Advanced Algorithm: Assignment #2

Due on Oct 10, 2018 at 12:00 pm

Professor Yitong Yin

Hao Li DZ1833013

Problem 1

Starting with a graph with n vertices and no edge, we consider the following process to build a random undirected simple graph. At each step, we choose an edge uniformly at random from the set of all remaining unchosen edges, and add it to the graph. I.e., in the first step, we choose an edge from the set of $\binom{n}{2}$ edges; in the second step, we choose from the remaining $\binom{n}{2-1}$ edges; and so on. Let X denote number of edges added until the graph is connected, give an upper bound for $E[X]$.

Solution
Let Y_k denotes the number of edges added while there are k connected components, to the time while there are $k-1$ connected components. Then clearly $X = \sum_{k=2}^{n} Y_k$.
When there are exactly k connected components, add a edge, let p_k denotes the probability that the number of connected components decreases. We have:

$$p_k \geq \frac{k-1}{n-1} \tag{1}$$

Equation 1 is due to: when there are k connected components, assume e is the edge that we are adding, the edge has 2 endpoint. The k connected components turns to be $k-1$ only if the 2 endpoint are not in the same connected component. Fix an endpoint, then there are $n-1$ vertices can be chosen and $k-1$ connected components. Then we can get the lower bound.
Next, we get:

$$E[Y_k] = \frac{1}{p_k} \leq \frac{n-1}{k-1} \tag{2}$$

Applying the linearity of expectations,

$$\begin{aligned}
E[X] &= E[\sum_{k=2}^{n} Y_k] \\
&= \sum_{k=2}^{n} E[Y_k] \\
&\leq \sum_{k=2}^{n} \frac{n-1}{k-1} \\
&= (n-1)\sum_{k=2}^{n} \frac{1}{k-1} \\
&\leq (n-1)(1 + ln(n-1)) = O(nlog\ n)
\end{aligned}$$

Problem 2

In Balls-and-Bins model, we throw m balls independently and uniformly at random into n bins. We know that the maximum load is $\Theta\left(\frac{\log n}{\log\log n}\right)$ with high probability when $m = \Theta(n)$. Then two-choice paradigm is another way to throw m balls into n bins: each ball is thrown into the least loaded of two bins chosen independently and uniformly at random(it could be the case that the two chosen bins are exactly the same, and the ball will be thrown into that bin), and break the tie arbitrarily. When $m = \Theta(n)$, the maximum load of two-choice paradigm is known to be $\Theta(\log\log n)$ with high probability, which is exponentially less than the maximum load when there is only one random choice. This phenomenon is called the *the power of two choices.*

Here are the questions:

- **Consider the following paradigm: we throw n balls into n bins. The first $\frac{n}{2}$ balls are thrown into bin independently and uniformly at random. The remaining $\frac{n}{2}$ balls are thrown into bins using the two-choice paradigm. What is the maximum load with high probability? You need to give an asymptotically tight bound (in the form of $\Theta()$).**
- **Replace the above paradigm to the following: the first $\frac{n}{2}$ balls are thrown into bins using the two-choice paradigm while the remaining $\frac{n}{2}$ balls are thrown into bins independently and uniformly at random. What is the maximum load with high probability in this case? You need to give an asymptotically tight bound.**
- **Replace the above paradigm to the following: assume all n balls are thrown in a sequence. For every $1 \leq i \leq n$, if i is odd, we throw $i-th$ balls into bins independently and uniformly at random, otherwise, we throw it into bins using the two-choice paradigm. What is the maximum load with high probability in this case? You need to give an asymptotically tight bound.**

Solution

Maybe there are some possible errors in the question stem. I mean "We know that the maximum load is $\Theta\left(\frac{\log n}{\log\log n}\right)$ with high probability when $m = \Theta(n)$." in the question item is "We know that the maximum load is $O\left(\frac{\log n}{\log\log n}\right)$ with high probability when $m = \Theta(n)$". And "When $m = \Theta(n)$, the maximum load of two-choice paradigm is known to be $\Theta(\log\log n)$ with high probability" should be "When $m = \underline{(}n)$, the maximum load of two-choice paradigm is known to be $O(\log\log n)$ with high probability".

