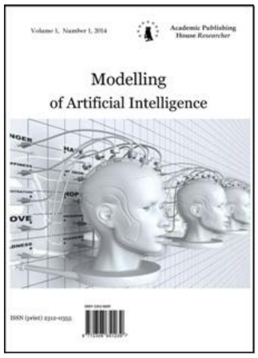
Published in the Russian Federation
Modeling of Artificial Intelligence
Has been issued since 2014.
ISSN: 2312-0355
E-ISSN: 2413-7200
Vol. 11, Is. 3, pp. 151-157, 2016

DOI: 10.13187/mai.2016.11.151
www.ejournal11.com



UDC 004

Expected Value and the Laplace Transform of the Maximum for I.I.D. Exponentials

Victor K. Ohanyan [a,*], Alexander Thomasian [b]

[a] Yerevan State University, Armenia
[b] College of Science and Engineering, American University of Armenia, Armenia

Abstract

We present three methods to compute the expected value of the maximum of n independent, identically distributed exponentials and also obtain their Laplace Stieltjes Transform-LST. The results are applicable in the following cases: (1) the time to data loss in disk arrays with n-way replication, where the time to disk failure is exponentially distributed; (2) the time to completion of n exponentially distributed parallel tasks; (3) an upper bound to the mean fork-join response time when arrivals are Poisson, service times are exponentially distributed, so that response times are exponentially distributed.

Keywords: combinatorial analysis, order statistics, performance analysis, queueing system, parallel processing, fork-join queueing system, reliability evaluation, disk arrays.

1. Introduction

The order statistics and especially the expected value of the maximum and minimum of independent random variables are discussed in most books on probability theory and statistics (Trivedi, 2002). We present three methods to compute the expected value and the Laplace Stieltjes Transform (LST) of the maximum in the case of random variables with a negative exponential distribution, whose coefficient of variation CV=1 is considered a midpoint between CV=0 and large CVs. The resulting closed form equations can be applied to the following problems.

Reliability Modeling: The reliability of an n-way parallel system, where it takes the failure of all n components with individual reliabilities $R_i(t),\ 1 \le i \le n$ for the system to fail is given as (Trivedi, 2002):

$$R_{parallel}\ (t)\ =\ 1-\prod_{i=1}^{n}[\,1-\ R_i(t)]\ .$$

It has been shown in numerous studies such as (Gibson, 1992; Schroeder, Gibson, 2007) that the exponential distribution is a sufficiently accurate approximation for the time to disk failure, so that: $R_i(t) = e^{-\lambda_i t}$. Most disk arrays have homogeneous disks so that $\lambda_i = \lambda$, for all i. This assumption has been subsequently used in reliability modeling studies of disk arrays, which additionally requires disk repair times to be exponentially distributed (Gibson, 1992; Trivedi, 2002). There is interest in the Mean Time to Data Loss - MTTDL, which is the time that it takes for

* Corresponding author
E-mail addresses: victo@aua.am (V.K. Ohanyan); alexthomasian@gmail.com (A. Thomasian)

all n disks to fail. The time to first failure is $R_{series}(t) = \prod_{i=1}^{n} R_i(t)$. In the case of the exponential distribution $R_{series}(t) = e^{-\Lambda t}$, where $\Lambda = \sum_{i=1}^{n} \lambda_i$ and it follows that the mean time to failure is $1/\Lambda$.

Parallelism: Parallel processing systems where a job consists of n tasks with i.i.d. exponentially distributed service times, which are executed in parallel. The interest is in the mean time to complete all tasks (Sun, Peterson, 2012).

