{36}$, as in Example 1. Let the random variables f, $t: S \to \mathbf{R}$ be defined by $f((i, j)) = 2i$ (twice the value of the first die) and $t((i, j)) = i + j$ (the total score). Both random variables have expectation 7. However, the value of f is usually much further removed from the expectation 7 than the value of t. To make this quantitative, one usually considers the *standard deviation*.

Let, generally, S be a sample space, with probability function P. Let $f: S \to \mathbf{R}$ be a random variable. The *standard deviation* of f is defined by

$$\sigma(f) = \left(\sum_{x \in S} (f(x) - E(f))^2 P(x) \right)^{1/2}.$$

Equivalently, if the random variable $h: S \to \mathbf{R}$ is defined by $h(x) = (f(x) - E(f))^2$, then

$$\sigma(f) = \sqrt{E(h)}.$$

So this is the square root of the expectation of the square of the deviation of f from its expected value. At first sight this definition seems unnecessarily complicated: why not simply take the expected value of the deviation itself? This can be done, and it yields $\tau(f) = \sum_{x \in S} |f(x) - E(f)| P(x)$ (the *mean deviation* of f). However, it turns out that $\sigma(f)$ is far simpler to use than $\tau(f)$, because it avoids the non-differentiable function $| \cdot |$.

The square of the standard deviation is called the *variance* of f, and denoted by $V(f)$:

$$V(f) = \sum_{x \in S} (f(x) - E(f))^2 P(x) = E(h).$$

This is the expectation of the squared deviation. A useful formula for the variance is:

$$V(f) = E(f^2) - E(f)^2.$$

Proof:
$$V(f) = \sum_{x \in S} (f(x) - E(f))^2 P(x)$$
$$= \sum_{x \in S} f(x)^2 P(x) - 2E(f) \sum_{x \in S} f(x) P(x) + E(f)^2 \sum_{x \in S} P(x)$$
$$= E(f^2) - 2E(f)E(f) + E(f)^2 = E(f^2) - E(f)^2.$$

Example 10 (continued). We calculate the variance of the function f. We have

$$E(f^2) = \frac{1}{6}(2^2 + 4^2 + 6^2 + 8^2 + 10^2 + 12^2) = \frac{364}{6} = 60\tfrac{2}{3},$$

and $E(f) = 7$, so

$$V(f) = 60\tfrac{2}{3} - 7^2 = 11\tfrac{2}{3}, \qquad \sigma(f) = \sqrt{11\tfrac{2}{3}} \doteq 3.416.$$

We can calculate the variance of t in a similar straightforward way, but it is easier to apply another formula.

Let, generally, S be a sample space with probability function P, and let f, $g \colon S \to \mathbf{R}$ be two *independent* random variables. Then we have

$$V(f + g) = V(f) + V(g).$$

The proof uses the relation $E(fg) = E(f)E(g)$, which we proved earlier for independent random variables:

$$
\begin{aligned}
V(f + g) &= E\big((f + g)^2\big) - E(f + g)^2 \\
&= E(f^2) + 2E(fg) + E(g^2) - E(f)^2 - 2E(f)E(g) - E(g)^2 \\
&= E(f^2) - E(f)^2 + E(g^2) - E(g)^2 = V(f) + V(g).
\end{aligned}
$$

Example 10 (continued). We use this formula to calculate the variance of t. We write $t = t_1 + t_2$, where t_1 is the value of the first die (so $t_1\big((i, j)\big) = i$) and t_2 is the value of the second die. From $t_1 = \frac{1}{2}f$ and the calculation we did for f we see that $V(t_1) = (\frac{1}{2})^2 V(f) = \frac{35}{12}$. Since t_1 and t_2 are independent, it follows that

$$V(t) = V(t_1) + V(t_2) = 5\tfrac{5}{6}, \qquad \sigma(t) = \sqrt{5\tfrac{5}{6}} \doteq 2.415.$$

Example 11. Let $p \in \mathbf{R}$, $0 < p < 1$, and let $S = \{0, 1\}$ be the sample space from Example 5 (so $P(1) = p$, $P(0) = q$). We saw in Example 5 that the inclusion map $f \colon S \to \mathbf{R}$ has expectation p. We have $f^2 = f$, so

$$V(f) = E(f^2) - E(f)^2 = p - p^2 = pq, \qquad \sigma(f) = \sqrt{pq}.$$

Example 12. Let S, f be as in Example 6; so $P(f = k) = \binom{n}{k}p^k q^{n-k}$. To calculate $V(f)$, we write $f = f_1 + f_2 + \cdots + f_n$. We now have the formula

$$V(f) = V(f_1) + V(f_2) + \cdots + V(f_n),$$

which is generally valid provided that any two of the f_i are independent (the proof is analogous to what we did above for the sum of *two* independent random variables). By Example 11, each f_i has variance $V(f_i) = pq$, so

$$V(f) = npq, \qquad \sigma(f) = \sqrt{npq}.$$

These formulas apply to any binomial random variable f. If we take $p = \frac{1}{2}$, then we see that the standard deviation of the size of a random subset of a set of n elements equals $\frac{1}{2}\sqrt{n}$.

