Analysis

Tongqing Qiu

November 21, 2007

1 t-smallest Algorithm Analysis

Lemma 1 *Let $X_1, X_2, \cdots, X_n$ be iid uniform $(0,1)$, for the jth order statistic, we can get that*

$$EX_{(j)} = \frac{j}{n+1}$$

and the variance is

$$Var_{(j)} = \frac{j(n-j+1)}{(n+1)^2(n+2)}$$

For "t-smallest" algorithm, $\hat{n} = \frac{t}{x_{(t)}}$. For convenience, we use x instead of $x_{(t)}$ from now on. Let $f(x) = \frac{1}{x}$. $p = EX = \frac{t}{n+1}$.

$$\begin{aligned} \hat{n} &= t\left(f(p) + (x-p)f'(p) + \frac{1}{2}(x-p)^2 f''(p) + \cdots\right) \\ &= t\left(\frac{1}{p} - \frac{x-p}{p^2} + \frac{(x-p)^2}{p^3} + \cdots\right) \end{aligned} \tag{1}$$

Take the expectation. The second term equals to zero.

$$\begin{aligned} E(\hat{n}) &\approx t\left(\frac{n+1}{t} + \frac{(n+1)^3}{t^3} \cdot \frac{t(n-t+1)}{(n+1)^2(n+2)}\right) \\ &= n+1+\frac{(n+1)(n-t+1)}{t(n+2)} \\ &\approx n + \frac{n+1}{t} \end{aligned} \tag{2}$$

Similarly, take variance

$$\begin{aligned} Var(\frac{\hat{n}}{n}) &\approx \frac{t^2}{p^4 \cdot n^2} Var(x-p) \\ &= \frac{t^2}{p^4 \cdot n^2} Var(x) \\ &= \frac{(n+1)^2(n-t+1)}{t(n+2)n^2} \\ &\approx \frac{n-t+1}{t(n+2)} \end{aligned} \tag{3}$$

2 Intersection Analysis

Given a random variable X, we denote $X - E[X]$ as X^c. X^c is often referred to as a *centered random variable*. It can easily be verified that $E[X^c] = 0$ and $Var[X] = Var[X^c]$.

Lemma 2

$$E(\hat{n}) = n + \frac{n+1}{t}$$

$$Var(\hat{n}) = \frac{n(n-t+1)}{t}$$

Lemma 3 *Let Y and Z be two sets of packets and $Y \bigcap Z = \emptyset$. Then we have (i) $E[U_Y^c U_Z^c] = 0$ and (ii) $E[U_{Y \bigcup Z}^c U_Y^c] \approx (1+\frac{1}{t})Var[U_Y]$.*

Proof. (■

i) Note that U_Y^c and U_Z^c are independent random variables since $Y \bigcap Z = \emptyset$. Therefore $E[U_Y^c U_Z^c] = E[U_Y^c]E[U_Z^c] = 0$. (ii)

$$\begin{aligned} E[U_{Y \bigcup Z}^c U_Y^c] &= \sum_{i=-\infty}^{\infty} E[U_{Y \bigcup Z}^c U_Y^c | U_Y^c = i] Pr[U_Y^c = i] \\ &= \sum_{i=-\infty}^{\infty} iE[U_{Y \bigcup Z}^c | U_Y^c = i] Pr[U_Y^c = i] \qquad (*) \end{aligned}$$

It suffices to prove that for any particular Y^* (a constant set) such that $U_{Y^*} - E[U_Y] = i$ and $Y^* \bigcap Z = \emptyset$,

$$\begin{aligned} E[U_{Y \bigcup Z}^c | U_Y^c = i] &= E[U_{Y \bigcup Z} | U_Y^c = i] - E[U_{Y \bigcup Z}] \\ &= E[U_{Y^* \bigcup Z}] - E[U_{Y \bigcup Z}] \end{aligned}$$

$$E[U_{Y^* \bigcup Z}] = (1+\frac{1}{t})\left((1+\frac{1}{t})n_Y + \frac{1}{t} + i + n_Z\right) + \frac{1}{t}$$

$$\begin{aligned} E[U_{Y \bigcup Z}^c | U_Y^c = i] &= E[U_{Y^* \bigcup Z}] - E[U_{Y \bigcup Z}] \\ &= (1+\frac{1}{t})\left((1+\frac{1}{t})n_Y + \frac{1}{t} + i + n_Z\right) + \frac{1}{t} - \left((1+\frac{1}{t})(n_Y + n_Z) + \frac{1}{t}\right) \\ &= (1+\frac{1}{t})(\frac{1}{t}n_Y + \frac{1}{t} + i) \end{aligned}$$

Note that $\sum iPr[U_Y^c = i] = E[U_Y^c] = 0$ and $\sum i^2 Pr[U_Y^c = i] = Var[U_Y^c]$, so

$$E[U_{Y \bigcup Z}^c U_Y^c] = (1+\frac{1}{t})Var[U_Y]$$

Similarly, $E[U_Y^c U_Z^c] = (1+\frac{1}{t})^2 Var[U_{Y \cap Z}]$ To simplify the notations in the following proof, we use $R, S, N, \hat{N}$. Recall that our estimator becomes $\hat{N} = D_R + D_S - D_{R \cup S}$. We have

$$\begin{aligned} Var(\hat{N}) =& Var(\hat{N}^c) \\ =& E[(D_R^c + D_S^c - D_{R\cup S}^c)^2] \\ =& E[(D_R^c)^2] + E[(D_S^c)^2] + E[(D_{R\cup S}^c)^2] \\ &+ 2E[D_R^c D_S^c] - 2E[D_R^c D_{R\cup S}^c] \\ &- 2E[D_S^c D_{R\cup S}^c] \end{aligned}$$

$n_{R\cup S} = n_R + n_S$ we have

$$E[(D^c_R)^2] = Var(D_R) = \frac{n_R - t + 1}{t(n_R + 2)} \quad (4)$$

$$E[(D^c_S)^2] = Var(D_S) = \frac{n_S - t + 1}{t(n_S + 2)} \quad (5)$$

$$\begin{aligned} E[(D^c_{R\cup S})^2] &= Var(D_{R\cup S}) \\ &= \frac{(n_{R\cup S} - t + 1)}{t(n_{R\cup S} + 2)} \end{aligned} \quad (6)$$

$$\begin{aligned} Var(\hat{N}) =& Var[D_R] + Var[D_S] + Var[D_{R\cup S}] + 2(1 + \frac{1}{t})^2 Var[D_{R\cap S}] \\ &- 2(1 + \frac{1}{t})Var[D_R] - 2(1 + \frac{1}{t})Var[D_S] \\ =& Var[D_{R\cup S}] + 2(1 + \frac{1}{t})^2 Var[D_{R\cap S}] - (1 + \frac{2}{t})\,[var[D_R] + Var[D_S]] \end{aligned} \quad (7)$$