Office hour: 4:00 p.m.-6:00 p.m.

One problem of the exam would be the same or little variation of one of the homework problems

Exams would take usually 3 hours, but usually extended one hour

The midterm not only includes the first 3 chapters but also the first two sections of chapter 5 (Markov)

Final will also include first couple of sections of chapter 8 and couple of sections of chapter 7.

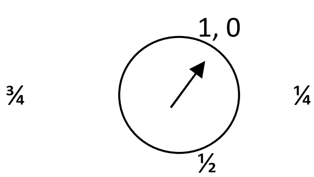
Chapter 4 is important but we will skip it although we may use the parts in chapter 5

- Objective of the course: Formulation and analysis of probability model/ stochastic models
- Real word problem that is random: complicated and not well defined
  - Examples: internet traffic, demand modeling, market share, stock price
- Model that will be generated from abstraction process: Put down some assumptions and hypothesis that try to capture the essence of the problem
  - Model could be used for empirical analysis
  - Models could be part of a larger model
  - The process of going from real word to model could be iterative
  - Is like description
- Example: rolling of single die
  - 6 **possible outcomes**: 1,…,6
  - **Subset of interest**: Grouping outcomes based on your interest (e.g. even vs. odd): {1,2,3} vs. {4,5,6} or {2,4,6} vs. {1,3,5}
  - **Probability**: Equally likely, experiment for the state. You try to roll very large number of times (The concept is called **the law of large numbers**: **if the number of times you do the experiment goes to infinity the probability will be calculated**) since we are not able to go to infinity then you assume the probability, but then you need to try multiple times and test your hypothesis. We assign the probability to outcomes, and this is axiom approach, in comparison to empirical that you must validate and reassure. P{[i]}=1/6; i=1,2…,6
- Formal framework of the probability theory(Ω, F, P)
  - **Sample space**: Ω: a set of all possible outcomes
    - Trajectory could be infinite number and you are picking from that and this set could be a function
    - Points in the set are abstract objects
  - **Events (F)**: a collection of subsets of Ω satisfying several properties
    - **Union**: $\cup$
    - **And: intersection :** $\cap$
    - Complement: "C" would be used
      - Closure: when you are dealing with an event you should be interested in the complemented

- If you are interested in two events you must be interested in intersection and union, otherwise your theory would not make sense
- Events must satisfy following three conditions (they are called **sigma algebra**): F collection of all sets including Ω
  - $E \in F$ => $E^c \in F$
  - $\Omega, \emptyset \in F$
  - E1,…,En∈ F => $\cup E_i \in F$
- **Probability** , P{.}: is a real value function defined on a given set of events
  - P{E}>=0 ∀ E∈ F
  - P{Ω}=1
  - Let E1,E2,… be disjoint events (intersection is empty; $E_i \cap E_j = \emptyset \; \forall i, j$ => P ($\cup E_i$) $= \sum P(E_i)$ called **countable additively**

- Regarding the probability it is usually subjective and sometimes it is common sense.
- Examples of "F":
  - $\{\emptyset, \Omega, \{1,2,3\}, \{4,5,6\}\} = F$
  - $\{\emptyset, \Omega, \{2,4,6\}, \{1,,3,5\}\} = F$
  - $\{\emptyset, \Omega, \{2,4,6\}\} = F3$→ is not Sigma algebra so the probability could not be defined over
  - $\{\emptyset, \Omega, \{1\}, \{2\}, \ldots \{1,2\}, \{1,3\}, \ldots, \{1,2,3,4,5,6\}\} = F4$→ power set

Computation of the probability example:



Ω=[0,1]

If we ask what is the probability of one value e.g. 1, ½, … that would be Zero. Alternative approach here is to look at the intervals here.

The probability should be allocated to the building blocks, as in the previous example was intervals.

- You start with elementary elements and you extend up to the sigma algebra set.
- You assign probability to building blocks set. You then extend your assignments up, and the extension would be conceptual.
- You make assumption for example over how the price would be evolved and then extend it to compute the probabilities
- **Uniform distribution**: means the probability of any probability interval would be the same

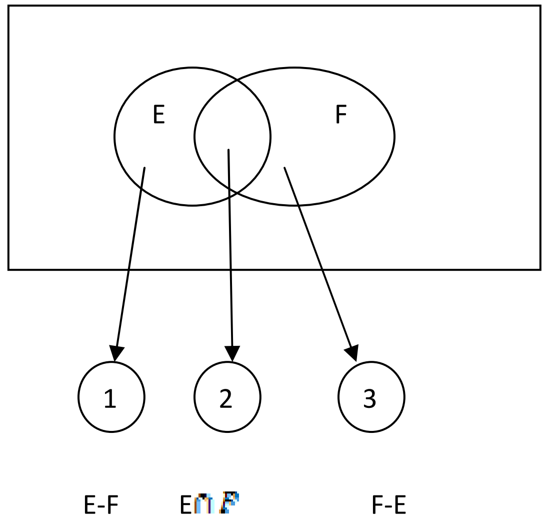
Next topic: Computational probability:

- Given P{E}, P{Ec}=?

Proof: Ω=E ∪Ec , both are disjoint => 1= P{Ω}=P{ E ∪Ec}= P{E}+P{Ec}=> P{Ec}=1-P{E}

- P(EUF)=P(E)+P(F)-P(E∩F)

  Proof: Van Diagram



P(EUF)= $\underline{P(E-F)}$+P(F-E)+$\underline{P(E\cap F)}$= P(E)+$\underline{P(F-E)+P(E\cap F)}$-P($E\cap F$)=P(E)+P(F)-P(E∩$F$)

- Extension-Inclusion-Exclusion principles

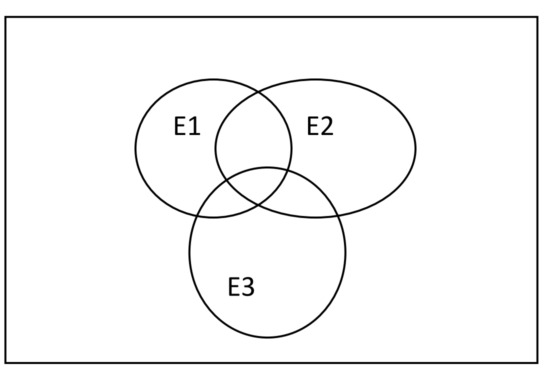
P(E1 $\cup E2 \cup) = P(E1) + P(E2) + P(E3) - P(E1\cap E2) - P(E2\cap E3) - P(E1\cap E3) + P(E1\cap E2\cap E3)$

P($\cup\, Ei) = \sum_{i=1}^{n} P(Ei) - \sum_{i=1}^{n}\sum_{j=1}^{n} P(Ei\cap Ej) + \sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} P(Ei\cap Ej\cap Ek)$+
$(-1)^{n+1}P(\cap_{i=1}^{n} E\, i)$

We use induction to prove this.
$\bigcup_{i=1}^{n+1} Ei = P\left(\bigcup_{i=1}^{n+1} Ei \cup Ei+1\right) = P\left(\bigcup_{i=1}^{n+1} Ei\right) + P(E\, i+1) - P\left(\bigcup_{i=1}^{n+1} Ei \cap Ei+1\right) = P\left(\bigcup_{i=1}^{n+1} Ei\right) +$
$P(E\, i+1) - P\left(\bigcup_{i=1}^{n+1}(Ei\cap Ei+1)\right) = \sum_{i=1}^{n} P(Ei) - \sum_{i=1}^{n}\sum_{j=1}^{n} P(Ei\cap Ej) + \sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} P(Ei\cap$
$Ej\cap Ek)$
+ $(-1)^{n+1}P(\cap_{i=1}^{n} E\, i)$

Second proof: Combinatorial proof:

Consider a arbitrary point in $x \in \bigcup_{i=1}^{n} Ei$



Suppose without loss of Generality x belongs to E1, E2, and Ek, where 1<=k<=n

Inclusion: how many times the point is counted when we go on inclusion

Excluded: how many times the point as we go to exclusion

Term	Inclusion/Exclusion	Count
1st	Inclusion	k
2nd	Exclusion	$-\binom{k}{2}$
..		..
k th		$(-1)^{n+1}\binom{k}{k}$
…		0

Total count= k-$\binom{k}{2}$+….+ $(-1)^{n+1}\binom{k}{k}$= 1-(1-k+$\binom{k}{2}$+….+ $(-1)^{n+1}\binom{k}{k}$) $= 1 - 0$

Binomial expansion:

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

$\leq$

- Two Sunday from today (23rd Sep) at 13:00 we will have the makeup class
- Assignment is put on the website that it is better that you attempt them

<Conditional Probabilities>

Let A& B be two events. P(A|B) is the probability of A occurs given that B occurred.

= $P(A \cap B)/P(B)$; provided that $P(B)>0$

This can also be written: $P(A \cap B)=P(A|B).P(B)$

Concept: As we vary A, P(A|B). i.e. "Probability"

(i) $P(\phi|B)=0$

(ii) $P(\Omega|B)=1$

(iii) $P(A1 \cup A2|B)= P((A1 \cup A2) \cap B)/P(B)= (P(A1 \cap B)+P(A2 \cap B))/P(B)$

Example:

1. You have 52 cars, and you take one randomly. I know that it is an Ace.

   A: ace id drawn

   B: spade ace

   $P(B|A)=P(B \cap A)/P(A)=1/52/4/42=1/4$

2. Multiple choice exam:

   m: # of possible choices

   k: the student knows the answer

   k': the student does not know

   assumptions:

   (i) P(k)=p

   P(k ')=1-p

   (ii) If the student does not know the answer s/he will choose randomly

   Let C= the student answered a question correctly

$P(k|c)= P(k\cap c)/P(C)= p/ (P(k\cap C)+P(k'\cap C))=p /(p+(1-p)1/m)$

P(c): we try to make two disjoint events

Computing probability conditioning

Partition Ω to n sets, E1, E2, …, En; these sets should be mutually exclusive.

$\cap E\, i = \Omega$

$P(F)= P(\bigcup (F \cap E\, i))= \Sigma\, P(F \cap E\, i)$

$= \Sigma P(F\, |E\, i).\, P(E\, i)$

- this part is called unconditioning).
- It is sort of divide and conquer

- You can choose E i (design partitions) so that all the smaller pieces would be simpler to calculate. If you divide terribly then you end up just making it more complex.
- **Bayes formula:** let E1, …, En be the partitions of Ω and F be an event of interest; kind of backtrack to find out whether the student knew the answer correctly or not. It is kind of go back in time, so you can attach time stamp (t) to it.

$$P\,(E\, i\, |F)= P(F|E\, i).P(Ei)/\, \Sigma P(F\, |E\, i).\, P(E\, i)$$

Condition above had at the left had reverse logical order, and at the right side it would be forward logical order.

Independence:

Let A & B be two events:

Earlier we said: $P(A\cap B)=P(A|B).P(B)= P(B|A).P(A)$

A & B are independent if

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

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

Combining two ill result in $P(A\cap B)=P(A).P(B)$

The above condition also is called factorization that you ignore information about the other guy

- For more events for any subcollections the probability should factorize the product of the individual probability

Events E1, E2, …, En are independent if P($\bigcap$ (E i, i=1, L))= $\prod$ (P (E i, i=1, L)

Chapter 1 is finished. You can now work on the problems

<Random variables>

Chance that will come from sample space

Example. Two dice are rolled (independently). Let X i = i th outcome. Consider (x1=3, x2=5)

For this experiment:

Ω={(1,1), (1,2),…., (6,6)} : 36 possible outcome

∃ = all possible subsets of Ω

P: { (i,j)}=1/6.1/6=1/36

- We could be interested in lots of "quantities" related to this experiment.

Example:

(i) x1+x2→ (i, j)➔ i+j

(ii) x1→ (i, j)→i

(iii) x1-x2→ (I, j)→ i-j

(iv) 2x1-x2→ (i, j)→ 2i-j

..

Definition: A random variable (r. v.) is a real value function defined on sample space.

