Solutions to Exercise 1

Random linear equations

First we would indeed show that with probability more than $\frac{1}{2}$ there exists no non-zero solution for the random system over $(\mathbb{Z}_2)^n$. Let $\underline{x} = (x_1, \ldots, x_n) \in (\mathbb{Z}_2)^n \setminus \{\underline{0}\}$ be a non-zero vector. For every j, the probability that $\sum_{j=1}^{n} a_{i,j}x_j = 0$ is exactly $\frac{1}{2}$ (it is exactly the probability that $\sum_{x_j=0} a_{i,j}$ is odd). Since there are $n+1$ such equations, all chosen independently, it means that the probability that $\underline{x}$ is a solution to the entire system is exactly 2^{-n-1}. Since there are $2^n - 1$ nonzero vectors in $(\mathbb{Z}_2)^n$, by the union bound this gives a probability of more than $\frac{1}{2}$ that none of them is a solution to this equation system.

Now to show that the above implies what we need, we will show that any non-zero solution over $\mathbb{R}^n$ implies a non-zero solution over $(\mathbb{Z}_2)^n$. If there is a non-zero solution over $\mathbb{R}^n$, then there exists such a solution that involves only rational numbers, because the equation system itself has only rational coefficients (in fact only 0 and 1). Let $\underline{u}$ be such a solution. Then, by multiplying it by a common multiple of all the denominators, we get a vector $\underline{v}$ which is a solution involving only integer numbers (the system is homogeneous so a multiple of a solution is also a solution). Next, let $k \geq 0$ be the maximal integer such that 2^k divides all the coordinates of $\underline{v}$. Then $\underline{w} = 2^{-k}\underline{v}$ is also a solution involving only integers, and additionally it has at least one odd coordinate. Finally, set $\underline{x}$ by letting x_i be the parity of w_i, i.e. 1 if w_i is odd and 0 if w_i is even. It is not very hard to see now that $\underline{x}$ is a non-zero solution of the system over $(\mathbb{Z}_2)^n$.

Games with envelopes

Denote $a_k = \Pr[X = k]$ for every k. Assuming that the opened envelope holds the amount k for an even k, the conditional expectation of the amount in the second envelope is equal to $(2ka_k + \frac{1}{2}ka_{k/2})/(a_k + a_{k/2})$. This is because the conditional probability of the other envelope having $2k$ is $(\frac{1}{2} \cdot a_k)/(\frac{1}{2} \cdot a_k + \frac{1}{2} \cdot a_{k/2}) = a_k/(a_k + a_{k/2})$, while the conditional probability of the other envelope having $\frac{1}{2}k$ is $a_k/(a_k + a_{k/2})$. This means that for an even k, it is desirable to switch only if $(2ka_k + \frac{1}{2}ka_{k/2})/(a_k + a_{k/2}) > k$, i.e. $a_k > \frac{1}{2}a_{k/2}$ (it is still not undesirable to switch also if this inequality is not made strict).

- Let r be such that $a_r > 0$ (there must be such an r if X indeed takes integer values over a probability space). If it is not undesirable to switch for every k that can be found in an envelope, then in particular for every $l > 0$ we have $a_{r2^l} \geq \frac{1}{2}a_{r2^{l-1}}$, and by induction $a_{r2^l} \geq 2^{-l}a_r$. This allows us to bound the expectation from below to contradict the finite expectation assumption: $\mathrm{E}[X] = \sum_{k=1}^{\infty} ka_k \geq \sum_{l=0}^{\infty} r2^l a_{r2^l} \geq \sum_{l=0}^{\infty} ra_r = \infty$.
- Take for example the probability distribution over the integers that yields $a_k = 0$ if k is not a power of 2 and $a_{2^l} = (\frac{2}{3})^{l+1}$ for all $l \geq 0$. This clearly is a probability distribution (the sum of all probabilities is 1) and satisfies $a_k > \frac{1}{2}a_{k/2}$ for every k that has a non-zero probability to appear in any of the envelopes. Note: a variation of this construction can

satisfy the extra condition that all amounts have a non-zero probability to appear in an envelope. Try finding that one for yourselves.

