MCMT Homework 9

Shun Zhang

Exercise 9.1

Let Y_t denote the number of coordinates that have been selected after t rounds. Define the distinguishing statistic $f : \{0,1\}^n \to R$ by $f(x) = \sum_{i=1}^{n} x_i$.

1. The probability that one coordinate is not selected in t steps is $p = (1-\frac{1}{n})^t$. The expectation of the number of coordinates that are not selected in t steps is $\mathrm{E}(n - Y_t) = np = n(1-\frac{1}{n})^t$. So $\mathrm{E}(Y_t) = n - n(1-\frac{1}{n})^t$.

2. Let $I_i(t)$ be the indicator that i-th coordinate has been selected at t-th step.

   $\mathrm{E}(I_i(t)I_j(t)) = 2(1-(1-\frac{1}{n})^t) - (1-(1-\frac{2}{n})^t) = 1 - 2(1-\frac{1}{n})^t + (1-\frac{2}{n})^t$.

   $\mathrm{E}(I_i(t))\mathrm{E}(I_j(t)) = (1-(1-\frac{1}{n})^t)^2$.

   $\mathrm{Cov}(I_i(t), I_j(t)) = \mathrm{E}(I_i(t)I_j(t)) - \mathrm{E}(I_i(t))\mathrm{E}(I_j(t)) = (1-\frac{2}{n})^t - (1-\frac{1}{n})^{2t} \leq 0$.

   $\mathrm{Var}(Y_t) = \mathrm{Var}(\sum_i I_i(t)) = \sum_i(\mathrm{Var}(I_i(t))) + \sum_{i\neq j}\mathrm{Cov}(I_i(t), I_j(t)) \leq \sum_i(\mathrm{Var}(I_i(t))) \leq np(1-p) \leq \frac{n}{4}$.

3. $\mathrm{E}_0(X_{ti}|Y_t) = \frac{Y_t}{n}\frac{1}{2}$. This is the probability that x_i is chosen, and times the probability that it is set to be 1.

   $\mathrm{E}_0(f(X_t)|Y_t) = \sum_i \mathrm{E}_0(X_{ti}|Y_t) = \frac{Y_t}{2}$.

   $\mathrm{E}_0(f(X_t)) = \sum_y \mathrm{E}_0(f(X_t)|Y_t = y)P(Y_t = y) = \sum_y \frac{y}{2}P(Y_t = y) = \frac{\mathrm{E}(Y_t)}{2}$.

4. $\mathrm{Var}_0(f(X_t)|Y_t) = Y_t\frac{1}{2}(1-\frac{1}{2}) = \frac{1}{4}Y_t$, as $f(X_t)$ given Y_t is a Binomial distribution. $\mathrm{E}_0\mathrm{Var}_0(f(X_t)|Y_t) = \frac{1}{4}\mathrm{E}_0(Y_t)$.

   $\mathrm{E}_0(f(X_t)|Y_t) = \frac{Y_t}{2}$. $\mathrm{Var}_0\mathrm{E}_0(f(X_t)|Y_t) = \mathrm{Var}_0\frac{Y_t}{2} = \frac{1}{4}\mathrm{Var}_0(Y_t)$.

   $\mathrm{Var}_0(f(X_t)) = \mathrm{E}_0\mathrm{Var}_0(f(X_t)|Y_t) + \mathrm{Var}_0(\mathrm{E}_0(f(X_t)|Y_t)) = \frac{1}{4}(\mathrm{E}(Y_t) + \mathrm{Var}(Y_t))$.

   By substituting by the results we have, $\mathrm{E}(Y_t) + \mathrm{Var}(Y_t) = n - n(1-\frac{1}{n})^t + (n^2-n)cov + n(1-\frac{1}{n})^t(1-(1-\frac{1}{n})^t) = n + (n^2-n)cov - n(1-\frac{1}{n})^t < n$, where cov is $\mathrm{Cov}(I_i(t), I_j(t))$ for $i \neq j$, which we have shown to be negative.

   So, $\mathrm{Var}_0(f(X_t)) < \frac{1}{4}n$.

5. f on π is a Binomial distribution. So $E_\pi f = \frac{n}{2}$, $\mathrm{Var}_\pi f = n\frac{1}{2}(1-\frac{1}{2}) = \frac{n}{4}$.

6. Consider two distributions $P^t(0,\cdot)$ and π.

$\sigma^2 \leq (\frac{n}{4}+\frac{n}{4})/2 = \frac{n}{4}$.

$\Delta = |\frac{EY_t}{2} - \frac{n}{2}| = \dfrac{n - EY_t}{2} = \dfrac{n(1-\frac{1}{n})^t}{2}$.

By Lemma 9.7, $||P^t(0,\cdot) - \pi||_{TV} \geq \dfrac{\Delta^2}{4\sigma^2+\Delta^2} = 1 - \dfrac{4\sigma^2}{4\sigma^2+\Delta^2} \geq 1 - \dfrac{n}{n + \frac{n^2(1-\frac{1}{n})^{2t}}{4}} = 1 - 8\dfrac{1}{8+2n(1-\frac{1}{n})^{2t}} = 1 - 8\exp\{-\log(8+2n(1-\frac{1}{n})^{2t})\}$.

We know $d(t) \geq ||P^t(0,\cdot) - \pi||_{TV}$. Let $t = \frac{1}{2}n\log n - cn$.

$d(\frac{1}{2}n\log n - cn)$
$\geq 1 - 8\exp\{-\log(8+2n(1-\frac{1}{n})^{n\log n - 2cn})\}$
$\geq 1 - 8\exp\{-\log 2n - \log((1-\frac{1}{n})^{n\log n - 2cn})\}$
$= 1 - 8\exp\{-\log 2n - (n\log n - 2cn)\log(1-\frac{1}{n})\}$
$\geq 1 - 8\exp\{-\log 2n - (n\log n - 2cn)(-\log 2 - \frac{1}{n})\}$
$\geq 1 - 8\exp\{-2c - O(n\log n)\}$
$\geq 1 - 8\exp\{-2c+1\}$.