Università
degli Studi di
Messina
DIPARTIMENTO DI INGEGNERIA

Computer System Analysis

Probability and Random variables

Probability models

- The study of random phenomena
  - Their future behavior is not predictable in deterministic fashion
- Such phenomena are capable of mathematical descriptions
- Ideal probabilistic model of real world situation
  - Lists of all possible outcomes and associated probabilities

Probability theory

- Mathematical tool for managing experiment whose outcome is not known a priori
- Introduced in the late 17^{th} century in the context of gamblings
- Notable mathematicians in probability theory were Pascal, Fermat, DeMoivre, Laplace, Bernoulli, Kolmogorov, Markov

An example

- The number of request arrivals to a WEB server in a fixed time interval *(0, t]*
- The number of arrivals is assumed to have a specific distribution (*Poisson*)
- We replace a complex physical situation with a simple model
  - Only one parameter
  - We can predict the future

Random experiment

- An experiment whose outcome is uncertain
- Random experiment
- Outcomes
  - Possible results an experiment can produce
  - Discrete
  - Continuous
  - Mixed

Sample space

The set of all the possible outcomes of a random experiment

- We will denote it with ***S***
- ***S*** depends on the experiment, but also on the purpose the experiment is carried out
- ***S*** can be either finite or infinite

Events

- An ***event*** is a subset of S
- A statement of conditions defining a subset is equivalently an event
  - The set of outcomes such that the statement is true
- A trial is a single performance of the experiment
- Only one outcome can occur on any trial

Definitions

- The sample space S is said ***universal*** event
- The null set is said ***null*** or ***impossible*** event
- The event {s}, $s \in S$, is said ***elementary*** event

Algebra of events

- Let E_1 be an event of S
- The complement of E_1 is

$$\overline{E}_1 = S - E_1 = \{s \in S | s \notin E_1\}$$

- Intersection of two events:

$$E_3 = E_1 \cap E_2 = \{s \in S | s \in E_1 \ and \ s \in E_2\}$$

- Union of two events:

$$E_3 = E_1 \cup E_2 = \{s \in S | s \in E_1 \ or \ s \in E_2\}$$

Algebra of events

- Two events are ***mutually exclusive*** or ***disjoint*** events when

$$A \cap B = \varnothing$$

- It is not possible for both of them to occur in the same trial
- All the definitions above can be easily extended to more events

Algebra of events

- Classical relationships hold
- Commutative laws: $A \cup B = B \cup A,\ A \cap B = B \cap A$
- Associative laws:

$$A \cup (B \cup C) = (A \cup B) \cup C,\ A \cap (B \cap C) = (A \cap B) \cap C,$$

- Distributive laws:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

- Identity laws: $A \cup \varnothing = A,\ A \cap S = A$
- Complementarity laws: $A \cup \overline{A} = S,\ A \cap \overline{A} = \varnothing$

Graphical representation

- Events could be represented with **Venn diagrams**

- It is the usual notation for managing sets

Example: dice roll

- $S = \{1, 2, 3, 4, 5, 6\}$
- "The outcome is 1" is an event (elementary)
- "The outcome is either 1 or 3 or 4" is an event
- "The outcome is an even number" (equivalent to "The outcome is 2, 4 or 6") is an event

Probability

- We assign probabilities to events in the sample space
- Probability represents the "tendency" of an event to be the outcome
- We will denote the probability of an event $A \in S$ with $P(A)$
- Interpretation: the *relative frequency*

Kolmogorov's axioms

- $\forall A \in S, P(A) \geq 0$
- $P(S) = 1$
- $\forall A, B \in S : A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

- If $A_j \cap A_k = \emptyset, \; \forall j \neq k$ then

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A)$$

Properties

- $\forall A \in S,\ P(\overline{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Pascal/Laplace definition

- $S = \{E_1, E_2, \ldots, E_N\}$
- $P(E_1) = P(E_2) = \ldots = P(E_N)$
- $A = E_1 \cup E_2 \cup \ldots \cup E_{N_A}$
  then
  $$P(A) = \frac{N_A}{N}$$
- The probability is evaluated as the ratio of favorable events over possible events

