CS225: Probability and Computing Homework 2

Zihao Ye

July 16, 2016

PROBLEM 1

(a)

$$\exp\left(-\Psi_X^*(t)\right) = \exp\left(\inf_{\lambda\geq 0}(-\lambda t + \Psi_X(\lambda))\right) = \min_{\lambda\geq 0}\frac{E[e^{\lambda|X|}]}{e^{\lambda t}}$$

此式即 *Chernoff Bound* 的形式.

由于 $\Psi_X(\lambda)$ 在 $\lambda > 0$ 时为严格凸函数, 因此二阶导数大于零, 因此 $(\lambda t - \Psi_X(\lambda))$ 的极值在 $t - \Psi'_X(\lambda) = 0$ 时取到, 即 $\Psi'_X(\lambda) = t$.

(b) 注意到特征函数 $\phi_X(\lambda)$ 与 $\Psi_X(\lambda)$ 之间的关系:

$$\Psi_X(\lambda) = \ln\left(\phi_X(-i\lambda)\right)$$

当 $X \sim N(\mu,\sigma)$ 时, $\phi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$, 由此可得

$$\Psi_X(\lambda) = \ln\left(e^{\mu\lambda + \frac{1}{2}\sigma^2\lambda^2}\right) = \mu\lambda + \frac{1}{2}\sigma^2\lambda^2$$

$\Psi'_X(\lambda) = \mu + \lambda\sigma^2 = t$ 即 $\lambda = \frac{t-\mu}{\sigma^2}$ 时 $\Psi_X^*(t)$ 取到极值 $\frac{(t-\mu)^2}{2\sigma^2}$(注意到 λ 需要非负, 因此该不等式仅在 $t \geq \mu$ 时成立).

$$Pr\{X \geq t\} \leq \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}$$

当 $t < \mu$ 时, $Pr\{X \geq t\} \leq 1$.

(c)

$$\Psi_X(\lambda) = \nu\left(e^{\lambda} - 1\right)$$

$\Psi'_X(\lambda) = \nu\exp\{\lambda\} = t$ 即 $\lambda = \ln t - \ln\nu$ 时 $\Psi_X^*(t)$ 取到极值.

$$Pr\{X \geq t\} \leq \exp\{t-\nu\}\nu^t/t^t$$

(d)

$$\Psi_X(\lambda) = 1 - p + pe^{\lambda}$$

$\Psi'_X(\lambda) = t$ 即 $\lambda = \ln\left(\frac{t(1-p)}{p(1-t)}\right)$ 时 $\Psi^*_X(t)$ 取得极值.

$$\Psi_X(\lambda) = (1-t)\ln\frac{1-t}{1-p} + t\ln\frac{t}{p}$$

(e) 由于 $X = \sum_{i=1}^n X_i$, 且 X_i 独立同分布.

$$\Psi_X(\lambda) = \ln E[e^{\lambda X}] = \ln\prod_{i=1}^n E[e^{\lambda X_i}] = \sum_{i=1}^n \Psi_{X_i}(\lambda) = n\Psi_{X_i}(\lambda)$$

固定 t, 设 $\lambda t - \Psi_X(\lambda)$ 在 λ^* 取极值, 即 $\Psi'_X(\lambda^*) = t$, 则 $\Psi^*_X(t) = \lambda^* t - \Psi_X(\lambda^*)$, 由上式可得

$$\sum_{i=1}^n \Psi'_{X_i}(\lambda^*) = t, \Psi'_{X_i}(\lambda^*) = \frac{t}{n}$$

因此 $\lambda\frac{t}{n} - \Psi_{X_i}(\lambda)$ 在 λ^* 取极值, 即

$$\Psi^*_{X_i}\left(\frac{t}{n}\right) = \lambda^*\frac{t}{n} - \Psi_{X_i}(\lambda^*)$$

得出 $\Psi_X(t) = n\Psi^*_{X_i}\left(\frac{t}{n}\right)$.