7. Chebyshev's inequality.

Chebyshev's inequality expresses that it is not likely that a random variable assumes a value that is far from its expected value. The precise formulation is as follows.

Let S be a sample space with probability function P, and let $f \colon S \to \mathbf{R}$ be a random variable. Then for any positive real number r we have

$$P(|f(x) - E(f)| \geq r) \leq \frac{V(f)}{r^2}.$$

Here the left hand side is the probability of the event $A = \{x \in S : |f(x) - E(f)| \geq r\}$, as usual.

Proof. Let A be as just defined. We have

$$V(f) = \sum_{x \in S} \big(f(x) - E(f)\big)^2 P(x)$$
$$= \sum_{x \in A} \big(f(x) - E(f)\big)^2 P(x) + \sum_{x \in S - A} \big(f(x) - E(f)\big)^2 P(x).$$

The second sum is obviously ≥ 0. Each term in the first is $\geq r^2 P(x)$, because of the definition of A. Therefore

$$V(f) \geq \sum_{x \in A} r^2 P(x) = r^2 P(A).$$

This implies Chebyshev's inequality.

8. Generating functions.

Let S be a sample space, and $f: S \to \mathbf{R}$ a random variable with the property that the image of f is contained in the set of non-negative integers. The *generating function* g of f is defined by

$$g(z) = \sum_{k=0}^{\infty} P(f = k) z^k.$$

Here z is a real variable with $|z| \le 1$. From

$$\sum_{k=0}^{\infty} P(f = k) = 1$$

it follows that the series converges, and that $g(1) = 1$.

The generating function can be used to calculate the expectation $E(f)$ and the variance $V(f)$ of f. Namely, suppose that the radius of convergence of the power series defining g is larger than 1. Then the derivative $g' = dg/dz$ is given by

$$g'(z) = \sum_{k=1}^{\infty} k P(f = k) z^{k-1},$$

for $|z| \le 1$. Putting $z = 1$ we find the following formula for the expectation:

$$E(f) = g'(1).$$

Similarly, one can prove

$$V(f) = g''(1) + g'(1) - (g'(1))^2,$$

see Exercise 22.

Example 13. Let p be a real number, $0 < p < 1$, and n a positive integer. We let the sample space $S = \{0, 1\}^n$ be as in Example 6; so an element $x \in S$ that has exactly k 1's has probability $p^k q^{n-k}$, where $q = 1 - p$. Let the random variable f also be as in Example 6; so $f(x)$ is the number of 1's in x. We have

$$P(f = k) = \binom{n}{k} p^k q^{n-k},$$

so the generating function g is given by

$$g(z) = \sum_{k=0}^{n} \binom{n}{k} (pz)^k q^{n-k} = (pz + q)^n.$$

10

Notice that indeed $g(1) = 1$. We have

$$g'(z) = np(pz + q)^{n-1}, \qquad g''(z) = n(n-1)p^2(pz + q)^{n-2}$$

and therefore

$$E(f) = g'(1) = np,$$
$$V(f) = g''(1) + g'(1) - (g'(1))^2 = n(n-1)p^2 + np - n^2p^2 = npq,$$

in agreement with what we found in Example 6 and Example 12.

Exercises.

1. Andrew, Beatrix and Charles are playing with a crown. If Andrew has the crown, he throws it to Charles. If Beatrix has the crown, she throws it to Andrew or to Charles, with equal probabilities. If Charles has the crown, he throws it to Andrew or to Beatrix, with equal probabilities.

At the beginning of the game the crown is given to one of Andrew, Beatrix and Charles, with equal probabilities. What is the probability that, after the crown is thrown once, Andrew has it? that Beatrix has it? that Charles has it?

2. From a shuffled deck of 52 playing cards one card is drawn. Let A be the event that an ace is drawn, B the event that a diamond is drawn, and C the event that a red card (heart or diamond) is drawn. Which two of A, B, and C are independent? What if a joker is added to the deck?

3. Let S be a sample space, and let A, $B \subseteq S$ be two independent events. Let $A' = S - A$ and $B' = S - B$. Prove that A' and B' are independent events. Are A and B' independent? And A and A'?