Fork/join – F/J queueing systems: We consider F/J queueing systems where each job spawns n tasks, which are processed in parallel on n servers. Jobs arrive according to a Poisson arrival process with rate λ and service times are exponentially distributed with rate μ and mean $\bar{x} = \frac{1}{\mu}$, so that each one of the n servers constitutes an M/M/1 queueing system (Kleinrock, 1975). Given that the utilization factor of the servers: $\rho = \lambda \bar{x} = \lambda/\mu$ is less than one, the mean task response time is $R(\rho) = \bar{x}/(1-\rho) = (\mu - \lambda)^{-1}$ and the task response time is exponentially distributed: $F(t) = 1 - e^{-t/R(\rho)}$ (see Kleinrock, 1975). That the mean fork-join response time $R_n^{F/J}(\rho)$ equals the maximum of n response time $R_n^{max}(\rho)$ is not true, but rather the latter is an upper bound to the former in the case of the exponential distribution (Nelson, Tantawi, 1988). There is an exact solution for $R_n^{F/J}$ only for $n = 2$: $R_2^{F/J}(\rho) = (1.5 - \rho/8))R(\rho)$ (Nelson, Tantawi, 1988), which is less than $R_2^{max}(\rho) = H_2 R(\rho)$ where $H_n = \sum_{k=1}^{n} 1/k$ is the Harmonic sum. An approximation to $R_N^{F/J}$ is obtained in (Nelson, Tantawi, 1988):

$$R_n^{F/J}(\rho) \approx \left[\frac{H_K}{H_2} + \left(1 - \frac{H_K}{H_2}\right)\frac{4\,\rho}{11}\right] R_2^{F/J}(\rho), \quad 2 \leq n \leq 32.$$

In a RAID level 5 (RAID5) disk loads are balanced via striping, i.e., partitioning large files into data strips which are placed round-robin across the n disks of the array, with one strip per row (stripe) dedicated to a parity strip which holds the eXclusive OR- XOR of the corresponding bits at the other strips in the row (Thomasian, Blaum, 2009). According to the left symmetric organization parity strips are placed in repeating right to left diagonals to balance the parity update load for *OnLine Transaction Processing* (OLTP) applications generate read/write requests to randomly placed small disk blocks. Such disk requests are expensive to process, since in addition to transfer time they incur both seeks and rotational delays (Thomasian, Fu, Han, 2007). When a single disk fails the rate of read accesses to the n-1 surviving disks is doubled, which is due to F/J requests for reconstructing missing blocks in addition to disk's own read requests to the disks. There is a smaller load increase for writes. Although disk service times are not exponentially distributed, such an assumption was used in (Menon, 1994). The performance of RAID5 and RAID6 arrays with an OLTP workload in normal and degraded modes is analyzed with general service times, i.e., an M/G/1 queueing model, in (Thomasian, Fu, Han, 2007). Rebuild processing in RAID5 systematically reconstructs the contents of the failed disk on a spare disk, so that according to read redirection, read requests directed to the failed disk are processed directly from the spare disk provided the requested data block has already been reconstructed on the spare disk (Thomasian, Blaum, 2009). As rebuild progresses the fraction of F/J read requests drops from 50% of read requests to the disk to zero. It can be observed from simulation results in (Thomasian, Tantawi, 1994) that with Poisson arrivals and general service times $R_n^{max}(\rho)$ remains an upper bound $R_n^{F/J}(\rho)$. For smaller fractions of F/J accesses $R_n^{F/J}(\rho) \approx R_n^{max}(\rho)$, where the latter is much easier to compute. The paper is organized as follows. In Section 2 we first provide the formula for calculating the expectation of the maximum of independent exponentially distributed random variables. We obtain the formula for the maximum of n i.i.d. exponentials using two different methods in Sections 3 and 4, while a third method is given in Appendix I.

The LST for the maximum is derived in Section 5. In Appendix II we provide the LST for exponentials with different rates based on the analysis in (Harrison, S. Zertal, 2007). Conclusions appear in Section 6.