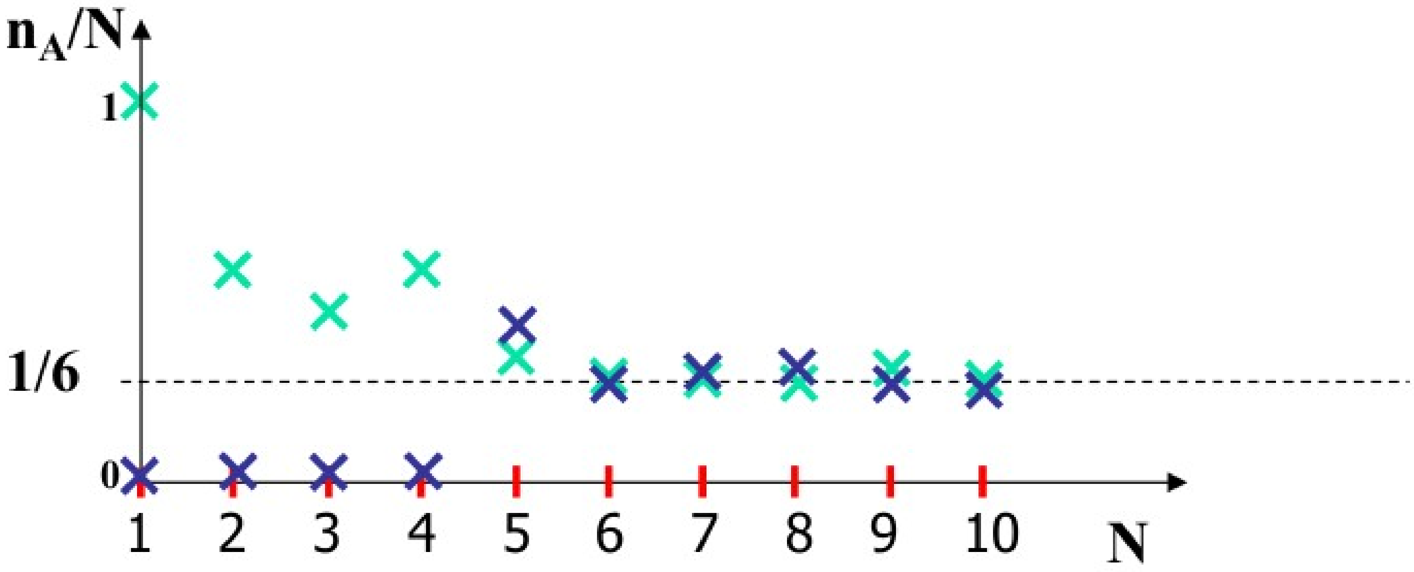
Relative frequency

- N trials of a random experiment
- N_A number of trials with A as outcome

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

Example: roll dice

- *A*={2}



Generic problem

1. Identify the sample space S
2. Assign probabilities
3. Identify the events of interest
4. Compute desired probabilities

Conditional probability

- Let us suppose we know that the outcome of an experiment is contained in $B \in S$
- $P(B) \neq 0$
- Considering a sample point

$$P(s|B) = \begin{cases} \frac{P(s)}{P(B)} & if \;\; s \in B \\ 0 & if \;\; s \notin B \end{cases} .$$

- The sample space is restricted to B

Conditional probability

- Probability of an event A may change if we have added information
- Conditional probability is the probability of A given that B occurs: $P(A|B)$

$$
\begin{aligned}
P(A|B) &= \sum_{s\in A} P(s|B) \\
&= \sum_{s\in A\cap\overline{B}} P(s|B) + \sum_{s\in A\cap B} P(s|B) \\
&= \sum_{s\in A\cap B} P(s|B) \\
&= \sum_{s\in A\cap B} \frac{P(s)}{P(B)} = \frac{P(A\cap B)}{P(B)}, \qquad P(B)\neq 0
\end{aligned}
$$

Summarizing

- Let us consider two events, A e B, whose single probability is not null
- The conditioned probability is:

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

where P(A,B) is their joint probability

- Intuition: the probability the event A happens considering it only if the event B happens also