Stochastic process helps you understand how things will behave

Stochastic process is sequence of random variable

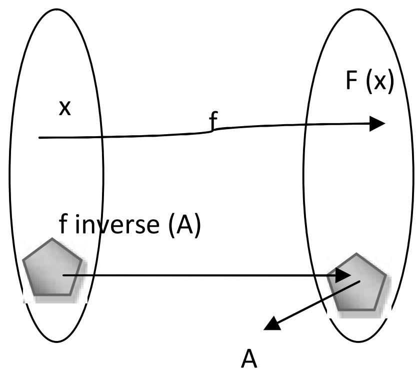
Description of random variable:

Sometimes we need to go back to the sample space

Distribution says what is the probability that something is lower or equal to something (ranging)

P{x1+x2=4}= p{{1,3}, {3,1}, {2,2}}=P({1,3})+P({3,1})+P({2,2})=1/36+1/36+1/36=3/36

$\Omega \rightarrow$ Range (real number)



Let X be a function defined on Ω.

ω:point in $\omega \in \Omega$

$\omega \rightarrow X(\omega)$

Let A be the event| subset of interest in the range specified by X

P{X$\in$A}=?= P{ X^{-1} (A)}

P {$\omega \in \Omega$: X(ω) $\in$ A}: probability of <u>inverse image</u>

Inverse of (A)={ $\omega \in \Omega$: X(ω) $\in$ A}

- Non-measurability in case there were no inverse image, since they do not collect into a union of sets. Usually happens with real values.
- Suppose the range space contains random numbers. In this case we will consider the form A is equal to $(-\infty, x]$ only.
- In other words we want to specify P{x$\in (-\infty, x]$}=p(X<=x}
- X is function of f(x) and it is called cumulative distribution function

Definition. The cumulative distribution function (cdf) X I denotes by F x(.) is defined by F X (x)=P(X<=x) where x is a real number.

- Function of random variable would be still a random variable

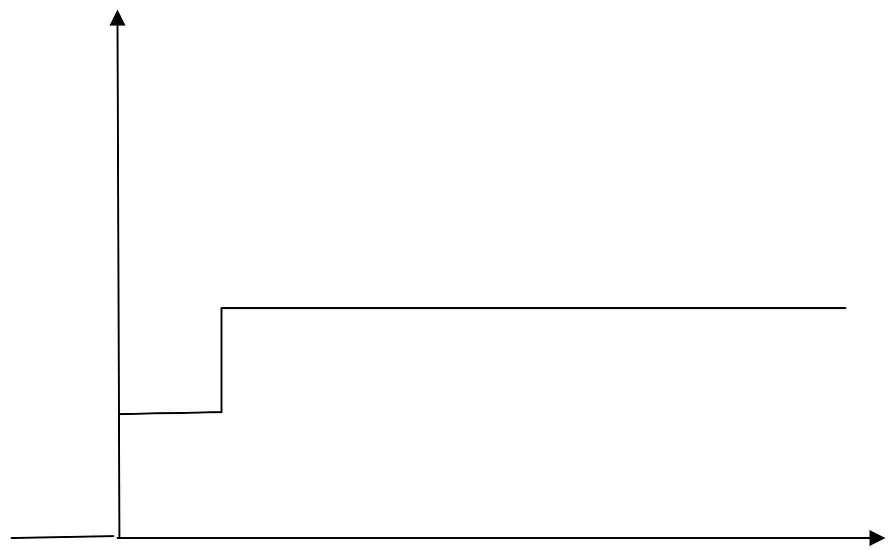
Example:

1. Coin flip:

   X={ H=1; T=0}

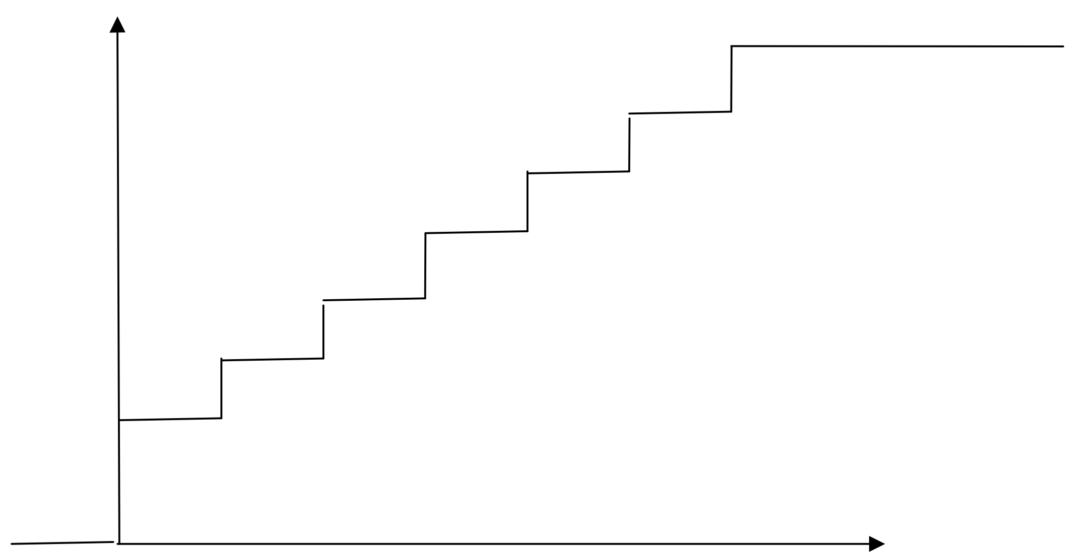
   Suppose biased P(H)=p; P(T)=1-p

   Form of the CDF would be the following



2. Die

   X={1, 2, 3, 4, 5, 6} with equal probability of 1/6



These two were discrete functions.

Basic Properties:

(1) Lim (F x(x), -∞)=0

(2) Lim (F x(x), ∞)=1

(3) F X (x) it would be non decreasing(↑)

(4) F X (x) is right continuous

(5) Lim (F X (x), x+ →a) = fx(a)

Next time we will over discreet and continuous variables, and two and more variable showing at the time

$\leq$

Example:

Discrete:

1. Bernoulli random variable (coin flip)

$$x = \begin{cases} 1(success) : p(probabilitymassat1) \\ 0(failure) : 1-p(probabilitymassat0) \end{cases}$$ (0<+p<=1)

Notion: X ~Bernoulli (P)

~: is distributed as

P: parameter

f X (1)=p

f X (0)=1-p

2. Binomial Random Variable:

   Repeat Bernoulli experiment.

   Let x1, x2,…, xn be Bernoulli(p) and they are" independent". Define X= $\sum_{i=1}^{n} X_i$ = total no. of success out of n independent Bernoulli trials

   Then: X~ Binomial (n, p)

   f: probability mass function will be noted by this (density function): derivative of cumulative distribution function

   F: Cumulative distribution function

   Probability mass function of is: f X (i)=p{X=i}= $p^i (1-p)^{n-i}$

3. Geometric Random variable:

   X= # of trials necessary to observe the first success in a sequence of independent Bernoulli trials.

   X~ Geoemtric (p)

   $f_x(i) = (1-p)^{i-1} p^i$ i>=1

$$\infty \sum_{i=1} f_k(i) = p\frac{1}{1-(1-p)} = 1$$

4. Poisson Rand variable:

   X ~ Poisson (λ) if

   $$f_x(i) = e^{-\lambda}\lambda^i / i!, i = 0,1,2,...$$

   Motivation: Poisson random variable is a random variable often used to model events that are randomly over time (e.g. click stream, shocks to the stock market, telephone traffic coming to UTD): usually things that occur often could be described with this; 1. rates on the intervals should be constant and 2. intervals should be disjoint and independent.

   If you keep doubling the intervals and make it smaller and smaller (n interval division where n is large), where in each we will have either zero or no event, would be similar to binomial that they would be independent.

   Total count so ~ Bin (n, p)

   np=λ

   Expected value (Mean) of Bernouli would be p

   For binomial the expectation would be np

   This means we are trying to fix mean number here

   Claim X n ~ Bin(n,p), with np=λ

   Then $\lim_{n->\infty} \{X_n = i\} = e^{-\lambda}\frac{\lambda^i}{i!}$

Continuous Random variables:

Concept of density function: $f_x(x) = \lim_{h->0}\frac{F(x+h)-F(x)}{h} = \frac{d}{dx}F_x(x)$

Derivative: rate of change of the function

In this case, we have

P{a<X<b}= $\int_a^b f_x(x)dx$

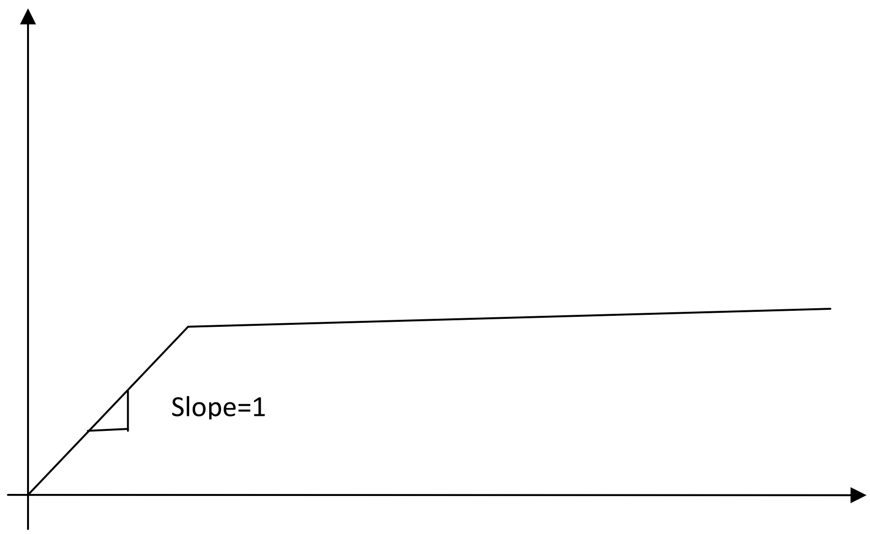
In general, X is <u>continuous</u> if there exists a function $f_x(.)$ so that: P{ $X \in B$ }= $\int_B f_x(x)dx$

1. Uniform (0,1]

X~ Unif (0,1)

$$f_x(x) = \begin{cases} 1 & if\, 0 < x < 1 \\ 0 if & otherwise \end{cases}$$

$$F_x(x) = \int_{-\infty}^{x} f_x(x)dx = x \qquad if \qquad x \in (0,1)$$

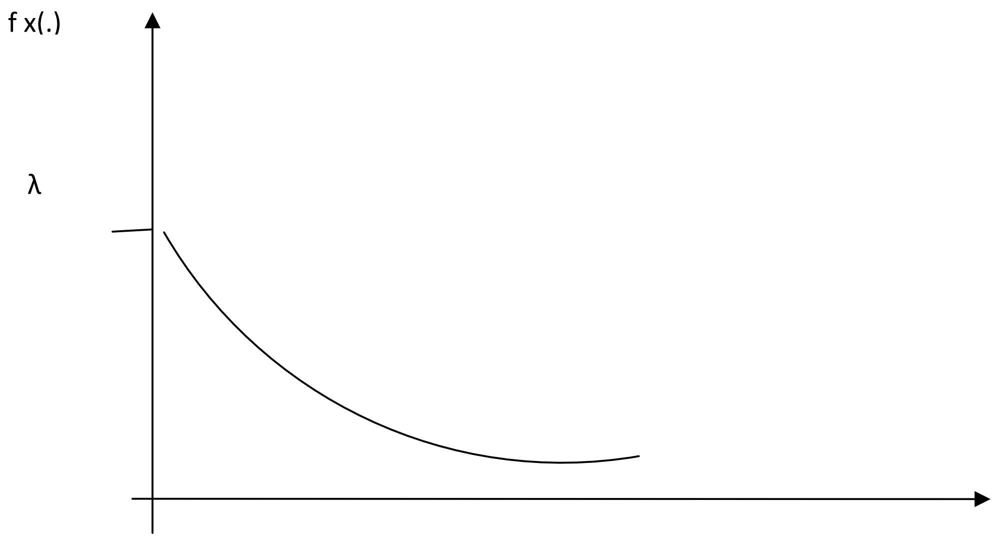


Using sequence of any random variable you can make the other random variables. Excel has the function Rand() which generates pseudo random variable, and it is deterministic random variable.

