MathTools HW 2

As in last week, do not be alarmed by the length of this exercise sheet; most of it is stories/context/hints.

1. **Stirling's approximation.** The following is an *extremely classical* estimate for the factorial:

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n\left(1 + O\left(\frac{1}{n}\right)\right).$$

It is illustrative to take the logarithm of this quantity[1]:

$$\log(n!) = n\log(n/e) + \frac{1}{2}\log(2\pi n) + o(1).$$

Prove the somewhat less precise estimate $|\log(n!) - n\log(n/e)| = O(\log n)$.

Guidance: Use the bounds

$$\int_{k-1}^{k} \log(x)dx \le \log k \le \int_{k}^{k+1} \log(x)dx,$$

whose correctness should be obvious to you.

2. **Binomial coefficients.** Recall that for $0 \le k \le n$ integers, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. It counts the number of distinct subsets $S \subset [n]$ of size exactly $|S| = k$.

   (a) Prove that for $0 \le k \le k+1 \le n/2$, $\binom{n}{k} < \binom{n}{k+1}$. That is, the mapping $k \mapsto \binom{n}{k}$ is increasing on $0 \le k \le n/2$.

   (b) Prove the upper and lower bounds (assuming $1 \le k \le n$):

$$\left(\frac{n}{k}\right)^k \le \binom{n}{k} \le \left(\frac{en}{k}\right)^k,$$

   Hint: for the upper bound, you might need to show that $k^k/k! \le e^k$; recall the Taylor expansion of e^x (an *extremely* useful thing to remember in itself!).

   (c) Fix a number $0 < \delta < 1/2$. Use Stirling's approximation to show that

$$\binom{n}{\lfloor \delta n \rfloor} = e^{n(h(\delta)+o(1))} \quad \text{as } n \to \infty,$$

   where $h(x) = x\log\frac{1}{x} + (1-x)\log\frac{1}{1-x}$ is the *binary entropy* function.[2]

[1] log denotes the logarithm with respect to the natural basis e.

[2] Recall that $\lfloor x \rfloor$, the *floor* of x, is the largest integer $\le x$. E.g, $\lfloor 1.76 \rfloor = 1$.

(d) As in (c), show that

$$\sum_{k=0}^{\lfloor \delta n \rfloor} \binom{n}{k} = e^{n(h(\delta)+o(1))} \quad \text{as } n \to \infty\,.$$

Remark: The estimates in (c)-(d) are important (and elementary) in the theory of error correcting codes. Perhaps we will say something more about this later on (but no promises).

3. **Poisson limit theorem.** In TA 1 we studied the behavior of a binomial random variables $S_n \sim \mathrm{Binom}(n, \alpha_n)$, where we let the parameter α_n vary with n. We've seen that: (i) When $\alpha_n n \to \infty$, then with overwhelming probability $S_n = \alpha_n n(1 + o(1))$ (see TA notes for precise statement); (ii) When $\alpha_n n \to 0$, we have $\Pr(S_n = 0) \to 1$.

   In this exercise we consider the regime $\alpha_n n = \Theta(1)$, specifically, set $\alpha_n = \lambda/n$ for a fixed $\lambda > 0$. A random variable Y has a Poisson distribution with intensity λ, denoted $Y \sim \mathrm{Pois}(\lambda)$, if its gets values in $0, 1, 2 \ldots$ (non-negative integers) with

   $$\Pr(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{for all } k = 0, 1, \ldots\,.$$

   Make sure you see that these probabilities indeed sum to 1!

   (a) Let $Y \sim \mathrm{Pois}(\lambda)$. Show that $\mathbb{E}[Y] = \lambda$.

   (b) As in (a), show that $\mathrm{Var}(Y) = \lambda$.