4. A random number is drawn from $\{1, 2, 3, \ldots, 899, 900\}$; each number has probability $\frac{1}{900}$. Are the events "the number is divisible by 3" and "the number is divisible by 5" independent? What is the probability that the number has no factor in common with 15?

5. In a lottery 10,000 tickets are sold for \$ 1 each. There are five prizes: \$ 5,000 (once), \$ 700 (once), \$ 100 (three times). What is the expected value of a ticket?

6. Four boxes contain red balls, white balls and blue balls:
    the first box contains 13 red balls, 11 white balls and 6 blue balls;
    the second box contains 13 blue balls, 11 red balls and 6 white balls;

11

the third box contains 13 white balls, 11 blue balls and 6 red balls;

the fourth box contains 10 red balls, 10 white balls and 10 blue balls.

One performs the following experiment: first a box is drawn at random (giving equal probability to all four boxes). Next 12 balls are drawn at random from this box.

(a) Prove that the expected number of red balls that are drawn is equal to the expected number of white balls that are drawn.

(b) What is the expected number of blue balls that are drawn?

Hint: avoid doing any calculations.

7. Let $S = \{1, 2, 3, \ldots, 899, 900\}$, and let each element of S have probability $\frac{1}{900}$. Define the random variables f, $g: S \to \mathbf{R}$ by

$$f(x) = \text{ remainder of } x \text{ upon division by 3},$$
$$g(x) = \text{ remainder of } x \text{ upon division by 4}.$$

(a) Prove that f and g are independent.

(b) Calculate $E(f + g)$, $E(fg)$ and $V(f + g)$.

8. For this exercise, you have to know that Martians come in three sexes: male, emale and female.

A Martian triple is going to have nine children. At each birth, the probability for each sex is $\frac{1}{3}$. What is the expected number of emales among the children? What is the variance?

9. The *Collected Novels* of Mrs. Kip Vandegril have been published in ten volumes. Someone checks out five of these volumes from the local library, chosen at random, without knowing that Mrs. Vandegril wrote only five novels, each occupying two volumes. Calculate, for each $i = 0, 1, 2$, the probability that he can read i complete novels. What is the expected number of complete novels he can read? And the variance?

10. Let p be a real number, $0 < p < 1$, and $q = 1 - p$. Consider, as in Example 8, the sample space $S = \{h, mh, mmh, mmmh, \ldots\}$, where the n-th element has probability $q^{n-1}p$. Let further the random variable $f: S \to \mathbf{R}$ map the n-th variable to n. In Example 8 we saw that $E(f) = 1/p$. Prove that $V(f) = q/p^2$.

11. Let f be a binomial random variable, as in Example 6. Calculate the probability that the value assumed by f is even, and the probability that the value assumed by f is odd.

12. Fifty-two chocolate letters, two for each letter of the alphabet, are distributed at random among fifty-two children, one each. Can you calculate the expected number of children whose first name starts with the letter that they get, without knowing the names of the children? And the variance?

13. Prove that at least 75% of all subsets of a set of 100 elements have cardinalities lying between 41 and 59 (inclusive).

14. Let S be a sample space, and $f: S \to \mathbf{R}$ a random variable. Define the random variable $g: S \to \mathbf{R}$ by $g(x) = |f(x) - E(f)|$. Prove that $V(g) = \sigma(f)^2 - \tau(f)^2$, and that $\sigma(f) \geq \tau(f)$. Under which conditions is $\sigma(f) = \tau(f)$?

15. Can the equality sign hold in Chebyshev's inequality?

16. Let S be a sample space, $A \subseteq S$ an event, and $f: S \to \mathbf{R}$ the corresponding random variable; so $f(x) = 1$ for $x \in A$ and $f(x) = 0$ for $x \notin A$. Let the *variance* $V(A)$ of A be defined as the variance of f. Prove that $V(A) = P(A) \cdot P(A')$, where $A' = S - A$.

17. Twenty distinct numbers are arranged in a list in a random order, such that all 20! orderings are equally likely. Going down the list, one marks every number that is larger than all earlier numbers on the list. (In particular, the first number of the list is marked.) Prove that, for each $i = 1, 2, \ldots, 20$, the probability that the i-th number on the list is marked is $1/i$. What is the expectation of the number of marked numbers on the list?

