TOPICS IN PROBABILITY. PART I: CONCENTRATION

Exercise sheet 5: Entropy and probabilistic method

1. Entropy and Variance

Exercise 1 (Warm-up: Uniform distribution as a maximal entropy distribution). *Consider a discrete probability space $(S, \mathcal{P}(S), \mathbb{P})$ where S is finite. The entropy of the distribution is given by $-\sum_{s\in S}\mathbb{P}[\{s\}]\log\mathbb{P}[\{s\}]$. Prove that among all probability distributions on $\{1,\dots,n\}$ the uniform distribution has maximal entropy.*

Exercise 2 (Variational characterization of variance). *Let X be a square-integrable random variable on some probability space. Prove that*

$$\mathrm{Var}[X] = \sup_Y (2\mathrm{Cov}(X,Y) - \mathrm{Var}[Y]),$$

where supremum is taken over all square-integrable random variables Y on the same probability space.

Exercise 3 (Entropy bounds variance). *Show that for any non-negative random variable Z, it holds*

$$\mathrm{Var}[Z] \le \mathrm{Ent}[Z^2].$$

Find a counterexample, which shows that the above inequality is not necessarily true if Z is not required to be non-negative.

Hint for the first part: for $p \in [1,2)$, consider $\Psi_p(Z) = \mathbb{E}[Z^2] - (\mathbb{E}[Z^p])^{2/p}$. Show that $\lim_{p\uparrow 2}\frac{\Psi_p(Z)}{2-p} = \mathrm{Ent}[Z^2]/2$ and that $p \mapsto \frac{\Psi_p(Z)}{1/p-1/2}$ is non-decreasing [1].

2. Probabilistic method: Random objects used to prove deterministic statements

Exercise 4 (Warm-up: Orthogonal projection in expectation). *Recall the $m \times N$-matrix P from the proof of Johnson-Lindenstrauss lemma, which has i.i.d. standard Gaussian entries. Extend this matrix to an $N \times N$-matrix $\tilde{P}$ by filling the new rows with zeroes. Show that $\mathbb{E}[\tilde{P}^2] = I_m$, where I_m is a diagonal matrix with first m values on the diagonal being 1's and 0 otherwise. Note that I_m is an orthogonal projection from $\mathbb{R}^N$ onto m-dimensional subspace $\mathbb{R}^m$ of $\mathbb{R}^N$.*

Remark: for a different insight on P being close to a projection read discussion in Section 1.1 in the following notes.

Exercise 5 (Erdös theorem, 1959). *Given $k \ge 3, g \ge 3$, there exists a graph with girth at least g and chromatic number $\chi(G)$ at least k. Recall that the girth of a graph is the length of its shortest circle; and the chromatic number of graph G is the least amount of colors necessary to color the vertices such that no two adjacent vertices share the same color.*
To prove this theorem proceed as follows:

[1] Extra hint: show for example that $\alpha(t) = t\log\mathbb{E}[Z^{1/t}]$ for $t \in (1/2, 1]$ is convex and rewrite the above function in terms of α.

- *Instead of working with $\chi(G)$ directly (as it is difficult), consider independence number $\alpha(G)$, which is the size of the largest set of mutually non-adjacent vertices. Show that $\chi(G) \geq \frac{|V(G)|}{\alpha(G)}$.*
- *Let n be large and consider an Erdös-Renyi graph on these vertices with probability p of edge being present. Choose p small enough such that the expected number of short cycles (of length less than g) is small, but large enough such that the existence of independent sets is unlikely. More precisely, adjust p such that with high probability there are at most $n/2$ cycles of length less or equal than g and G contains no independent sets of size greater or equal to $n/(2k)$.*[2]
- *Remove a vertex in each cycle of the graph from Step 2. Argue that the resulting graph with high probability is the desired one.*

$\star$ **For those who know or would like to learn something about packings and coverings.**

Exercise 6 (Tightness of Johnson-Lindenstrauss)**.** *Johnson-Lindenstrauss lemma proven in the lecture shows that for any k points in a Hilbert space H (or $\mathbb{R}^N$ for simplicity) can be mapped into $\mathbb{R}^m$ with $m \gtrsim \log k$ while distorting the distances between them by at most a constant factor. Show that $m \gtrsim \log k$ is a necessary condition. Assume that the mapping considered in the statement of the lemma are exclusively linear maps.*
Hint: show that the image of k orthonormal vectors $x_1, \ldots, x_n$ in H (or $\mathbb{R}^N$ as in the lecture) under a map $T : H \to \mathbb{R}^m$ that nearly preserves distances (corresponds to $\frac{P}{\sqrt{m}}$ in the lecture notes) is a packing of a ball in $\mathbb{R}^m$.

Literature on packings and coverings: see Section 5.2 in "Probability in High Dimension", Ramon van Handel or Section 4.2 in "High-Dimensional Probability", Roman Vershynin.

[2]For example, $p = n^{-1+1/(2g)}$, Markov inequality might be useful.