Independent events

- Two event are independent when

$$P(A|B) = P(A)$$

meaning that B does not affect A

- Also

$$P(A,B) = P(A)P(B|A) = P(B)P(A|B)$$

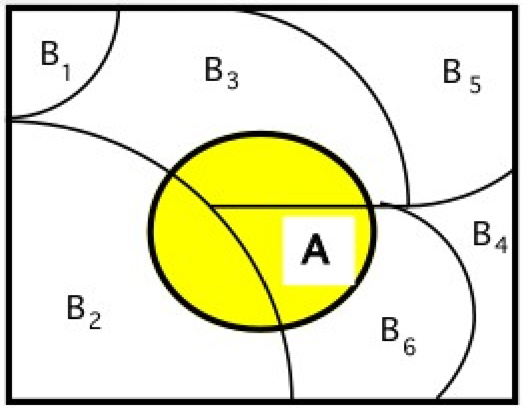
$$P(A|B) = \frac{P(A,B)}{P(B)}$$

$$P(A,B) = P(A)P(B)$$

Total probability theorem

$$B_i \cap B_k = \emptyset$$

$$B_1 \cup B_2 \cup \ldots \cup B_N = S$$



$$P(A) = \sum_{i=1}^{N} P(A|B_i)\, P(B_i)$$

Example

- We are waiting for an incoming phone call between 8:00 p.m and 9:00 p.m.
- Compute the probability the phone rings between 8:20 p.m and 8:30 p.m.

Example

- Let us consider two coins *m1* and *m2*
- Compute the following probability

$$P(m1=heads \mid m2=heads)$$

Example

- The random experiment: to roll two dice
- Estimate the probability to have the outcome d1="1" and d2="3"

Example

- The random experiment: to roll two dice
- Let us consider two events:

$$A \equiv \{(d_1, d_2) : d_1 + d_2 \leq 4\}$$

$$B \equiv \{(d_1, d_2) : d_1 \leq 4\}$$

How much is P(A|B) ?

Example

- The random experiment: to roll two dice
- The first trial produces "6" as outcome
- How much is the probability to have "6" again in the second trial?

Bayes' theorem

If

$$\bigcup_{i=1}^{n} B_i = S; \quad B_i \cap B_k = \emptyset, \ \forall i, k$$

then

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B)}$$

Bayes' theorem (cont.)

- When an even A is casued by different events B_k
- $P(A/B_k)$ is the probability to have A due to B_i (*a priori probability*)
- $P(B_k/A)$ is the probability that the cause of A is B_k (*a posteriori probability*)

Bayes' theorem (cont.)

- The Bayes' theorem allows to re-evaluate assumptions (B_k) on the basis of experimental data
- Given an event A, it gives the probability B_k is its cause

An example: software test

A software house has developed a method of software testing which is 100% positive when a bug exists, but also in 5% of correct software cases.

It is also seen that new versions of software are affected by bugs in 1% of cases.

What is the probability that in the event of a positive test on a new version of this software is really buggy?

Solution: software test (1)

We want to infer *a posteriori knowledge*, in specific conditions, given *a priori knowledge* on the conditions:

A priori knowledge

P(T|S)	Test (T)	State (S)
0.05	positive	correct
0.95	negative	correct
1	positive	bug
0	negative	bug

We know $P(T|S)$

We woud like to know $P(S=b|T=p)$

Software test (2)

Constrains

$$P(T=p|\ S=c) + P(T=n|\ S=c) = 1$$

$$P(T=p|\ S=b) + P(T=n|\ S=b) = 1$$

Experimental data

$$P(S=b)=0.01 \text{ e } P(S=c)=0.99$$

Software test (3)

$$P(S=b|T=p) = P(T=p|S=b)\,P(S=b) \,/$$

$$(P[T=p|S=b]\,P[S=b] + P[T=p|S=c]\,P[S=c]) =$$

$$= 1*0.01 \,/\, (1*0.01 + 0.05*0.96) =$$

$$= 0.168$$

Random variables: motivation

- Sample space is made of different kind of "objects"
  - Not necessarily numbers
- It is often useful to associate numbers to the outcomes
- Random variables create association and give a compact (abstract) representation of a random experiment
- Probabilistic computation is performed in terms of random variables

