Online Supplement to "Optimal Synchronization Policies for Data Warehouses"

Debabrata Dey
School of Business, University of Washington, Box 353200, Seattle, Washington 98195–3200, USA,
ddey@u.washington.edu

Zhongju Zhang
School of Business, University of Connecticut, Stamford, Connecticut 06901–2315, USA,
john.zhang@business.uconn.edu

Prabuddha De
Krannert School of Management, Purdue University, West Lafayette, Indiana 47907–2056, USA,
pde@mgmt.purdue.edu

Derivation of a Few Important Results

Proposition 1 $\int_0^I dt_n \int_0^{t_n} dt_{n-1} \ldots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 \left(\sum_{i=1}^{n} t_i^m\right) = \dfrac{I^{m+n}}{(m+1)(n-1)!}$, *for all* $m \geq 0$.

Proof: We can prove this easily using induction. Note that, when $n = 1$, $\int_0^I t_1^m dt_1 = \frac{I^{m+1}}{m+1}$. Now, suppose that the above is true for n; we will show that it is true for $n+1$.

$$\begin{aligned}
&\int_0^I dt_{n+1} \int_0^{t_{n+1}} dt_n \ldots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 \left(\sum_{i=1}^{n+1} t_i^m\right) \\
&= \int_0^I dt_{n+1} \left[\int_0^{t_{n+1}} dt_n \ldots \int_0^{t_2} dt_1 \left(\sum_{i=1}^{n} t_i^m\right)\right] + \int_0^I t_{n+1}^m dt_{n+1} \int_0^{t_{n+1}} dt_n \ldots \int_0^{t_2} dt_1 \\
&= \int_0^I dt_{n+1} \left[\frac{(t_{n+1})^{m+n}}{(m+1)(n-1)!}\right] + \int_0^I t_{n+1}^m dt_{n+1} \left[\frac{(t_{n+1})^n}{n!}\right] \\
&= \frac{1}{(m+1)(n-1)!} \frac{I^{m+n+1}}{m+n+1} + \frac{1}{n!} \frac{I^{m+n+1}}{m+n+1} = \frac{I^{m+n+1}}{(m+1)n!}.
\end{aligned}$$

This completes the proof. ■

Proposition 2 $\int_0^\infty dt_n \int_0^{t_n} dt_{n-1} \ldots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 e^{-\lambda t_n} \left(\sum_{i=1}^{n} t_i\right) = \dfrac{n(n+1)}{2\lambda^{n+1}}$.

Proof:

$$\begin{aligned}
&\int_0^\infty dt_n \int_0^{t_n} dt_{n-1} \ldots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 e^{-\lambda t_n} \left(\sum_{i=1}^{n} t_i\right) \\
&= \int_0^\infty dt_n e^{-\lambda t_n} \left[\int_0^{t_n} dt_{n-1} \ldots \int_0^{t_2} dt_1 \left(\sum_{i=1}^{n-1} t_i\right) + t_n \int_0^{t_n} dt_{n-1} \ldots \int_0^{t_2} dt_1\right] \\
&= \int_0^\infty dt_n e^{-\lambda t_n} \left[\frac{(t_n)^n}{2(n-2)!} + t_n \frac{(t_n)^{n-1}}{(n-1)!}\right] \qquad \text{(by Proposition 1)} \\
&= \frac{n+1}{2(n-1)!} \int_0^\infty e^{-\lambda t_n} (t_n)^n dt_n = \frac{n+1}{2(n-1)!} \frac{n!}{\lambda^{n+1}} = \frac{n(n+1)}{2\lambda^{n+1}}.
\end{aligned}$$

■

Proposition 3 $\int_0^\infty dt_n \int_0^{t_n} dt_{n-1} \ldots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 e^{-\lambda t_n} \left(\sum_{j=1}^{n} (t_n - t_j) \right) = \frac{n(n-1)}{2\lambda^{n+1}}$.

Proof:

$$\begin{aligned}
&\int_0^\infty dt_n \int_0^{t_n} dt_{n-1} \ldots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 e^{-\lambda t_n} \left(\sum_{j=1}^{n} (t_n - t_j) \right) \\
&= \int_0^\infty dt_n \int_0^{t_n} dt_{n-1} \ldots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 e^{-\lambda t_n} \left(n t_n - \sum_{j=1}^{n} t_j \right) \\
&= n \int_0^\infty t_n e^{-\lambda t_n} dt_n \int_0^{t_n} dt_{n-1} \ldots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 - \frac{n(n+1)}{2\lambda^{n+1}} \qquad \text{(by Proposition 2)} \\
&= \frac{n}{(n-1)!} \int_0^\infty (t_n)^n e^{-\lambda t_n} dt_n - \frac{n(n+1)}{2\lambda^{n+1}} = \frac{n(n-1)}{2\lambda^{n+1}}.
\end{aligned}$$

■

Distributions Used in Simulation Experiments

Here, we discuss the various distributions of inter-arrival times used in the simulation experiments described in Section 4 of the paper.

