THE EXPONENTIAL DISTRIBUTION AND THE APPLICATION TO MARKOV MODELS
Usman Yusuf Abubakar
Department of Mathematics/Statistics Federal University of Technology, Minna, Nigeria.
Email: abbkruy@yahoo.com

Abstract
This paper discusses the characteristics of the exponential distribution and the related distribution functions including gamma, weibull and lognormal then relates some of their properties to the application of Markov models. One of the major properties is forgetfulness, the consequence of this is that Markov and stationarity assumptions imply that the times between events must be negative-exponentially distributed. To make a decision on the application of Markov model to any process in real life situation, it is advised that it should be fitted to the form of the negative exponential density functions which implies that the most likely times are close to zero, and very long times are increasingly unlikely. That is, the most likely values are considered to be clustered about the mean, and large deviations from the mean are viewed as increasingly unlike. If this characteristic of the negative exponential distribution seems incompatible with the application one has in mind then a Markov model may not be appropriate.

Keywords; Exponential distribution, Random variables, Memory-less, Markov models, Stationarity assumption, Application.

Introduction
In making mathematical models for a real-world phenomenon it is always necessary to make certain simplifying assumptions so as render the mathematics tractable. One simplifying assumption that is often made when modeling with Markov principle is to assume that certain random variables are exponentially distributed. The reason for this is that exponential distribution is both relatively easy to work and is often a good approximation to the actual distribution. An important simplifying assumption in making Markov chain models is that the time it takes to make a transition (random variable) be described by negative-exponential distribution. In some of the applications Abubakar(2007) utilized both the exponential and Weibull respectively to describe the waiting time in the states of semi-Markov model for leprosy treatment. Also Abubakar (2010) considered as a random variable the time it takes for sahel savannah to be transformed to sahel savannah using Weibull distribution function in a semi-Markov model for desertification. It is therefore of interest in this paper to examine the exponential distribution function and its application to modeling Markov processes.

The exponential distribution
The probability density function of the random variable T having the exponential distribution is
\[
f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}
\]
Kohlas(1982). The distribution has a parameter \( \lambda \), which also determines the shape of the distribution. The mean \( \mu \) of the exponential distribution is
\[\mu = E(T) = \int_0^\infty t \lambda e^{-\lambda t} dt = \int_0^\infty t \lambda e^{-\lambda t} dt\]
Substituting \( w = \frac{t}{\lambda} \), \( t = w/\lambda \) and \( dw = \frac{1}{\lambda} dt \) in the integrand gives
\[\mu = \int_0^\infty w e^{-w} dw = \int_0^\infty \frac{1}{\lambda} \lambda e^{-w} dw = \frac{1}{\lambda} \int_0^\infty e^{-w} dw = \frac{1}{\lambda} \int_0^\infty e^{-w} dw = \frac{1}{\lambda}.
\]
Consider
\[E(T^2) = \int_0^\infty t^2 \lambda e^{-\lambda t} dt\]
Let \( w = \frac{t}{\lambda} \), \( t = \lambda w, \) and \( dw = \frac{1}{\lambda} dt \) so that
\[E(T^2) = \int_0^\infty \lambda^2 e^{-\frac{1}{\lambda} w} dw = \frac{w^2}{\lambda} e^{-w} = \frac{2}{\lambda} \int_0^\infty e^{-w} dw = \frac{2}{\lambda}.
\]
The variance \( \sigma^2 \) of the exponential distribution is therefore given by
\[\sigma^2 = E(T^2) - (E(T))^2 = 2 - \frac{1}{\lambda^2}.
\]
The mean is \( 1/\lambda \), and the variance is \( (1/\lambda)^2 \). Thus the mean and variance are not separately adjustable, as one may frequently desire.
The survivor function $S(t)$ is given by
$$S(t) = P(T > t)$$
and it is the probability that an equipment has survived up to time $t$.

