Probability Theory Summary – Appendix A

Three viewpoints on the formulation of Probability Theory

a. A Priori approach – our intuitive feel for equally likely events like the roll of unbiased/not loaded dice. Events that are equally likely may not be obvious.
b. Relative-Frequency approach – repeating an experiment n times in an identical fashion/environment. Requirement is that the experiment be repeated an infinite number of times. This execution (simulation) approach has 'opened' up new areas where complexity may have been an initial problem, e.g. network performance theory (queuing theory, how the World Wide Web operates).
c. Axiomatic Definitions – entire theory is built on axioms and if done properly all the properties are well defined. Theory must be related to the real-world problems in order for it to be useful. One can use the other two methodologies to assist in understanding the problems in an axiomatic fashion.

Axiomatic Approach – foundation is three fundamental axioms of probability

1. $0 \leq P(A) \leq 1$ The probability that an event A occurs is a number between 0 and unity (1).
2. $P(S) = 1$ The probability of an certain event (the entire sample space) is unity.
3. $P(A_1 + A_2) = P(A_1) + P(A_1)$ The probability of the union of two mutually exclusive events (A_1, A_2) is the sum of the probabilities.

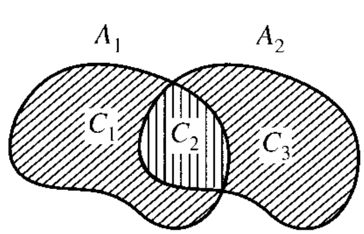
Set Theory – Venn Diagrams Georg Cantor (1845-1918) – Father of modern set theory

Union	$B = A_1 + A_2$ or $B = A_1 \cup A_2$	Boolean OR
Intersection	$B = A_1 A_2$ or $B = A_1 \cap A_2$	Boolean AND

Null Set Φ such that $P(\Phi) = 0$

Complement $A + \bar{A} = \cup$ The Universal Set

For events that are <u>not</u> mutually exclusive, use a strategy of splitting events into unions of other mutually exclusive events. Using Figure A1 as a visual aid:



$A_1 + A_2$ are not mutually exclusive (not disjoint) but

$A_1 = C_1 + C_2$ and $A_2 = C_2 + C_3$ which breaks A_1 and A_2 into two other disjoint events
$P(A_1 + A_2) = P(C_1 + C_2 + C_2 + C_3) = P(C_1) + P(C_2) + P(C_3)$ since $C_2 + C_2 = C_2$
Adding C_2 and subtracting C_2 from the equation for the union along with the probability of the union of two mutually exclusive events (A_1, A_2) results in

$$P(A_1 + A_2) = [P(C_1) + P(C_2)] + [P(C_2) + P(C_3)] - P(C_2)$$

$$= P(A_1) + P(A_2) - P(A_1A_2) \text{ since } C_2 = \text{the intersection of } A_1 \text{ and } A_2$$

You can apply this result to three or more events (Eq. A11) to have a way of dealing with events that are not mutually exclusive (most events in the real world are not mutually exclusive/disjoint, with any number of usually hidden dependencies).

<u>Conditional Probabilities</u> (or dependent probabilities)

$P(A_1A_2) = P(A_1)\,P(A_2 \mid A_1)$ where $P(A_2 \mid A_1)$ is the probability of A_2 knowing that A_1 has occurred.

If the events are **independent,** then $P(A_1A_2) = P(A_1)\,P(A_2)$ or $P(A_1 \mid A_2) = P(A_1)$

Example of a conditional probability $P(A_1 \mid A_2)$:

$P(A_1)$ = prob of a drawing the four of clubs from a deck of 52 cards ($S = 52$)
$P(A_1) = 1/52$

$P(A_1 \mid A_2)$ = prob of drawing the four of clubs knowing (A_2),
that a club has been drawn
$P(A_1 \mid A_2) = 1/13$ (in this case $S = 13$ made up from the 2,3,4,5,6,7,8,9,10,J,Q,K,A)

Generalizing Conditional Probability:

Conditional probability for the intersection of three or more events $P(A_1A_2A_3 \ldots A_n)$
is the joint product of one independent probability $P(A_1)$ and n -1 dependent probabilities

$$P(A_1A_2A_3 \ldots A_n) = P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1A_2)\,P(A_4 \mid A_1A_2A_3) \cdots P(A_n \mid A_1A_2 \cdots A_{n-1})$$

Random Variables (**X**)

Discrete - finite
Continuous – infinite (no quantization)

X is a random variable if we associate each value of **X** with an element in event **A** defined on sample space **S**.

Density and Distribution Functions – mathematical tools to predict the future

$P(x_i)$ – probability function for the *discrete* random variable **X**.

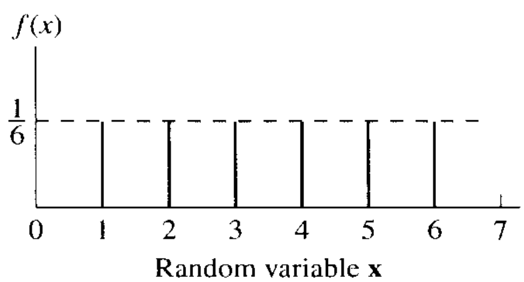
$P(\mathbf{x} = x_i) = P(x_i) = f(x_i)$ probability **density** function (PDF)

$P(\mathbf{x} \leq x) \equiv F(x) = \sum_{\mathbf{x} \leq x} f(x)$ cumulative **distribution** function (CDF)

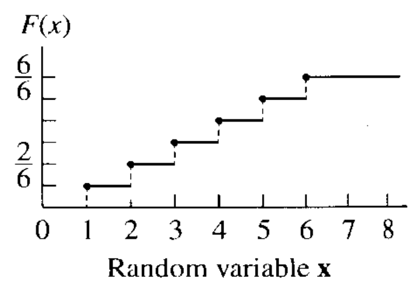
F(x) is a function defined in terms of the probability that $\mathbf{x} \leq x$

The two figures below show a good graphical representation of f(x), the density function and F(x), the distribution function for a single die.

PDF



$F(x) = P(\mathbf{x} \leq x)$ CDF



The domain of the random variable is x = 1, 2, 3, 4, 5, 6 where using the equally-likely events hypothesis, the probability of rolling any number from 1 → 6 is

$P(x = i) = 1/6$ for $i = 1 \rightarrow 6$ thus $f(x) = 1/6$ a constant density function as shown

The distribution function $P(\mathbf{x} \leq x) \equiv F(x) = x/6$ for $1 \leq x \leq 6$ An example can be seen as

the probability of rolling a 2 or less, $P(x \leq 2) = 2/6$ OR
the probability of rolling a 6 or less, $P(x \leq 6) = 1$

Which is a fancy way of saying $P(S) = \sum_{\text{Over all } i} f(x_i) = 1$ The sum of all possibly outcomes in the sample space S is 1 and can never be greater than 1. Counter example $P(\Phi) = 0$

<u>Binomial Distribution</u> (discrete probability model for the binomial density function)

Also called the Bernoulli Distribution

Associated with events that can occur (success) or not occur (failure) and thus have outcomes that can be associated with:

$\boldsymbol{p}$ = probability of success on one trial
1 – $\boldsymbol{p}$ = probability of failure on one trial
n = the number of independent trials
r = the number of successes in **n** trials

independent events that can occur or not occur in other words success (p) or failure (1 – p)
population (n) must remain constant for each trial to be independent

$$B(r; n, p) = \begin{bmatrix} n \\ r \end{bmatrix} p^r (1-p)^{n-r}$$

Author uses B(r; n, p) instead of f(x) for the density function

[expressed as *n choose r*]