2. Exponential Random variable

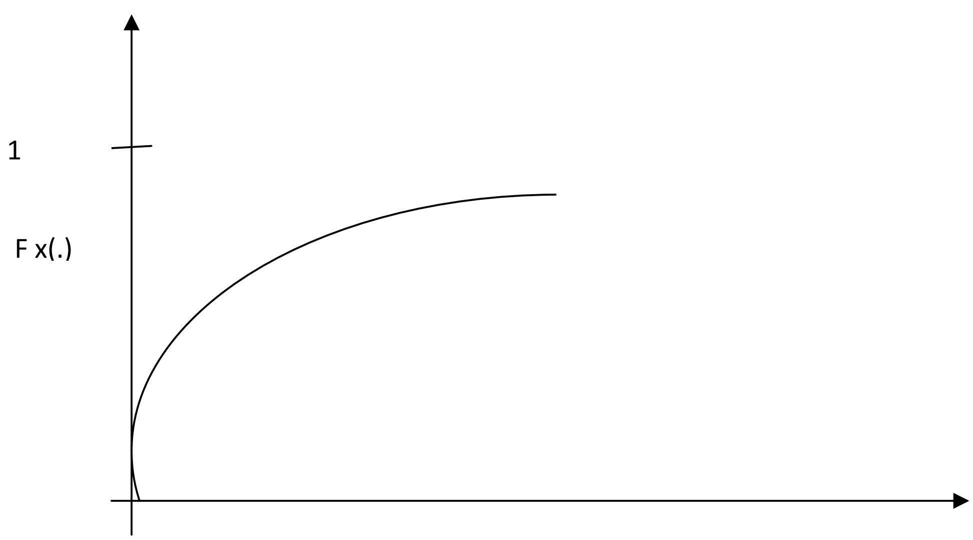
X~Exp(λ) if

$$f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x >= 0 \\ 0 & otherwise \end{cases}$$



Key property: it has no “Memory”

It is determinant of distribution of lifetime saying that the life of old table would have the same variation as the life of the new table.



$$F_x(x) = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x} \qquad x >= 0$$

3. Gamma | Erlang Random Variable

X ~ Gamma (n, λ) if

$$f_x(x) = \begin{cases} e^{-\lambda x} \dfrac{(\lambda x)^{n+1}}{(n+1)!} \lambda & if \quad x >= 0 \\ 0 \quad otherwise \end{cases}$$

Interpretation: Let x1, x2, …, xn be n independent exponential random variable with parameter λ

Definition:

$X = \sum_{i=1}^{n} X_i$ then X~ Gamma (n, λ)

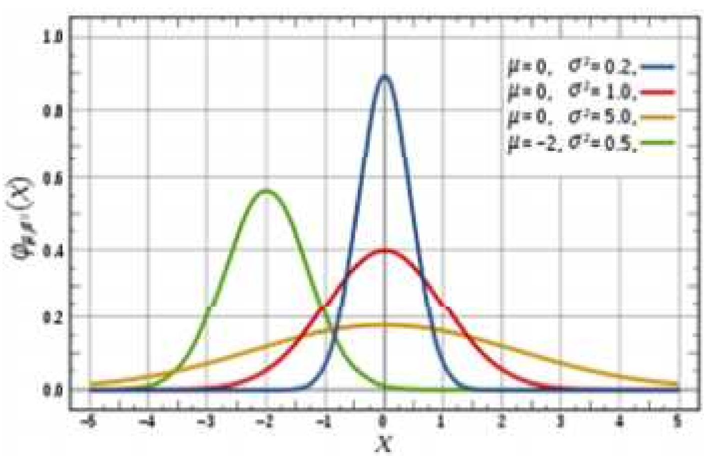
So it would be distribution of life of five tables.

4. Normal Random variable:

X~N(μ, σ^2) if

-∞<x<∞

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-1/2\ (\frac{x-\mu}{\sigma})^2}$$



Interpretation: Average of large number of independent random variables.

Expectation: Let X be a random variable.

The expectation of X, E(x) is defined by

$$E(x) = \begin{cases} \sum_{all x} x f_x(x) if \quad x \quad is \quad discrete \\ \int_{-\infty}^{+\infty} x f_x(x) dx \, if \quad x \quad is \quad contineous \end{cases}$$

Notation = $\int_{-\infty}^{\infty} x dF_x(x)$

Stiglar integral, Raymond formula

Example:

1. X~Bionomial (p)

   E(x)=1.p+0.(1-p)=p

2. X~Bin(n,p)

   E(x)= $\sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = np$

3. X ~ Exponential (n,p)

   E(x)= $\sum_{i=0}^{n} ip(1-p)i-1 = 1/p$

4. X ~ Poisson (λ)

   E(x)= $\sum_{i=0}^{n} \frac{ie^{-\lambda}\lambda^i}{i!} = \lambda$

5. X~Unif(0,1)

   E(x)= $\int_0^1 x.1.dx = 1/2$

6. X~Exp(λ)

   E(x)= $\int_0^{\infty} x.\lambda e^{-\lambda x}.dx = 1/\lambda$

7. X~Gamma(n,λ)

   E(x)= $\int_0^{\infty} \frac{x.e^{-\lambda x}.(\lambda x)^{n-1}}{(n-1)!}.dx = n/\lambda$

   X~N(μ, σ^2)

$$E(x)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma}}e^{-1/2\,(\frac{x-\mu}{\sigma})^2}dx=\mu$$

Interpretation:

E(x): log run average of sequence of independent observed value of x (sampl/realizations)

Function of X, g(x)

$$E[g(x)]=\begin{cases}\sum_{allx} g(x)f_x(x) \\ \int_{-\infty}^{\infty} g(x)df_X(x)dx \quad \int_{-\infty}^{\infty} g(x)dF_X(x)\end{cases}$$

Examples:

1. E(X^n)→ nth movement
2. $E[(x-\mu_x)^2]=Var(x)$: Variance

   Long run average of squared deviation from the mean

   Spread, variability

   Square smoothes the abstract value and therefore is used here

   Short cut: = $E\{X^2-2XE(X)+E(x)^2\}=E(x^2)-2E(x)^2-E(x)^2=E(x^2)-E(x)^2$

   Def.

   $$\sigma_x=\sqrt{Var(x)}$$

   Coefficient of Variation (nonnegative):

   $$X=\begin{cases}10+1 & w.p. \quad 1/2 \\ 10-1 & w.p. \quad 1/2\end{cases}$$

   $$Y=\begin{cases}10{,}000+1 & w.p. \quad 1/2 \\ 10{,}000-1 & w.p. \quad 1/2\end{cases}$$

   $$C_v=\sigma_x/E(x)$$

   For exponential distribution is one.

3. Ratio= $\frac{E(x)-r_f}{\sigma_x} = E(\frac{x-r_f}{\sigma_x})$

   If you take a risky decision you will deviate from the mean. For investment in mutual fund it will be used.

   r f is the reference that you are comparing to for example risk free market, and E(x), and σ are related to the mutual fund's performance during last years.

$\leq$

Joint distribution:

Let x, y (or (x,y) be the random variables.

Joint distribution cumulative of (x,y) is defined by $F_{X,Y}$ (x, y)= {X<=x, Y<=y} for all x,y

Marginal Distribution:

$F_X(x) = \lim_{x \to \infty} F_{X,Y}(x, y)$

$F_Y(y) = \lim_{y \to \infty} F_{X,Y}(x, y)$

Independence:

Definition, X & Y are said to be <u>independent</u> if

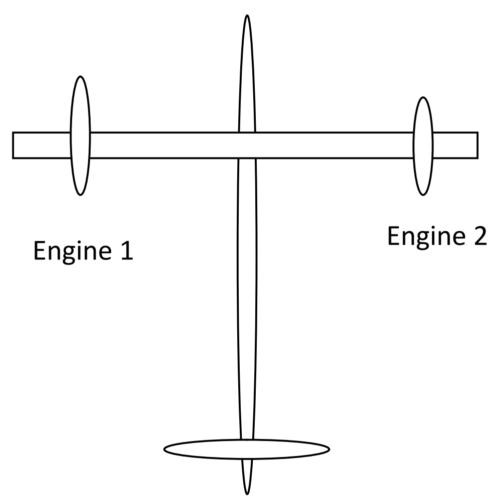
$$F_{X,Y}(x,y) = F_X(x)\, F_Y(y) \text{ for all } x,y$$

In terms of probability mass/ density function

This is equivalent to

$$f_{X,Y}(x,y) = f_X(x).f_Y(y) \quad \forall x, y$$

Example:



Type 1 shock: knock out engine 1

Type 2 shock: knocks out engine 2

Type 3 shock: knocks out engine 1&2

X_i= time of occurrence of the type-i shock

i=1,2,3

Assumption: $X_i \sim$ Exp (λ_i), i=1,2,3

X_1, X_2, X_3 are independent (all mutually)

Define:

Y_1= failure time of engine 1

=min(X_1, X_3)

Y_2= failure time of engine 1

=min(X_2, X_3)

Question: (Y_1, Y_2)=?

Tail: 1-F is called tail since it talks about the variable greater than something (*complementary*, or *survival* probability/function could also be called)

In this case it is easier to work with tail

A: p{Y_1>u, Y_2>v}= p{X_1>u, X_3>u, X_2>v, X_3>v}= p{ X_1>u, X_3>max(u, v), X_2>v}= p{ X_1>u} p{X_3>max(u, v)} p{X_2>v}= $e^{-\lambda_1 u}$ $e^{-\lambda_2 v}$ $e^{-\lambda_3 \max(u,v)}$

Question 2: are Y_1 & Y_2 independent?

A2 p{Y_1 >u}=$e^{-(\lambda_1+\lambda_3)u}$

That is $Y_1 \sim$ Exp ($\lambda_1+\lambda_3$)

Since in this case we have non negative the lower bound will go to zero

Similarly p(Y_2>v}=$e^{-(\lambda_2+\lambda_3)v}$

⇨ $e^{-\lambda_1 u}$ $e^{-\lambda_2 v}$ $e^{-\lambda_3 (u+v)}$

Since this is not equal to $e^{-\lambda_1 u}$ $e^{-\lambda_2 v}$ $e^{-\lambda_3 \max(u,v)}$ so they are not independent

Expectation of g(x,y)

E[g(x,y)]= $\sum_x \sum_y g(x,y) f_{x,y}(x,y)$

$$\int \int g(x,y) f_{x,y}(x,y)dxdy$$

Examples:

(i) f(x,y)=x+y

E(X+Y)=

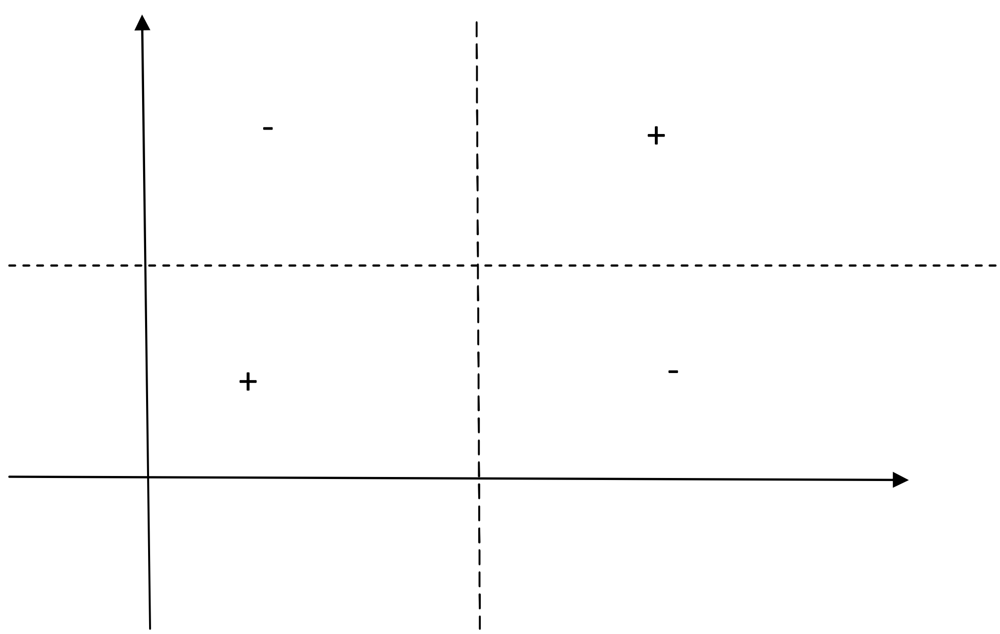
$\int \int (x+y) f_{x,y}(x,y)dxdy = \int \int x f_{x,y}(x,y)dxdy + \int \int y f_{x,y}(x,y)dxdy = \int x f_x(x)dx +$

$\int y f_x(x)dx$ =E(x)+E(y)

(ii) g(x, y)=[X-E(x)][Y-E(y)]

E{[XE(x)][Y-E(y)]}=Cov [x, y]: covariance

For random vector of X, Y, and deviation from the mean, you will get the long run average of product of deviations from the mean



Covariance will tell you *how* variables <u>move together</u> on average

Cov [x,y]= E{XY-XE(Y)-E(X)Y+E(X)E(Y)}=E(XY)-E(X)E(Y)-E(X)E(Y)+E(X)E(Y)=E(XY)-E(X)E(Y)

E(XY)= $\int \int xy f_{x,y}(x,y)dxdy$

Basic properties of covariance:

- Cov[cX, Y]= c. Cov [X,Y]
- Cov [X,X]= Var (X)
- X <u>independent</u> of Y => *Cov [X,Y]=0*

Proof:

Important fact: X, Y are *independent* => *E(g(x)h(y)]=E[g(x)] . E[h(y)]*

If the variables the distribution of the expectation would be ***finite*** reverse of above could be hold.