Hyperexponential Distribution

Let x_1 and x_2 be two exponential random variables with rate λ_1 and λ_2, respectively. Then, $x = \alpha x_1 + (1-\alpha)x_2$ is a hyperexponential random variable if $0 < \alpha < 1$ and $\lambda_1 \neq \lambda_2$. In our simulation, to obtain a hyperexponential random variable with rate λ (i.e., with mean $\frac{1}{\lambda}$), we set:

$$\lambda_1 = \frac{\lambda}{1+\gamma}, \lambda_2 = \frac{\lambda}{1-\gamma}, \text{ and } \alpha = 0.5.$$

The CV for this random variable is given by: CV $= \sqrt{(1+2\gamma^2)}$. Thus, by varying γ, we could produce a hyperexponential random variable with different CV's. The exponential random variables x_1 and x_2 were generated by drawing two uniform random numbers u_1 and u_2 in the range $[0, 1]$ (using a standard random number generator), and by setting:

$$x_1 = -\frac{\ln u_1}{\lambda_1}, \text{ and } x_2 = -\frac{\ln u_2}{\lambda_2}.$$

Exponential Distribution

To obtain an exponential random variable x with rate λ, we generated a uniform random number $u \in [0, 1]$ and set $x = -\frac{\ln u}{\lambda}$. The CV of x is 1.

Lognormal Distribution

If a random variable Z follows the normal distribution with mean μ and variance σ^2, then a random variable $x = e^Z$ is said to have a lognormal distribution, with mean $e^{\mu + \frac{\sigma^2}{2}}$ and variance $e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}$. Given two uniform random numbers u_1 and u_2 in $[0, 1]$, a standard normal random variable z can be generated by using the Box-Muller transformation (Ross 1997): $z = \sqrt{-2 \ln u_1} \cos(2\pi u_2)$. A normal variable $Z(\mu, \sigma^2)$ can then be obtained from $Z = \mu + \sigma z$, which can, in turn, be used to generate the lognormal variable $x = e^Z$. In our experiments, to keep the rate at λ, we set $\mu = -\ln(\lambda\sqrt{2})$ and $\sigma^2 = \ln 2$, so that the CV of x was 1.

Pareto Distribution

A random variable x following the Pareto distribution has the following cumulative distribution function (Mood et al. 1974):

$$F(x) = 1 - \left(\frac{\gamma}{x}\right)^{\theta}, \text{ where } \gamma, \theta > 0.$$

After drawing a uniform random number $u \in [0, 1]$, we can generate a Pareto random variable x by setting $x = \gamma(1-u)^{-\left(\frac{1}{\theta}\right)}$. In our experiments, we set $\gamma = \frac{1}{\lambda}\frac{\sqrt{2}}{1+\sqrt{2}}$ and $\theta = 1 + \sqrt{2}$ in order to keep the rate at λ and CV=1.

Power Distribution

A random variable x following a one-sided power-k distribution has the following cumulative distribution function (van Dorp and Kotz 2002):

$$F(x) = 1 - \left(\frac{b-x}{b}\right)^{k}.$$

Note that the triangular distribution is a special case of the above when $k = 2$. In our experiments, we set $b = \frac{k+1}{\lambda}$ to keep the rate at λ. Given a uniform random number $u \in [0, 1]$, a power-k random variable x is generated from $x = \frac{k+1}{\lambda}[1 - (1-u)^{\frac{1}{k}}]$. It can be shown that the CV of x is $\sqrt{\frac{k}{k+2}}$. Thus, by varying k, one can generate power random variables with different CV's.

Uniform Distribution

After drawing a uniform random number $u \in [0, 1]$, a uniform random variable x with rate λ can be obtained by setting $x = \frac{2u}{\lambda}$. The CV of x is $\frac{1}{\sqrt{3}}$.

Erlang Distribution

Let $x_1, x_2, \ldots, x_k$ be i.i.d. exponential random variables, each with a rate of $k\lambda$. Then, an Erlang-k random variable x with rate λ is obtained from $x = \sum_{i=1}^{k} x_k$. The CV of an Erlang-k random variable is $\frac{1}{\sqrt{k}}$. Thus, one can generate Erlang variables with different CV's by varying k.

References

Mood, A.M., F.A. Graybill, D.C. Boes. 1974. *Introduction to the Theory of Statistics.* McGraw-Hall, New York, NY.

Ross, S. 1997. *Simulation.* Academic Press, San Diego, CA.

van Dorp, J.R., S. Kotz. 2002. A novel extension of the triangular distribution and its parameter estimation. *The Statistician* **51** 63–79.