Suppose $F(t)$ is the cumulative distribution function of the random variable $T$. Then
$$S(t) = 1 - F(t)$$
For the exponential distribution
$$F(t) = P(T \leq t) = \int_0^t f(s)ds$$
$$= \int_0^t \lambda e^{-\lambda s} ds$$
$$= 1 - e^{-\lambda t}$$
Therefore
$$S(t) = e^{-\lambda t}$$
The failure rate or the hazard rate, ‘$h$’ associated with the random variable $T$ is given by
$$h(t) = \frac{f(t)}{S(t)}$$
For the exponential distribution, the hazard rate is given by
$$h(t) = \lambda$$
This explains that the failure rate for an exponential random variable is constant. Let us consider the conditional probability $P(t < T < t + \Delta t / T > t)$. That is, the probability that the equipment will fail during the next $\Delta t$ time units, given that it survived at time $t$. Using the definition of conditional probability, we have
$$P(t < T < t + \Delta t / T > t) = \frac{P(t < T < t + \Delta t)}{P(T > t)}$$
$$= \frac{\int_0^{t+\Delta t} f(s)ds}{p(T > t)}$$
$$= \frac{\Delta t f(\xi)}{s(t)}$$
Where $t < \xi < t + \Delta t$.

Figure 1 plots this function for three values of $\lambda$. Notice that the function intercepts the vertical axis at $\lambda$, that it diminishes monotonically to zero (asymptotically), and that the rate of convergence is proportional to $\lambda$. The total area under the curve is, of course, always equal to 1, as it must be for any density function.

Most applications are based on its ‘memory-less’ property, when the measurement variable $T$ has a time dimension. This property refers to the phenomenon in which the history of the past events does not influence the probability of occurrence of present or future events. According to Ross (1989) A random variable $X$ is said to be without memory, or memory-less, if
P\{X > s + t | X > t\} = p\{X > s\} \text{ for all } s, t \geq 0 \quad (1)

If we think of X as being the lifetime of an instrument, then Equation (1) states that the
probability that the instrument lives for at least s + t hours given that it has survived t hours
is the same as the initial probability that it lives for at least s hours. In other words, If the instrument is alive
at time t, then the distribution of the remaining amount
of time that it survives is the same as the
original lifetime distribution, that is, the instrument does not remember that it has already been in use for
a time t. The condition in Equation (1) is equivalent to
\[
P\{X > s + t | X > s\} = P\{X > s\}\]

Or
\[
P\{X > s + t\} = P\{X > s\} P\{X > t\} \quad (2)
\]

Since Equation (2) is satisfied when X is
exponentially distributed for
\[
\frac{e^{-x(t+s)}}{e^{-x(t+t)}} = e^{-x(s+t)},
\]
follows that exponentially
distributed random variables are memory-less and
forgettable. Following Feller (1971),” We shall refer to
this lack of memory as the Markov property of the
exponential distribution”.

Feller (1971) has identified the following as a
characteristic property of the exponential distribution.
Let \((X_1, X_2) \in \Omega \text{ and } (Y_1, Y_2) \in \Omega \text{ and } \Omega = \{0, 1, 2\} \text{ are nonempty regions, and let } X_1 \text{ and } X_2 \text{ be two independent random variables with densities}
\]

and \(f_2\) and denote the density of their sum \(S = X_1 + X_2\) by \(g\). The pairs \((X_1, S)\) and \((X_2, X_2)\) are related by their linear transformation \(X_1 = x_1, x_2 = S - X_1\) with
determinant 1 and Since the events \((X_1, X_2, S) \in \Omega \text{ and } (Y_1, Y_2) \in \Omega \text{ are identical it is seen that the joint
density of } (Y_1, Y_2)\) is given by
\[
g(y_1, y_2) = f(a_{11}y_1 + a_{12}y_2 + a_{21}y_1 + a_{22}y_2) \Delta.
\]

Where \(\Delta = a_{11}a_{22} - a_{12}a_{21} > 0 \quad (3)\)

the joint density of the pair \((X_1, S)\) is given by
\[
f_1(x)f_{2s}(s - x_1)\text{ Integrating over all } x \text{ we obtain the marginal density } g \text{ of } S. \text{ The conditional density } u_2 \text{ of } X_1 \text{ given that } S = s \text{ satisfies}
\]
\[
u_2(x) = f_{1}(x) f_{2s}(s - x_1) g(s) \Delta(1_f(x) - a e^{-ax})\]