Part 1

For the first half of the balls, that is $\frac{n}{2}$ balls. If we have thrown this half balls using the one-choice strategies, let the max load at this time is T, then we have:

$$Pr[T > t] \leq \binom{m}{t}(\frac{1}{n})^t \tag{3}$$

$$\leq (\frac{em}{tn})^t \tag{4}$$

$$= (\frac{e}{2t})^t \tag{5}$$

Let $t = \frac{2\log n}{\log\log n}$, then we obtain:

$$Pr[T > t] = (\frac{e}{2t})^t \tag{6}$$

$$\leq \frac{1}{n^2} \tag{7}$$

Assume all bins are $\frac{2\log n}{\log\log n}$ balls, then we throw the remaining half part. So we have:

$$Pr[T > (\frac{2\log n}{\log\log n} + \log\log n)] < \frac{1}{n^2} + \frac{1}{n^2} \tag{8}$$

$$= \frac{2}{n^2} \tag{9}$$

And there are not all bins have $\frac{2\log n}{\log\log n}$ balls, so T_{res} must be less than this result. So:

$$Pr[T_{res} > (\frac{2\log n}{\log\log n} + \ln\ln n)] < Pr[T > (\frac{2\log n}{\log\log n} + \log\log n)] \tag{10}$$

$$< \frac{2}{n^2} \tag{11}$$

We can always have a means to make $\frac{2}{n^2}$ to $\frac{1}{n^2}$. So we can get the upper bound.
If the question item is "Θ", the lower bound is also $o(\frac{\log n}{\log\log n})$. For the first part, we have:

$$Pr[T < o(\frac{\log n}{\log\log n})] < \frac{1}{n} \tag{12}$$

so we can easily get the lower bound $o(\frac{\log n}{\log\log n})$. Then the bound is tight.

Part 2

This part is much easier than the first part.
Same as the first part, we take two step to consider this problem. Then the answer is $\Theta(\log\log n) + \Theta(\frac{\log n}{\log\log n}) = \Theta(\frac{\log n}{\log\log n})$.

Part 3

First give the lower bound. Consider if i is even, do not throw balls. So the lower bound is the lower bound of throw $\frac{n}{2}$ balls into n bins, what is $o(\frac{\log n}{\log\log n})$. And when using two-choice if i is even, the bound cannot be better..
Consider a sequence number $A = \{2, 4, \dots, 2k\}$, where $k = \lfloor \frac{n}{2} \rfloor + 1$. When $i = a_j \in A$. If we use one choice strategy, the probability that the max load of all bins will increase is p_j. If we use two-choice strategy, the probability that the max load of all bins will increase is q_j. Then if we prove that:

$$\forall a_j \in A,\ p_j \geq q_j \tag{13}$$

then we can prove the upper bound is $o(\frac{\log n}{\log\log n})$.
For the one choice strategy, the max load increase only if we throw a ball into a max load. So the number of max load bins is np_i. So:

$$q_i = \frac{\binom{np_i}{2}}{\binom{n}{2}} \tag{14}$$

$$= \frac{np_i(np_i - 1)}{n(n-1)} < p_i \tag{15}$$

So $\Theta(\frac{\log n}{\log\log n})$ is tight.

Problem 3

Let X be a real-valued random variable with finite $\mathbb{E}[X]$ and finite $\mathbb{E}\left[\mathrm{e}^{\lambda X}\right]$ for all $\lambda \geq 0$. We define the log-moment-generating function as
$\Psi_X(\lambda) := \ln \mathbb{E}[\mathrm{e}^{\lambda X}]$ **for all $\lambda \geq 0$, and its dual function:**
$\Psi_X^*(t) := \sup\limits_{\lambda \geq 0}(\lambda t - \Psi_X(\lambda))$.
Assume that X is NOT almost surely constant. Then due to the convexity of $\mathrm{e}^{\lambda X}$ with respect to λ, the function $\Psi_X(\lambda)$ is strictly convex over $\lambda \geq 0$.