Definition

A random variable (r.v.) X on a sample space S is a function

$$X : S \to \Re$$

that assigns a real number $X(s)$ to each sample point $s \in S$

- Such that $\forall x \in \mathbb{R}, X(s) \leq x$ is an event

- *A r.v. partitions S into a mutually exclusive and exhaustive set of events*

Random variable

- Function (measurable) from the outcomes of the experiment to a real number
- Ex: toss a coin
  - Sample space -> *S={head, tail}*
  - r.v. $X : S \to \{0, 1\}$
- Discrete
- Continuous
- Mixed

Discrete r.v.

- A r.v. defined over a discrete sample space is a discrete r.v.
  - A r.v. over a continuous sample space could be either continuous or discrete

Example - Discrete random variable

- Toss a coin

$$S = \{head, tail\}$$
$$X(head) = 1, \quad X(tail) = 0$$

$$P\{X = 0\} = 0.5$$
$$P\{X = 1\} = 0.5$$

- Constrain

$$\sum_i P(X = x_i) = 1.0$$

Example - Continuous random variable

- Uniform generation of a real random number in [0,1]

$$S = [0,\ 1]$$
$$X(x_i) = x_i$$

$$P\{X = x_i\} = 0.0$$

$$P\{X \leq x\} = \frac{x - 0.0}{1.0 - 0.0} = x$$

The probability mass function

- An event in terms of r.v. is defined as

$$A_x = \{s \in S | X(s) = x\}$$

thus

$$P(A_x) = P([X \equiv x]) = \sum_{X(s)=x} P(s)$$

- The probability mass function of X is

$$p_X(x) = P(X = x) = \sum_{X(s)=x} P(s)$$

Properties

1) $0 \le p_X(x) \le 1$

2) $\sum_i p_X(x_i) = 1$

Probability of events

- We are interested in computing probabilities of subsets A
- Notation: $[X \in A] \quad P(X \in A)$
- Other notation:
  - $A = (a, b), \quad P[X \in (a, b)] = P(a < x < b)$
  - $A = (a, b], \quad P(a < x \leq b)$
  - $A = (-\infty, x], \quad P(X \leq x)$
- In general: $P(X \in A) = \sum_{x_i \in A} p_X(x_i)$

Cumulative distribution Function (CDF)

$$F_X(t) = P(X \leq t) = \sum_{x \leq t} p_X(x)$$

$$\lim_{t \to -\infty} F_x(t) = 0 \, , \lim_{t \to +\infty} F_x(t) = 1$$

$$F_x(t) : monotonic \, , non\ decreasing\ function$$

$$0 \leqslant F_x(t) \leqslant 1 \, ; -\infty < t < +\infty$$

$$P(a < X \leqslant b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

Toss a coin

- S={head, tail}
- X: S->{0, 1}
  - X(head) = 0
  - X(tail) = 1

$$F_X(t) = \begin{cases} 0 & t < 0 \\ q & 0 \leqslant t < 1 \\ 1 & t \geqslant 1 \end{cases} \qquad q = 0.5$$

We are interested to r.v. where the parameter is the time, thus t>0

Examples

- Bernoulli probability mass function
- Binomial probability mass function

Continuous random variable

- The sample space is not enumerable
- It is required that the definition of $X(s)$ is such that $X(s) \leq x$ is an event
- It is a function that assigns a real number to each sample point in S such that

$$\forall x \in \Re, \{s | X(s) \leq x\}$$

  is an event
- Definition of $F_X(t)$ is extended

Probability distribution function (pdf)