- Take the distribution for which $a_k = 0$ if k is even, and $a_{2l-1} = 2^{-l}$ for all $l \geq 1$. This clearly has finite expectation, but when finding an odd number in an envelope it is always better to switch, because in this case the other envelope would contain twice the amount with conditional probability 1.

Solutions to Exercise 2

Satisfying part of a CNF

First, without loss of generality we can assume that the CNF satisfies all of the following (note that we allow the same clause to appear and be counted more than once all throughout the following).

- No clause is of the form $\neg x_i$. The reason is that if there exist such a clause, then we can replace x_i with its negation all throughout the CNF. This will not create new clauses of the type $\neg x_i$ because the CNF does not contain both clauses $\neg x_i$ and x_i.
- Every clause that contains a positive literal consists of only this positive literal. For clauses containing positive literals that have more than one literal, we just replace them with the lowest index positive literal. Any assignment to the variables that satisfies the new clause will also satisfy the old clause, so by this action we did not increase the maximum number of clauses satisfiable at once.
- Every clause without positive literals consists of exactly two negative literals. By the above we already know that such a clause has to contain at least two literals, and clauses containing more than two can be replaces with clauses containing only the lowest index two negative literals.

When the CNF satisfies all the above, an expectation argument is used. Every variable x_i is independently chosen to be 1 with probability $\frac{1}{2}(\sqrt{5}-1)$, and 0 with probability $\frac{1}{2}(1-\sqrt{5})$. A clause of the type x_j will be satisfied with probability $\frac{1}{2}(\sqrt{5}-1)$, while a clause of the type $\neg x_j \vee \neg x_k$ will be satisfied with probability $1-(\frac{1}{2}(1-\sqrt{5}))^2 = \frac{1}{2}(\sqrt{5}-1)$. By the linearity of expectation the expected total number of satisfied clauses is $\frac{1}{2}(1-\sqrt{5})m$, which means that there exists an assignment as required.

Isolating multisets

This proof generally follows the proof for the isolating lemma of sets, only here we keep for every $a \in A$ not two but $r+1$ values $W_{0,a}, \ldots, W_{r,a}$, where $W_{s,a}$ is defined as the minimum weight among all members of $\mathcal{F}$ that contain a exactly s times.

Now a is called ambiguous if there exists $s < t$ for which $W_{s,a} = W_{t,a}$. Proving that if there are no ambiguous members of A then there is a unique member of $\mathcal{F}$ achieving a minimum is done in much the same way as the proof done in class. Defining $V_{s,a} = W_{s,a} - s \times w(a)$ we note that $V_{s,a}$ is fully determined by the weights of the members of $A \setminus a$. Now $W_{s,a} = W_{t,a}$ if and only if $V_{t,a} - V_{s,a} = (t-s)w(a)$, and (as in class) this happens for at most one value of $w(a)$ and hence with probability at most $\frac{1}{c}$.

By a simple union bound over s, t and a the above would give us already a $1 - \binom{r+1}{2}\frac{n}{c}$ lower bound on the probability that there is a unique minimal weight member of $\mathcal{F}$, but we want

more. For this we look at any fixed weight function for $A \setminus a$, and prove that there are only at most r corresponding possible values for $w(a)$ which would make a ambiguous.

This follows immediately from the following claim: Fixing the weight on all but a, we say that i distinguishes s for a, if for some $t > s$ the value $w(a) = i$ makes $W_{s,a}$ and $W_{t,a}$ both equal and minimum among $W_{1,a}, \ldots, W_{r,a}$. The claim is that for every $s < r$, there is at most one i that distinguishes it for a (and clearly r itself is never distinguished). Assume on the contrary that both i and j distinguish s for a, where $i < j$. If we set x_k to be the value of $W_{k,a}$ where $w(a) = i$, then for $w(a) = j$ we would have $W_{k,a} = x_k + (j - i)k$, and hence for every $t' > s$ we have now $W_{t',a} - W_{s,a} = x_{t'} - x_s + (t' - s)(j - i) > x_{t'} - x_s$. Since $x_{t'} - x_s \geq 0$ (we used here that s was distinguished by i and hence x_s is minimal among the x_k) this gives that there exists no t' that can make j distinguish s for a.