- **Prove the following Chernoff bound:**
  $\Pr[X \geq t] \leq \exp(-\Psi_X^*(t))$. **In particular if $\Psi_X(\lambda)$ is continuously differentiable, prove that the supreme in $\Psi_X^*(t)$ is achieved at the unique $\lambda \geq 0$ satisfying $\Psi'_X(\lambda) = t$**
  where $\Psi'_X(\lambda)$ denotes the derivative of $\Psi_X(\lambda)$ with respect to λ.
- **Normal random variables. Let $X \sim \mathrm{N}(\mu, \sigma)$ be a Gaussian random variable with mean μ and standard deviation σ. What are the $\Psi_X(\lambda)$ and $\Psi_X^*(t)$? And give a tail inequality to upper bound the probability $\Pr[X \geq t]$.**
- **Poisson random variables. Let $X \sim \mathrm{Pois}(\nu)$ be a Poisson random variable with parameter ν, that is, $\Pr[X = k] = \mathrm{e}^{-\nu}\nu^k/k!$ for all $k = 0, 1, 2, \ldots$. What are the $\Psi_X(\lambda)$ and $\Psi_X^*(t)$? And give a tail inequality to upper bound the probability $\Pr[X \geq t]$.**
- **Bernoulli random variables. Let $X \in \{0, 1\}$ be a single Bernoulli trial with probability of success p, that is, $\Pr[X = 1] = 1 - \Pr[X = 0] = p$. Show that for any $t \in (p, 1)$, we have $\Psi_X^*(t) = D(Y \| X)$ where $Y \in \{0, 1\}$ is a Bernoulli random variable with parameter t and $D(Y\|X) = (1-t)\ln\frac{1-t}{1-p} + t\ln\frac{t}{p}$ is the Kullback-Leibler divergence between Y and X.**
- **Sum of independent random variables. Let $X = \sum\limits_{i=1}^{n} X_i$ be the sum of n independently and identically distributed random variables $X_1, X_2, \ldots, X_n$. Show that $\Psi_X(\lambda) = \sum\limits_{i=1}^{n} \Psi_{X_i}(\lambda)$ and $\Psi_X^*(t) = n\Psi_{X_i}^*(\frac{t}{n})$. Also for binomial random variable $X \sim \mathrm{Bin}(n, p)$, give an upper bound to the tail inequality $\Pr[X \geq t]$ in terms of KL-divergence.**
  Give an upper bound to $\Pr[X \geq t]$ when every X_i follows the geometric distribution with a probability p of success.

Solution

Part 1

Part 1.1

We can get $Pr[X \geq t] \leq e^{-\Psi^*_X(t)}$ using these equations:

$$
\begin{aligned}
Pr[X \geq t] &= Pr[e^{\lambda X} \geq e^{\lambda t}],\ for\ all\ \lambda \geq 0 &(16)\\
&\leq \frac{E[e^{\lambda X}]}{e^{\lambda t}} &(17)\\
&= e^{\ln(\frac{E[e^{\lambda X}]}{e^{\lambda t}})} &(18)\\
&= e^{-(\lambda t - \Psi_X(\lambda))} &(19)
\end{aligned}
$$

Problem 3 continued on next page. . .

Equation (17) makes sense due to generalized Markov's inequality.
Then for all $\lambda \geq 0$, $Pr[X \geq t] \leq e^{-(\lambda t - \Psi_X(\lambda))}$. So:

$$\begin{aligned} Pr[X \geq t] &\leq \min e^{-(\lambda t - \Psi_X(\lambda))} \\ &= e^{-\sup_{\lambda \geq 0}(\lambda t - \Psi_X(\lambda))} \\ &= e^{-\Psi_X^*(t)} \end{aligned}$$

Then we proof the Chernoff bound.

Part 1.2

Define a function $f(\lambda) = \lambda t - \Psi_X(\lambda)$, the function reaches its extreme value when $f'(\lambda) = 0$.
That is:

$$\begin{aligned} f'(\lambda) &= t - \Psi_X^{'}(\lambda) \\ &= 0 \end{aligned}$$

Then:

$$\Psi_X'(\lambda) = t \tag{20}$$

That's the proof.

Part 2: Normal random variables.