2. Maximum of Exponential Independent Random Variables

Let $X_1, \dots, X_n$ be n independent random variables with exponential distribution $\varepsilon(\lambda_i)$ for $i = 1, \dots, n$. We want to study

$$X(n) = max_{1 \leq i \leq n}(X_i).$$

First, we need the distribution function F_n of $X(n)$, which by independence is:

$$F_n(t) = P(X(n) \leq t) = \prod_{i=1}^{n} P(X_i \leq t) = \prod_{i=1}^{n} \left(1 - e^{-\lambda_i t}\right) I_{t\geq 0}, \qquad \forall t \in R,$$

where $I_{t\geq 0}$ is the indicator function of the event $t \geq 0$. The product can be expanded:

$$\prod_{i=1}^{n} \left(1 - e^{-\lambda_i t}\right) = 1 + \sum_{k=1}^{n-1} (-1)^k \sum_{|E|=k}^{n-1} e^{-t \sum_{l\in E} \lambda_l}$$

The sum $\sum_{|E|=k}$ is a sum for all subsets of $\{1,\dots,n\}$ with k elements. It means

$$\prod_{i=1}^{n} (1 - e^{-\lambda_i t}) = 1 - \sum_{i=1}^{n} e^{-t\lambda_i} + \sum_{1\leq i<j\leq n} e^{-t(\lambda_i+\lambda_j)} - \dots + (-1)^n e^{-t\sum_{i=1}^{n}\lambda_i}.$$

Now, we are able to calculate the expectation of $X(n)$:

$$E[X(n)] = \int_{0^-}^{+\infty} \left(1 - F(t)\right) dt = \int_{0^-}^{+\infty} \left(1 - \prod_{i=1}^{n} (1 - e^{-\lambda_i t})\right) dt =$$

$$= \sum_{k=1}^{n+1} (-1)^k \sum_{|E|=k} \int_{0^-}^{+\infty} e^{-t\sum_{l\in E}\lambda_l} dt = \sum_{k=1}^{n+1} (-1)^k \sum_{|E|=k} \frac{1}{\sum_{l\in E}\lambda_l}.$$

Therefore

$$E[X(n)] = \sum_{i=1}^{n} \frac{1}{\lambda_i} - \sum_{1\leq i<j\leq n} \frac{1}{\lambda_i + \lambda_j} + \sum_{1\leq i<j<k\leq n} \frac{1}{\lambda_i + \lambda_j + \lambda_k} - \dots + (-1)^{n-1} \frac{1}{\sum_{i=1}^{n}\lambda_i}$$

3. The First Proof for the Maximum of n i.i.d. Exponentials

Now, we assume that the variables $(X_i)_{i\in N}$ are i.i.d with distribution $\varepsilon(\lambda)$. In this particular case, the expectation can be greatly simplified. For each k, there are $\binom{n}{k}$ subsets of $\{1,\dots,n\}$ with k elements. So we have

$$E[X(n)] = \frac{-1}{\lambda} \sum_{k=1}^{n} \binom{n}{k} (-1)^k \frac{1}{k}.$$

This sum hides a well known sum: the harmonic sum.

LEMMA:

$$\sum_{i=1}^{n} \frac{1}{i} = \sum_{i=1}^{n} \frac{1}{i} (-1)^{i-1} \binom{n}{i} \qquad (1)$$

holds for any natural n.

PROOF: We prove it by the method of mathematical induction. For $n = 1$, the identity is true: 1=1. To complete the inductive step of a proof using the principle of mathematical induction, we assume that (1) is true for an arbitrary positive integer n and show that under this assumption,

$$\sum_{i=1}^{n+1} \frac{1}{i} = \sum_{i=1}^{n+1} \frac{1}{i} (-1)^{i-1} \binom{n+1}{i}$$

must also be true (that is we assume that it is true for a natural n and prove it for $n + 1$). Since

$$\sum_{i=1}^{n+1} \frac{1}{i} = \sum_{i=1}^{n} \frac{1}{i} + \frac{1}{n+1},$$

using inductive step, we get

$$\sum_{i=1}^{n+1} \frac{1}{i} = \sum_{i=1}^{n} \frac{1}{i} (-1)^{i-1} \binom{n}{i} + \frac{1}{n+1}.$$