For the counter example, we will show a family of multisets over $A = \{a, b\}$ for which every weight function $w : A \to \{1, 2, 3\}$ will have two sets of equal weights (noting that the formula for simple sets would have given $1 - \frac{2}{3} > 0$). For simplicity we denote by (i, j) the multiset with i copies of a and j copies of b. The family:

$$\mathcal{F} = \{(13, 0), (10, 1), (8, 2), (5, 4), (4, 5), (2, 8), (1, 10), (0, 13)\}$$

This construction is not magical: All possible ratios $w(a)/w(b)$ were ordered in a descending order, and then the multisets were constructed so that the difference between each two consecutive multisets (in the presented order) has the corresponding ratio in the ordered list. Then for every possible w indeed the two consecutive multisets corresponding to $w(a)/w(b)$ are the two multisets with minimal weight.

Solutions to Exercise 3

Fixed points in a permutation

This question is not too hard to solve by an exact calculation of the probability for no fixed points using the inclusion-exclusion principle. Also, there was some confusion in the question because of grammar (too many negations in a sentence). Still, it is instructive to see how the probability for no fixed points can be bounded by using the second moment in a non-standard way.

Let $X = f(\sigma)$ denote the random variable of the number of fixed points. $X = \sum_{i=1}^{n} X_i$, where X_i is the indicator variable for the event that $\sigma(i) = i$. Hence $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i] = n \cdot \frac{1}{n} = 1$. Also, $\mathrm{V}[\mathrm{X_i}] = \frac{1}{n} - \frac{1}{n^2}$ and $\mathrm{Cov}[X_i, X_j] = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}$, and so $\mathrm{V}[X] = n(\frac{1}{n} - \frac{1}{n^2}) + n(n-1)(\frac{1}{n^2(n-1)}) = 1$.

Now we prove that these imply that $\Pr[X = 0] < \frac{19}{20}$. For convenience denote $p_i = \Pr[X = i]$, and assume on the contrary that $p_0 \geq \frac{19}{20}$. This means that $p_1 + 2p_2 \leq \frac{1}{10}$ (because $\sum_{i=0}^{\infty} p_i = 1$), and so using $1 = \mathrm{E}[X] = \sum_{i=0}^{\infty} i p_i$ we get $\sum_{i=3}^{\infty} i p_i \geq \frac{9}{10}$. Finally we write

$$\mathrm{V}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = (\sum_{i=0}^{\infty} i^2 p_i) - 1 \geq (\sum_{i=3}^{\infty} i^2 p_i) - 1 \geq 3(\sum_{i=3}^{\infty} i p_i) - 1 \geq \frac{17}{10} > 1,$$

a contradiction.

A martingale inequality

For the equality, we write by the linearity of expectation:

$$\sum_{i=1}^{n} \mathrm{E}[(X_i - X_{i-1})^2] = \mathrm{E}[X_n^2] - \mathrm{E}[X_0^2] + 2\sum_{i=1}^{n} \mathrm{E}[(X_{i-1} - X_i)X_{i-1}]$$

Now to finish this part we need to prove for every i that $\mathrm{E}[(X_{i-1} - X_i)X_{i-1}] = 0$. This follows from the martingale's lack of memory. Let us do it using a basic method:

$$\begin{aligned}
\mathrm{E}[(X_{i-1} - X_i)X_{i-1}] &= \sum_{\Pr[X_{i-1}=a]>0} \mathrm{E}[(X_{i-1} - X_i)X_{i-1} | X_i - 1 = a]\Pr[X_{i-1} = a] \\
&= \sum_{\Pr[X_{i-1}=a]>0} a(a - \mathrm{E}[X_i | X_i - 1 = a])\Pr[X_{i-1} = a] \\
&= 0
\end{aligned}$$