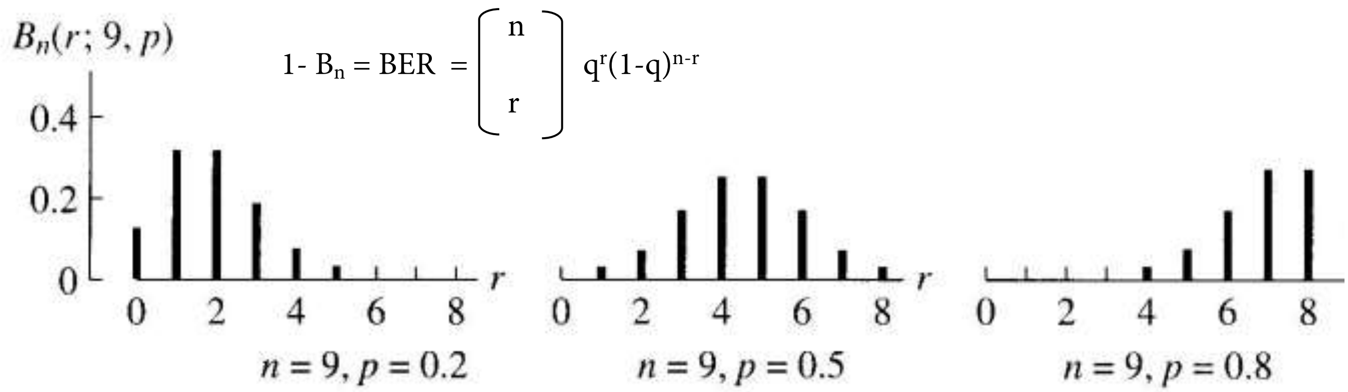
where $\begin{bmatrix} n \\ r \end{bmatrix} = \dfrac{n!}{r!\,(n-r)!}$ = the number of combinations of **n** items taken **r** at time

Distribution Example: The probability of rolling exactly two 4's in six throws of a die. (n = 6 trials, r = 2 successes where p = 1/6) Remember 0! = 1 by definition. We'll see Binomial Distributions in the reliability of M-of-N Systems.

For a communications example, we can use the bit error rate (BER = q = probability of 1 bit in error), substitute q (failure) for p (success) and ***change*** the definition of r from the number of successes to the number of failures to obtain the probability of failure rather than the probability of success shown above. In words the Binomial Distribution for ***errors*** is the number of combinations of bit failures for n bits times the probability of failure for 1 bit (BER) raised to the number of bits in error (in this case r = failures) times the probability of no errors (1 – BER) raised to the number of bits **not** in error (n – r).

Figure A3 Binomial Density Function for fixed n = 9 Shown is the probability of success B(r) for n = 9 bits with decreasing bit error rates 1 – p = q = 0.8, 0.5 and 0.2

$$1 - B_n = BER = \begin{bmatrix} n \\ r \end{bmatrix} q^r (1-q)^{n-r}$$



Expected # of successes for transmitting 9 bits in one trial: np = 1.8 np = 4.5 np = 7.2

Hypergeometric Distribution - a distribution that can be used for non-independent trials or events without replacement (each trial changes when population changes w/o replacement)

$$\Pr(X = k) = \frac{\binom{n}{k}\binom{N-n}{j-k}}{\binom{N}{j}}$$

k = number of desired objects in the sample
j = number of objects in the sample
n = number of desired objects in the population
N = number of objects in the population

see Equation A21a, page 394 for author's term definitions

Poisson Distribution (discrete probability model for the Poisson density function)

At this stage we'll take the derivation of the Poisson density function as a special limit of the binomial density function where **p is very small and n is very large**. Section A8.2 takes another far more formal derivation of the Poisson distribution as the governing probability law for a Poisson Process, a process that represents the number of events that have occurred up to time t, which is nothing more than a counting process. The Poisson Distribution is a special kind of Markov process, a discrete-state continuous-time process/model. Mother Nature (the real world) does a lot of things that follow a Poission Distribution.

For our interests in events which occur in time, the Poisson distribution can take on a special limiting form where we'll define the rate of occurrences as the constant

λ = occurrences per unit of time (this will end up being called the failure rate)

where the most probable number of occurrences (the graphs of np shown in Figure A3)

$$np = \lambda t$$

Taking the limiting form of these terms (a large sample space) plugged into the binomial distribution (binomial density) results in a useful form of the Poisson Distribution:

$$f(r; \lambda, t) = \frac{(\lambda t)^r e^{-\lambda t}}{r!}$$

This Poisson Distribution/Poisson Density Function is the limiting form of the binomial distribution
r = # of successes λ is a constant

Continuous Random Variables for events where we have no reason to believe that the random variable takes on discrete (countable) values

The distribution function (CDF) for the discrete case was previously defined as:

Cumulative **Distribution** Function (CDF) $P(\mathbf{x} \le x) \equiv \mathbf{F(x)} = \sum_{\mathbf{x} \le x} f(x)$

If the spacings between the discrete values of the random variable **x** are Δx and we let $\Delta x \to 0$ then the discrete variable becomes the continuous variable and the summation becomes an integration (fundamental theorem of calculus).

Thus the cumulative distribution function of a continuous random variable **x** with the probability that **x** takes on all values less than or equal to a specified value x is

$$P(\mathbf{x} \le x) = \int_a^x f(x)\, dx = F(x) \quad \text{the CDF for } a < \mathbf{x} \le b \qquad f(x) = \text{probability density function - PDF}$$

which shows that the integral of the **PDF** (probability density function) is the **CDF** (cumulative distribution function) or stated another way that the derivative of the **CDF** is the **PDF** → $dF(x) / dx = f(x)$

An example of a more useful form of this relates the probability of failure with a specified period of operation

$$P(a \le \mathbf{x} \le b) = \int_a^b f(x)\, dx = F(b) - F(a)$$ for example the probability of a failure within the first month of operation where a = 0 and b = 1 month.

When we deal with continuous probability, it makes sense to talk of the probability that **x** is within an interval rather that at one point. Where's the PDF? The density function f(x) only has a value when integrated over some finite interval, normally a time interval.

Sections A6.2 through A6.6 take a look at a number of continuous variable distributions which are of major interest since they relate how numerous features of the real world operate or how to describe the operations of Mother Nature (lol). In summary:

<u>Rectangular or Uniform Distribution</u>

This model predicts a uniform probability of occurrence in an interval between *a* and *b*

$$P(x < \mathbf{x} \le x + \Delta x) = \Delta x / (b - a) \qquad f(x) = 1 / (b - a) \text{ for } a < x < b$$

<u>Exponential Distribution</u> (continuous density function where now the random variable is switched to time **t** and defined as the failure time of the item in question.)

$f(t) = \lambda e^{-\lambda t}$ for $t > 0$ PDF where λ = **constant** failure rate (failures/hour)

This distribution appears heavily in reliability work. A justification is that the time between failures for a system made up of many components, none of which has a high probability of failure, approaches an exponential distribution.

$$\text{CDF} = F(t) = 1 - e^{-\lambda t}$$

Geometric interpretation of the CDF: F(t) is the area under the curve of f(t) or as previously stated F(x) is the integral of f(x) → the CDF is the integral of the PDF.

To define reliability, we know that the probability of failure as a function of time is

$P(\mathbf{t} \leq t) = F(t)$ which is the definition of the failure distribution function