Proof: Left hand side=

$$\int \int g(x)h(y)f_{x,y}(x,y)dxdy = \int \int g(x)h(y)f_x(x)f_y(y)dxdy = \int g(x)f_x(x)dx \int h(y)f_y(y)dy =$$

E[g(x)]E[h(y)]

In particular, X independent of Y implies E(X, Y)=E(X)E(Y)

Therefore X independent of Y implies that Cov[X,Y]=0

- Covariance equal to zero *<u>does not imply</u>* that both of them are independent, since covariance is kind of *<u>average function</u>*, and things could be *<u>cancelled nicely</u>* but they would not be independent.
- When you *<u>average</u>* something you *<u>are losing information</u>*.

Example is:

$$x = \begin{cases} 1 & 1/3 \\ 0 & 1/3 \\ -1 & 1/3 \end{cases}$$

$Y=X^2$

The covariance between two is zero; however, they are *not independent*.

The information regarding the *<u>marginal</u>* would <u>not</u> be <u>sufficient</u>, unless the variables would be <u>independent</u>.

Related concept: the *<u>scale of each variable</u>* will have effect on covariance in *<u>linear manner</u>* which is not desirable as a result the covariance is *divided* by the *<u>standard deviations</u>*. So:

$\rho_{X,Y}$ = Correlation coefficient between X & R

= $\dfrac{Cov[X,Y]}{\sigma_x \sigma_y}$

-1<= $\rho_{X,Y}$ <=1

(iii) Var(X+Y)=E[(X+Y)-E(X+Y)]^2=E{[X-E(X)]+[Y-E(Y)]}^{2}=E{[X-E(X)]^2+[Y-E(Y)]^2+2(X-E(X))(Y-E(Y))}=Var(X)+Var(Y)+2Cov[X,Y]

Last part 2Cov[X, Y] =0 if X and Y would be independent

This has implication in finance, and you want to *diversify*, and to reduce the variability you need the *Cov[X,Y] be negative*.

You need to invest in *bond*, and *stock* together since they *go in separate direction*, and by that *variability* and *risk* would be reduced

General Formula:

Var($\sum_{i=1}^{m} c_i X_i$)= $\sum_{i=1}^{m} c_i^2 Var(X_i) + \sum_{i \neq j}^{m} \sum c_j c_i Cov(X_i, X_j)$

Var($\sum_{i=1}^{m} c_i X_i \sum_{i=1}^{m} d_i Y_i$)= $\sum_i \sum_j c_j c_i Cov(X_i, Y_j)$

Convolution:

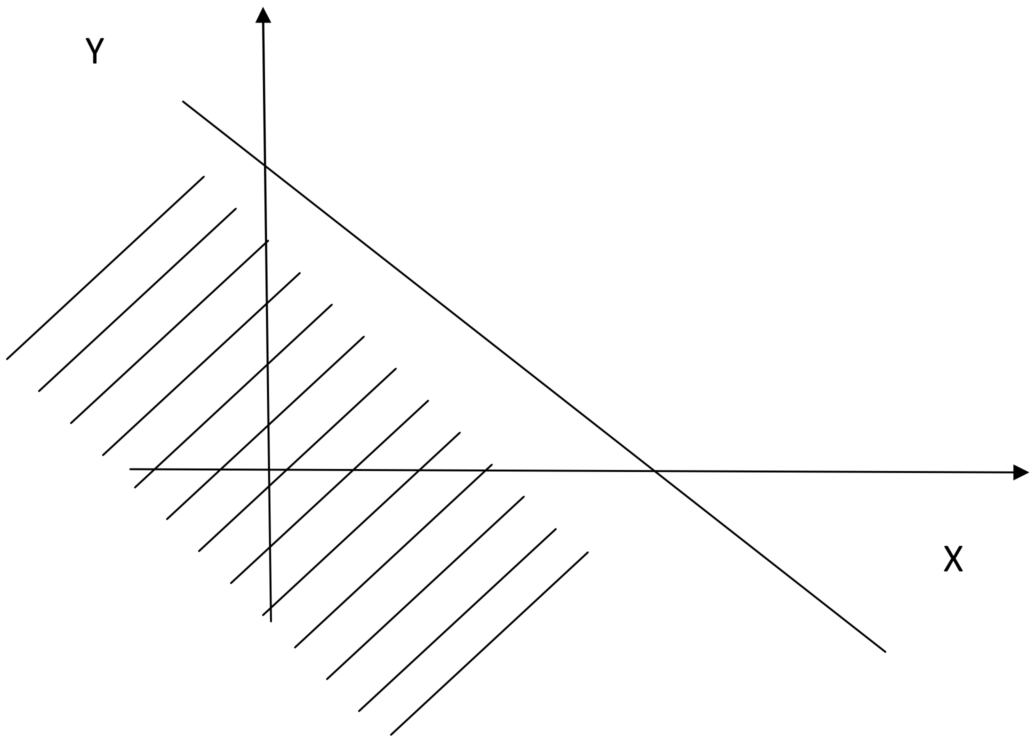
Suppose that X & Y are independent

What is the distribution of X+Y?

Answer: *p{X+Y<=t}* = $\int \int_{x+y<t} f_{x,y}(x,y)dxdy$ = $\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f_{x,y}(x,y)dy \right] dx$ = $\int_{-\infty}^{+\infty} F_Y(t-x) f_X(x)dx$

$$f_{X+Y}(t) = \int_{-\infty}^{+\infty} f_Y(t-x) f_X(x)dx$$

Since both are independent were able to do this



Example:

(i) X is Bin(n,p), y~Bin(m,p)

We chose same p for both binomial distributions

X+Y~?

P{X+Y=k}= $\sum_{i=0}^{k} p\{X=i, Y=k-i\} = \sum_{i=0}^{k} p\{X=i\} p\{Y=k-i\} =$

$\sum_{i=0}^{k} \binom{n}{i} p^{i}(1-p)^{n-i} \binom{m}{k-i} p^{ki}(1-p)^{(m-k+i)} = \binom{m+n}{k} p^{k}(1-p)^{m+n-k}$, k=0,…, m+n

Assumption was that they were independent.

Direct interpretation:

Think of this as m+n Bernoulli trials with success probability p each.

(ii) Suppose X is Poisson with λ_1, Y Poisson with λ_2, and they are independent

F $_{X+Y}$

(n)=p{X+Y=n}= $\sum_{i=0}^{n} p\{X=i\}P\{Y=n-i\} = \sum_{i=0}^{n} e^{-\lambda_1}\lambda_1^{\,i}/i!\,e^{-\lambda_2}\lambda_2^{\,n-i}/(n-i)! =$

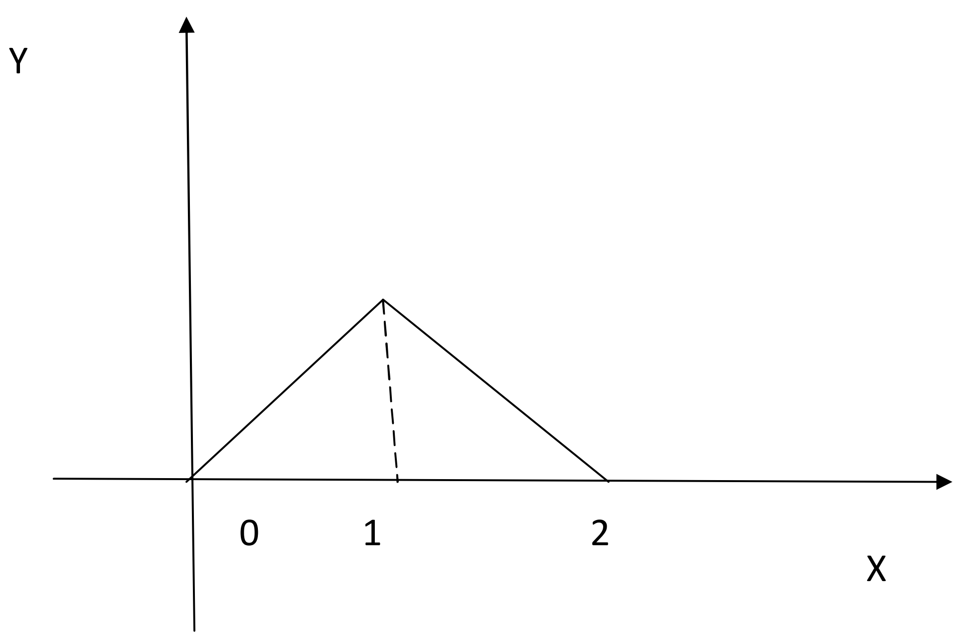
$e^{-(\lambda_1+\lambda_2)}/n!\sum_{i=0}^{n}\lambda_1^{\,i}\lambda_2^{\,n-i}.n!/i!(n-i)! = e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n/n!$

(iii) X~Exp(λ), Y~Exp(λ), and X and Y are independent

$f_{X+Y}(t) = \int_0^{+\infty} f_Y(t-x)f_X(x)dx = \int_0^t \lambda e^{-\lambda(t-x)}\lambda e^{-\lambda x}dx = \lambda^2 e^{-\lambda t}t = e^{-\lambda t}\frac{(\lambda t)^{2-1}}{(2-1)!}\lambda$: Gamma Distribution with (2, λ)

(iv) X~Uniform(0,1), Y~Uniform(0,1), and X is independent of Y

$f_{X+Y}(t) =$



Moment generating functions

Let X be random variable the MGF of X, denoted by

$\varphi_X(t)$ is defined by

$\varphi_X(t) = E(e^{tx})$

For all t so that the expectation is well defined.

Properties:

If $\frac{d}{dt}\varphi_X(t) = E(x)$ first movement

$$\frac{d^n}{dt^n}\varphi_X(t) = E(x^n)$$

- MGF characterizes the distribution of the random variable

Sometimes it is *easier* to work with moment generating function that accumulative distribution

Examples.