We know:

$$X \sim \mathrm{N}(\mu, \sigma)$$

that is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

where $f(x)$ is the probability distribution of X.
So we have these equations:

$$\begin{aligned} E[e^{\lambda X}] &= \int_{-\infty}^{+\infty} e^{\lambda x} f(x) dx \qquad (21) \\ &= \int_{-\infty}^{+\infty} e^{\lambda x} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx \qquad (22) \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp(\lambda x - \frac{(x-\mu)^2}{2\sigma^2}) dx \qquad (23) \end{aligned}$$

Let $F(x) = \lambda x - \frac{(x-\mu)^2}{2\sigma^2}$ in equation (23), then we have:

$$\begin{aligned} F(x) &= -[\frac{(x-\mu)^2}{2\sigma^2} - \lambda x] \qquad (24) \\ &= -\frac{(x-\mu)^2 - 2\sigma^2\lambda x}{2\sigma^2} \qquad (25) \\ &= -\frac{x^2 - 2\mu x + \mu^2 - 2\sigma^2\lambda x}{2\sigma^2} \qquad (26) \\ &= -\frac{(x-(\mu+\sigma^2\lambda))^2 - (\mu+\sigma^2\lambda)^2 + \mu^2}{2\sigma^2} \qquad (27) \\ &= -[\frac{(x-(\mu+\sigma^2\lambda))^2}{2\sigma^2} + \frac{-(\mu+\sigma^2\lambda)^2 + \mu^2}{2\sigma^2}] \qquad (28) \end{aligned}$$

Then we obtain:

$$\exp(F(x)) = -\exp[\frac{(x-(\mu+\sigma^2\lambda))^2}{2\sigma^2} + \frac{-(\mu+\sigma^2\lambda)^2+\mu^2}{2\sigma^2}] \tag{29}$$

$$= -\exp[\frac{(x-(\mu+\sigma^2\lambda))^2}{2\sigma^2}]\exp[\frac{-(\mu+\sigma^2\lambda)^2+\mu^2}{2\sigma^2}] \tag{30}$$

Return to equation (23), that is:

$$E[e^{\lambda X}] = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp(\lambda x - \frac{(x-\mu)^2}{2\sigma^2})dx \tag{31}$$

$$= \int_{-\infty}^{+\infty} -\frac{1}{\sqrt{2\pi}\sigma}\exp[\frac{(x-(\mu+\sigma^2\lambda))^2}{2\sigma^2}]\exp[\frac{-(\mu+\sigma^2\lambda)^2+\mu^2}{2\sigma^2}]dx \tag{32}$$

$$= \exp[\frac{-(\mu+\sigma^2\lambda)^2+\mu^2}{2\sigma^2}]\int_{-\infty}^{+\infty} -\frac{1}{\sqrt{2\pi}\sigma}\exp[\frac{(x-(\mu+\sigma^2\lambda))^2}{2\sigma^2}]dx \tag{33}$$

In the equation (33), we have this equation:

$$\int_{-\infty}^{+\infty} -\frac{1}{\sqrt{2\pi}\sigma}\exp[\frac{(x-(\mu+\sigma^2\lambda))^2}{2\sigma^2}]dx = 1 \tag{34}$$

for the reason that if

$$X \sim \mathrm{N}(\mu+\sigma^2\lambda, \sigma) \tag{35}$$

equation (34) is the probability integration of X.
Then we get:

$$E[e^{\lambda X}] = \exp[\frac{-(\mu+\sigma^2\lambda)^2+\mu^2}{2\sigma^2}] \tag{36}$$

Then:

$$\Psi_X(\lambda) = \ln(E[e^{\lambda X}]) \tag{37}$$

$$= \frac{-(\mu+\sigma^2\lambda)^2+\mu^2}{2\sigma^2} \tag{38}$$

In part 1.2, we know the supreme value obtained when $\Psi'_X(\lambda) = t$. What is:

$$\Psi'_X(\lambda) = \frac{-2(\mu+\sigma^2\lambda)\sigma^2}{2\sigma^2} \tag{39}$$

$$= -(\mu+\sigma^2\lambda) \tag{40}$$

$$= t \tag{41}$$

That is:

$$\lambda = -\frac{t+\mu}{\sigma^2} \tag{42}$$

Next, return to the dual function:

$$\Psi^*_X(t) := \sup_{\lambda\geq 0}(\lambda t - \Psi_X(\lambda)). \tag{43}$$

In short,

$$\Psi_X(\lambda) = \frac{-(\mu+\sigma^2\lambda)^2+\mu^2}{2\sigma^2} \tag{44}$$

$$= -\mu\lambda - \frac{\sigma^2\lambda^2}{2} \tag{45}$$

Then:

$$\lambda t - \Psi_X(\lambda) = \lambda t + \mu\lambda + \frac{\sigma^2\lambda^2}{2} \tag{46}$$