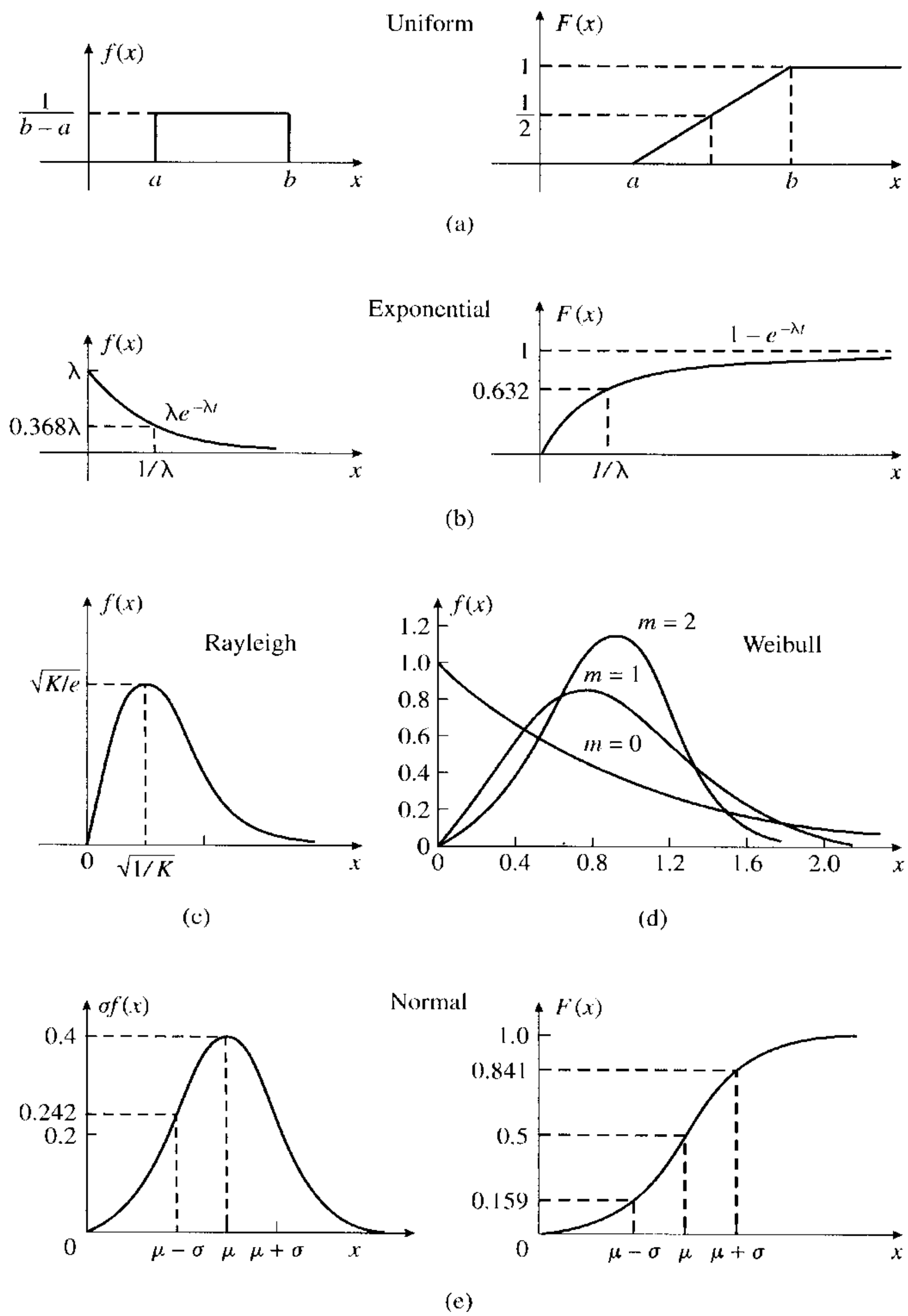
The definition of reliability is the probability of success (no failures) in terms of $F(t)$ as

$R(t) = p(\mathbf{t} \geq t) = 1 - F(t)$ = the probability of **not** observing any failure before time t

Thus for the exponential distribution: $p(\mathbf{t} > t) = R(t) = 1 - F(t) = e^{-\lambda t}$

R(t) is the probability of no failure (or the probability of success) in the interval from 0 to t

Distributions (details to follow):



Hazard Function **z(t)** – a function that measures the probability that a product will fail during a particular instant of time under the condition that it did not fail before that instant of time (t).

$z(t) = f(t) / 1 - F(t)$ = probability of failure at time t / probability of no failure prior to t

For an exponential distribution $z(t) = \lambda e^{-\lambda t} / 1 - (1 - e^{-\lambda t}) = \lambda$ The Hazard rate is a constant which equals the constant failure rate λ.

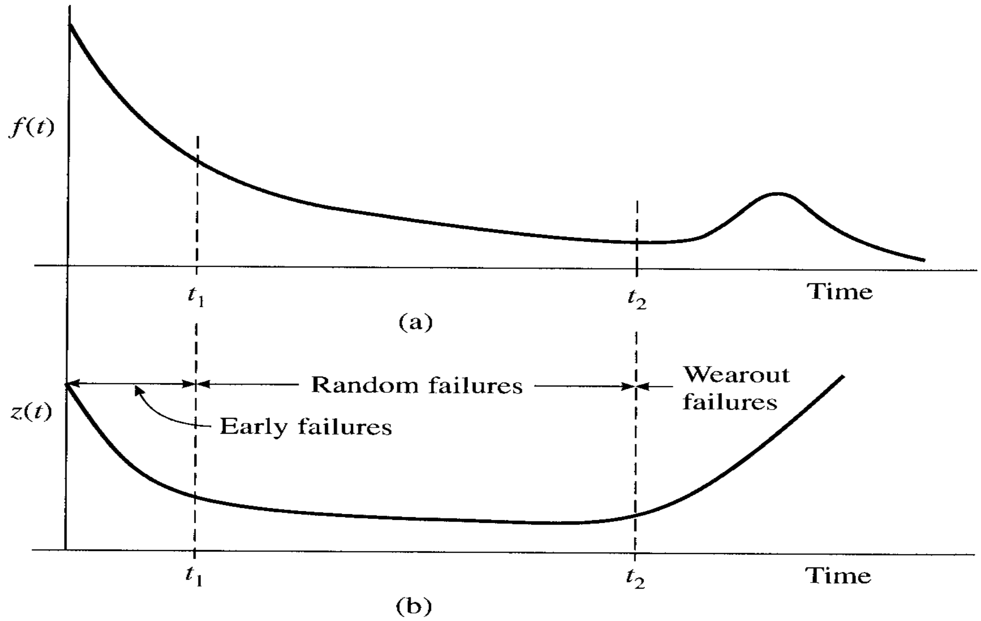


Figure B6 General form of failure curves. (a) Failure density; (b) hazard rate.

Figure B6(b) is the **bathtub curve** and represents the three regions of a systems hazard rate over its operational life: infant mortality (decreasing failure rate), constant failure rate (between t_1& t_2) and wearout or old age (increasing failure rate). Thus for this situation during an item's operational life, the time between t_1 & t_2 , we can assume a constant failure rate → exponential distribution.

PROBLEM:

Equipment failure, constant failure rate with λ = 0.1 failures per year.
Thus we know $f(t) = \lambda e^{-\lambda t}$ and $F(t) = 1 - e^{-\lambda t}$ CDF = ∫ PDF

What is the probability of equipment failure during the first year?

$$pdf = f(x) = \lambda e^{-\lambda x} \qquad x \geq 0 \qquad \lambda = 0.1 \text{ per year}$$

$$\Pr(0 \leq X \leq 1 \text{year}) = \int_{0=x}^{1=x} \lambda e^{-\lambda t}\, dt$$

$$= \lambda \int_0^1 e^{-\lambda t}\, dt$$

$$= \lambda \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^1$$

$$= -e^{-\lambda t}\Big|_0^1$$

$$= -e^{-(0.1)(1)} - \left(-e^{-(0.1)(0)} \right)$$

CDF

$$= -e^{-0.1} + 1 = \left(1 - e^{-\lambda x}\right)$$

$$= -.9048 + 1 = 0.095162$$

QUESTION: WHAT IS DIFFERENCE between this solution AND USING

$$\Pr\{k \text{ failures in time } [0, t]\} = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \quad \text{when } k = 1$$

~~PG 214~~ OLD TEXT

$$\Pr\{1 \text{ failure in } [0,1]\} = 0.09048$$

MY ANSWER: If we CALCULATE 2, 3, 4 ...., ∞ failures in one year by above and add, we reach the integral value eg as $k \to \infty$ $\frac{e^{-\lambda t}(\lambda t)^k}{k!} \to \int_0^1 \lambda e^{-\lambda t} dt$

See lecture slide Page 14 of 29 where k = r = # of occurences/failures, the Poisson Distribution for one and only one failure

Which is the hard way of answering the question when you could easily formulate:

$P(0 \leq \mathbf{t} \leq 1) = \text{CDF} = F(t) = 1 - e^{-\lambda t} = 1 - e^{-0.1} = 1 - 0.9048374 = 0.09516258$

Rayleigh Distribution

Single parameter density function, handles an increasing failure rate

$$f(x) = K\,x\,e^{-Kx^2/2} \quad \text{for } 0 < x \le \infty$$

$$F(x) = 1 - e^{-Kx^2/2}$$

The distribution is found in wireless communications multipath models, software manpower estimation, etc.