Therefore we have to prove the following equivalent identity:

$$\sum_{i=1}^{n} \frac{1}{i} (-1)^{i-1} \binom{n}{i} + \frac{1}{n+1} = \sum_{i=1}^{n+1} \frac{1}{i} (-1)^{i-1} \binom{n+1}{i}$$

or

$$\frac{1}{n+1} = \sum_{i=1}^{n} \frac{1}{i} (-1)^{i-1} \left[\binom{n+1}{i} - \binom{n}{i}\right] + (-1)^n \frac{1}{n+1}$$

It is not difficult to see that the difference of two combinations has the form:

$$\binom{n+1}{i} - \binom{n}{i} = \frac{n!\ i}{i!\,(n-i)!\,(n-i+1)}.$$

Therefore, we can rewrite (1) to the form:

$$\frac{1}{n+1} - (-1)^n \frac{1}{n+1} = \sum_{i=1}^{n} (-1)^{i-1} \frac{n!}{i!\,(n-i)!\,(n-i+1)!}$$

or

$$\frac{1}{n+1} - (-1)^n \frac{1}{n+1} = \frac{1}{n+1} \sum_{i=1}^{n} (-1)^{i-1} \frac{n!}{i!\,(n-i+1)!}.$$

Canceling by $\frac{1}{n+1}$ we come to the following equivalent identity:

$$1 + (-1)^{n+1} = \sum_{i=1}^{n} (-1)^{i-1} \binom{n+1}{i}$$

or

$$\sum_{i=0}^{n+1} (-1)^{i-1} \binom{n+1}{i} = (1-1)^{n+1} = 0$$

The proof is complete.

4. The Second Proof for the Maximum of *n* i.i.d Exponentials

This proof is inspired by the way of calculating $\sum_{k=0}^{n} \binom{n}{k} (-1)^k \frac{1}{k+1}$. The goal is to find a function whose derivative or primitive integral is linked to the sum we want to calculate.

We introduce $g_n(x) = \frac{(1-x)^n - 1}{x}$ and $G_n(x) = \int_{0^-}^{x} g_n(t)\, dt$. The domain of g can be extended by continuity in zero and we have :

$$g_n(x) = \sum_{k=1}^{n} \binom{n}{k} (-1)^k x^{k-1} \quad and \quad G_n(x) = \sum_{k=1}^{n} \binom{n}{k} (-1)^k \frac{x^k}{k}$$

So $\mathrm{E}[\xi_n] = \frac{-1}{\lambda} G_n(1)$. As we know how to factorize $a^n - b^n$, we have:

$$g_n(x) = \frac{(1-x) - 1}{x} \sum_{k=0}^{n-1} (1-x)^k = - \sum_{k=0}^{n-1} (1-x)^k .$$

And $\int_{0^-}^{1} (1-t)^k\, d\,t = \frac{1}{k+1}$, so we have the result:

$$E[X(n)] = \frac{1}{\lambda} \sum_{i=1}^{n} \frac{1}{i}.$$

5. The LST for the Maximum of *n* Exponentials

Another interesting characterization of $X(n)$ is its LST. For η a nonnegative random variable, we define its Laplace's transform by the following formula

$$L_\eta :\ s \mapsto E\left[e^{\{-s\eta\}}\right]$$

If η has a density function f_η , we have

$$\forall\, s \in R_+, \qquad L_\eta(s) = \int_{0^-}^{+\infty} e^{-sx} f_\eta(x) d\,x$$

In the case of $X(n)$, we have

$$\forall\, s \in R_+, \qquad L_{\xi_n}(s) = \int_{0^-}^{+\infty} e^{-sx} f_\eta(x) d\,x$$

$$= \left[e^{-sx}\left(F_\xi(x) -\right)\right]_0^{+\infty} + s \int_{0^-}^{+\infty} e^{-sx} \left(1 - F_\xi(x)\right) dx$$

$$= -s\int_{0^-}^{+\infty}\left(e^{-sx}\sum_{k=1}^{n}(-1)^k\sum_{|E|=k}e^{-x\sum_{l\in E}\lambda_l}\right)dx = -s\sum_{k=1}^{n}(-1)^k\sum_{|E|=k}(-1)^k\frac{1}{\sum_{l\in E}\lambda_l+s}$$