Put $\lambda = -\frac{t+\mu}{\sigma^2}$ in, then:

$$\begin{aligned}\Psi_X^*(t) &= (-\frac{t+\mu}{\sigma^2})(t+\mu) + \frac{\sigma^2}{2}(-\frac{t+\mu}{\sigma^2})^2 &(47)\\ &= -\frac{(t+\mu)^2}{\sigma^2} + \frac{(t+\mu)^2}{2\sigma^2} &(48)\\ &= -\frac{(t+\mu)^2}{2\sigma^2} &(49)\end{aligned}$$

Finally, we obtain:

$$\begin{aligned}Pr[X \geq t] &\leq \exp(\Psi_X^*(t)) &(50)\\ &= \exp(-\frac{(t+\mu)^2}{2\sigma^2}) &(51)\end{aligned}$$

Conclude, the answer of the problem is:

$$\Psi_X(\lambda) = -\mu\lambda - \frac{\sigma^2\lambda^2}{2} \tag{52}$$

$$\Psi_X^*(t) = -\frac{(t+\mu)^2}{2\sigma^2} \tag{53}$$

$$Pr[X \geq t] \leq \exp(-\frac{(t+\mu)^2}{2\sigma^2}) \tag{54}$$

part 3: Poisson random variables.
In similar as part 2, I write this part in short.

$$\begin{aligned}E[e^{\lambda X}] &= \sum_{k=1}^{+\infty} \exp(\lambda k)\frac{exp(-v)v^k}{k!} &(55)\\ &= \exp(-v)\sum_{k=1}^{+\infty}\frac{\exp(\lambda k)v^k}{k!} &(56)\\ &= \exp(-v)\sum_{k=1}^{+\infty}\frac{[v\exp(\lambda)]^k}{k!} &(57)\\ &= \exp(-v)\exp(ve^{\lambda}) &(58)\\ &= \exp(v(e^{\lambda}-1)) &(59)\end{aligned}$$

The general Poisson distribution is $P[X=k] = \frac{exp(-\lambda)\lambda^k}{k!}$, but λ is a variable in our question, so I use v to represent the variable in the Poisson distribution, so we get equation (55).
We all know $\exp(\lambda k) = (\exp(\lambda))^k$, so we get equation (57).
We can use **Taylor's formula** to get equation (58).
Then:

$$\begin{aligned}\Psi_X(\lambda) &= \ln(\exp(v(e^{\lambda}-1))) &(60)\\ &= v(e^{\lambda}-1) &(61)\end{aligned}$$

Similarly, we have:

$$\begin{aligned}\Psi'_X(\lambda) &= ve^{\lambda} && (62)\\ &= t && (63)\end{aligned}$$

Then:

$$\lambda = \ln(\frac{t}{v}) \tag{64}$$

So:

$$\begin{aligned}\Psi^*_X(t) &= t\ln(\frac{t}{v}) - v(e^{\lambda} - 1) && (65)\\ &= t\ln(\frac{t}{v}) - t + v && (66)\end{aligned}$$

Finally:

$$Pr[X \geq t] \leq \exp(t\ln(\frac{t}{v}) - t + v) \tag{67}$$

Conclude, the answer of this part is:

$$\Psi_X(\lambda) = v(e^{\lambda} - 1) \tag{68}$$

$$\Psi^*_X(t) = t\ln(\frac{t}{v}) - t + v \tag{69}$$

$$Pr[X \geq t] \leq \exp(t\ln(\frac{t}{v}) - t + v) \tag{70}$$

Part 4: Bernoulli random variables.
In similar as part 23, I also write this part in short.

$$E[e^{\lambda X}] = pe^{\lambda} + (1 - p) \tag{71}$$

We obtain equation (71) due to the two Bernoulli random distribution equation:

$$Pr[X = 1] = p \tag{72}$$

$$Pr[X = 0] = 1 - p \tag{73}$$

Then:

$$\begin{aligned}\Psi_X(\lambda) &= \ln(pe^{\lambda} + (1 - p)) && (74)\\ & && (75)\end{aligned}$$

Similarly, we have:

$$\begin{aligned}\Psi'_X(\lambda) &= \frac{pe^{\lambda}}{(pe^{\lambda} + (1 - p))} && (76)\\ &= t && (77)\end{aligned}$$