Weibull Distribution

A two parameter distribution that handles increasing, constant or decreasing failure rates. The second parameter is a shape factor ***m*** that determines the shape of the distribution (for a given λ). The Rayleigh Distribution where m = 1 and the Exponential Distribution where m = 0 are special cases of a Weibull Distribution (see Figure A5).

$$f(t) = m\,\lambda\,(\lambda t)^{m-1}\,e^{-(\lambda t)^m} \quad \text{for } t > 0,\ m > 0,\ \lambda > 0$$ (later in the notes we substitute $\alpha = m$)

The mean E(x) shown in Table A1 (e.g., MTTF) uses the Gamma function which is a table look-up value that can be found at http://www.efunda.com/math/gamma/findgamma.cfm)

Normal Distribution

A highly utilized two-parameter distribution, sometimes called a Gaussian Distribution.

1. Continuous and symmetrical about the y-axis.
2. Total probability under the curve is one.
3. The probability of a single point occurring on the continuous distribution is zero. Therefore in normal distribution we describe the probability of a single point falling within some specified range.
4. The binomial distribution approaches the normal distribution for large n. It can also be shown that when a random variable is the sum of many other random variables, the variable will have a normal distribution in most cases.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2} \qquad \text{(A32)}$$

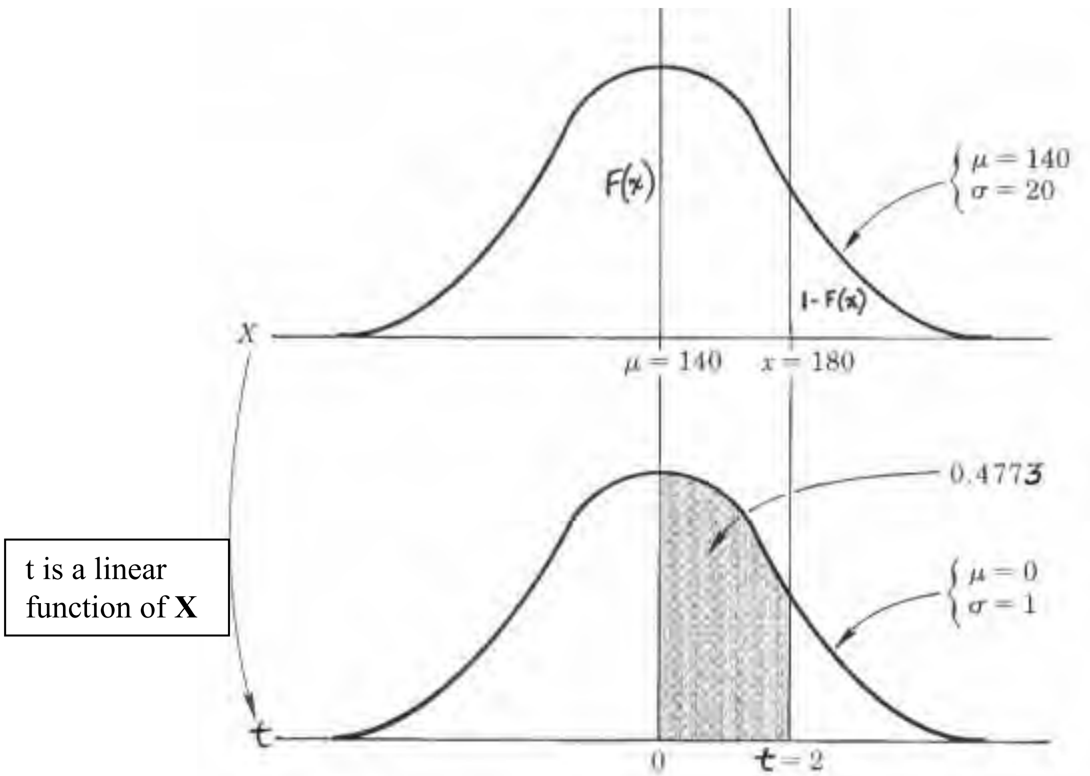
μ = mean $\quad \sigma$ = standard deviation

The area under the f (x) density curve between a and b is the area of interest since it represents the probability that **x** is within the interval $a < \mathbf{x} \le b$.

For tabular convenience, the areas under the normal distribution have been transformed to the standard normal distribution where $\mu = 0$ and $\sigma = 1$. This is done by a change in scale by expressing a given range in standard deviation units given the random variable **x**:

$$t = (\mathbf{x} - \mu) / \sigma \qquad \text{(t is a linear function of } \mathbf{x}\text{)}$$

Problem: What is the probability that the classes' weights fall between 140 and 180 pounds given that the standard deviation of the classes' weights is $\sigma = 20$ pounds, the mean $\mu = 140$ pounds and we assume that weights are normally distributed?



First compute the translation to the standardized normal distribution (The textbook Appendix A uses t as the transformed variable)

$$t = (\mathbf{x} - \mu) / \sigma = (180 - 140) / 20 = 2 \quad \text{(two standard deviations or two sigma } \sigma\text{)}$$

From the CRC Normal Distribution Tables (which uses x as the transformed variable), the entry in the standardized normal distribution table shows the area from $t = -\infty$ to $2 = F(2) = 0.9773$ The area from $t = 2$ to $+\infty = 1 - F(t) = 1 - 0.9773 = 0.0227$ Thus the area under the normal curve from $t = 0$ to 2 or $x = 140$ to 180 pounds $=$ $(0.9773 - 0.0227) / 2 = 0.4773$ (since the curve is symmetrical about the y axis) which is the probability that a student's weight in this class falls between 140 and 180 pounds.

Moments

Weighted integrals of the density function f(x) which describe various geometrical properties of the associated function, a kind of short-story descriptive of the function.

1st Moment - Expected Value or Mean E(**x**)

2nd Moment – Variance of **x** or var **x**

σ^2 = var **x** = E(**x**) - μ

TABLE A1 Mean and Variance for Several Distributions

Distribution	$E(\mathbf{x})$	var $\mathbf{x}$
Binomial	np	$np(1-p)$
Poisson	μ	μ
Exponential	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Rayleigh	$\sqrt{\frac{\pi}{2K}}$	$\frac{0.4292}{K}$
Weibull	$\left(\frac{K}{m+1}\right)^{1-\varepsilon}\Gamma(\varepsilon)$ $\varepsilon \equiv \frac{m+2}{m+1}$ $\Gamma \equiv$ the gamma function	$\left(\frac{K}{m+1}\right)^{1-\delta}\Gamma(\delta) - [E(\mathbf{x})]^2$ $\delta \equiv \frac{m+3}{m+1}$
Normal	μ	σ^2

Markov Models

Markov models are functions of two random variables – the state of the system **x**
the time of the observation **t**

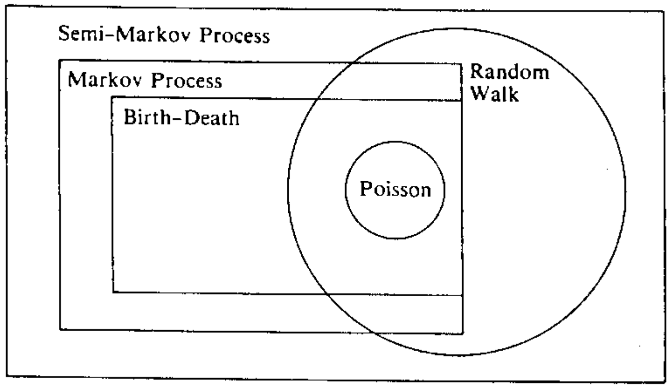
Four models can result since **x** and **t** may be either discrete (state changes are integers) or continuous (transitions between states may take place at any instant of time).
Any Markov model is defined by a set of probabilities p_{ij} which define the probability of moving from any state **i** to any state **j**.

Markov Chain – discrete-state discrete-time. A set of random variables $\{\mathbf{X}_n\}$ forms a chain if the probability that the next state $\{\mathbf{X}_{n+1}\}$ depends only on the current state and not any previous state. No memory or memoryless.

Markov Property – past history summarized in the specification of the current state.

STOCHASTIC PROCESSES - A FAMILY OF RANDOM VARIABLES $X(t)$ WHERE THE RANDOM VARIABLES ARE INDEXED BY TIME (t). THE # OF PEOPLE IN A MOVIE TICKET LINE IS A STOCHASTIC PROCESS

- MARKOV MODELS

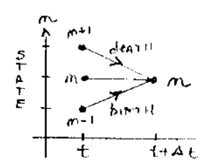


1. Semi-Markov Process
   This stochastic process can have an arbitrary distribution of time between state changes and any new state is possible given the current state (i.e. the probability $P[S_k \mid S_j]$ is arbitrary).