Now for the inequality. To prove it we shall prove that $\mathrm{E}[X_n^2] \geq \mathrm{E}[X_0^2]$. For this we prove for every a for which $\Pr[X_0 = a] > 0$ that $\mathrm{E}[X_n^2 | X_0 = a] \geq a^2$. We know from arguments we have seen in class that $\mathrm{E}[X_n | X_0 = a] = a$. Now what we need is a direct consequence of e.g. $\mathrm{E}[(X_n - a)^2 | X_0 = a] \geq 0$.

Exposing a permutation

The lemma in class will not work here because a permutation chosen uniformly at random is not expressible as choosing each value $\sigma(i)$ independently of the others. We will directly assume that a permutation $\tilde{\sigma}$ was already chosen, and used for setting the values of the martingale $X_i = \mathrm{E}[c(\sigma)|\sigma(1) = \tilde{\sigma}(1), \ldots, \sigma(i) = \tilde{\sigma}(i)]$.

We note that the probability space for a uniformly random choice of σ conditioned on the values $\sigma(1) = \tilde{\sigma}(1), \ldots, \sigma(i-1) = \tilde{\sigma}(i-1)$ is identical to the following: Take a uniformly random choice of σ conditioned on $\sigma(1) = \tilde{\sigma}(1), \ldots, \sigma(i) = \tilde{\sigma}(i)$, and after that swap $\sigma(i)$ with $\sigma(k)$ where $i \leq k \leq n$ is chosen uniformly among the $n - k + 1$ possible choices.

To complete the proof we will show that a swap of $\sigma(i)$ with any $i \leq k \leq n$ results in a change of $c(\sigma)$ by no more than 1. There are three cases.

- If $k = i$ then there is no change in σ, and so the change in $c(\sigma)$ is 0.
- If in the decomposition to disjoint cycles of σ, the nodes i and k are different but lie in the same cycle, then after the swap this cycle will be split into two cycles, and all other cycles of the decomposition will remain the same. Thus $c(\sigma)$ increases by 1.
- If in the decomposition to disjoint cycles of σ, the nodes i and k lie in different cycles, then after the swap these cycles will be replaced by one cycle spanning the union of their vertices. All other cycles again remain the same, and thus $c(\sigma)$ decreases by 1.

Solutions to Exercise 4

Kleitman for multisets

This will be proved by an application of the FKG theorem over the set $S' = S \times \{1, \ldots, r\}$. Given a multiset C over S with up to r copies of each element, we say that the set $C' \subseteq S'$ represents C if the following occurs: For each $a \in S$, setting $0 \leq i \leq r$ to be the number of its appearances in C, C' contains the elements $(a,1), (a,2), \ldots, (a,i)$, and C' contains no other elements.

Note that any set $D \subseteq S'$ represents a multiset C if and only if every $(a,i) \in D$ implies $(a,j) \in D$ for every $j < i$ and $a \in S$, and that moving to the representing set is a one to one correspondence. Moreover, if A' represents A and B' represents B then $A' \subseteq B'$ if and only if $A \subseteq B$.

Set $\delta : \mathcal{P}(S') \to \mathbb{R}^+$ so that $\delta(D) = 1$ if D represents some multiset C over S, and $\delta(D) = 0$ otherwise. Note that this is a log-super-modular function because if A and B represents multisets over S, then so do $A \cap B$ and $A \cup B$.

Now set $f(C) = 1$ if C contains any $D \subseteq S'$ that represents a member of $\mathcal{A}$ and $f(C) = 0$ otherwise, and set $g(C) = 1$ if C contains any $D \subseteq S'$ that represents a member of $\mathcal{B}$ and $g(C) = 0$ otherwise. These functions are easily shown to be monotone nondecreasing over $\mathcal{P}(S')$. In addition, if $f(C) = 1$ and C represents some multiset then it represents a member of $\mathcal{A}$, because $\mathcal{A}$ is monotone nondecreasing and moving to representations (as commented above) preserves containments. The same holds for g with respect to $\mathcal{B}$.