18. Let the sample space be as in Example 7.
(a) What is more likely: that no man gets his own hat back, or that exactly one man gets his own hat back? Does your answer change if the number of men is different from 10?
(b) What is the probability that exactly nine men get their own hats back?
(c) Let A_{ij} be the event that the i-th man and the j-th man get their own hats back. What is the probability of A_{ij}, for $i \neq j$? Is the event that the i-th man gets his own hat back independent from the event that the j-th man gets his own hat back?
(d) In Example 7 we saw that the expected number of men that get their own hats back equals 1. Prove that the standard deviation of the number of men that get their own hats back is also 1.

19. Let f be a random variable assuming only non-negative integer values, and denote by A_k the event that $f(x) \geq k$. Prove that $E(f) = \sum_{k=1}^{\infty} P(A_k)$.

20. A street performer hides your coin under one of three cups. You want to get it back, but you don't have the faintest idea under which cup it is. The rules of the game require that, first, you point at one of the cups; next, to help you, the performer lifts one of the other two cups, showing that there is no coin under it; finally, you must determine your choice: stick to the cup you chose initially, or change to the last one. What is your strategy?

21. Klaas owes Karel an amount of x Dutch guilders, where $0 \le x \le 1$; but all he has is one Dutch guilder, and neither has change. They decide to settle the debt by playing the following game.

(i) First they make sure that $0 \le x \le \frac{1}{2}$. If this is not the case, it is achieved by letting the coin change hands, so that the debt x is changed into $1 - x$ (owed by the *other* person). If now $x = 0$ then the game is over, else they pass to (ii).

(ii) The person having the coin flips it. If Queen Beatrix comes up, he pockets the coin, and the game is over. If the other side comes up, his debt is doubled (i. e., x is replaced by $2x$). Now the game is started all over again (at (i)).

Prove that this is a fair game, in the sense that the expected amount of money that Karel receives is x guilders, for the initial value of x. (You may suppose that the coin is fair; this is guaranteed by Queen Beatrix.)

   Hint: write x in binary.

22. Let f be a random variable with the property that the image of f is contained in the set of non-negative integers, with generating function g. Prove the formula $V(f) = g''(1) + g'(1) - (g'(1))^2$. (You may assume that the radius of convergence of g is larger than 1.)

23. Let p be a real number, $0 < p < 1$, and $q = 1 - p$. Consider, as in Example 8 and Exercise 10, the sample space $S = \{h,\, mh,\, mmh,\, mmmh,\, \ldots\}$, where the n-th element has probability $q^{n-1}p$. Let the random variable $f \colon S \to \mathbf{R}$ map the n-th variable to n. What is the generating function of f? Verify the result of Example 8 that $E(f) = 1/p$, and the result of Exercise 10 that $V(f) = q/p^2$.

24. Let S be a sample space, and $f_1,\, f_2 \colon S \to \mathbf{R}$ two independent random variables. Assume that the images of f_1 and f_2 are contained in the set of non-negative integers, and denote by $g_1,\, g_2$ the generating functions of $f_1,\, f_2$. Let f be the random variable $f_1 + f_2$. Prove that the generating function g of f is given by

$$g(z) = g_1(z)g_2(z).$$

25. Let the sample space S be as suggested in Exercise 17 (so $\#S = 20!$), and denote, for $1 \leq i \leq 20$, by A_i the event that the i-th number on the list is marked.

(a) Prove that the events A_1, A_2, ..., A_{20} are independent in the sense that

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

for every subset I of $\{1, 2, \ldots, 20\}$.

(b) Calculate the variance of the random variable that indicates how many numbers are marked by the procedure described in Exercise 17.

(c) Prove that the generating function of the random variable from (b) is given by

$$\frac{1}{20!} \cdot \prod_{i=0}^{19} (z + i).$$

26. Each box of the cereal *Flakey* contains, inside the package, a small plastic letter. Only the six letters A, E, F, K, L, Y occur, and with each purchase of a box they all have the same probability $\frac{1}{6}$.

(a) Prove that the expectation of the number of boxes of *Flakey* that one needs to buy until one has two distinct letters equals $1 + \frac{6}{5}$.

(b) Prove that the expectation of the number of boxes that one needs to buy until one has all six letters equals $6 \cdot (1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}) = 14.7$.

(c) Calculate the generating function of the random variable from (b).

27. Let the sample space S_n and the random variable $f_n \colon S_n \to \mathbf{R}$ be as in Example 7 and Exercise 18, but with n men instead of 10.

(a) Let $0 \leq k \leq n$. Prove that

$$P(f_n = k) = \frac{1}{k!} \cdot \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}.$$

(b) Denote the generating function of f_n by g_n. Prove that for any two real numbers z, t with $|t| < 1$ one has

$$\sum_{n=0}^{\infty} g_n(z)t^n = \frac{e^{t(z-1)}}{1 - t}.$$