In the special case of exponential densities
\[
f_1(x) = f_{2}(x) = a e^{-ax} \text{ (where } x > 0) \text{ we get } u_2(x) = \frac{1}{a} \text{ for } 0 < x < s \text{. in other words, given that } X_1 + X_2 = s \text{ the variable } X_4 \text{ is uniformly distributed over the interval } (0, s) \text{. intuitively speaking, the knowledge that } S = s \text{ gives us no clue as to the possible position of the random point } X_4 \text{ within the interval } (0, s). \text{ This result conforms with the notion of complete randomness inherent in the exponential distribution. Also cited from Feller (1971) on the randomness of the exponential distribution. Let } X_1, \ldots, X_n \text{ be independent with the common density} \frac{ax^{a-1} e^{-ax}}{e^{-ax}} \text{ for } x > 0. \text{ Put } S_j = X_1 + \ldots + X_j. \text{ Then } (S_1, S_2, \ldots, S_n + s) \text{ is obtained from } \left(\frac{X_1}{X_2}, \ldots, X_n\right) \text{ by a linear}
transformation of the form
\]

where determinant \(\Delta = a_1(2a - a_1:a_2 > 0)\) with determinant 1. Denote by \(\Omega\) the “octant” of
points \(x_j > 0\). The density of \((X_1, \ldots, X_n)\) is
concentrated on \(\Omega\) and is given by
\[
g(x_1, \ldots, x_n) = g_1(x_1) g_2(x_2) \ldots g_n(x_n) \Delta.
\]

If \(x_j > 0\). The variables \(S_1, \ldots, S_n\) map \(\Omega\) onto
the region \(\Omega\) defined by \(0 < s_1 < s_2 < \ldots < s_n < s\) and see (3) within \(\Omega\). The density of
\((S_1, \ldots, S_n + s)\) is given by
\[
g_1(x_1) g_2(x_2) \ldots g_n(x_n) \Delta.
\]

The marginal density of \(S_n\) is
known to be the gamma density
\[
g(x_n) = \frac{1}{\Gamma(s_n) \Gamma(s_n)} x^{s_n - 1} e^{-x} / n!\]
and hence the conditional density of the \(n\) \text{ tuple} \((S_1, \ldots, S_n + s)\) given that \(S_n + s = s\) equals \(n! \Gamma(s_n)\) for \(0 < s_1 < s_2 < \ldots < s_n < s\) (and zero elsewhere). In other words, given that \(S_n + s = s\) the
variables \(S_1, \ldots, S_n\) are uniformly distributed over their possible range. We may say that given \(S_n + s = s\), the variables \(S_1, \ldots, S_n\) represent \(n\) points chosen independently and at random in the interval
\([0, s)\) numbered in their natural order from left to
right.

Following Tijms(1988), Let \(X\) be a positive random variable with the probability distribution function \(F(x)\)
having finite mean \(E(x)\) and finite standard deviation
\(\sigma(X)\). The random variable \(X\) may represent the
lifetime of some item or the time to complete some
task. The coefficient of variation of the positive
random variable \(X\) is defined by
\[
c_X = \frac{\sigma(X)}{E(X)}
\]

In applications one often works with the squared
coefficient of variation \(c_X^2\) rather than with \(c_X\). The
(squared) coefficient of variation is a measure of the
variability of the random variable \(X\). For example, the
deterministic distribution has \(c_x^2 = 0\), the exponential
distribution has \( c_x^2 = 1 \) and the Erlang-k distribution has the intermediate value \( c_x^2 = \frac{1}{k} \).

A random variable with increasing (decreasing) failure rate has a property that its coefficient of variation is smaller(larger) than or equal to 1. The failure rate is a concept that enables us to discriminate between distribution functions on physical considerations.

We now discuss the gamma, lognormal and Weibull distributions for a positive random variable \( T \) and identify the relevant properties of these distributions to the negative-exponential distribution.