由上一问的结论, 当 $X \sim Bin(n,p)$ 时:

$$\Psi^*_X(t) = n\left((1-t)\ln\frac{1-t}{1-p} + t\ln\frac{t}{p}\right)$$

$$Pr\{X \geq t\} \leq \left(\left(\frac{1-t}{1-p}\right)^{1-t}\left(\frac{t}{p}\right)^t\right)^{-n}$$

当 $X_i \sim Ge(1,p)$ 时:

$$\Psi_{X_i}(t) = \log\left(\frac{p}{1-qe^{\lambda}}\right)$$

$$\Psi^*_{X_i}(t) = t\ln\frac{t}{q(t+1)} - \ln p(t+1)$$

则 $\Psi^*_X(t) = n\Psi^*_{X_i}(\frac{t}{n}) = t\ln\frac{t}{q(t+n)} - n\ln\frac{p(t+n)}{n}$.

$$Pr\{X \geq t\} \leq \frac{p^n(1-p)^t(t+n)^{t+n}}{n^n t^t}$$

PROBLEM 2

(a)

$$Pr[|X| \geq \delta] \leq \min_{t\geq 0} \frac{E[e^{t|X|}]}{e^{t\delta}} = \min_{t\geq 0} \frac{\sum\limits_{k=0}^{\infty} \frac{t^k}{k!} E[|X|^k]}{\sum\limits_{k=0}^{\infty} \frac{t^k}{k!} \delta^k}$$

由于 X 固定, 设

$$t^*(\delta) = \arg_t \left\{ \frac{\sum\limits_{k=0}^{\infty} \frac{t^k}{k!} E[|X|^k]}{\sum\limits_{k=0}^{\infty} \frac{t^k}{k!} \delta^k} \right\}$$

为方便起见, 用 t^* 简写 (下同).

当 $t^* \neq 0$ 时, 设

$$c = \min_{t\geq 0} \frac{E[e^{t|X|}]}{e^{t\delta}}$$

$$a_k = \frac{{t^*}^k}{k!} E[|X|^k]$$

$$b_k = \frac{{t^*}^k}{k!} \delta^k$$

则 $\{a_k\}, \{b_k\}$ 均为正项级数, $c = \frac{\sum_{k=0}^{\infty} a_k}{\sum_{k=0}^{\infty} b_k}$.

若 $a_k > cb_k$ 恒成立, 则有 $\sum_{k=0}^{\infty} a_k > \sum_{k=0}^{\infty} cb_k$, 得

$$c < \frac{\sum\limits_{k=0}^{\infty} a_k}{\sum\limits_{k=0}^{\infty} b_k}$$

矛盾, 因此必存在 k 使得 $\frac{e[|X|^k]}{\delta^k} = \frac{a_k}{b_k} \leq c$, 对应的 k 阶矩比 *Chernoff Bound* 更优.

当 $t^* = 0$ 时, $c = 1$, 这个情况是 trivial 的.

(b) 一般情况下 $E[e^{t|X|}]$ 比 $E[|X|^k]$ 计算难度较低. 而且不需要去具体分析应该使用几阶矩, 通用性更好.

PROBLEM 3

(a) 容易证明 $(1-p)^n = (1-\frac{1}{n})^n \geq e^{-1} (n = 1$ 是个 special case, 例外).

$$Pr\{R \cap S = \emptyset, R \cap T \neq \emptyset\} = (1-p)^n \left(1-(1-p)^n\right) \geq e^{-1} - e^{-2}$$

(b) 用 $x_1, x_2, \cdots, x_n$ 表示 $s_1, s_2, \cdots, s_n$ 是否被选, $y_1, y_2, \cdots, y_n$ 表示 $t_1, t_2, \cdots, t_n$ 是否被选.
令 $X = \sum_{i=1}^{n} x_n$, $Y = \sum_{i=1}^{n} y_n$, 由于 x_i, y_i 满足 pairwise independent, 可以证明, $E[XY] = E[X]E[Y]$.