Therefore, we obtain

$$L_{\xi_n}(s) = s\left(\sum_{i=1}^{n}\frac{1}{\lambda_i+s} - \sum_{1\le i<j\le n}\frac{1}{\lambda_i+\lambda_j+s} + \cdots + (-1)^{n-1}\frac{1}{\sum_{i=1}^{n}\lambda_i+s}\right)$$

In the particular case of i.i.d. random variables, we get

$$L_{\xi_n}(s) = s\sum_{k=1}^{n}\binom{n}{k}(-1)^{k+1}\frac{1}{k\lambda+s}.$$

The above equation can be used to obtain the moments of $X(n)$. The first and second moment for $n=2$ taking the first and second derivatives of the LST is $1.5/\lambda$ and $3.5/\lambda^2$, respectively (Kleinrock, 1975), so that the variance is $1.25/\lambda^2$.

Noting that the exponentials are mutually independent the variance of the maximum can be obtained directly (Trivedi, 2002):

$$Var[X(n)] = \sum_{i=1}^{n}VarX_i = \frac{1}{\lambda^2}\sum_{i=1}^{n}\frac{1}{i^2}.$$

The analysis in (Harrison, S. Zertal, 2007) for the Laplace transform for exponentials with different rates is given in Appendix II.

6. Conclusion

We have presented three methods to compute the expected value of the maximum of n i.i.d. exponentials and their Laplace transform, which can be used to compute higher moments of the maximum.

Appendix I: The third Proof for the Maximum of n i.i.d Exponentials

The method described here to obtain $E[X]$ where $X = max(X_1, X_2, \dots, X_n)$ and $F_{X_i} = 1 - e^{-\lambda_i t}$

$$F_X(t) = \prod_{i=1}^{n}F_{X_i}(t)\ .$$

Let $R_{X_i}(t) = 1 - F_{X_I}(t)$, we have

$$R_X(t) = 1 - \prod_{i=1}^{n}[1 - R_{X_i}(t)]\ .$$

Given $\lambda_i = \lambda,\ 1 \le i \le n$

$$E[X] = \int_{0^-}^{\infty}[1 - (1-e^{-\lambda t}]\,dt\ .$$

Let $u = 1 - e^{-\lambda t}$ so that $-\lambda t = ln(1-u)$ and $dt = [1/\lambda(1-u)]\,du$

$$E[X] = \frac{1}{\lambda}\int_{0^-}^{1}\frac{1-u^n}{1-u}du$$

$$E[X] = \frac{1}{\lambda}\int_{0^-}^{1}\sum_{i=1}^{n}u^{i-1}\,du = \frac{1}{\lambda}\sum_{i=1}^{n}\int_{0}^{1}u^{i-1}\,du = \int_{0^-}^{1}u^{i-1}\,du = \frac{u^i}{i}|_0^1 = \frac{1}{i}$$

$$E[X] = \frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i} = \frac{H_n}{\lambda}$$

Appendix II: The LST for the Maximum of n Exponentials with Different Rates

According to (Harrison, S. Zertal, 2007) the maximum of n r.v.'s with negative exponential distribution with parameters $\underline{\lambda} = (\lambda_1, \dots, \lambda_n)$ can be expressed as the recurrence for $1 \le m \le n$ (where $\backslash j$ indicates the exclusion of λ_j)

$$\left(s+\sum_{j=1}^{m}\lambda_j\right)L_m\,(\underline{\lambda},s\,)=\sum_{j=1}^{m}\lambda_j\;L_{m-1}\,(\underline{\lambda}_{\setminus j},s).$$

The k^{th} moment of the maximum of K exponentially distributed i.i.d. r.v. is then

$$M_n\,(\underline{\lambda},k)=\frac{k\,M_n\left(\underline{\lambda},\;k-1\right)}{\sum_{j=1}^{K}\lambda_j}+\frac{\sum_{j=1}^{K}M_{n-1}(\underline{\lambda}_{\setminus j},k)}{\sum_{j=1}^{K}\lambda_j}.$$

Acknowledgments

The research of the first author was partially supported by the Mathematical Studies Center of Yerevan State University. The research of the second author was partially supported by the Fulbright Association scholarship.

