Computer System Fundamentals HW #2

Quan Zhou

Feb 10th, 2016 (update)

Problem 1

(a) N = 20; Pr(X) = 0.96; $Pr(\bar{X}) = 1 - Pr(X) = 0.04$
so the event when 20 connections to the web server where no connections will be refused is a binomial distribution. Therefore we have: $Pr(X = 20) = \binom{20}{0}(Pr(X = 1))^{20}(Pr(X = 0))^{0} = 0.4420$

(b) N = 20, $Pr(X = 19) = \binom{20}{1}(Pr(X = 1))^{19}(Pr(X = 0))^{1} = 0.3683$

(c) N = 20, $Pr(X > 17) = \binom{20}{0}(Pr(X = 1))^{20}(Pr(X = 0))^{0} + \binom{20}{1}(Pr(X = 1))^{19}(Pr(X = 0))^{1} + \binom{20}{2}(Pr(X = 1))^{18}(Pr(X = 0))^{2} + \binom{20}{3}(Pr(X = 1))^{17}(Pr(X = 0))^{3} = 0.9561$

(d) $Pr(X \geq 4) = 1 - Pr(\bar{X} = 1) - Pr(\bar{X} = 2) - Pr(\bar{X} = 3) = 1 - 0.04 - (0.04)^{2} - (0.04)^{3} = 0.9583$

(e) The event when some number of requests that will be successfully served before a connection is refused is geometric distribution. $\mathbb{E}(X = 1) = \frac{Pr(X)}{Pr(\bar{X})} = \frac{0.96}{0.04} = 24$

(f) For the event of a sample size of 50 requests being processed through the web server, let's assume the distribution is Gaussian (Normal). so $\mathbb{E}(X) = 50*0.96 = 48$

(g) For a very large number of requests being processed through the web server, we could assume the number of successful requests through the server is normally distributed variable. Then we are able to find:

$$z \geq \frac{x - \mu}{\frac{\sigma}{\sqrt{(n)}}}$$

$$= \sqrt{(50)}(\frac{x}{\mu} - 1)(\frac{\mu}{\sigma})$$

$$= \sqrt{(50)}(\frac{x}{\mu} - 1)(\frac{\frac{1-p}{p}}{\frac{\sqrt{(1-p)}}{p}})$$

$$= \sqrt{(50)}(\frac{x}{\mu} - 1)(\sqrt{(1 - p)})$$

$$= \sqrt{(50)}(\frac{0.95}{0.96} - 1)(\sqrt{(0.96)})$$

$$= -0.071$$

By looking up the Z-table, we found the probability is 1- 0.4721 = 0.5278

(h) None. For all the questions are focused on what the odds are for successful or failing connection, and the variables are not function of time(for we assume the probability of successful or failing connection is constant or not a function of time). These questions have to do with availability of the web server, not reliability. Rather, if the questions were asked about the probability of the server is still up over a period of time, then it is a questions of reliability.

Problem 2

I (a) Since the arrival rate is 200 packets per second, so the parameter $\lambda = \frac{200packets}{second} \cdot \frac{10second}{1000} = 2$. We have the following probability mass function (pmf) with Possion distribution:

$$f(x) = \left(\frac{2^x}{x!}\right) e^{-2}$$

(b) The probability of more than 3 packets are received over the 10 msec period is given by the following equation:

$$
\begin{aligned}
Pr(X > 3) &= 1 - Pr(X = 0) - Pr(X = 1) - Pr(X = 2) - Pr(X = 3) \\
&= 1 - (1 + 2 + 2 + \frac{4}{3})e^{-2} \\
&= 0.1429
\end{aligned}
$$

(c) Now the probability of inter-arrival time of packets is a exponentially distributed random variable. λ is actually the rate of arrival packets at the Ethernet adapter, or 200 packet/msec, and we get:

$$g(T) = 200e^{-200T}$$

(d) The probability of two packets will be separated by more than 8 msec is given by the cumulative distribution function (cdf) G:

$$G(T) = 1 - e^{-0.2T}$$

so:

$$
\begin{aligned}
Pr(T > 8msec) &= 1 - G(T = 8msec) \\
&= 1 - (1 - e^{-0.2T} \\
&= e^{-1.6} \\
&= 0.2019
\end{aligned}
$$

(e) The variance of an exponential distribution is given by λ^{-2}, or standard deviation by λ^{-1}, so we derive the standard deviation of the inter-arrival time of packet to be 5.