(i) X~ Bin (n,p)

$$\varphi_X(t) = \sum_{i=0}^{n} e^{ti}\binom{n}{i}p^i(1-p)^{n-i} = \sum_{i=0}^{n} \binom{n}{i}(pe^t)^i(1-p)^{n-i} = (e^t p + 1 - p)^n$$

Y~ Bin(m,p)

$\varphi_y(t) = (e^t p + 1 - p)^m$; Independence assumption

$$\varphi_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX}e^{tY}) = \varphi_X(t)\varphi_Y(t) = (e^t p + 1 - p)^{m+n}$$

$\leq$

- Sample exams of two previous years will be posted on e-learning next week

MGF:

$\Phi_x(t) = E(e^{tx})$

A continuous example:

X=Exp(λ)

$$E(e^{tx}) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda - t}\int_0^\infty (\lambda - t)\lambda e^{-(\lambda - t)x}dx = \frac{\lambda}{\lambda - t}$$

X~ Gamma(n,λ)

$$E(e^{tx}) = (\frac{\lambda}{\lambda - t})^n$$

(ii) X~N(μ, σ^2)

$$\Phi_x(t) = E(e^{tx}) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Several Related concepts:

- Probability Generating Function (PGF)➔ $E(z^x)$, where x is a nonnegative integer value random variable (0<=z<1)
    - $E(z^x) = \sum_{j=0}^{\infty} z^j p\{x = j\} = p\{x = 0\} + zp\{x = 1\} + z^2 p\{x = 2\} + ...$
    - $1/n!.\frac{d^n}{dz^n}E(z^x)(z \to 0) = p\{x = n\}$
    - Note. $z^x = e^{(\ln z)x}$: now it is close to *moment generating function*
- Laplas transformation ➔ $E(e^{-tx})$ for nonnegative x

  We tried to make it finite in contrast to the moment generating function

Example: X~ poisson (λ)

$$E(z^i) = \sum_{j=0}^{\infty} z^j e^{-\lambda} \frac{\lambda^j}{j!} = e^{-\lambda} \sum_{j=0}^{\infty} z^j \frac{\lambda^j}{j!} = e^{-\lambda(1-z)}$$

Multivariate extension:

Let X=(x_1,…, X_n)

The multivariate MGF is $E(z^{t^T X})$

Example: let z_1, z_2, .., z_m be i.i.d (independent and identically distributed) N(0,1) random variables.

Consider $X_1=a_{11}z_1, a_{12}z_2+\ldots+a_{1n}z_n+\mu_1$

$F_2(z)$ $X_2=a_{21}z_1, a_{22}z_2+\ldots+a_{2n}z_n+\mu_2$

…

$F_n(z)$ $X_n=a_{n1}z_1, a_{n2}z_2+\ldots+a_{nn}z_n+\mu_n$

X=A.Z+μ

X is said to have *the multivariate normal distribution*.

Time to failure for engine 1, time to failure for engine 2, and ….

The life time of each engine would be multivariate random variable

The MGF of X is:

$$E(z^{t^T X}) = e^{\sum_{i=1}^{m} t_i \mu_i + 1/2 \sum_i \sum_j t_i t_j Cov(x_i, x_j)}$$ (Since, MGF of normal variable: $e^{\mu t + \frac{\sigma^2 t^2}{2}}$)

$$E(t^T X) = \sum_{i=1}^{m} t_i \mu_i$$

$$Var(t^T X) = \sum_i \sum_j Cov(t_i x_i, t_j x_j) = \sum_i \sum_j t_i t_j Cov(x_i, x_j)$$

Generally since the function gets all the **inputs** they should be dependents; however, we will try to *prove that they are independent*.

In general zero covariance does not imply independence; however, in multivariate it implies that.

Note. $\text{cov}[x_i, x_j] = 0 \forall i \neq j$ then are independent.

Limit Theorem:

Interested in $\hat{X}(n) = \frac{1}{n}\sum_{j=1}^{n} X_i$ where x_1, x_2, ... is i.i.d.

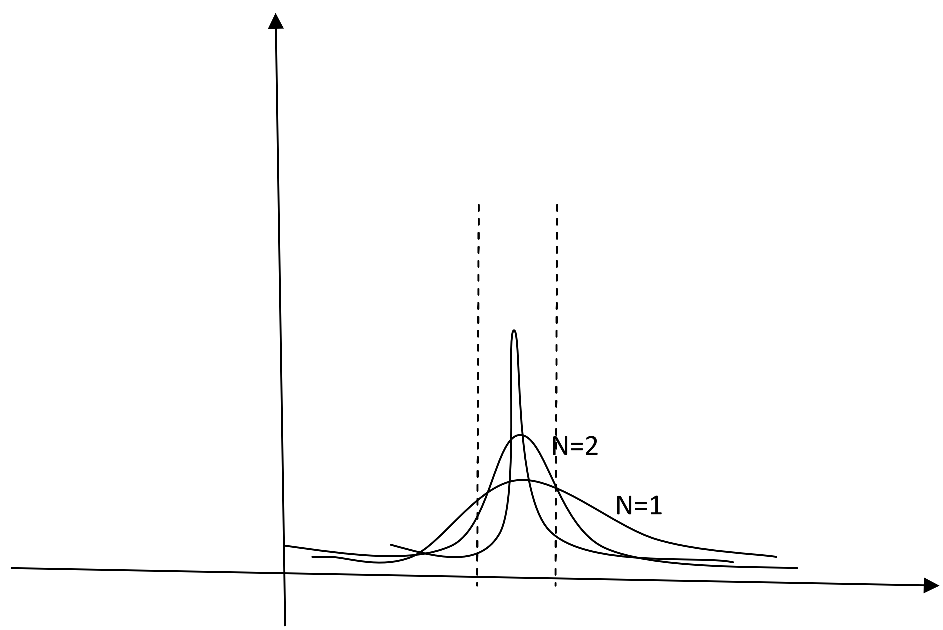
(1) Weak law of large numbers (WLLNs)

(2) Strong law of large numbers (SLLNs)

(3) Central limit theorem (CLT)

(1) E($\hat{X}(n)$)= μ

Var($\hat{X}(n)$)= $\frac{1}{n}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$



$\lim_{n\to\infty} P\{\left|\hat{X}(n) - \mu\right| > \varepsilon\} = 0$ WLLNs

Proof: For X>=0

P{X>t}<=E(X)/t (Markov)

$$E(X) = \int_0^\infty x dF(x) = \int_0^t x dF(x) + \int_t^\infty x dF(x) >= \int_t^\infty t dF(x) = tP\{X > t\}$$

Chebyshev:

$$P\{|X-\mu| > \varepsilon\} \le \frac{Var(x)}{\varepsilon^2} \Leftrightarrow P\{|X-\mu|^2 > \varepsilon^2\} \le \frac{Var(x)}{\varepsilon^2}$$

The latter expression is the same as the Markov relation just you need to replace these.

$$\lim_{n\to\infty} P\{|\hat{X}(n) - \mu| > \varepsilon\} \le \frac{Var(x)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

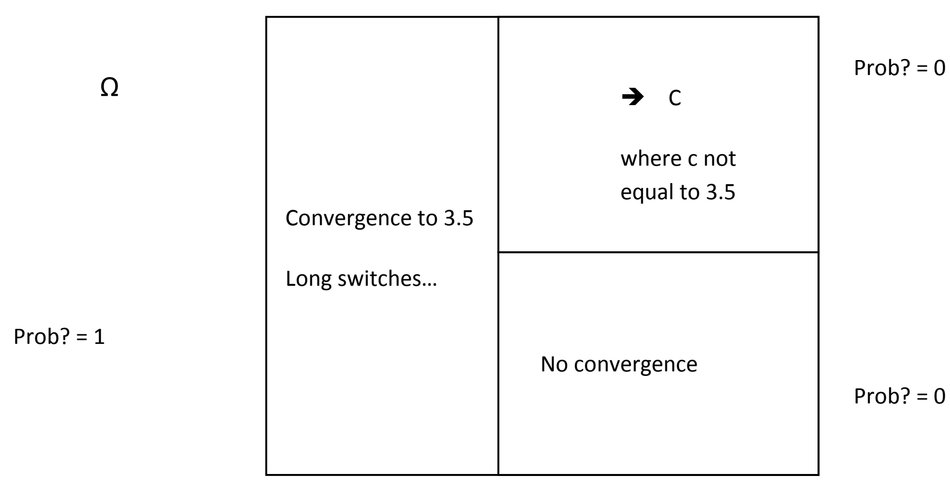
(ii) SLLN: As $n \to \infty$

$\hat{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$ with probability 1 → with respect to 1, (almost surely)

Consider $x_1, x_2, \ldots$, where the X_i's are i.i.d are outcomes of independent rolls of die (balanced)

Roll the dice infinite time, infinite samples of numbers 1 to 6.

$$\Omega = enire\ list(111,\ldots,1213,\ldots,5656\ldots,\ldots.)$$



Implies that: Average of X_i's would be in the cluster and not outside

The weak law in contrast said the probability of being outside is Zero.

(ii) CLT: (Central Limit Theorem)

Talks about stretching, means mean will appear but as you push you will see the distribution:

$$\lim_{n\to\infty}\left\{\frac{\hat{X}(n)-\mu}{\sigma/\sqrt{n}} \le x\right\} = \frac{1}{\sqrt{2\pi}.1}\int_{-\infty}^{x} e^{-y^2/2}dy$$: cdf. + N(0,1)

Proof (outline)

Complete the MGF of

$$MGF_n \to e^{t^2/2}$$

Stochastic process:

Definition: A stochastic process is collection of random variables.

Market share, total sales in the day, Bernouli random variable, stock prices are all examples of stochastic processes

Notation: $\{X(t), t \in T\}$: t is typically related to time, but it is not mandatory and it could be order that you are doing

Index set: T=[0,∞) → continuous time

T=0, 1, 2, … → discrete time

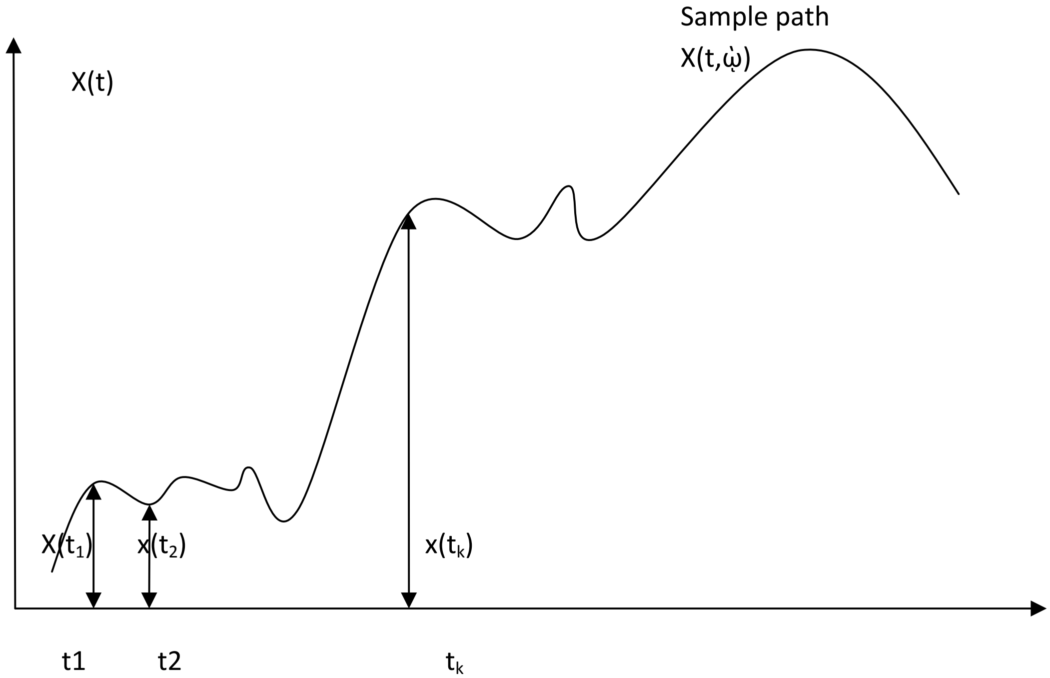
X is State space that could be continuous or discrete

Most sophisticated is continuous time and continuous space, but it is the same as sum and integral

{X(t), t>=0}

X(t, Ω) that Ω is sample space

How do we describe a stochastic process (s. p.)?



Finite dimensional distribution:

We need to specify the joint distribution of $(X(t_1), X(t_2), \ldots, X(t_k))$ for all $0<=t_1<t_2<\ldots.<t_n<\infty$ for any $k>=1$.

Chapter 3:

Conditional distribution:

Recall that if E & F are two events then $P(E \mid F) = \dfrac{P(E \cap F)}{P(F)}$

Example: Let (X_1,X_2) be the outcome of the two independent rolls of dies

$P\{X_1=1|X_1+X_2=5\}= P\{X_1=1,X_1+X_2=5\}/ P\{X_1+X_2=5\}=1/36/4/36=1/4$

Conditional random variable: $(X_1=1|X_1+X_2=5)$~?