2. Random Walk
   This stochastic process has an arbitrary distribution of time between events. However, the next-state probabilities depend on the current position. The restriction can be described as $P[S_k \mid S_j] = P_{k-j}$. In other words, the transition probabilities depend only on the distance from the current position.

3. Markov Processes (Chains)
   This stochastic process restricts the time between events to be memoryless, but the next-state probabilities $P[S_k \mid S_j]$ are still arbitrary.

4. Birth-Death Processes
   This stochastic process restricts the time between events to be memoryless and further restricts the next-state probabilities to be nonzero for only the nearest-neighbor states (i.e., $|k - j| > 1 \Rightarrow P[S_k \mid S_j] = 0$).



Poisson Process   special case of a Binomial distribution where n is very large

For the development of Markov processes, the textbook derives the Poisson Process using the Poisson distribution as the governing law. The process focus is a discrete-state continuous-time model based on three constraints:

1. Occurrences from n to n + 1 in time Δt occur with a probability of $\lambda\,\Delta t$
   λ is a constant. No negative number of occurrences.
2. Occurrences are independent.
3. Two or more occurrences in the interval Δt are negligible; a second-order effect.

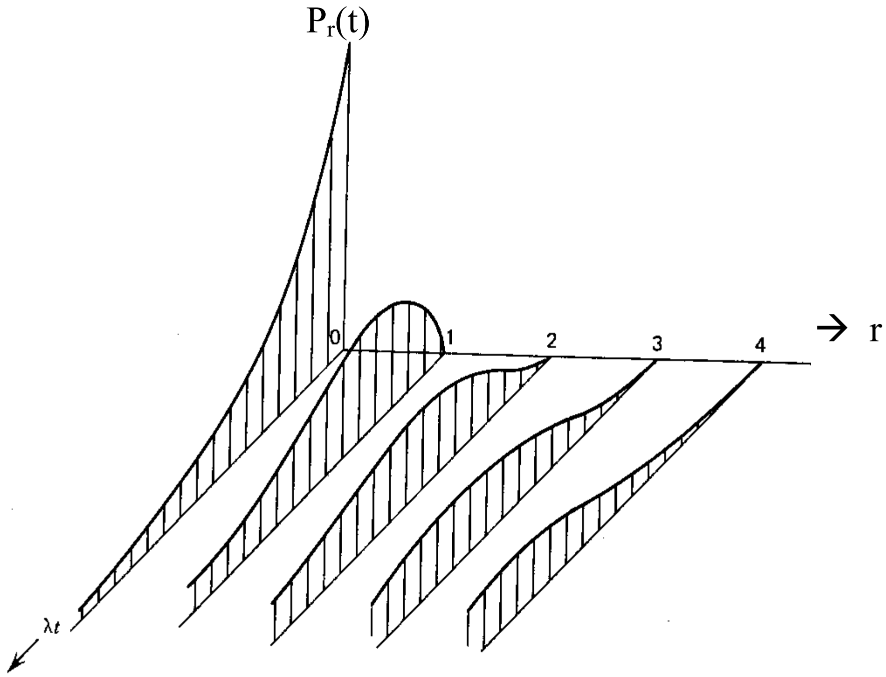
Textbook pages 405 – 407, difference equations → differential equations solved using undetermined coefficients (assume a solution of the form $Ae^{-\lambda t}$)

$P_o(t) = e^{-\lambda t}$ where $P_o(t)$ is the probability of no occurrences.

$P_1(t) = \lambda t\, e^{-\lambda t}$ solving $P_n(t)$ for n = 2, 3, ····· generates the Poisson distribution

$$P_r(t) = \frac{(\lambda t)^r e^{-\lambda t}}{r!} = f(r; \lambda, t)$$

(same as the Poisson Distribution formula A23)
one can view r as the number of failures



Graph of various Poisson Distributions (for r = 0, 1, 2, 3, 4)

Key outcome of the Poisson Process?

With t_o being the time of first occurrence, the probability of no occurrences

$$P_o(t) = P(t < t_o) = 1 - P(t_o < t)$$

But this is the cumulative distribution function F(t) for the random variable **t** (the time of occurrences) from which the density function (PDF) follows as

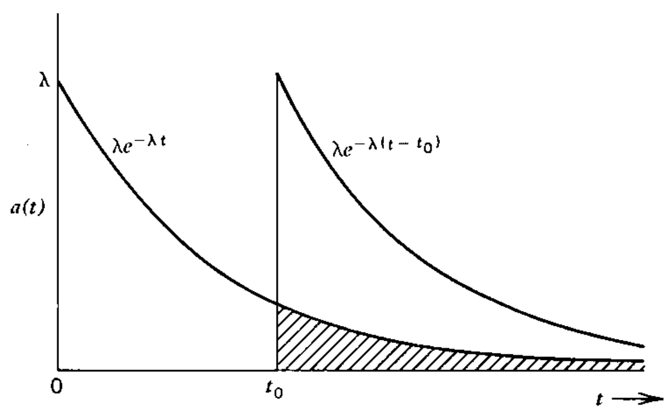
$$f(t) = dF(t) / dt = \frac{d}{dt}(1 - e^{-\lambda t}) = \lambda e^{-\lambda t}$$

The conclusion is that the time of first occurrence is **exponentially distributed** and since each occurrence is independent, it also means that the time between any two occurrences is exponentially distributed (exponential interarrival times).

The same mathematical development is used in queueing theory for network performance. One develops (Kleinrock's *Queueing Systems*) a Poisson process with a constant arrival rate λ and shows that the times between arrivals is exponentially distributed.

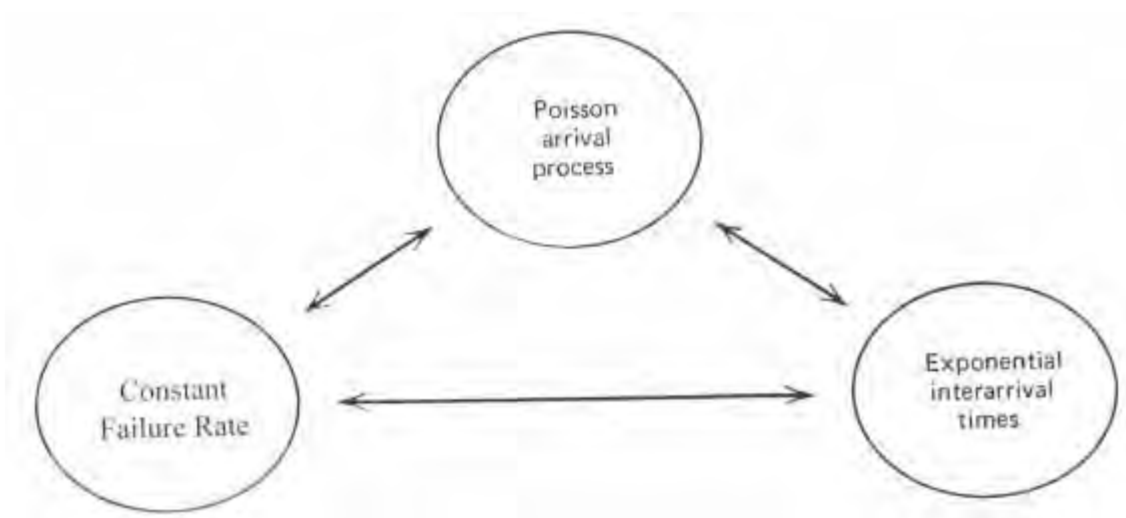
The future of an exponentially distributed random variable is independent of the past history of that variable and thus the distribution remains constant in time. The time until a future arrival occurs is independent of how long it has been since the last arrival.

Only an exponential distribution (in the continuous domain) or a geometric distribution (in the discrete domain) has this property.



For a new event at t_0, the probability of arrival requires that you normalize the cross-hatched area under the graph since the $\sum p = 1$. This tail possesses the exact same shape as the PDF everywhere.

The Memoryless Triangle (showing the interdependency of attributes)



The Poisson Paradox

You are at a bus stop where the bus arrivals are Poisson distributed (exactly how nature normally works) and therefore the time between bus arrivals is exponentially distributed and λ = bus arrival rate (busses/hour).

When you walk up to the bus stop, how long do you have to wait until the next bus?