II (a) For one packet in the system, $T_s = 4.8$ msec. Therefore, the capacity this Ethernet Adapter is $\frac{1000}{4.8}$ or 208.33 packets per sec.

(b) Assuming the Ethernet adaptor is M/M/1 system (Poisson process arrival, exponential distribution for servicing and one server), the probability of having x packets in the system is given by:

$$f(x) = (1 - \rho)(\rho^x) = 0.04(0.96^x) \tag{1}$$

where the utilization $\rho = \lambda T_s = \frac{0.2}{msec} 4.8$ msec $= 0.96$

(c) To find w, by Little's Law we have:

$$w = \lambda T_w \tag{2}$$

$T_w = T_q - T_s = 5 - 4.8 = 0.2$ msec Therefore:
$w = \lambda(0.2 \text{ msec}) = \frac{200 \text{ packets}}{1000 \text{ msec}}(0.2 \text{ msec}) = 0.04$ packets

(d) As found from part (c), $T_w = 0.2$ msec

(e) Slowdown $= \frac{1}{1-\rho} = \frac{1}{1-0.96} = \frac{1}{0.04} = 25$

Problem 3

(a) By definition, the exponential distribution has the following pdf:

$$f(x) = \lambda e^{-\lambda x} \tag{3}$$

In this problem, we will look at the inverse of above equation:

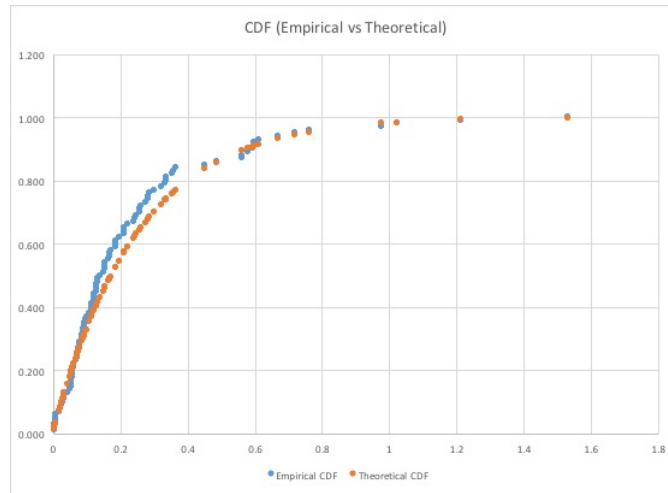$y = \lambda e^{-\lambda x}$

$\frac{y}{\lambda} = e^{-\lambda x}$

$ln(\frac{y}{\lambda}) = -\lambda x$

$x = -\frac{ln(\frac{y}{\lambda})}{\lambda}$

Define $\frac{y}{\lambda}$ as a random variable $x\prime$, and x being $g(x)$, we have:

$$g(x) = \frac{ln(x\prime)}{\lambda} \tag{4}$$

See Java code "RandomNumGenerator"

CDF (Empirical vs Theoretical)

They pretty much match with one another. The reason is that because the 100 random numbers are exponentially distributed as indicated in the problem. The more generated random numbers, the closer the curve fits the exponential CDF.

(b) First, we find the end time in seconds for each query. Then, we set up the histogram table for each-second interval to account for the number of queries being served during that interval. Last, we plot the normalized histogram (number of queries divided by the total 100 queries).