The gamma distribution: The density \( f(t) \) given by

\[
f(t) = \frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}, \quad t \geq 0,\]

where the shape parameter \( \alpha \) and the scale parameter \( \lambda \) are both positive. Here \( \Gamma(\alpha) \) is the complete gamma function defined by

\[
\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} \, dx, \quad \alpha > 0.
\]

And having the property \( \Gamma(\alpha+1) = \alpha \Gamma(\alpha) \) for any \( \alpha > 0 \). The probability distribution function \( F(t) \) may be written as

\[
F(t) = \frac{1}{\Gamma(\alpha)} \int_0^t e^{-x} x^{\alpha-1} \, dx, \quad t \geq 0.
\]

The latter integral is known as the incomplete gamma function.

If the shape parameter \( \alpha \) is a positive integer \( k \), the gamma distribution is the well-known Erlang-k \( (E_k) \) distribution for which

\[
f(t) = \frac{\lambda^k}{k!} t^{k-1} e^{-\lambda t} \quad \text{and} \quad F(t) = 1 - \sum_{j=0}^{k-1} \frac{(-\lambda t)^j}{j!}, \quad t \geq 0.
\]

The Erlang-K distribution has a very useful interpretation. A random variable with an Erlang-K distribution can be represented as the sum of \( k \) independent random variables having a common exponential distribution.

The mean and the squared coefficient of variation of the gamma distribution are given by

\[
E(X) = \frac{\alpha}{\lambda} \quad \text{and} \quad c_x^2 = \frac{1}{\lambda}.
\]

Since \( E(X) \) and \( c_x^2 \) can assume arbitrarily positive values for the gamma density, a unique gamma distribution can be fitted to each positive random variable with given first two moments. To characterize the shape and the failure rate of the gamma density, we distinguish between the cases \( c_x^2 < 1 \) (\( \alpha > 1 \)) and \( c_x^2 \geq 1 \) (\( \alpha \geq 1 \)).

The gamma density is always unimodal, that is the density has only one maximum. For the case of \( c_x^2 < 1 \) the density first increases to the maximum at \( t = (\alpha - 1)/\alpha \lambda > 0 \) and next decreases to zero as \( t \to \infty \), whereas for the case of \( c_x^2 \geq 1 \) the density has its maximum at \( t = 0 \) and thus decreases from \( t = 0 \). The failure rate function is increasing from zero to \( \lambda \) if \( c_x^2 < 1 \) and is decreasing from infinity to zero if \( c_x^2 \geq 1 \). The exponential distribution \( (c_x^2 = 1) \) has a constant failure rate \( \lambda \) and is a natural boundary between the cases \( c_x^2 < 1 \) and \( c_x^2 > 1 \).

The lognormal distribution: The density \( f(t) \) is given by

\[
f(t) = \frac{1}{\alpha \sqrt{2\pi} t} e^{-[(\ln(t)-\lambda)^2]/2\alpha^2}, \quad t > 0,
\]

where the shape parameter \( \alpha \) is a positive and the scale parameter \( \lambda \) may assume each real value. The probability distribution function \( F(t) \) equals

\[
F(t) = \phi \left( \frac{\ln(t) - \lambda}{\alpha} \right) t > 0,
\]

where \( \phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} \, du \) is the standard normal probability distribution function. The mean and the squared coefficient of variation of the lognormal distribution are given by

\[
E(X) = e^{\lambda + \alpha^2/2} \quad \text{and} \quad c_x^2 = e^{2\alpha^2} - 1.
\]

Thus a unique lognormal distribution can be fitted to each positive random variable with given first two moments.

The Weibull distribution: The density \( f(t) \) is given by

\[
f(t) = \alpha \lambda \left( \frac{t}{\lambda} \right)^{\alpha-1} e^{-\alpha t}, \quad t > 0.
\]

With the shape parameter \( \alpha > 0 \) and scale parameter \( \lambda > 0 \). It is observed that the Weibull distribution reduces to the exponential distribution when the shape parameter is unity.