$$Pr\{X > 0 \wedge Y > 0\} = Pr\{XY \geq 1\} \leq E[XY] = E[X]E[Y] = p^2 n^2$$

易证 $E[X] = np$, $E[X^2] = (np)^2 + np(1-p)$.
一个可以估计 $Pr\{X > 0\}$ 的优美方法是使用 *Cauchy Inequality*(via 游宇榕):

$$E[X^2]Pr\{X > 0\} \geq (E[X])^2 \Longrightarrow Pr\{X > 0\} \geq \frac{np}{1-p+np}$$

原题要求的是

$$\begin{aligned} Pr\{X = 0 \wedge Y > 0\} &= Pr\{Y > 0\} - Pr\{X > 0 \wedge Y > 0\} \\ &\geq \frac{np}{1-p+np} - (np)^2 \\ &\geq \frac{np}{1+np} - (np)^2 \end{aligned}$$

当 $np = 0.3$ 时, 上式取 max≈ 0.14 为一个常数.
因此只需要取 $p = 0.3/n$ 即可.
Note: 另一种估计 $Pr\{X > 0\}$ 的方法是尽量糅合 X 的一阶矩与二阶矩 (via 刘志健),

$$E[X] = \sum_{i=1}^{k} i Pr\{X = i\} + \sum_{i=1}^{k} \frac{i^2}{i} Pr\{X = i\} \leq kPr\{X > 0\} + E[X^2]k^{-1}$$

PROBLEM 4

设 $|U| = m$, U 中元素可列为 $u_1, u_2, \cdots, u_m$, 则可以写出如下的 Linear Program:

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{m} c_j \\ \text{subject to} \quad & 0 \leq c_j \leq 1 \qquad \forall j \in [m] \\ & \sum_{j=1}^{m} M_{ij} c_j \geq 1 \quad \forall i \in [n] \end{aligned}$$

容易证明, 原问题的解一定满足上述 LP 的约束, 因此
按照如下算法进行 rounding, 产生 km(其中 k 是个待定的数) 个 mutually independent 的均匀分布在 [0,1] 中的随机变量 $\{x_{ij}\}$. u_j 被选当且仅当存在某个 i, 使得在第 i 个 trial 中 $x_{ij} \leq c_j$. 设一共选了 T 个元素, 第 i 个 trial 选择了 T_i 个元素.

$$E[T] = E\left[\sum_{i=1}^{k} T_i\right] = \sum_{i=1}^{k} E[T_i] = k\left(\sum_{j=1}^{m} c_j\right)$$

因此这个算法得到解的期望 $E[T]=k\cdot OPT$.
下面计算这个算法得到的解合法的概率:

$$
\begin{aligned}
Pr\{\text{所有集合均被 touch}\} &= 1-Pr\{\text{存在未被 touch 的集合}\}\\
&\geq 1-\sum_{i=1}^{n} Pr\{\text{集合 } i \text{ 未被 touch}\}\\
&= 1-\sum_{i=1}^{n}\prod_{j=1}^{m}\left(1-M_{ij}c_j\right)\\
&\geq 1-\sum_{i=1}^{n} e^{-\left(\sum_{j=1}^{m} M_{ij}c_j\right)}\\
&\geq 1-n\cdot e^{-k}
\end{aligned}
$$

令 $k=2\log(n)$, 则 $1-n\cdot e^{-k}=1-\frac{1}{n}$, 此时算法 w.h.p. 将得到一个原问题的 $2\log(n)=O(\log(n))$ 的近似.

PROBLEM 5

以下两问均使用了 $x_1,x_2,\cdots,x_n$mutually independent 的条件, 而这个条件在原题中并未给出.