Answer A: Average time between buses is $1/\lambda$, your arrival is random so you would expect to wait ½ λ on the average for the next bus.

Answer B: For any time that you walk up to the bus stop, the distribution is exactly the same as the original distribution thus the average time one has to wait for the bus, independent of any time you arrive, is $1/\lambda$.

Answer: Given that the bus arrivals are Poisson distributed, the answer is B (the paradox).

Matrix Nomenclature (for later use in Appendix B with Markov Processes)

For Markov Models based on a homogeneous process, which is an outcome of the mathematical development for a Poisson Process as previously outlined, the process has constant coefficients. We can write a convenient set of differential equations with a set of constants represented by a state-transition rate matrix (transition probability matrix)

For two states (0 and 1) the transition matrix would have the following elements:

$$\mathbf{P} = \begin{matrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{matrix}$$

For four states (0, 1, 2, 3)

$$\mathbf{P} = \begin{matrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ P_{30} & P_{31} & P_{32} & P_{33} \end{matrix}$$

where P_{ii} is the probability the system will remain in the same state during one transition and P_{ij} is the probability that the system will move from state i to state j.

The transition probabilities of leaving & arriving for the state must balance. For example, the first row of **P** are all of the transitional probabilities for state 0.

The process is ergodic if every state can be reached from any other state with a positive **P**

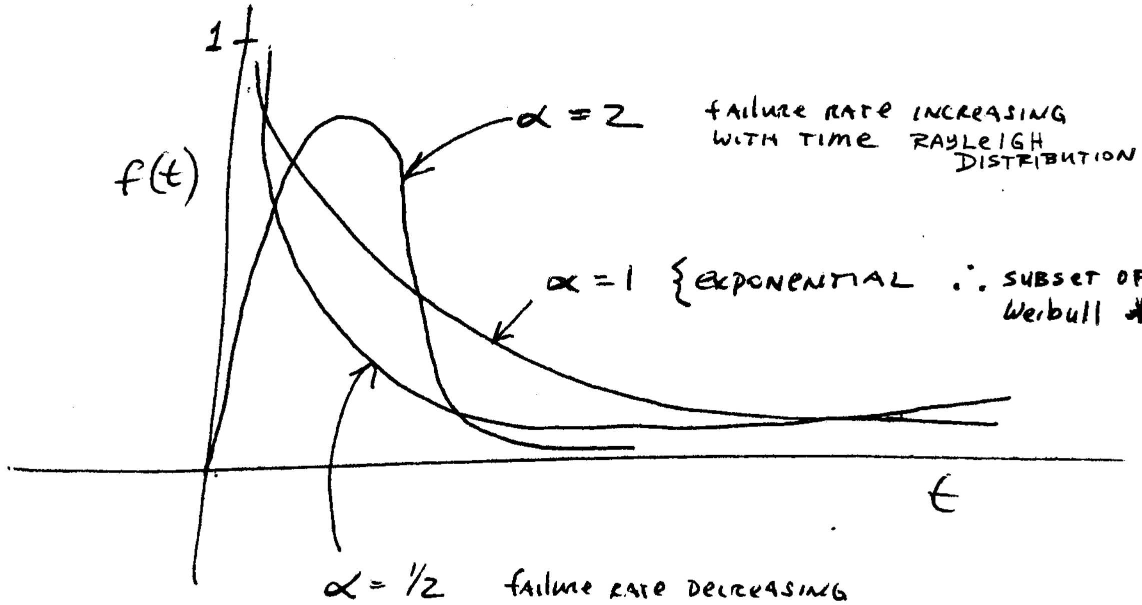
An absorbing state (trapping state) is a state from which it is not possible to reach any other state (the end of the process). In the transition matrix, any column having only a single entry is an absorbing state.

All of this lends itself to a very convenient graphical means of handling Markov Models which will be detailed in Appendix B6 – Markov Reliability and Availability Models.

FAILURE RATE NOT CONSTANT FOR PERIOD OF CONSIDERATION

Weibull handles failure rates either increasing or decreasing with time   another degree of freedom α shape factor
(PARAMETER TO BE ESTIMATED

pdf $\quad f(t) = \alpha\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^{\alpha}} \qquad t>0,\ \alpha>0,\ \lambda>0$



FAMILY OF Weibull Curves for $\lambda = 1$

failure rate $= (\lambda t)^{\alpha}$

$$\text{CDF} = F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

$$\text{Reliability} = R(t) = e^{-(\lambda t)^{\alpha}}$$

$$\text{HAZARD} = z(t) = \alpha\lambda(\lambda t)^{\alpha-1}$$

Mean TIME TO FAILURE $\mu = \Gamma((\alpha+1)/\alpha)/\lambda$ EQ. 20 where Γ is the GAMMA FUNCTION

NOTE Weibull functions relationship to time which is always (λt)

GAMMA FUNCTION Web Site WWW.EFUNDA.COM/MATH/GAMMA/FINDGAMMA.CFM

Weibull Distribution

How do we estimate λ and α ?

ANALYTICAL METHOD involves the solution of a system of TRANSCENDENTAL EQUATIONS

Graphical technique — Transform Weibull distribution into a LINEAR FUNCTION OF $\ln t$

LINEAR REGRESSION ANALYSIS

$$R(t) = \epsilon^{-(\lambda t)^{\alpha}}$$

$$\ln_{\epsilon} R(t) = -(\lambda t)^{\alpha} \quad \text{OR} \quad \ln_{\epsilon} \frac{1}{R(t)} = (\lambda t)^{\alpha}$$

GET RID OF MINUS SIGN

$$\ln \ln \frac{1}{R(t)} = \ln (\lambda t)^{\alpha} = \ln \lambda^{\alpha} + \ln t^{\alpha}$$

$$= \underbrace{\alpha \ln \lambda}_{\text{constant}} + \alpha \ln t$$

$$y = mx + b$$

$$\ln \ln \frac{1}{R(t)} = \alpha \ln t + \alpha \ln \lambda$$

slope $m = \alpha$ shape FACTOR

y-intercept $b = \alpha \ln \lambda$

thus $\ln \lambda = b/\alpha$

$$\epsilon^{b/\alpha} = \lambda$$

Since we normally collect j^{th} unit failure at time t

thus we are obtaining CDF $F(t_j) = 1 - R(t_j)$

Time is usually over an interval

NORMALLIZING THE DATA # OF UNIT / TOTAL # UNITS

Estimator for $F(t_j) = \dfrac{j - 0.5}{N}$ which estimates the midpoints of the intervals (steps) of the cummulative distribution

Example 1. Data are collected from the file system of a time-sharing system about the transient faults in 8 disk drives in an effort to discover whether the time between transient errors follows an exponential distribution. The estimated value of λ is 0.1344 (time in minutes) corresponding to a MTBF of about 7 minutes. The total number of observed errors is 877 in a 5-day interval. Table 2–12a shows the observed errors by division into time categories and the expected number of errors in each time category according to an exponential distribution. For instance, the first row in the table means that 548 errors were observed with times between errors of 0 to 5 minutes, while an exponential distribution with $\lambda = 0.1344$ gives the expected number of errors in that range as 429.20 (given that the total number of failures is 877). The remaining categories have to be pooled until no E_i is smaller than 5. The result of this operation is shown in Table 2–12b. The number of degrees of freedom is $m = 8 - 1 - 1 = 6$ because there are eight different categories, and one parameter (λ) has been estimated from the data. For 6 degrees of freedom, $\chi^2_{0.05} = 12.592$. Since $\chi^2 > \chi^2_{0.05}$, the hypothesis that the time between errors has an exponential distribution must be rejected.