   (c) Prove the Poisson limit theorem. Namely, if $S_n \sim \mathrm{Binom}\left(n, \frac{\lambda}{n}\right)$, then for every **fixed** k,

   $$\lim_{n \to \infty} \Pr\left(S_n = k\right) = \frac{\lambda^k}{k!} e^{-\lambda}\,.$$

4. K_4 **count in a random graph.** In class, you've shown that $f(n) = n^{-2/3}$ is the threshold for the appearance of K_4 in a random graph. In particular, when $p = \omega(n^{-2/3})$, G contains a copy of K_4 with probability tending to 1.

   For what follows, assume that $p = \omega(n^{-2/3})$. Let X be the number of K_4-s in G. Show that w.h.p, $X = (1 \pm o(1))\mathbb{E}[X]$. In other words, show that one can choose some $\epsilon_n = o(1)$ such that

   $$\lim_{n \to \infty} \Pr\left(|X - \mathbb{E}[X]| \geq \epsilon_n \mathbb{E}[X]\right) = 0\,.$$

5. Consider the following graph property $\mathcal{P}$: every edge of G is contained in a triangle.

   Prove the following. There exists a constant $\alpha > 0$ such that the following holds: if $p = \alpha\sqrt{\frac{\log n}{n}}$, then $G \sim \mathcal{G}(n, p)$ satisfies the property with high probability.

   Hint: Perhaps the following, more verbose, description of this property can help: for every indices $u, v \in [n]$, **if** the edge $\{u, v\}$ belongs to G, **then** there is some vertex $w \notin \{u, v\}$ such that the edges $\{u, w\}$ and $\{v, w\}$ are also contained in G.

6. **Appearance of a subgraph in $\mathcal{G}(n,p)$.** Let H be a *fixed* graph, with v vertices and e edges. The **density** of H is the ratio $d = e/v$.

   (a) Prove that if $p = o(n^{-1/d})$ then w.h.p, $G \sim \mathcal{G}(n,p)$ doesn't contain a copy of H.

   Example: $H = K_4$, so $v = 4$ and $e = 6$. Here $d = 3/2$, and indeed you proved in class that $n^{-1/d} = n^{-2/3}$ is the threshold for the appearance of K_4 in G.

   (b) Find a counter example showing that in general, $f(n) = n^{-1/d}$ doesn't have to be a threshold for the appearance of H in G. More explicitly, construct a graph H and sequence $p = p_n$ such that $p = \omega(f(n))$, so that

   $$\liminf_{n\to\infty} \Pr\left(G \sim \mathcal{G}(n,p) \text{ contains a copy of } H\ \right) < 1\,.$$

   Hint: Build H starting from K_4, adding more vertices and edges.

   Remark: We say that G contains a copy of H iff there are v vertices of G, call them $i_1,\ldots,i_v \in [n]$, and a mapping $\phi : \{i_1,\ldots,i_v\} \mapsto [v]$, such that $\{i_k, i_\ell\}$ is an edge in G (for all $k,\ell \in [v]$) if and only if $\{\phi(i_k), \phi(i_\ell)\}$ is an edge in H.

7. **Monotone graph properties.** Let $\mathcal{P}$ be a monotone graph property, and let $p_1 \le p_2$. Prove that

   $$\Pr\left(G \sim \mathcal{G}(n,p_1) \text{ satisfies } \mathcal{P}\right) \le \Pr\left(G \sim \mathcal{G}(n,p_2) \text{ satisfies } \mathcal{P}\right)\,.$$

   Hint: Here is a way to sample a graph from $\mathcal{G}(n,p_2)$: sample two graphs $G_1 \sim \mathcal{G}(n,p_1)$ and $G_2 \sim \mathcal{G}(n,q)$, for some appropriate choice of q. Construct the graph G by including every edge that appears in either G_1 or G_2 - in other words, set $G = G_1 \cup G_2$). The graphs G_1, G_2, G will live in the same probability space, so that G_1 and G are dependent, and have the "correct" marginal distributions.

   In probability, this type of argument is sometimes called a **coupling** (in this case, between the distributions $\mathcal{G}(n,p_1)$ and $\mathcal{G}(n,p_2)$).