Then:

$$\lambda = \ln \frac{t(1-p)}{p(1-t)} \tag{78}$$

So:

$$\begin{aligned}
\Psi_X^*(t) &= t\lambda - \ln(pe^{\lambda} + (1-p)) && (79)\\
&= t\ln\frac{t(1-p)}{p(1-t)} - \ln(p \times \frac{t(1-p)}{p(1-t)} + 1 - p) && (80)\\
&= t\ln\frac{t(1-p)}{p(1-t)} - \ln\frac{(1-p)}{(1-t)} && (81)\\
&= t\ln(\frac{t}{p} \times \frac{(1-p)}{(1-t)}) - \ln\frac{(1-p)}{(1-t)} && (82)\\
&= t\ln\frac{t}{p} + t\ln\frac{(1-p)}{(1-t)} - \ln\frac{(1-p)}{(1-t)} && (83)\\
&= (t-1)\ln\frac{(1-p)}{(1-t)} + t\ln\frac{t}{p} && (84)\\
&= (1-t)\ln\frac{(1-t)}{(1-p)} + t\ln\frac{t}{p} && (85)\\
&= D(Y||X) && (86)
\end{aligned}$$

Then we prove it.

Part 5:Bernoulli random variables.

Part 5.1

We know that binomial random distribution is the sum of n independently and identically Bernoulli random distribution. For Bernoulli random distribution X_i, we have:

$$\Psi_{X_i}^*(t) = D(Y_i||X_i) \tag{87}$$

Then we have:

$$\Psi_{X_i}^*(\frac{t}{n}) = (1-\frac{t}{n})\ln\frac{(1-\frac{t}{n})}{(1-p)} + \frac{t}{n}\ln\frac{\frac{t}{n}}{p} \tag{88}$$

Then:

$$\begin{aligned}
\Psi_X^*(t) &= n\Psi_{X_i}^*(\frac{t}{n}) && (89)\\
&= (n-t)\ln\frac{n-t}{n(1-p)} + t\ln\frac{t}{np} && (90)\\
&= nD(p||fractn) && (91)
\end{aligned}$$

Then:

$$Pr[X \geq t] \leq \exp(nD(p||\frac{t}{n})) \tag{92}$$

Part 5.2

In short:

$$Pr[X = k] = (1-p)^{k-1}p \tag{93}$$

Then:

$$E[\exp(\lambda X)] = \sum_{k=1}^{+\infty} \exp(\lambda k) \times (1-p)^{k-1} p \tag{94}$$

$$= \frac{p}{1-p} \sum_{k=1}^{+\infty} (e^{\lambda}(1-p))^k \tag{95}$$

Let $a = e^{\lambda}(1-p)$, then $\sum_{k=1}^{+\infty}(e^{\lambda}(1-p))^k = \sum_{k=1}^{+\infty} a^k$ which is the sum of a Geometric progression and $0 < a < 1$. Then, we get:

$$E[\exp(\lambda X)] = \frac{p}{1-p} \sum_{k=1}^{+\infty} (e^{\lambda}(1-p))^k \tag{96}$$

$$= \frac{p}{1-p} \times \frac{(1-p)e^{\lambda}}{1-(1-p)e^{\lambda}} \tag{97}$$

$$= \frac{pe^{\lambda}}{1-(1-p)e^{\lambda}} \tag{98}$$

Then:

$$\Psi_X(\lambda) = \ln(\frac{pe^{\lambda}}{1-(1-p)e^{\lambda}}) \tag{99}$$

$$\tag{100}$$

Then:

$$\Psi'_X(\lambda) = 1 + \frac{(1-p)e^{\lambda}}{1-(1-p)e^{\lambda}} \tag{101}$$

$$= t \tag{102}$$

So:

$$\Psi^*_{X_i}(t) = t\lambda - \ln(\frac{pe^{\lambda}}{1-(1-p)e^{\lambda}}) \tag{103}$$

$$= t\lambda - p \ln \frac{t-1}{1-p} \tag{104}$$

$$= t \ln \frac{t-1}{t(1-p)} - p \ln \frac{t-1}{1-p} \tag{105}$$

Then:

$$n\Psi^*_{X_i}(\frac{t}{n}) = n(\frac{t}{n} \ln \frac{\frac{t}{n}-1}{\frac{t}{n}(1-p)} - p \ln \frac{\frac{t}{n}-1}{1-p}) \tag{106}$$

$$= t \ln \frac{t-n}{t(1-p)} - np \ln \frac{t-n}{n(1-p)} \tag{107}$$

Then:

$$\Psi^*_X(\frac{t}{n}) = t \ln \frac{t-n}{t(1-p)} - np \ln \frac{t-n}{n(1-p)} \tag{108}$$