Without any information x_1 could be from 1 to 6, with the condition information x_1 values scope changed to 1,2,3,4

In general:

$f_{X|Y}(x|y)=f_{X,Y}(x,y)/f_Y(y)$

we can sometimes define sequentially:

$X1 \sim F$

$X_2 \sim F_{X2|X1}$

$X_3 \sim F_{X3|X2,X1}$

Another example: $X_i \sim$Poisson (λi), i=1,2

X_1 & X_2 are independent

$(X_1|X_1+X_2=n)\sim ?$

$P\{X_1=i|X_1+X_2=n\}= P\{X_1=i,X_2=n-i\}/P\{ X_1+X_2=n\}=$

$$\frac{e^{-\lambda_1}\frac{\lambda_i^{\,i}}{i!}e^{-\lambda_{21}}\frac{\lambda_2^{\,n-i}}{(n-i)!}}{e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_i+\lambda_2)^n}{n!}}=\frac{n!}{i!(n-i)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^i\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-i}$$

$(X_1|X_1+X_2=n)\sim$Bin (n, $\frac{\lambda_2}{\lambda_1+\lambda_2}$)

A continuous Example:

$$f(x,y)=\begin{cases}6xy(2-x-y) & ,0<x<1, y<1\\ 0 & otherwise\end{cases}$$

$$f_{X|Y}(x|y)=f_{X,Y}(x,y)/f_{|Y}(y)=\frac{6xy(2-x-y)}{\int_0^1 6xy(2-x-y)dx}$$

Conditional Expectation:

$$E(X|Y=y)=\begin{cases}\sum_{all\quad x} xf_{X|Y}(x|y)\\ \int xf_{X|Y}(x|y)dx\end{cases}$$

Example: Poisson: $E(X_1|X_1+X_2=n)=n\frac{\lambda_2}{\lambda_1+\lambda_2}$

Continuous: $E(X_1|X_1+X_2=n)= \int x\frac{6xy(z-x-y)}{\int_0^1 6xy(z-x-y)dx}dx$

Computing Expectation by Conditioning:

$$E_Y[E(X|Y)] = E_Y[f(Y)] = E(X)$$

Calculate conditional expectations and then do the unconditioning and you do the divide and conquer principles.

Read ahead chapter 3, and study the problems to be able to reach the exam

Y　　X|Y

≤

E(X)=E [E(X|Y)]

Step 1: Compute E(X|Y=y) $\forall y$: Conditioning

Step 2: Compute: $\int E(X|Y=y)dF_Y(y)$: Unconditioning

This yields E(X)

Simple Example: 3 coins with success probabilities p1, p2, p3, return value 1 for success

Let X be the return value (when we choose a coin randomly and flip “toss” it) (condition: which point is chosen); weights would be 1/3 since we chose randomly each of the coins

E(X| I=1) = p1 w1=1/3

E(X| I=2)= p2 1/3

E(X | I=3)=p3 1/3

E(X)=p1.1/3+p2+1/3+p3.1/3=1/3(p1+p2+p3)

Finding the condition should be done by you by your ***creativity and cleverness*** to make things easy instead of making it difficult (so choose it judiciously)

Proof:

E(E(X|Y)]=

$$\int E(X|Y=y)f_y(y)dy = \int\int xf_{X|Y}(x|y)dxf_y(y)dy = \int\int xf_{X,Y}(x,y)/\cancel{f_y(y)}dx\cancel{f_y(y)}dy =$$

$$\int\int xf_{X,Y}(x,y)dydx = \int f_x(x)dx$$ =E(x)

Example:

1) X~ Geometric (p)

Let Y= outcome of the first trial

We will use recursive algorithm

Step 1: E(X|Y=1)= 1 → p

E(X|Y=0)= E(1+X’)=1 + E(X) → 1-p ,since X’~X (same distribution)

Step 2: E(X)= 1.p + [1+ E(X)].(1-p) => E(X)=1/p

- We learn by doing in this conditioning, and you condition on different information, and via trial and error you will find the trial and error

2) Variance (x) – $E(X^2)$-$[E(XP]^2$

$E(X^2)$=?

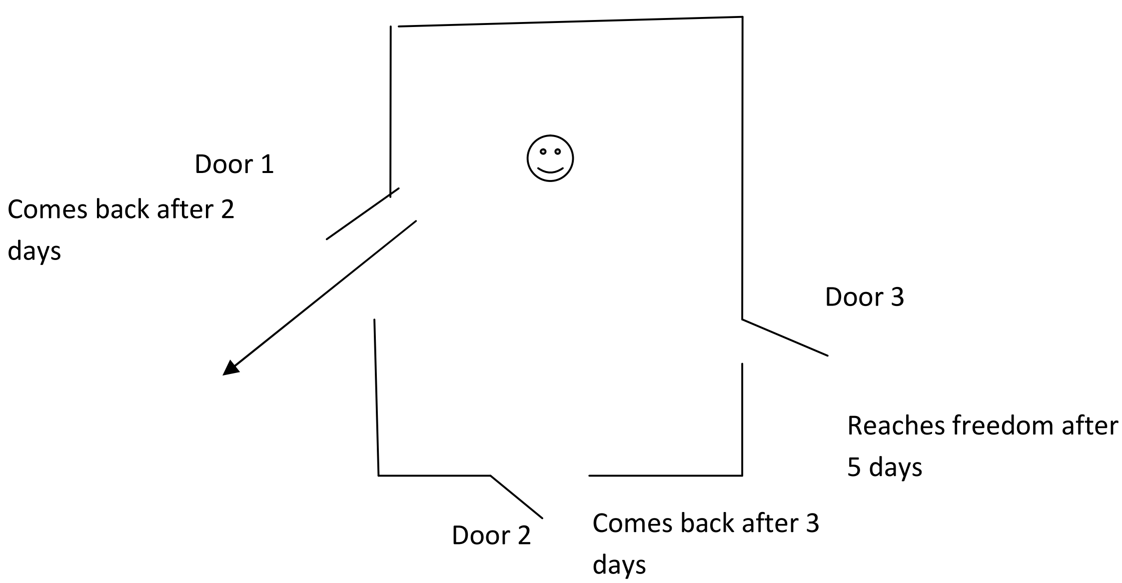
$E(X^2|Y=1)=1$

$E(X^2|Y=0)= E[(1+X')^2]$: /* it is important that you keep in mind that the square should /* be inside, and not outside, since it is convex function

$= E[1+2X'+X'^2]= 1+ 2E(X)+ E(X^2) \rightarrow 1-p$

$E(X')= 1-p+ [1+2 E(X)+ E(X^2)] \rightarrow$ Solve for second moment

3) Prison's problem:



The assumption is that it is memoryless

Let X = total # of days necessary to reach freedom

E(X)=?

By definition: Messy

By conditioning:

Let Y = door chosen

$E(X|Y=1) = 2 + E(X)$

$E(X|Y=2) = 2 + E(X)$

$E(X|Y=3)=5$

Each has probability of 1/3

$E(X) = [2+E(X)]1/3 + [3+E(X)]1/3 + 5.1/3$; Solve for E(X): $E(X)=10$

Stopping time: at some time you will stop and it will depend on the activity that you have done previously could provide another way of the calculation; will be discussed later

4. Variance $(X) = 1/3\{ (4 + 4E(X)+ E(X^2)) + 9 + 6E(X)+ E(X^2) + 25\}$

$E(X^2|Y=1)= E[(2+X')^2]$ 1/3

$E(X^2|Y=2)= E[(3+X')^2]$ 1/3

$E(X^2|Y=3)= 5$ 1/3

Solve for $E(X^2)$

5. Let $X_1, X_2, \ldots$ be i.i.d

Let N (nonnegative integer values) be independent of the X_i 's. (independence example: insurance; damage to the airplane, demand that comes for given day)

Consider:

$$\sum_{i=1}^{n} X_i$$ Random Sum

$E(X)=?$

What piece of information will help you to do the calculation of expectation: here for example N would be helpful; since here N is random and is making problem so you try to fix it

$E(X|N=n)=nE(X)$ /* since X is sum of the variables

$\Leftrightarrow E(X|N)=N.E(X)$: this is only a function of N

$E[E(X_1|N)]=E[N.E(X_1)]= E(N).E(X1)$: if the variable would not be dependent we would not be able to take this (N assumed here to be independent of X's)

6. Var (X) $\rightarrow$ Cov [X,X]

Consider Cov[X,Y]

Conditional Covariance Formula:

Cov [X,Y]= E_z [Cov [X,Y|Z]] + Cov_z [E(X|Z), E(Y|Z)]

Proof: Cov [X,Y]= E(XY) - E(X)E(Y) = E[E(XY |Z)] – E[E(X|Z)]E[E(Y|Z)]= E[E(XY |Z) – E(X|Z)E(Y|Z)] + {E[E(X|Z)E(Y|Z)] -E[E(X|Z)]E[E(Y|Z)]}= E_z [Cov [X,Y|Z]] + Cov_z [E(X|Z), E(Y|Z)]

Expected covariance of expectation + Covariance of Expectations so will lead to this formula

Two midterms were released on the websites.

Var ($\sum_{i=1}^{n} X_i$)= Cov [$\sum_{i=1}^{n} X_i$, $\sum_{i=1}^{n} X_i$]= E_N [N.Var (X_1)] +Var_N [N. E(X_1)]= E(N).Var (X_1) + $[E(X_1)]^2$.Var (N) /* this formula needs X_i's to be independent

Another Example (Application)

Let Y & Z be independent

Define Mixture of two variables or distribution "+" is convolution operation: Either behaving like the first guy or second guy

$$X= \begin{cases} Y & p \\ Z & 1-p \end{cases}$$

Q: Cov[X,Y]=?

$$I= \begin{cases} 1 & X=Y \\ 0 & X=Z \end{cases}$$

Cov [X,Y|I=1]= Variance (Y) p

Cov [X,Y|I=0]=0 (Covariance b/w Y &Z) 1-p

1^{st} Term=Var(Y). p+0.(1-p)

E(X | I=1)= E(Y) E(Y|I=1)= E(Y)

E(X | I=0)= E(Z) E(Y|I=0)= E(Y)

Since covariance b/w any two constant is equal to zero, second term would be zero.

Computing probability by conditioning:

Let E be an Event Question is what is P(E)?

$$1_E = \begin{cases} 1 & E \quad occurs \\ 0 & otherwise \end{cases}$$

Fact $E(1_E)=P(E)$

$E[E(1_E|Y)]= E[P(E|Y)] \Rightarrow P(E)= E[P(E|Y)]$

For the prisoner choice you will condition on the previous conditions

Example:

1) P{X<Y} (X & Y are independent)

If you have value of one of them then you would be able to calculate it

$$= \int P\{X < Y \mid Y = y\} dF_y(y) = \int F_X(y) dF_y(y)$$

Special case X~Exp(λ)

Y~Exp(μ)

$$P\{X<Y\}=\int_0^\infty (1-e^{-\lambda y})\mu e^{-\mu y} dy = 1 - \frac{\mu}{\lambda+\mu}\int (\lambda+\mu)e^{-(\lambda+\mu)y} dy = \frac{\mu}{\lambda+\mu}.*$$

Extension: $X_1,...,X_n$ ~Exp(λ_i)

$$P\{\underbrace{X_1 = \min(X_1,...,X_n)}\} = \frac{\lambda_1}{\lambda_1 + .. + \lambda_n}$$

$$\min(X_1, \underbrace{\min(X_2,...,X_n)}_{Exp(\lambda_2 + .. + \lambda_n)})$$

This will be helpful for finding out which event happened first

2. Convolution {X+Y<=t} = $\int P\{\underbrace{X+Y < t \mid Y = y}_{P(X\le t-y)}\} dF_y(y)$

Due to independence

$$= \int F_X(t-y)dF_y(y)$$

3. Telephone Blocking

Let X be the number of telephone calls arriving in an hour. Poisson (λ)

Each call is blocked with probability p independently

Let $I_1, I_2, \ldots$ be i.i.d sequence of Bernoulli (p)