Now it remains to write and analyse the FKG inequation

$$\left(\sum_{C \subseteq S'} f(C)\delta(C)\right)\left(\sum_{C \subseteq S'} g(C)\delta(C)\right) \leq \left(\sum_{C \subseteq S'} f(C)g(C)\delta(C)\right)\left(\sum_{C \subseteq S'} \delta(C)\right).$$

The following facts finish the proof:

- $\left(\sum_{C \subseteq S'} f(C)\delta(C)\right) = |\mathcal{A}|$, because $\delta(C) = 1$ and $f(C) = 1$ if and only if C represents a multiset (by the definition of δ) which is a member of $\mathcal{A}$ (by the discussion following the definition of f).
- Similarly, $\left(\sum_{C \subseteq S'} g(C)\delta(C)\right) = |\mathcal{B}|$.
- $\left(\sum_{C \subseteq S'} f(C)g(C)\delta(C)\right) = |\mathcal{A} \cap \mathcal{B}|$ by what we know of f, g and δ.
- Finally, by the definition of δ and the number of possible multisets over S we have $\left(\sum_{C \subseteq S'} \delta(C)\right) = (r+1)^{|S|}$.

Walking on hypercubes

We define a joint distribution over $\underline{X} = X_0, X_1, \ldots$ and $\underline{Y} = Y_0, Y_1, \ldots$. We start by setting $X_0 = (0, \ldots, 0)$ with probability 1 while Y_0 is drawn uniformly from $\{0,1\}^n$. Now given $X_k = (a_1, \ldots, a_n)$ and $Y_k = (b_1, \ldots, b_n)$, we choose values for $X_{k+1} = (c_1, \ldots, c_n)$ and $Y_{k+1} = (d_1, \ldots, d_n)$ as per the following:

First we uniformly choose $1 \leq i \leq n$. For all $j \neq i$ we set $c_j = a_j$ and $d_j = b_j$. Now if $a_i = b_i$, then with probability $\frac{1}{2}$ we set $c_i = d_i = a_i$ and with probability $\frac{1}{2}$ we set $c_i = d_i = 1 - a_i$. If $a_i \neq b_i$, then with probability $\frac{1}{2}$ we set $c_i = d_i = a_i$ and with probability $\frac{1}{2}$ we set $c_i = d_i = b_i$.

Now conditioned on $X_0, \ldots, X_k$ and $Y_0, \ldots, Y_k$ (but not on Y_{k+1}), the distribution of X_{k+1} depends only on the value of X_k and is identical to a single step taken according to the distribution defined in the question. This means that $\underline{X}$ is indeed distributed like the random sequence of the question. Also, conditioned on $X_0, \ldots, X_k$ and $Y_0, \ldots, Y_k$ (but not on Y_{k+1}), the distribution of Y_{k+1} depends only on the value of Y_k and has the same transition probabilities as those of $\underline{X}$. Therefor the unconditional distribution of each Y_k is the uniform one, because it is the stationary distribution for this transition matrix.

Now given k we define c such that $k = n(\ln n + c)$, and show that if c is large enough then the variation distance between X_k and Y_k is smaller than ϵ (which is what we need). For $1 \leq i \leq n$ let E_i denote the event that i was chosen at least once in the transitions described above leading from (X_0, Y_0) to (X_k, Y_k). Now the probability of E_i not to occur is at most $(1 - \frac{1}{n})^k < e^{-k/n} = e^{-c}/n$. For c large enough this is smaller than ϵ/n. Now setting E as the conjunction of all events E_i, the probability of E not to occur is bounded (by the union bound) by ϵ. Finally, whenever E occurs we have $X_k = Y_k$, because once an i is chosen in step s, the i'th coordinate of X_t and Y_t will remain identical to each other for all $t > s$. This allows us to use what was shown in Assignment 0 to bound the distance between the two distributions by ϵ.