Discrete Probability

1. **Branching processes.** Let $(Z_n)_{n\in\mathbb{N}_0}$ be a branching process, where offspring is distributed according to an $\mathbb{N}_0$-valued random variable X. Assume $E(X) < \infty$.

   (a) Show $E(Z_n) = E(X)^n$.

   (b) If $E(X) < 1$ then $P(Z_n > 0) \le E(X)^n$.

   (c) If $E(X) < 1$ then the *total progreny* T satisfies $E(T) = \frac{1}{1-E(X)}$.

   (d) The process $Z_n/E(X)^n$ is a martingale.

2. **CLT for number of edges.** Prove that the number of edges in $G(n, \lambda/n)$ satisfies a central limit theorem with asymptotic mean and variance equal to $\lambda n/2$.

3. **Connectivity with given cluster size.** Show

$$P_\lambda\big(1 \not\longleftrightarrow 2 \,\big|\, |\mathcal{C}(1)| = \ell\big) = 1 - \frac{\ell - 1}{n-1}, \qquad 1 \le \ell \le n.$$

4. **Upper bound for mean cluster size.** Show that for $v \in [n]$,

$$E_\lambda(|\mathcal{C}(v)|) = 1 + (n-1)P_\lambda(1 \longleftrightarrow 2).$$

   Furthermore, if $\lambda < 1$,

$$E_\lambda(|\mathcal{C}(v)|) \le \frac{1}{1-\lambda}.$$

5. **Large deviations.** The following large deviation bound is fairly standard in probability theory (see Thm. 2.19 in [Hof] or elsewhere):

   Theorem. (Cramér's upper bound, Chernoff bound)
   Let $(X_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of random variables. Then for all $a \ge E(X_1)$,

$$P\Big(\sum_{i=1}^n X_i \ge na\Big) \le e^{-nI(a)} \qquad \text{if } a \ge E(X_1) \text{ and}$$

$$P\Big(\sum_{i=1}^n X_i \le na\Big) \le e^{-nI(a)} \qquad \text{if } a \le E(X_1)$$

   with "rate function"

$$I(a) = \sup_{t\in\mathbb{R}} \Big(ta - \log E(e^{tX_1})\Big).$$

Compute $I(a) = I_\lambda(a)$ for $(X_i)_{i\in\mathbb{N}}$ being independent Poisson random variables with parameter λ. Prove that $I_\lambda(a) > 0$ for $a \neq \lambda$.

Furthermore, if $(X_i)_{i\in\mathbb{N}}$ are independent Bernoulli(p) random variables, verify that

$$I(a) = a\log\left(\frac{a}{p}\right) + (1-a)\log\left(\frac{1-a}{1-p}\right)$$

whenever $a \in (p, 1]$ (where we adapt the convention that $x \log x = 0$ for $x = 0$). Prove that, in this case, $I(a) \geq I_\lambda(a)$ from the previous example (with $\lambda = p$). This suggests that the upper tail of a binomial random variable is "thinner" than the one of a Poisson random variable.

6. **Mean number of triangles.** We say that the distince vertices (i, j, k) form an occupied triange when the edges ij, jk, and ki are all occupied. Note that (i, j, k) is the same triangle as (i, k, j) and as any other permutation. Compute the expected number of occupied triangles in $G(n, \lambda/n)$.

7. **Clustering coefficient.** The clustering coefficient CC_G of a random graph G is defined as

$$\mathrm{CC}_G = \frac{E(\Delta_G)}{E(W_G)}$$

where

$$\Delta_G = \sum_{i,j,k\in V} \mathbb{1}\{ij, jk, ki \text{ occupied}\}, \qquad W_G = \sum_{i,j,k\in V} \mathbb{1}\{ij, jk \text{ occupied}\}.$$

Thus (since we are not restricting to $i < j < k$ in Δ_G and $i < j$ in W_G), Δ_G is six times the number of triangles in G, and W_G is two times the number of wedges in G, and CC_G is the ratio of the number of expected closed triangles to the expected number of wedges. Compute CC_G for $G(n, \lambda/n)$.