Define Y= $\sum_{i=1}^{X} I_i$ it means the total number of blocked calls

$$P\{Y=k\} = \sum_{n=0}^{\infty} \mathrm{P}\{\mathrm{Y}=\mathrm{k} \mid \mathrm{X}=\mathrm{n}\}\mathrm{P}\{\mathrm{X}=\mathrm{n}\} = \sum_{k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!}$$

$$= e^{-\lambda} \frac{\lambda p^k}{k!} \sum_{k}^{\infty} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!} = e^{-\lambda} \frac{\lambda p^k}{k!} e^{\lambda(1-p)}$$

Thus, Y~ Poisson (λp)

E(y)=λp

4. Let X & Y be independent (e.g. scores taken from different class)

Consider (X|X>Y) ">=" X

Comparisons of Random Variables:

Definition: $X \geq^{st} Y$ (Stochastically larger) if P{X>t} >= P{Y>t}

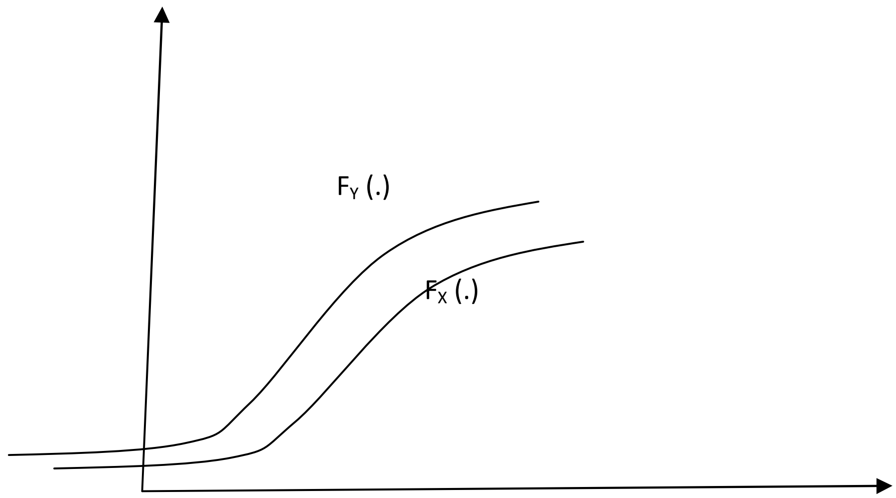
/* tail of X be higher than tail of Y */

1-P{X<=t} 1-{Y<=t}

$F_X(t) \leq F_Y(t) \forall t$

This will help to compare two populations

This is called First order dominance and second and third order dominance is also defined in finance



Claim $(X \mid X > Y) \geq^{st} X$

$$P\{X > t \mid X > Y\} = \frac{P\{X > t, X > Y\}}{P\{X > Y\}}^{st} \geq ?\, P\{X > t\} \Leftrightarrow$$

$$P\{X > t, X > Y\} \geq ?\, P\{X > t\} P\{X > Y\}$$

$$LHS = \int P\{X > t, X > Y \mid Y = y\} dF_Y(y) = \int P\{X > \max(t, y)\} dF_Y(y) \geq$$

$$P\{X > t\} \int P\{X > Y \mid Y = y\} dF_Y(y) = P\{X > t\}.P\{X > Y\}$$

This proves everything we had before

Chapter has more example in 3.6, 3.7 had examples and they are very lengthy, and if you have time you can go over them.

Try to do more homeworks first instead of 3.6, 3.7. Next week we will work on problems. We will also go over chapter 5.

The exam will have one problem that would be from homeworks and close book, but the other would be open book, but not open computer.

1.13:

Wins: 7, 11

Loses: 2, 3, 12

Rest (continuation case): 4, 5, 6, 8, 9, 10

If the sum is anything else, then she continues throwing until she either throws that number again (in which case she wins)

The probability of winning?

7: (1,6), (2,5)

4: P(4)/(P(4)+P(7))

5: P(5)/(P(5)+P(7))

Then you will calculate for all the continuation case and sum them up to get the answer.

1.20. Three dice are thrown. What is the probability the same number appears on exactly two of the three dice?

1/6(1-1/6)-> $\binom{3}{2}$

2.19. The answer for seventeen is:

$$\frac{n!}{x_1!\,x_2!\ldots x_r!}\,p_1^{x_1}p_2^{x_2}..p_r^{x_r}$$

Here sum of x_i's would be super type, call it call A, and the second variety would be Type B. Then we got binomial. As a result we will have the sum of pi as a new type.

2.27. A: k times

B: n-k times

H= # of heads

H=1: A: $\binom{k}{1}p.(1-p)^{k-1}$

B: $\binom{n-k}{1}p.(1-p)^{n-k-1}$

…

Sum: $\sum_{i=0}^{?}\binom{k}{i}p^i.(1-p)^{k-i}.\binom{n-k}{i}p^i.(1-p)^{n-k-1}=\sum_{i=0}^{?}\binom{k}{i}\binom{n-k}{i}p^{2i}.(1-p)^{n-2i}$;

p=1/2

$$Q:\sum_{i=0}^{?}\binom{k}{i}\binom{n-k}{i}=\binom{n}{k}$$

Hint: $(1+x)^{n+m}=(1+x)^n(1+x)^m$

$$\sum_{i=0}^{?}\binom{n+m}{i}x^{n+m-i}=\sum_{i=0}^{?}\binom{n}{i}x^{n-i}.\sum_{i=0}^{?}\binom{m}{i}x^{m-i}$$

2.42. m type of coupons

$$X=\sum_{i=0}^{?}x_i$$

X_1= 1

X_2= $Geom(\frac{m-1}{m})$: you have taken the first type and you are searching for new type (something different from the previous type)

X_3= $Geom(\frac{m-2}{m})$

2.47. P{x=3}=$p_1.p_2.p_3$

E(X)= $p_1+p_2+p_3=1.8$

Max $p_1.p_2.p_3$

s.t. $p_1+p_2+p_3=1.8$

Lagrange, dynamic programming, geometric or anything is okay.

2.48. $E[g(x)]=\int g(x).dF(x)$ solve by integration by parts: $\int udv=uv-\int vdu$; g(0)=0.

$$g(x)=\int_0^x g'(t).dt=g(x)-g(0)$$

$$E[g(x)]=EX\int_0^X g'(t).dt=EX\int_0^\infty I_{\{x,t\}}g'(t).dt=EX\int_0^w p\{x>t\}g'(t).dt(+g(0))$$: use of indicator

2.56. $X=\sum_{i=0}^{n}x_i$; x_i={1 if type I occurs; 0 otherwise}

$$E(X)=\sum_{i=0}^{n}E(x_i)=\sum_{i=0}^{n}(1-(1-p_i)^k)$$

$$Var(X)=\sum_{i=0}^{n}Var(x_i)+2\sum\ \sum_{j\neq i}^{n}Cov(x_i,x_j)=\sum_{i=0}^{n}Var(x_i)+2\sum\ \sum_{j\neq i}^{n}E(x_i,x_j)-E(x_i)E(x_j)$$

To calculate this $p\{X_iX_j=0\}=\{E_i \cup E_j\}= P(E_i)+ P(E_j)-P(E_i \cap E_j)=[1-(p_i+p_j)]^k$

2.68. $P\{X_1 + \cdots + X_{10} \geq 15\}$=> $P\{(X_1 + \cdots + X_{10})/10 \geq 15/10\}$. $\dfrac{\bar{x}-1}{\sqrt{1/10}}=\dfrac{15/10}{\sqrt{1/10}}$

70. $e^{-n}\sum_{k=0}^{n}\dfrac{n^k}{k!}$:$Poission(10):P\{X\leq n\}$ = $P\{\dfrac{\sum_{i=1}^{n}X_i}{n}\leq 1\}=P\{\overline{X}-1\leq 0\}$

73. a) $E(N_i)=np_i$

b) var $(N_i)=np_i(1-P_i)$

c)$Cov[N_i, N_j]=Cov[\sum_{i=1}^{n}X_i , \sum_{i=1}^{n}Y_i]$: indicator variable X_i , Y_j

$$\begin{cases}1 \textit{ if} \quad type \quad i \quad occurs\, in \quad the \quad kth \quad trial\\ \qquad 0 \quad otherwise\end{cases}$$

$$=\sum_{k}\ \sum_{l=1}^{n}Cov[X_k,Y_l]$$

K=l: $Cov[X_k,Y_k]=E(X_k,Y_k)-E(X_k).E(Y_k)=-p_i.p_j$

$k\neq l: Cov[X_k,Y_l]=0$: Since they are independent

Ans. = $-np_ip_j$

75) $N_1=0$

$N_2=1$

$N_3=0$

$N_4=3$

$N_5=1$

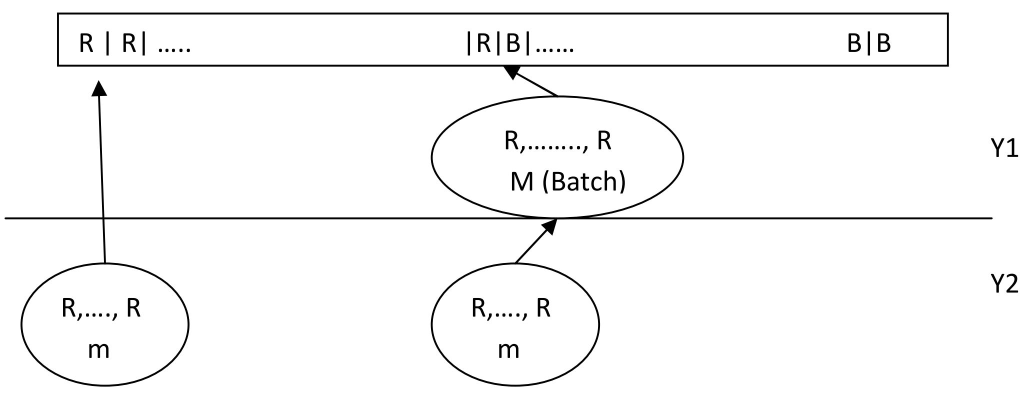
$$N_2 = \begin{cases} 1 & 1/2 \\ 0 & 1/2 \end{cases}$$

$$N_3 = \begin{cases} 2 & 1/3 \\ 1 & 1/3 \\ 0 & 1/3 \end{cases}$$

3.13. $f_{X|X>1}(x) = \dfrac{f_x(x)}{P\{X>1\}} = \dfrac{\lambda e^{-\lambda x}}{e^{-\lambda}} = \lambda e^{-\lambda(x-1)}, x > 1$

$$E(X \mid X>1) = \int_1^\infty \lambda e^{-\lambda(x-1)} dx = 1 + 1/\lambda$$

3.28.



Sum=X_k=$\sum_{i=1}^{k} Y_i$

$$Q: E(X_k) = \sum_{i=1}^{k} E(Y_i)$$

Total # of Balls=r+b+km

At the end of iteration k-1,

C_i= size of the ith ancestral line after k-1 iterations

$$\sum_{i=1}^{r+b} C_i = r + b + (k-1)m$$

=> $E(\sum_{i=1}^{r+b} C_i) = r + b + (k-1)m$ => $\sum_{i=1}^{r+b} E(C_i) = r + b + (k-1)m$ since they are identically distributed we will have

$$\sum_{i=1}^{r+b} E(C_i) == (r+b)E(C_i) = r + b + (k-1)m \Rightarrow E(C_1) = 1 + (k-1)m/(r+b)$$

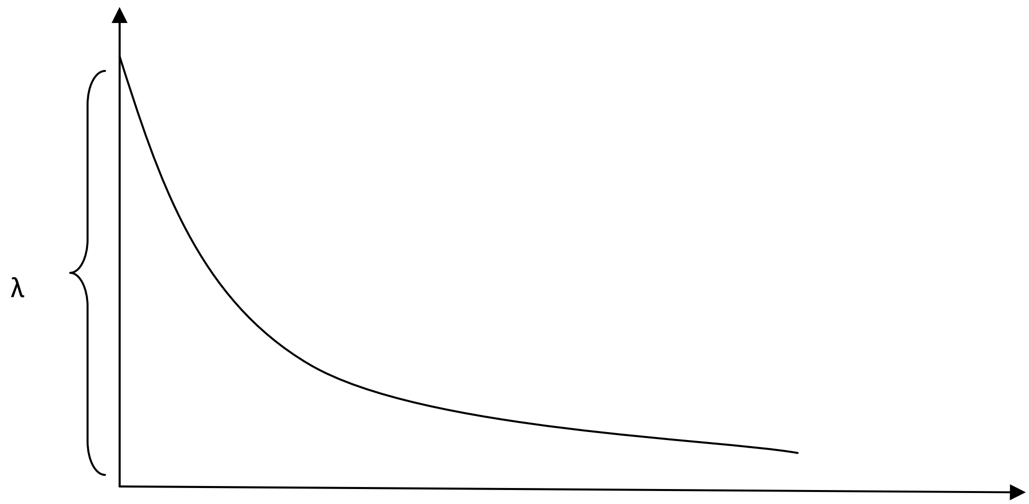
$$E(Y_k) = E[\frac{\sum_{i=1}^{r} C_i}{r + b + (k-1)m}] = \frac{rE(C_1)}{r + b + (k-1)m} = \frac{rE(C_1)}{(r+b)E(C_1)} = \frac{r}{r+b}$$

Chapter 5:

Poisson process:

Exponential distribution: X=Exp(λ) if

$$f_x = \begin{cases} \lambda e^{-\lambda t}, x > 0 \\ x = 0, otherwise \end{cases}$$



E(λ)=1/λ

Var(x)=1/λ^2

Coefficient of variation =1

Characteristics:

(i) Memoryless property

(ii) Failure rate

(i) Def. X is memoryless if P{X=s+t|X=s}=P{x=t}, $\forall s, t$.