(a) 用 I_i 表示从第 i 为从第 i 位开始到 $i+k-1$ 位全部为 1, 第 $i+k$ 位为 0(当 $i+k>0$ 时除外) 的 indicator. 则 $X_k=\sum_{i=1}^{n-k} I_i+I_{n-k+1}$, 由此得出:

$$E[X_k]=\sum_{i=1}^{n-k} E[I-i]+I_{n-k+1}=(n-k)2^{-(k+1)}+2^{-k}=(n-k+2)2^{-(k+1)}$$

(b) 设 $f(x_1,x_2,\cdots,x_n)$ 为 $x_1,x_2,\cdots x_n$ 组成的长度大于 k 的 turn 个数, 则有

$$|f(x_1,x_2,\cdots,x_i\cdots,x_n)-f(x_1,x_2,\cdots,y_i,\cdots,x_n)|\leq 1$$

满足 *Lipschitz Condition*, 则有:

$$Pr\{|X_k-E[X_k]|\geq t\}=Pr\left\{\left|f(\mathbf{X})-E[f(\mathbf{X})]\right|\geq t\right\}\leq 2\exp\left(-\frac{t^2}{2n}\right)$$

因此 $t=\sqrt{2n\log n}$ 是一个 *concentration bound*, 此时 $Pr\{|X_k-E[X_k]|\geq t\}=O(\frac{1}{n})$.

PROBLEM 6

(a) 令 $d_H^{(i)}(X,A)$ 为 $E[d_H(X,A)\mid X_1,X_2,\cdots,X_n]$, 则 $\{d_H^{(i)}(X,A)\}$ 是一个 *Doob sequence*.
设 $Y_i=d_H^{(i)}(X,A)$, 则易证:

$$|Y_i-Y_{i-1}|\leq 1$$

由 *Azuma's Inequality*, 有

$$Pr\{Y_n - Y_0 \geq -t\} \leq \exp\left\{-\frac{t^2}{2n}\right\}$$

注意到 $Y_0 = E[d_H(X,A)]$, $Y_n = d_H(X,A)$,

$$\mu(A) = Pr\{Y_n = 0\} \leq Pr\{Y_n - Y_0 \geq -Y_0\} \leq \exp\left\{-\frac{Y_0^2}{2n}\right\}$$

解得 $Y_0 \leq \sqrt{2n\ln\frac{1}{\mu(A)}}$.

(b) 原问题即证明 $\forall\epsilon, \exists\delta > 0, \mu(A) \geq 2^{-\delta n} \Longrightarrow \mu(A_{\epsilon n}) > 1 - 2^{-\delta n}$

由 *Azuma's Inequality* 易证明:

$$\begin{aligned}
\mu(A_{\epsilon n}) &= Pr\{d_H(X,A) \leq \epsilon n\} \\
&= Pr\{d_H(X,A) - E[d_H(X,A)] \leq \epsilon n - E[d_H(X,A)]\} \\
&\geq 1 - \exp\left\{-\frac{(\epsilon n - E[d_H(X,A)])^2}{2n}\right\}
\end{aligned}$$

令 $0 < \delta < \frac{\epsilon^2}{8\ln(2)}$, 则

$$\epsilon n - E[d_H(X,A)] \geq \epsilon n - \sqrt{2n\ln\frac{1}{\mu(A)}} \geq \left(\epsilon - \sqrt{2\delta\ln(2)}\right)n > \frac{\epsilon}{2}n > 0$$

.

由此:

$$1 - \exp\left\{-\frac{(\epsilon n - E[d_H(X,A)])^2}{2n}\right\} > 1 - \exp\left\{-\frac{\epsilon^2 n}{8}\right\}$$

而

$$1 - 2^{-\delta n} < 1 - 2^{-\frac{\epsilon^2}{8\ln(2)}n} = 1 - \exp\left\{-\frac{\epsilon^2 n}{8}\right\}$$

由此任取 $\delta \in \left(0, \frac{\epsilon^2}{8\ln(2)}\right)$ 即可.

ACKNOWLEDGEMENT

感谢贝小辉老师在近似算法课上介绍了 Problem 4 的 randomized rounding 的做法.
感谢游宇榕和刘志健同学在第 3 题中给予的启发.