3
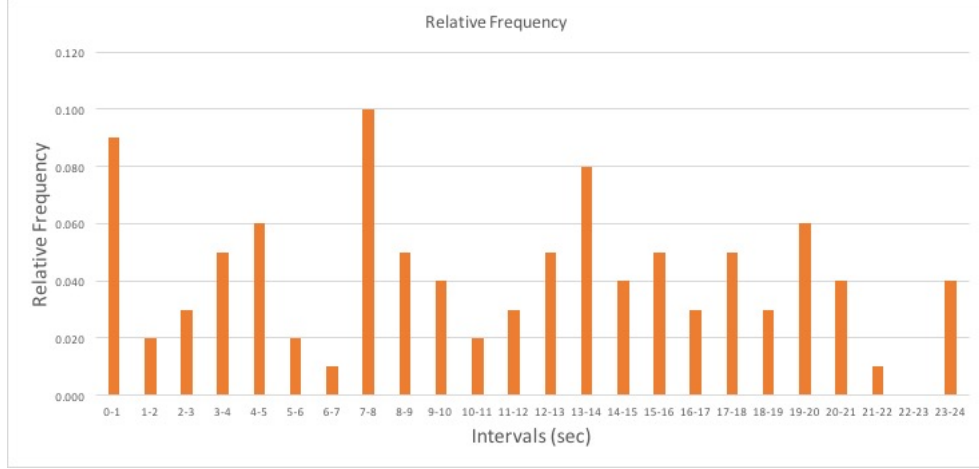
From intuition plus eyeballing, it is concluded that the distribution is Poisson as we assumed the "random" inter-arrival times.

The formula is written as below:

$$f(x) = \left(\frac{4^x}{x!}\right) e^{-4} \qquad (5)$$

Problem 4

(a) Assume the four-level cache system is M/M/1 and Given:

$Pr(X_1 = Hit) = \frac{1}{2}, Pr(X_2 = Hit) = \frac{3}{5}, Pr(X_3 = Hit) = \frac{7}{10}$ for L1, L2 and L3 cache
then we can derive the following: Probability of L1 hit $= Pr(X_1 = Hit) = 0.5$
Probability of L2 hit after L1 miss $= Pr(X_1 = Miss) * Pr(X_2 = Hit) = 0.5 * 0.6 = 0.3$
Probability of L3 hit after L2 miss $= Pr(X_1 = Miss)*Pr(X_2 = Miss)*Pr(X_3 = Hit) = 0.5*0.4*0.7 = 0.14$
Probability of accessing the last memory level $= Hit|\bar{X}_1 \cap \bar{X}_2 \cap \bar{X}_3) = \frac{1}{2} * \frac{2}{5} * \frac{3}{10} = \frac{3}{50}$
$T(L1)_s = 1, T(L2)_s = 2, T(L3)_s = 7, T(rest)_s = 20$
so PMF for random variable t is:

$$f(t) = \begin{cases} 0.5, & t = 1 \\ 0.3, & t = 3 \\ 0.14, & t = 10 \\ 0.06, & t = 30 \end{cases} \qquad (6)$$

(b) $\mu = \mathbb{E}(t) = (0.5)1 + (0.3)3 + (0.14)10 + (.06)30 = 4.6$ clock cycles

(c)

$$variance = \sum_{i=1}^{4}[(x_i - \mu)^2 p_i] = 0.5(4.6 - 1)^2 + 0.3(4.6 - 3)^2 + 0.14(4.6 - 10)^2 + 0.06(4.6 - 30)^2 = 50.04$$

so the standard deviation for random variable t is $\sqrt{(50.04)} = 7.07$ clock cycles

(d)

4

Problem 5

(a) By comparing both the empirical cdf and theoretical cdf (by finding the differences between the range/bin defined in the table, see excel), one can conclude both are similar and match to each other. As the central limit theorem states: the arithmetic mean of 100 independent random variables (which has the mean and variance defined to be uniformally distributed) will be normally distributed. This exercise just proves CLT.