TABLE 2–12 *Data on transient faults for the time-sharing file system (Example 1)*

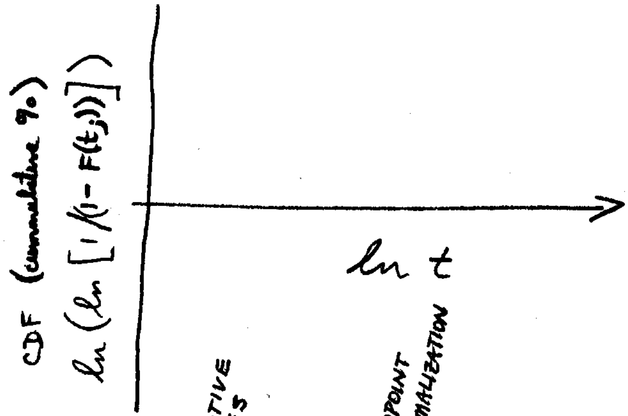
a. Collected Data

Time Category (mins)	Observed Errors, O_i	Expected Errors E_i	Time Category (mins)	Observed Errors O_i	Expected Errors, E_i
0–5	548	429.20	55–60	2	0.2639
5–10	148	219.15	60–65	1	0.1347
10–15	63	111.89	65–70	1	0.06881
15–20	35	57.13	70–75	1	0.03514
20–25	28	29.17	75–80	1	0.01794
25–30	18	14.89	80–85	1	0.009160
30–35	12	7.60	85–90	1	0.004690
35–40	6	3.88	90–95	1	0.002395
40–45	3	1.98	95–100	1	0.001215
45–50	1	1.01	100–105	1	0.000627
50–55	3	0.5178			

b. Pooled Categories

Time Category (mins)	O_i	E_i	$(O_i - E_i)^2/E_i$
0–5	548	429.20	32.88
5–10	148	219.15	23.10
10–15	63	111.89	21.36
15–20	35	57.13	8.57
20–25	28	29.17	0.04
25–30	18	14.89	0.64
30–35	12	7.60	2.53
35–∞	25	7.93	36.74
		Total χ^2 =	125.86

Problem 2 PAGE 77



Category	Assume: failure at end point of interval t (min)	ln (t)	Cumulative failures j	Midpoint normalization (j−0.5/N) $F(t_j)$	$\ln\{\ln[1/(1-F(t_j))]\}$ Transformed CDF
0–5	5	1.609	548	0.624	−0.0213
5–10	10	2.303	696	0.793	+0.4544
10–15	15	2.708	759	0.865	+0.6939
	20	2.996	794	0.905	0.8551
	25	3.219	822	0.937	1.0153
	30	3.401	840	0.957	
	35		852	0.971	
	40		858	0.978	
	45		861	.981	
	50		862	.982	
	55		865	.986	
	60		867		
	65		868		
	70		869		+1.5339
	75		870		
	80		871		
	85		872		
	90		873		
	95		874		
	100		875		
	105		876		
105–110	110	4.700	877 = N	0.9994	+2.0108

correlation coefficient $r^2 = 0.98383$

Y-intercept = −0.8838078

α = slope = +0.5808259874

$\lambda = 0.218353592$

Transformed CDF$(\ln t) = \alpha \ln t + \alpha \ln \lambda$

m = slope = α

b = Y-intercept = $\alpha \ln \lambda$

$\lambda = e^{b/\alpha}$

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}}$$

$$= 1 - e^{-(0.218t)^{0.58}}$$

Table 2-12 Page 57 *Reliable Computer Systems* textbook

Microsoft Excel - Problem 2 Page 57.xls

C27 {=LINEST(E3:E24,B3:B24,,TRUE)}

	A	B	C	D	E
1	t (hours)	ln(t)	j	F(tj)	ln{ln[1/(1 - F(tj))]}
2					
3	5	1.6094	548	0.624287343	-0.0213
4	10	2.3026	696	0.79304447	0.4544
5	15	2.7081	759	0.864880274	0.6939
6	20	2.9957	794	0.904789054	0.8551
7	25	3.2189	822	0.936716078	1.0153
8	30	3.4012	840	0.957240593	1.1481
9	35	3.5553	852	0.970923603	1.2635
10	40	3.6889	858	0.977765108	1.3366
11	45	3.8067	861	0.981185861	1.3796
12	50	3.9120	862	0.982326112	1.3952
13	55	4.0073	865	0.985746864	1.4471
14	60	4.0943	867	0.988027366	1.4873
15	65	4.1744	868	0.989167617	1.5097
16	70	4.2485	869	0.990307868	1.5339
17	75	4.3175	870	0.991448119	1.5606
18	80	4.3820	871	0.992588369	1.5902
19	85	4.4427	872	0.99372862	1.6237
20	90	4.4998	873	0.994868871	1.6625
21	95	4.5539	874	0.996009122	1.7091
22	100	4.6052	875	0.997149373	1.7682
23	105	4.6540	876	0.998289624	1.8518
24	110	4.7005	877	0.999429875	2.0108
25					
26			**slope**	**y-intercept**	
27			**0.580825987**	**-0.883807793**	
28			0.016649983	0.064874509	
29		**r**2 =**	**0.983830895**	0.062730877	
30			1216.926866	20	
31			4.788805552	0.07870326	
32					
33	**λ =**	**0.218353592**			

Sheet1 / Sheet2 / Sheet3

Ready

Expected (model) failures

$$E_{i_m} = \left[1 - e^{-(0.218t)^{0.58}}\right] 877 \; - F(t_{m-1})$$

Σ previous failures

Catagory	t (min)	θ_i		E_i (rounded to 2 decimal places)		$(\theta_i - E_i)^2 / E_i$
0-5	5	548		570.82		0.9119
5-10	10	148		124.45		4.4565
10-15	15	63		62.08		0.0136
15-20	20	35		36.36		0.0510
20-25	25	28		23.16		1.0127
25-30	30	18		15.56		0.3835
30-35	35	12		10.85		0.1209
35-40	40	6		7.79		0.4123
40-45	45	3		5.72		1.2946
45-50	50	1	4	4.28	7.53	1.6548
50-55	55	3		3.25		
55-60	60	2		2.50		
60-65	65	1	4	1.95	5.98	0.6556
67-70	70	1		1.53		
70-75	75	1		1.21		
75-80	80	1		0.97		
80-85	85	1		0.78		
85-90	90	1	8	0.63	5.16	1.5631
90-95	95	1		0.51		
95-100	100	1		0.42		
100-105	105	1		0.35		
105-110	110	1		0.29		
		$N = 877$		875.45		$\Sigma \chi^2 = 12.5306$

] ≡ Grouping for $E_i \geq 5$

Degrees of freedom (m)

$$m = 12 - 2 - 1 = 9$$

Number of observations grouped to $E_i \geq 5$

# of estimated parameters e.g. λ and α

From tables/calculation

$$\chi^2_{0.05} = 16.9$$

Thus $\chi^2 < \chi^2_{0.05}$ can accept Weibull distribution

another way : Accept hypotheses if $P(\chi^2) \leq 0.95$ when testing at 95% confidence level

$\mathrm{CDF}(\chi^2) = P(12.5306) = 0.815$ thus accept hypotheses

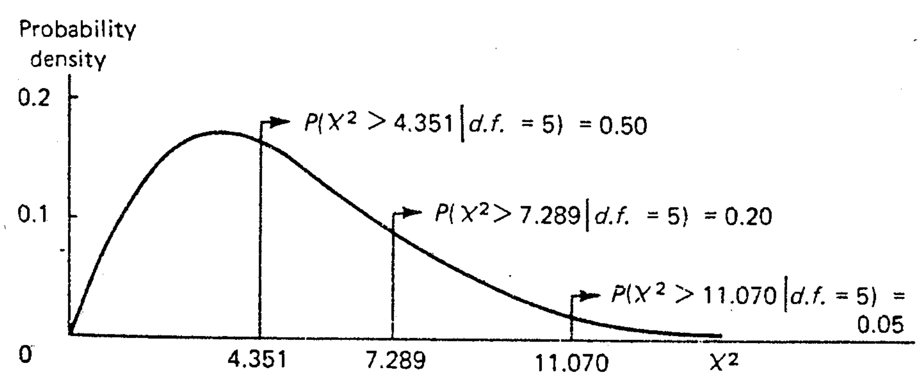
- The Chi-Square distribution (χ^2) is the most important of all distribution-free tests.
- It was introduced by Karl Pearson in 1900.
- Although the test is distribution free since it makes NO assumptions about the population from which the sample is drawn, there are four conditions that should be valid for the chi-square analysis to be applied to any kind of test.
  1. The sample of observations should be independent of one another and drawn from the target population.
  2. The data are usually of nominal measurement.
  3. The sample should contain at least 50 observations.
  4. There should be no fewer than five observations in any expected cell. [E_i must be equal to at least 5 and at times it may be necessary to pool categories.]