The probability distribution function \( F(t) \) equals

\[
F(t) = 1 - e^{-\alpha t}, \quad t \geq 0.
\]

The mean and the squared coefficient of variation of the Weibull density are given by

\[
E(X) = \frac{1}{\Gamma \left( 1 + \frac{1}{\alpha} \right)} \quad \text{and} \quad c_x^2 = \frac{\Gamma \left( 1 + 2/\alpha \right)}{\left( \Gamma \left( 1 + 1/\alpha \right) \right)^2} - 1.
\]

A unique Weibull distribution can be fitted to each positive random variable with given first two moments. For that purpose a non-linear equation in \( \alpha \) must be solved numerically. The Weibull density is
always unimodal with a maximum at $t = \frac{\lambda}{\lambda - 1}$ if $c_2^2 \geq 1 (\alpha > 1)$ and at $t = 0$ if $c_2^2 \leq 1 (\alpha < 1)$. The failure rate function is increasing from 0 to infinity if $c_2^2 \leq 1$ and is decreasing from infinity if $c_2^2 > 1$.

The gamma and weibull densities are similar in shape, and for $c_2^2 < 1$ the lognormal densities takes on shapes similar to the gamma and weibull densities. For the case of $c_2^2 \geq 1$ the gamma and weibull densities have their maximum value at $t = 0$ so that most outcomes will be small and very large outcomes occur only occasionally. The lognormal density tends to zero as $t \to 0$ faster than any power of $t$, and thus the lognormal distribution will typically produce fewer small outcomes than the other two distributions. This latter fact explains the popular use of the lognormal distribution in actuarial studies. The differences between the gamma, weibull and lognormal densities become most significant in their tail behavior. The densities for large $t$ go down like $\frac{1}{\lambda t}, \frac{1}{t^\alpha}$ and $\frac{1}{(\log t)^2/2\alpha^2}$. Thus for given mean and coefficient of variation the lognormal density has always the longest tail only if $\alpha > 1$, that is only if its coefficient of variation is less than one. In the figure 2, we illustrate these facts by drawing the gamma, weibull and lognormal densities for $c_2^2 = 0.25$, where $E(X) = 1$ is taken.

![Graph showing gamma, lognormal, and weibull densities](image)

**Fig 2** The gamma, lognormal and weibull densities with $E(x) = 1$ and $c_2^2 = 0.25$. Source, Tijms (1988).

We discuss some useful extensions of Erlangian (exponential) distributions. An Erlang-$k$ $(E_k)$ distributed random variable can be represented as the sum of $k$ independent exponentially distributed random variables with the same means. A generalized Erlangian distribution is one built out of sum or mixture of exponentially distributed components, or a combination of both. A particularly convenient distribution arises when these components have the same means.
Markov process. Following Korve(1993), we may give a mathematical definition of a Markov chain as a sequence $X_0, X_1, \ldots$ of discrete random variables with the property that the conditional probability distribution of $X_{n+1}$ given $X_0, X_1, \ldots, X_n$ depend only on the value of $X_n$ but not further on $X_0, X_1, \ldots, X_n$. That is for any set of values, $h, i, \ldots, j$ in the discrete state space,

$P(X_{n+1} = j | X_0 = h, \ldots, X_n = i) = P(X_{n+1} = j | X_n = i)$

The matrix $P$ whose entries are the $P_{ij}$'s is called the transition probability matrix for the process. The above chain is a first order Markov chain. In this process, the probability of making transition to a future state does not depend on the previous state but only depend on the present state. In other words, the probability of making a transition to a future state does not depend on the past history A Markov chain with discrete state and continuous time is referred to as Markov process.

Following Howard(1960) we let $a_{ij}$ represent the transition rate of the process from state $i$ to state $j$, $i \neq j$. In a short time interval $(t, t + \Delta t)$, the process currently in state $i$ will make a transition to state $j$ with probability $a_{ij} \Delta t$, $i \neq j$. If $X_t$ is the state of the process at time $t$, then we have

$P(X_{t+\Delta t} = j | x_t = i) = a_{ij} \Delta t.$

Suppose that the transition rate do no change with time (the $a_{ij}$'s are constants) stationarity. We describe the process by a transition rate matrix $A$ with components $a_{ij}$. After some manipulation, we have

$$\frac{dP_i(t)}{dt} = \sum_{j=1}^{n} P_i(t)a_{ij} \quad i, j = 1, 2, 3...$$

With the property

$$a_{ij} = -\sum_{k \neq j} a_{jk}, \quad i, j = 1, 2, 3$$

in matrix form, we have

$$\frac{d}{dt} P(t) = P(t)A.$$

And this is the general representation of

$$\frac{dP_j(t)}{dt} = \sum_{i=1}^{n} P_i(t)a_{ij} \quad (Chapman - Kolmogorov equation).$$

It is a linear, first-order differential equation with constant coefficients – the $a_{ij}$'s and these are the elements of the transition matrix $A$ which are interpreted as the mean of negative exponential distribution. That is, distribution of time spent in state $i$ when $j$ is the $r$ state.