$$f_x(t) = \frac{dF_x(t)}{dt}$$

$$F(t) = \int_0^t f(x)dx$$

$$\int_0^\infty f(x)dx = 1$$

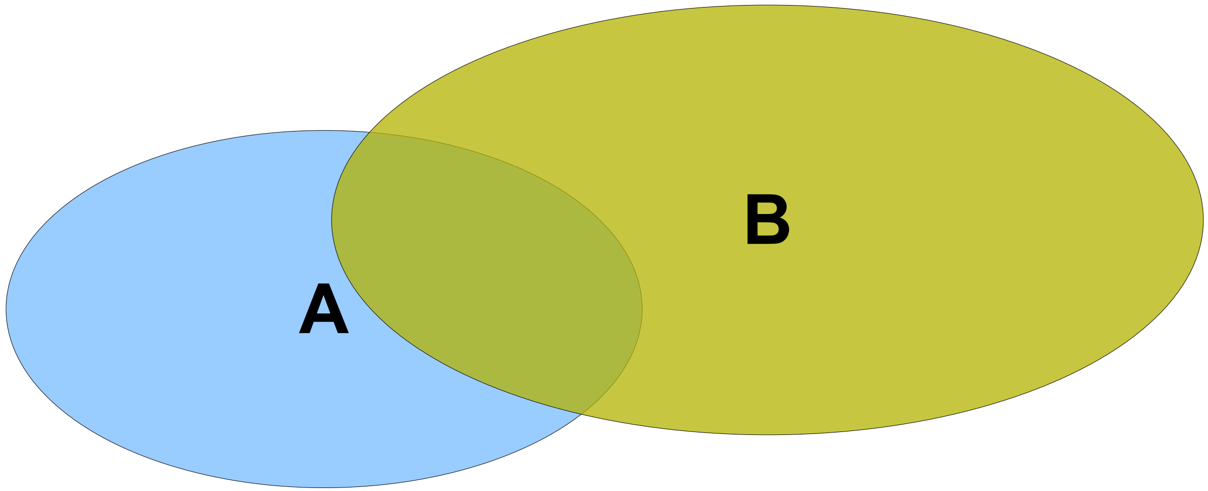
$$P(X \le t) = \int_0^t f(x)\,dx$$

$$\Pr(a \le x \le b) = \int_a^b f(x)dx = F(b) - F(a)$$

Sum of two events

Given two events A e B, where (A and B)≠Ø

$$P[A \text{ or } B] = P[A] + P[B] - P[A,B]$$

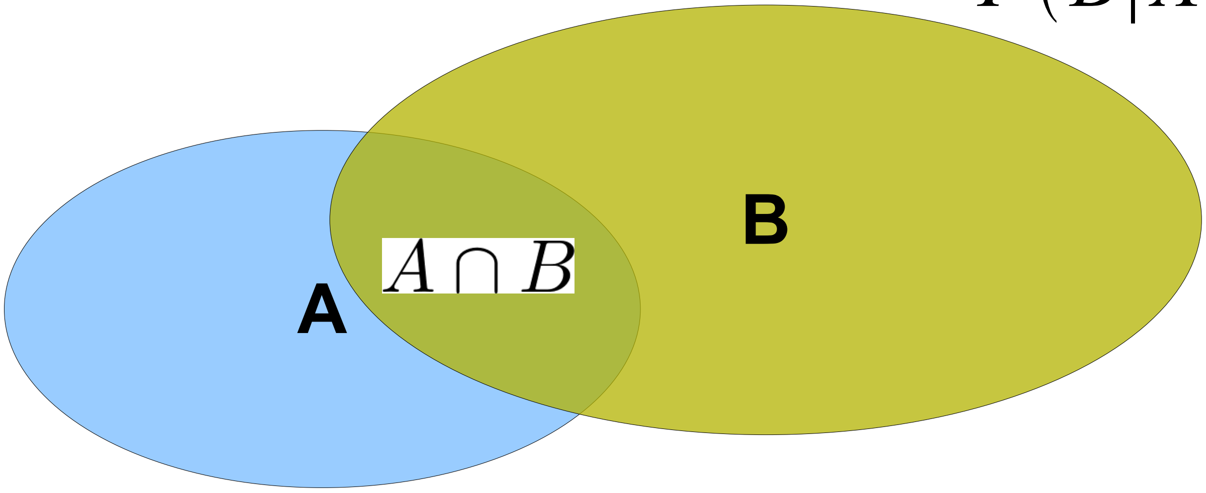


Joint event

The probability that two elementary events, *A* and B, happens during a random experiment

$$P(A \cap B) = P(A|B)\,P(B) =$$

$$= P(B|A)\,P(A)$$



Expected value

- It is a sort of “center of gravity” around which all the possible values are spread
- First moment

$$E[X] = \int_{0}^{+\infty} t\, f(t)\, dt$$

Operations

- Stochastic modeling tries to predict system behavior from the knowledge of its components
- The various parts are to be combined in some way