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)}{E_i}$$

where O_i is the observed frequency and
E_i is the expected or theoretical frequency

- The chi-square distribution is a different distribution with each change in the number of degrees of freedom. It is badly skewed to the right for a few degrees of freedom and approaches the shape of the normal distribution when the number of degrees of freedom is approximately 30.
- The number of degrees of freedom may be described as the number of observations that are free to vary after certain restrictions have been placed on the data. These restrictions are inherent in the organization of the data. For example, if a sample of 50 items is classified as effective or defective, the determination that 40 are effective automatically means that the remaining group of 10 are defective, thus the degrees of freedom d.f. = 2 – 1 = 1 Since the number of parts is known, when the total in one category is ascertained, the total number in the other category is determined.
- Chi-Square Distribution Table and example of the use of the chi-square table.

Distribution of Chi-Square for Five Degrees of Freedom



The table of chi-square values list the values of chi-square that give a specified area in the right-hand tail of the above distribution for 5 degrees of freedom. For example the area to the right of $\chi^2 = \cancel{7.289}$ is 0.20 or 20%
4.351
0.50 or 50%

PERCENTAGE POINTS, CHI-SQUARE DISTRIBUTION (Continued)

$$F(\chi^2) = \int_0^{\chi^2} \frac{1}{2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)}\, x^{\frac{n-2}{2}} e^{-\frac{x}{2}}\, dx$$

n \ F	.005	.010	.025	.050	.100	.250	.500	.750	.900	.950	.975	.990	.995
1	.0000393	.000157	.000982	.00393	.0158	.102	.455	1.32	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.103	.211	.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	.0717	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	.207	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	.412	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	.676	.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	.4.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
[illegible]	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
[illegible]	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

THE KOLMOGOROV-SMIRNOV ONE-SAMPLE TEST

The Kolmogorov-Smirnov one-sample test may be used as an alternative to a chi-square test for goodness of fit when one is concerned with the hypothesis that an observed sample distribution is drawn from a population with a given theoretical distribution. It has two advantages over chi-square:

1. The Kolmogorov-Smirnov test treats individual observations separately. Thus, it is not necessary to lose information by combining categories as is often required in chi-square. This makes it a more powerful test than chi-square.

2. When samples are very small, it may be possible to use the Kolmogorov-Smirnov test when chi-square would be impractical because of the requirement that each expected frequency be at least five and that the sample be at least 50.

The Kolmogorov-Smirnov test involves working with two cumulative frequency distributions. One is an observed cumulative frequency distribution secured from the sample, and the other is a theoretical cumulative frequency distribution assumed in the null hypothesis. The point at which these two distributions show the greatest divergence is determined, and a decision is made to accept or reject the null hypothesis depending on the probability that the observed difference would occur if the observations were really a random sample from the theoretical distribution.

If F_o represents the cumulative values of f_o (the same observed distribution used in chi-square) expressed as a proportion of the total, and if F_e represents the cumulative values of f_e (the expected distribution) shown as a proportion of the total, then D can be defined as the maximum absolute difference between F_o and F_e.

$$D = \text{maximum} \, |F_o - F_e| \qquad (9.3)$$

This maximum difference, D, can be compared with a known theoretical sampling distribution of D that is determined by the assumptions in H_o. Certain critical values of this known distribution are shown in Table 9.11. The table gives values of D for five different levels of significance for samples as large as 35. For samples larger than 35, the value of D can be computed using the fractions shown in the bottom row of the table. For example, if the sample size, n, is 49, $D = \frac{1.36}{\sqrt{49}} = 0.19$ for $\alpha = 0.05$ (two-tail test).

■ ***Example.*** A baker who plans to add fruitcake to the line of baked goods wishes to test a statement made by a competitor to the effect that, "If you want a popular fruitcake that will sell, put in lots of nuts."

The baker makes one fruitcake from each of six different recipes, differing only in the proportion of nuts. Cake A has the smallest proportion of nuts and Cake F has the largest proportion. The baker gives a slice of each of the cakes to each of 12 homemakers. The women taste the slices and designate the cake that each likes best and would be most likely to buy. The following shows the test results.

Fruitcake ranked by the proportion of nuts (A has the fewest nuts)	*Number of homemakers selecting each cake*
A	0
B	1
C	1
D	1
E	5
F	4
Total	12

The problem is to test at $\alpha = 0.05$ the hypothesis that the proportion of nuts is not important to the popularity of the fruitcake.

STEPS:

1. H_o: The observed distribution comes from a population that has a uniform distribution (*i.e.*, the proportion of nuts is not important).
2. H_a: The observed distribution comes from a population that does not have a uniform distribution (*i.e.*, the proportion of nuts is important).
3. An $\alpha = 0.05$ requires a value of $D = 0.375$ ($D_{\alpha=0.05}$ for $n = 12$).
4. Criterion: Reject H_o (accept H_a) if $D > 0.375$; do not reject H_o if $D \leq 0.375$, when $D = \text{maximum } |F_o - F_e|$.
5. Using the sample results, the value of D is computed in Table 9.12. The assumption underlying the expected distribution is that if the proportion of nuts is not important, each recipe for fruitcake should be chosen by two homemakers.

TABLE 9.12

Computation of *D* for a Kolmogorov-Smirnov One-Sample Test

Fruitcake ranked by proportion of nuts (*A* has fewest nuts)	Number chosen f_o	f_e	cumulative F_o	cumulative F_e	$\lvert F_o - F_e \rvert$
A	0	2	$\frac{0}{12}$	$\frac{2}{12}$	$\frac{2}{12}$
B	1	2	$\frac{1}{12}$	$\frac{4}{12}$	$\frac{3}{12}$
C	1	2	$\frac{2}{12}$	$\frac{6}{12}$	$\frac{4}{12}$
D	1	2	$\frac{3}{12}$	$\frac{8}{12}$	$\frac{5}{12} = D$ maximum
E	5	2	$\frac{8}{12}$	$\frac{10}{12}$	$\frac{2}{12}$
F	4	2	$\frac{12}{12}$	$\frac{12}{12}$	0
Total	12	12			

Since $D(\frac{5}{12} = 0.417) > D_\alpha(0.375)$, reject H_o and accept H_a. The proportion of nuts in the fruitcake does make a difference in its acceptance.

H_o - null hypothesis. The hypothesis that assumes that there is no significant difference between the value of the universe parameter being tested and the value of the statistic computed from a sample drawn from that universe. The null hypothesis assumes that the difference between the parameter designated in the hypothesis and the statistic is a sampling difference.

H_a - alternate hypothesis, the hypothesis which will be accepted if statistical testing leads to rejection of H_o.

TABLE 9.11

Table of Critical Values of D in the Kolmogorov-Smirnov One-Sample Test*

Sample size (n)	Level of significance for D = maximum $\lvert F_o - F_e \rvert$				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Over 35	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

* Adapted from Massey, F. J., Jr., 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, pp. 46, 70, with the kind permission of the author and publisher.

KOLMOGOROV - SMIRNOV TEST   EXAMPLE 1

TABLE 2-12

	f_o	f_e	CUMMULATIVE F_o	F_E	$\|F_o - F_e\|$	
0-5	548	429.2	$\frac{548}{877}$	$\frac{429.2}{877}$	0.135	← MAX
5-10	148	219.15	696/877	648.35/877	0.054	
10-15	63	111.89	759/877	760.24/877	0.001	
15-20	35	57.13	794/877	817.37/877	0.027	
20-25	28	29.17	822/877	846.54/877	0.028	
25-30	18	14.89	840/	861.43/	0.024	
30-35	12	7.60	852/	869.03/	0.019	
35-∞	25	7.41	877/	876.44/	10^{-4}	

$D =$ MAXIMUM $= 0.135$

$m = 877 \quad \alpha = 0.05 \quad D_\alpha = 1.36/\sqrt{m} = 0.046$

$D(0.135) > D_\alpha(0.046)$ REJECT EXPONENTIAL DISTRIBUTION

SOME DEFINITIONS

H_o - null hypothesis. The hypothesis that assumes that there is no significant difference between the value of the universe parameter being tested and the value of the statistic computed from a sample drawn from that universe. The null hypothesis assumes that the difference between the parameter designated in the hypothesis and the statistic is a sampling difference.

H_a - alternate hypothesis, the hypothesis which will be accepted if statistical testing leads to rejection of H_o.