Semi-Markov process. The Markov process discussed above has the property that state changes can only occur at the appropriate time instants. However, given the nature of some processes, transition may not actually occur at these time instants. We therefore consider a ‘situation’ where the time between transitions may be several of units of time and where the transition time can depend on the transition that is being made. This leads to a general form of Markov process called a semi-Markov process Howard(1971).

Suppose the process enters state $i$. Let $Y_i$ be the time he spent in state $i$ before moving out of the state $i$. Then $Y_i$ is called the waiting time in state $i$.

We let $w_i(\cdot)$ be the probability distribution function of $Y_i$. Then

$$w_i(m) = P(Y_i = m) = \sum_{j=1}^{n} P_{ij}f_j(m)$$

The cumulative probability distribution $W_i(\cdot)$ and the complimentary cumulative probability distribution $W_i(\cdot)$ for the waiting times are given as follows

$$W_i(m) = P(Y_i \leq m) = \sum_{m=1}^{n} W_i(m)$$

$$= \sum_{m=1}^{n} \sum_{j=1}^{n} P_{ij}f_j(m)$$

$$= \sum_{j=1}^{n} P_j f_j(n)$$

And

$$W_i(n) = P(Y_i > n) = 1 - W_i(n) = \sum_{m=n+1}^{\infty} W_i(m)$$

$$= \sum_{m=n+1}^{\infty} \sum_{j=1}^{n} P_{ij}f_j(m)$$

$$= \sum_{j=1}^{n} P_j f_j(n)$$

(4)
\( i,j=1,2,\ldots n \) denoting states and \( m=1,2,\ldots n \) representing time.

**Interval transition probabilities.** We define \( \phi_{ij}(n) \) to be the probability that the process will be in state \( j \) in year \( n \) given that he entered state \( i \) in year zero. This is called the interval transition probability from state \( i \) to state \( j \) in the interval \( (0,n] \). Then

\[
\phi_{ij}(n) = \delta_{ij} W_i(n) + \sum_{k=1}^{n} P_{ik} \sum_{m=1}^{n} f_{ik}(m) \phi_{kj}(n-m)
\]

\[
\delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}
\]

\( W_i(n) \) is as defined in (4). We observe that the quantity \( \sum_{m=1}^{n} f_{ik}(m) \) has to be described by negative exponential distribution perhaps gamma or Weibul. In some of the applications Abubakar (2007) utilized both the exponential and Weibull respectively to describe the waiting time in the states of semi-Markov model for leprosy treatment. Also Abubakar (2010) considered as a random variable the time it takes for Sudan savannah to be transformed to Sahel savannah and consequently expressed the time with Weibull distribution function in a semi-Markov model for desertification.

**Conclusion**

It is observed that the Markov and stationarity assumptions imply that the times between events must be negative-exponentially distributed. The parameters of these distributions, the \( \lambda_{ij} \)'s may be dependent on the state occupied, \( i \), and the next state, \( j \), but all of the distributions must be of the negative exponential form. No other distribution family can even be considered as a candidate for describing the times between events.

It was mentioned that in many applications, the times between events are most naturally conceived of as having a density function of the general form shown in Fig2 (perhaps a gamma or weibull or lognormal).

That is, one tends to think in terms of some nominal value, the mean, plus or minus some relatively minor variation. Or, put another ways, the most likely values are considered to be clustered about the mean, and large deviations from the mean are viewed as increasingly unlike. However, the form of the negative exponential density functions implies that the most likely times are close to zero, and very long times are increasingly unlikely. If this characteristic of the negative exponential distribution seems incompatible with the application you have in mind, perhaps a Markov model is inappropriate. This is an important understanding to be able to distinguish between those processes which might properly be modeled as stationary Markov process and those which should not.

**References**