So:

$$Pr[X \geq t] \leq \exp(t \ln \frac{t-n}{t(1-p)} - np \ln \frac{t-n}{n(1-p)}) \tag{109}$$

That's all.

Problem 4

A boolean code is a mapping $C : \{0,1\}^k \to \{0,1\}^n$. Each $x \in \{0,1\}^k$ is called a message and $y = C(x)$ is called a codeword. The code rate r of a code C is $r = \frac{k}{n}$. A boolean code $C : \{0,1\}^k \to \{0,1\}^n$ is a linear code if it is a linear transformation, i.e. there is a matrix $A \in \{0,1\}^{n \times k}$ such that $C(x) = Ax$ for any $x \in \{0,1\}^k$, where the additions and multiplications are defined over the finite field of order two, $(\{0,1\}, +_{\text{mod } 2}, \times_{\text{mod } 2})$.
The distance between two codeword y_1 and y_2, denoted by $d(y_1, y_2)$, is defined as the Hamming distance between them. Formally, $d(y_1, y_2) = \|y_1 - y_2\|_1 = \sum_{i=1}^{n} |y_1(i) - y_2(i)|$. The distance of a code C is the minimum distance between any two codewords. Formally, $d = \min_{\substack{x_1, x_2 \in \{0,1\}^k \\ x_1 \neq x_2}} d(C(x_1), C(x_2))$.

Usually we want to make both the code rate r and the code distance d as large as possible, because a larger rate means that the amount of actual message per transmitted bit is high, and a larger distance allows for more error correction and detection.

- **Use the probabilistic method to prove that there exists a boolean code $C : \{0,1\}^k \to \{0,1\}^n$ of code rate r and distance $\left(\frac{1}{2} - \Theta\left(\sqrt{r}\right)\right) n$. Try to optimize the constant in $\Theta(\cdot)$.**
- **Prove a similar result for linear boolean codes.**

Solution
Part 1
Let's take two steps to solve the problem. Step one:
Define $D(y_i, y_j)$ denotes the Hamming distance between codeword $y_i = C(x_i)$ and $y_j = C(x_j)$ where $y_i, y_j \in \{0,1\}^n$, $x_i, x_j \in \{0,1\}^k$.
Let X is random variable represent the distribution of $D(y_i, y_j)$.
We can obviously get:

$$Pr[X = x] = \frac{C_n^x}{2^n} \tag{110}$$

Then:

$$\begin{align} E[X] &= \sum_{x=1}^{n} x Pr[X = x] \tag{111} \\ &= \sum_{x=1}^{n} \frac{x C_n^x}{2^n} \tag{112} \\ &= \sum_{x=1}^{n} \frac{n C_{n-1}^{x-1}}{2^n} \tag{113} \\ &= \frac{n}{2^n} \sum_{x=1}^{n} C_{n-1}^{x-1} \tag{114} \\ &= \frac{n}{2^n} \times 2^{n-1} \tag{115} \\ &= \frac{n}{2} \tag{116} \end{align}$$

Then we obtain:

$$Pr[X \leq (1 - \sigma) \times \frac{n}{2}] \leq (\frac{e^{-\sigma}}{(1 - \sigma)^{1-\sigma}})^{\frac{n}{2}} \tag{117}$$

due to Chernoff bound(the lower tail). Step one is based on the distribution of C. Then step two is based on the selection of x_i and x_j.
Step two:
We all know, X depends on the selection of x_i and x_j, and there are $\binom{2^k}{2}$ selection from the message $\{0,1\}^k$. We define X_i denotes one selection. Also we have:

$$Pr[X_i \leq (1-\sigma) \times \frac{n}{2}] \leq (\frac{e^{-\sigma}}{(1-\sigma)^{1-\sigma}})^{\frac{n}{2}} \tag{118}$$