Joint distribution

- Let us consider two r.v. X_1 e X_2 defined over $[0, +\infty)$
  - Ex: failure time of two hard disks of a RAID 0 system storage
- Their joint distribution is defined as

$$F_{X_1,X_2}(x_1, x_2) = P[(X_1 \leq x_1)\ and\ (X_2 \leq x_2)]$$

- X_1 and X_2 are said *independent* r.v. iff

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)\ F_{X_2}(x_2)$$

Maximum of two r.v.

- X1 and X2 are independent and in the range [0,+∞)
- Y_{max} their maximum r.v. and $F_{max}(t)$ the CDF

$$
\begin{aligned}
F_{max}(t) = P\{Y_{max} \leq t\} = \\
P\{X_1 \leq t \wedge X_2 \leq t\} = \\
F_{x_1, X_2}(t, t) = \\
F_1(t) \;\; F_2(t)
\end{aligned}
$$

- Ex: X_1 and X_2 the failure time of two components => F_{max} is the probability that both are failed

The Maximum of r.v.: an example

- Two threads run concurrently with completion time given by $F_1(t)$ and $F_2(t)$
- Process completion time?

$$F_1(t) = 1 - e^{-4t}; \quad F_2(t) = 1 - e^{-5t}$$

$$F_{max}(t) = F_1(t)F_2(t) = 1 - \mathrm{e}^{-4t} - \mathrm{e}^{-5t} + \mathrm{e}^{-9t}$$

- The maximum of two independent, exponentially distributed r.v. is **not** an exponentially distributed r.v.

Minimum of two r.v.

- X1 and X2 are independent r.v. defined in the range [0,+∞)
- Y_{min} their minimum and $F_{min}(t)$ its CDF

$$
\begin{aligned}
F_{min}(t) = P\{Y_{min} \le t\} = \\
P\{X_1 \le t \vee X_2 \le t\} = \\
1 - P\{X_1 > t \vee X_2 > t\} = \\
1 - F_{X_1, X_2}(t, t) = \\
1 - (1 - F_1(t))(1 - F_2(t))
\end{aligned}
$$

Minimum of two r.v.: an example

- Two independent processes execute in parallel with completion time $F_1(t)$ ed $F_2(t)$
- Which is the completion time of one of them?

$$F_1(t)=1-e^{-4t};\ \ F_2(t)=1-e^{-5t}$$

$$\begin{aligned}
&F_{min}(t)=1-(1-F_1(t))(1-F_2(t))=\\
&=1-(\mathrm{e}^{-4\mathrm{t}}\mathrm{e}^{-5\mathrm{t}})=\\
&=1-\mathrm{e}^{-9\mathrm{t}}
\end{aligned}$$

- The minimum of two independent, exponentially distributed r.v. **is** an exponentially distributed r.v.

Summation of r.v.

- *X1* and *X2* two independent r.v.
- *X1* followed by *X2*
- $F_s(t)=P[X_1+X_2\leq t]$

$$F_s(t)=\iint_A f(x,y)\,dx\;\;dy$$

$$A=\{(x,y)|x+y\leqslant t\}$$

Summation of r.v. (cont.)

$$F_s(t) = \iint_A f(x,y)\,dx\,dy = \int_0^t \int_0^{t-x} f(x,y)\,dy\;\;dx$$

by posing $y=z-x$

$$F_s(t) = \int_0^t \int_x^t f(x,z-x)\,dz\;\;dx = \int_0^t \int_0^z f(x,z-x)\,dx\;\;dz$$

$$F_s(t) = \int_0^t \int_0^z f(x,z-x)\,dx\;\;dz$$

Summation of r.v. (cont.)