References

Gibson, 1992 – Gibson G.A. (1992). Redundant Disk Arrays - Reliable, Parallel Secondary Storage. MIT Press.

Harrison, S. Zertal, 2007 – Harrison P. G. and Zertal S. (August 2007), Queueing models of RAID systems with maxima of waiting times. Performance Evaluation 64 (7-8): 664-689.

Kleinrock, 1975 – Kleinrock L. (1975). Queueing Systems, Vol. I: Theory, Wiley-Interscience.

Menon, 1994 – Menon J. (1994). Performance of RAID5 disk arrays with read and write caching, Distributed and Parallel Databases 2 (3): 261 - 293.

Nelson, Tantawi, 1988 – Nelson R. D. and Tantawi A.N. (June 1988). Approximate analysis of fork-join synchronization in parallel queues. IEEE Trans. Computers C-37(6): 736-743.

Schroeder, Gibson, 2007 – Schroeder B. and G. A. Gibson (2007). Understanding disk failure rates: What does an MTTF of 1,000,000 hours mean to you? ACM Trans. on Storage -TOS 3(3): Article No. 8.

Sun, Peterson, 2012 – Sun J. and Peterson G. D. (Nov. 2012). An effective execution time approximation method for parallel computing. IEEE Trans. Parallel and Distributed Systems-TPDS 24 (11): 2025-2032.

Thomasian, Tantawi, 1994 – Thomasian A. and Tantawi A.N. (1994), Approximate solutions for M/G/1 fork/join synchronization. Proc.26th Winter simulation Conf.-WSC, Lake Buena Vista, FL,Dec., 361-368.

Thomasian, Fu, Han, 2007 – Thomasian A., Fu G. and Han C. (June 2007). Performance of two-disk failure-tolerant disk arrays. IEEE Trans. Computers 56 (6): 799 – 814.

Thomasian, Blaum, 2009 – Thomasian A. and Blaum M. (November 2009). Higher reliability redundant disk arrays: Organization, operation, and coding. ACM Trans. on Storage - TOS 5(3): Article No. 7.

Thomasian, 2014 – Thomasian A. (August 2014). Analysis of fork/join and related queueing systems. ACM Computing Surveys 47(2): 17:1-17:71.

Trivedi, 2002 – Trivedi K. S. (2002). Probability and Statistics with Reliability, Queueing, and Computer Science Applications, 2nd ed., Wiley-Interscience.

УДК 004

Среднее значение максимума и преобразование Лапласа для независимых и одинаково распределенных экспонент

Виктор Кароевич Оганян [a, *], Александр Томасян [b]

[a] Ереванский Государственный Университет, Армения
[b] Американский университет Армении, Ереван, Армения

[*] Корреспондирующий автор
Адреса электронной почты: victo@aua.am (В.К. Оганян); alexthomasian@gmail.com (А. Томасян)

Аннотация. Мы приводим три метода для вычисления среднего значения для максимума n независимых и одинаково распределенных экспонент, а также вычисляем его преобразование Лапласа-Стильтеса. Результат применяется в следующих случаях: (1) время потери данных в дисковых массивах с-полосной репродукцией, где время до выхода из строя диска имеет экспоненциальное распределение; (2) время окончания n экспоненциально распределенных параллельных задач; (3) верхняя граница среднего времени ожидания для параллельного соединения, когда поступления распределены по Пуассону, времена обслуживания экспоненциальны, так что времена ожидания экспоненциально распределены.

Ключевые слова: Комбинаторный анализ, порядковая статистика, анализ эксплуатационных качеств, система очередей, параллельный процесс, не параллельные системы массового обслуживания, оценка надежности.