Then we get:

$$\begin{aligned} Pr[d \geq (1-\sigma) \times \frac{n}{2}] &= 1 - Pr[\exists i, j D(y_i, y_j) < (1-\sigma) \times \frac{n}{2}] && (119) \\ &= 1 - Pr[\bigcup_{i=1}^{\binom{2^k}{2}} X_i \leq (1-\sigma) \times \frac{n}{2}] && (120) \\ &\geq 1 - \binom{2^k}{2}(\frac{e^{-\sigma}}{(1-\sigma)^{1-\sigma}})^{\frac{n}{2}} && (121) \end{aligned}$$

Let return back to the problem, the problem wanna us to prove that there exists a boolean code C make the distance $\left(\frac{1}{2} - \Theta\left(\sqrt{r}\right)\right) n$. Then I want to calculate the probability

$$Pr[d < \left(\frac{1}{2} - \Theta\left(\sqrt{r}\right)\right) n] \tag{122}$$

We can get:

$$\begin{aligned} Pr[d \geq \left(\frac{1}{2} - \Theta\left(\sqrt{r}\right)\right) n] &= Pr[d \geq \left(1 - 2\Theta\left(\sqrt{r}\right)\right) \frac{n}{2}] && (123) \\ &\geq 1 - \binom{2^k}{2}(\frac{e^{-2\Theta\left(\sqrt{r}\right)}}{(1-2\Theta\left(\sqrt{r}\right))^{1-2\Theta\left(\sqrt{r}\right)}})^{\frac{n}{2}} && (124) \end{aligned}$$

using the equation (121).
The rest is to prove equation (124) is more than $1 - \frac{1}{n}$ because the probability method want:

$$Pr[d \geq \left(\frac{1}{2} - \Theta\left(\sqrt{r}\right)\right) n] = 1 - \frac{1}{n} \tag{125}$$

And we wanna equation (124) tends to be $\frac{1}{n}$, that is what I want. And $\frac{1}{n}$ can ensure the tight bound of the constant in part 1.2.
Let $x = 1 - 2\Theta\left(\sqrt{r}\right)$, then equation (124) transform to:

$$1 - \binom{2^k}{2}(\frac{\frac{1}{e}e^x}{x^x})^{\frac{n}{2}} \geq 1 - \frac{1}{n} \tag{126}$$

Then we wanna to prove:

$$\binom{2^k}{2}(\frac{1}{e}(\frac{e}{x})^x)^{\frac{n}{2}} \leq \frac{1}{n} \tag{127}$$

then:

$$(\frac{1}{e}(\frac{e}{x})^x)^{\frac{n}{2}} \leq \frac{2^{1-2k}}{n} \tag{128}$$

then:

$$\frac{n}{2}(-1+x\ln\frac{e}{x}) \leq (1-2k)ln2 - \ln n \tag{129}$$

then:

$$(-1+x\ln\frac{e}{x}) \leq \frac{2(1-2k)ln2}{n} - \frac{2\ln n}{n} \tag{130}$$

$$(-1+x\ln\frac{e}{x}) \leq \frac{-4kln2}{n} \tag{131}$$

that is:

$$(-1+x\ln\frac{e}{x}) \leq -4rln2 \tag{132}$$

Then:

$$x\ln\frac{e}{x} + 4rln2 \leq 1 \tag{133}$$

Then we can easily use **Squeeze theorem**, that is, $x\ln\frac{e}{x} \sim -x$. And $x = 1-2\Theta\left(\sqrt{r}\right)$. Then the problem can be easily proved.

Part 1.2

Then we wanna to optimize the constant.

Let $x = 1-2t\sqrt{r}$, in what t is the constant. Then we want:

$$4\ln 2(\frac{1-x}{t})^2 \leq 1 \tag{134}$$

That is:

$$t^2 \geq \ln 2(1-x)^2 \tag{135}$$

then:

$$t^2 \geq \ln 2 \tag{136}$$

then:

$$t \geq \sqrt{\ln 2} \tag{137}$$

Then I optimized the constant.

Part 2

When the code is linear boolean code, we cannot use the strategy above because the distance between codeword depends on the distance between message.

I cannot prove a similar result, but I guess the distance is also $(\frac{1}{2}-\Theta(\sqrt{r}))n$.