- By definition

$$F_s(t) = \int_0^t f_s(z)\, dz$$

thus

$$f_s(z) = \int_0^z f(x, z-x)\, dx = \int_0^z f_{X1}(x) \cdot f_{X2}(z-x)\, dx$$

i.e.

$$f_s(t) = f_{X1}(t) \times f_{X2}(t)$$

$$F_s(t) = F_{X1}(t) \times F_{X2}(t)$$

Summation of two r.v.: an example

$$F_1(t)=1-e^{-4t};\quad F_2(t)=1-e^{-5t}$$

$$F_1(t)\times F_2(t)=\int_0^t\int_0^t\left(1-e^{-5(t-x)}\right)4e^{-4x}dx=$$

$$=\int_0^t[4e^{-4t}-4e^{-5t}e^{(5-4)x}]dx=$$

$$=\int_0^t 4e^{-4x}dx-4e^{-5t}\int_0^t e^x dx=$$

$$=[-e^{-4x}]_0^t-4e^{-5t}[e^x]_0^t=$$

$$=\left(-e^{-4t}+1\right)-e^{-5t}\left(e^t-1\right)=$$

$$=1-e^{-4t}-4e^{-4t}+4e^{-5t}=$$

$$=1+4e^{-5t}-5e^{-4t}$$

- Two stages Erlang distribution

Event generator source

- Discrete events: one event into an infinitesimal time interval dt at most
- Constant probability to generate an event in *dt*: $\lambda\, dt$
- Independent events

Stationary Poisson process

Poisson process

- Probability that *n* events are generated in *Δt*

$$P[X(\Delta t)=n]\equiv p_n(\Delta t)=\frac{(\lambda\Delta t)^n}{n!}\mathrm{e}^{-\lambda\Delta t}$$

Poisson process

- Many natural phenomena (counting) are represented by Poisson process
  - Number of accidents in a given time interval
  - Number of shooting stars spotted on August 15th in 20 min.
  - Number of photons emitted by a radioactive source

The observer

- The observer of a Poisson process is interested in the time that elapses between two emissions (r.v.)
- Probability to have no emission in [0, t[

$$e^{-\lambda t}$$

- Probability to have one emission only in (t,t+dt[

$$\lambda dt$$

- Thus, the probability density function is

$$\lambda e^{-\lambda t} dt$$

The exponential distribution

$$F_x(t) = 1 - e^{-\lambda t} \qquad f_x(t) = \lambda e^{-\lambda t}$$

- It is widely used in system modeling
  - Telematics (packets inter-arrival time)
  - Electronics (components fault time)
  - Computer engineering (jobs process time)

The *memoryless* property

- Given the r.v. X, representing the component life time
- the r.v. $Y=X-t$ is the remaining life time in t

The *memoryless* property

$$F_t(y) = P(Y \le y \mid X > t) = P(X - t \le y \mid X > t)$$

$$= P(X \le y + t \mid X > t) = \frac{P(X \le y + t, X > t)}{P(X > t)} =$$

$$= \frac{P(t < X \le y + t)}{P(X > t)} = \frac{\int^{y+t} f(t)dt}{\int^{\infty} f(t)dt} =$$

$$= \frac{\int^{y+t} \lambda e^{-\lambda t} dt}{\int^{\infty} \lambda e^{-\lambda t} dt} = \frac{e^{-\lambda t}(1 - e^{-\lambda y})}{e^{-\lambda t}} = 1 - e^{-\lambda y}$$

- The remaining life time does not depend on the life already spent

The expected value

- The r.v. X such that $F_X(t)=1-e^{-\mu t}$
- Its expected value is

$$E[X] = \int_0^{+\infty} t \cdot f_X(t)dt = \mu \int_0^{+\infty} t \cdot e^{-\mu t} dt$$

- Applying the integration by parts method

$$E[X]=\mu\left[-\frac{t \cdot e^{-\mu t}}{\mu}\right]_0^{+\infty} + \mu \int_0^{+\infty} \frac{e^{-\mu t}}{\mu} dt = \frac{1}{\mu}$$