"old" with age s "new"

(X-s|x>s): residual random variable at time s(incremental size start at time s) = X_s =(X-s|X>s)= X

What is the probability of residual of x at time s (given that s has survived t); it would be equal the “new” X

There are related concepts that you can write stochastically less or stochastically greater. Since in most time variables the residuals tend to be shorter than. You can say that “new” better than “used”, and you can go over reverse stochastically less, and that notion means “new” worse than “used”.

This is equivalent to LHS (left hand side)= $\frac{P\{X > s+t, X = s\}}{P\{X > s\}} = P\{X > t\}$ or

$$P\{X > s+t\} = P\{X > s\}P\{X > t\} \forall s,t \geq 0$$

$$\overline{F}_X(s+t) = \overline{F}_x(s)\overline{F}_x(t)$$

Fact: If X~Exp(λ), then X is memoryless

$$P\{X > s+t\} = e^{-\lambda(s+t)}$$

$$P\{X > s\} = e^{-\lambda s}$$

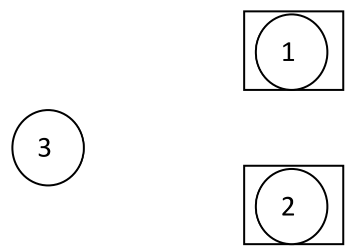
$$P\{X > t\} = e^{-\lambda t}$$

Slight extension: (X-S|X>S) $\overset{st}{=}$ X

Where S is an independent random variable

Pf. $$\frac{P\{X - S > t, X > S\}}{P\{X > S\}} = \frac{\int_0^\infty P\{X > S+t, S = s\} dF_S(s)}{\int_0^\infty P\{X > S \mid S = s\} dF_S(s)} = \frac{\int_0^\infty e^{-\lambda(s+t)} dF_S(s)}{\int_0^\infty e^{-\lambda s} dF_S(s)} = e^{-\lambda t} = RHS$$

Application (Simple scheduling problem): suppose you have two jobs and two machines, and machines are working in parallel. These jobs will take (time) X_1~exp(λ_1), X_2~exp(λ_2), X_3~exp(λ_3), and they are independent. What is the probability that these three job 3 to be the last one to depart?



Depends on which of the two jobs will finish first. If X_1 is shorter job 3 goes in and if X_2 finishes faster job 3 will go in.

Q: P{ probability that these three job 3 to be the last one to depart }=?

Answer: P{X3>X2-X1|X1,X2}P{X1<X2} + P{X3>X1-X2|X1,X2}P{X2<X1}= $\frac{\lambda_2}{\lambda_3+\lambda_2}\frac{\lambda_1}{\lambda_1+\lambda_2}+\frac{\lambda_1}{\lambda_3+\lambda_1}\frac{\lambda_2}{\lambda_1+\lambda_2}$

This is due to the oct 3rd proof that said for exp. Distribution we have: P{ $X_1=\min(X_1,\ldots,X_n)$} = $\frac{\lambda_1}{\lambda_1+..+\lambda_n}$

Q: P{Job 2 to finish last}= $\frac{\lambda_1}{\lambda_1+\lambda_2}\frac{\lambda_3}{\lambda_3+\lambda_2}$

Result: Exponential distribution is only continuous distribution with the memoryless property

Proof:

(i) Exp => Memoryless (Done Earlier)

Fact: If X~Exp(λ), then X is memoryless

$$P\{X>s+t\}=e^{-\lambda(s+t)}$$

$$P\{X>s\}=e^{-\lambda s}$$

$$P\{X>t\}=e^{-\lambda t}$$

Now if F is memoryless, then we need to show that F is exponential. [Geometric is also memoryless; however in discrete variables]

Let g(t) (tail) be a continuous function satisfied g(s+t)=g(s)g(t): $\forall s,t\geq 0$

Goal: show that g(t) must be of the form g(t)= $e^{-\lambda t}$ for some λ (This is exponential tail so it is exponential distribution)

We are going to work with rational t. Suppose t is rational, i.e. t=m/n for some integers m, n. (it is definition), if it works for all rational numbers by continuity you will have it for all real numbers.

$g(m/n)=\overbrace{g(1/n+1/n+\ldots+1/n)}^{\text{m times}} = [g(1/n)]^m=[g(1)]^{m/n}=e^{(m/n)\ln(g(1))}=e^{-[-\ln g(a)]m/n}$ where $[-\ln g(a)]=\lambda$

$g(1)=\overbrace{g(1/n+..+1/n)}^{\text{n times}}= [g(1/n)]^n$

This holds for al real t, by continuity

(ii) Failure rate function

P{t<x<=t+δ|X>t}

$$X \in [t, t+\delta]$$

Define failure rate function of X: $r_X(t)$=

$$\lim_{\delta\to 0}\frac{P\{t<X<t+\delta \mid X>t\}}{\delta}=\lim_{\delta\to 0}\frac{\dfrac{F(t+\delta)-F(t)}{1-F(t)}}{\delta}=\lim_{\delta\to 0}\frac{1}{\overline{F}(t)}\frac{F(t+\delta)-F(t)}{(t+\delta)-t}=\frac{f(t)}{\overline{F}(t)}$$

Example: X~Exp(λ)

$r_X=\dfrac{\lambda e^{-\lambda t}}{e^{-\lambda t}}=\lambda$

In general failure function could be increasing or decreasing. Bath tub shape is also possible: applies when the child is born and there is higher likelihood for failure and then it decreases and as she gets old it the failure rate will increase. We are going to show that if the random variable has the constant failure rate it is then exponential.

Result

$$\overline{F}(t)=e^{-\int_0^t r(y)dy}, t\geq 0$$

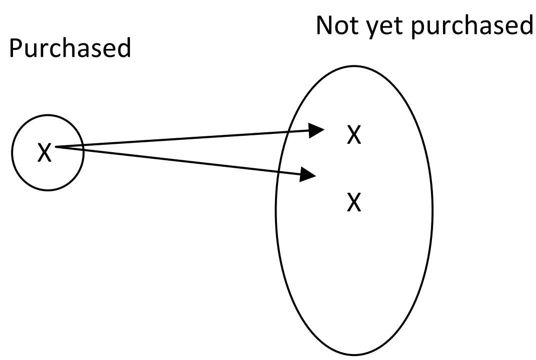
Proof: $\int_0^t r(y)=\int_0^t \frac{f(y)}{\overline{F}(y)}dy=-\ln\overline{F}(y)\Big|_0^t=-\ln\overline{F}(t)$

This is a useful formula, since once you have failure function then you integrate the failure function and then you will get the tail function.

Application (Frank Bass model): New product purchase model. If individual has not purchased the product until now, then time to purchase could be described in terms of failure function.

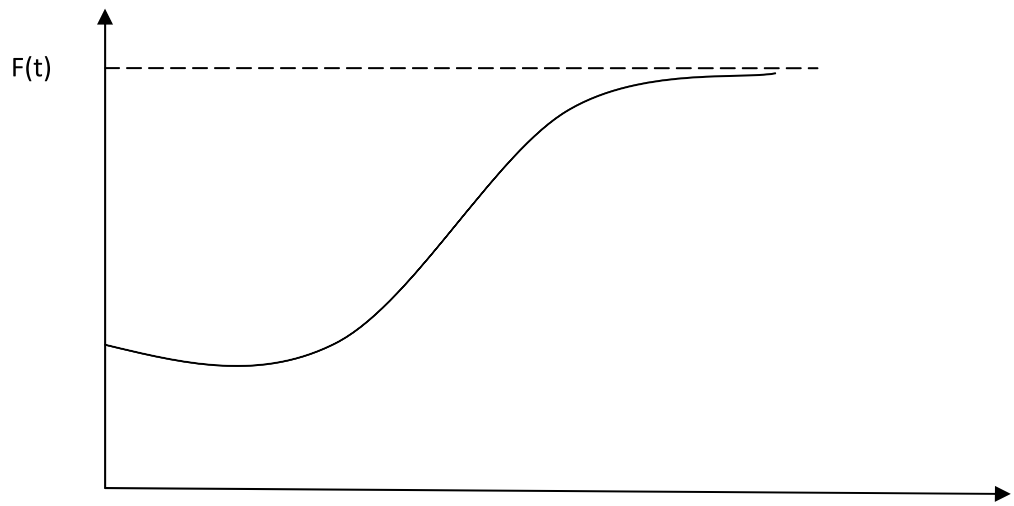
Let T = time to purchase of a new product (a given consumer). T~F(.).

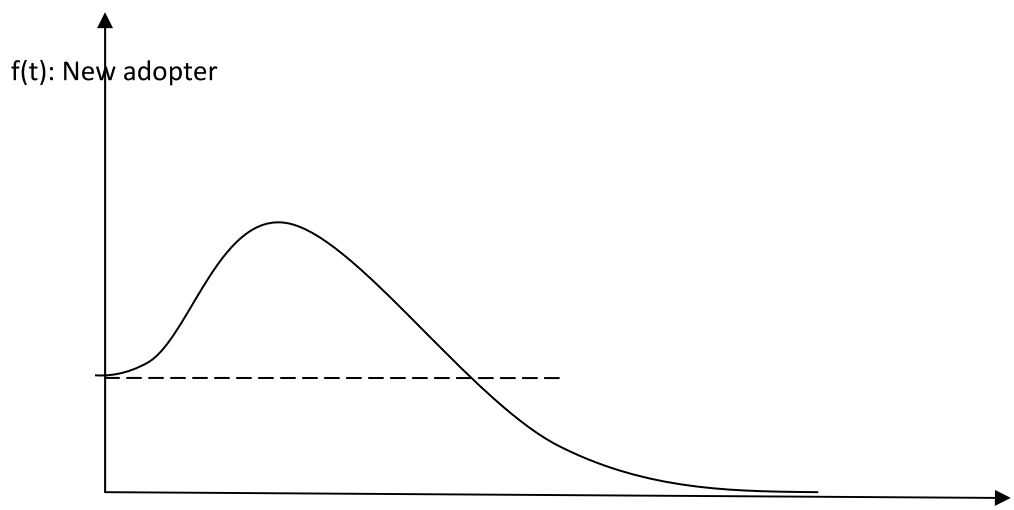
Assumption. $\frac{f(t)}{\overline{F}(t)} = p$ if you assumed that it is completely independent and it is not true, so you need to add another term to it. You divide the population into two group of purchase that one group will have an affect on the people who has not purchased.



$$\frac{f(t)}{\overline{F}(t)} = p + q.F(t)$$

The solution of this: $F(t) = \dfrac{1-e^{(p+q)t}}{1+q/q.e^{-(p+q)t}}$





Poisson process:

Goal: We want to model "events" that occur randomly over time.

For finance for example news released randomly over time.

A process can be building block of all stochastic process could be deemed as Poisson

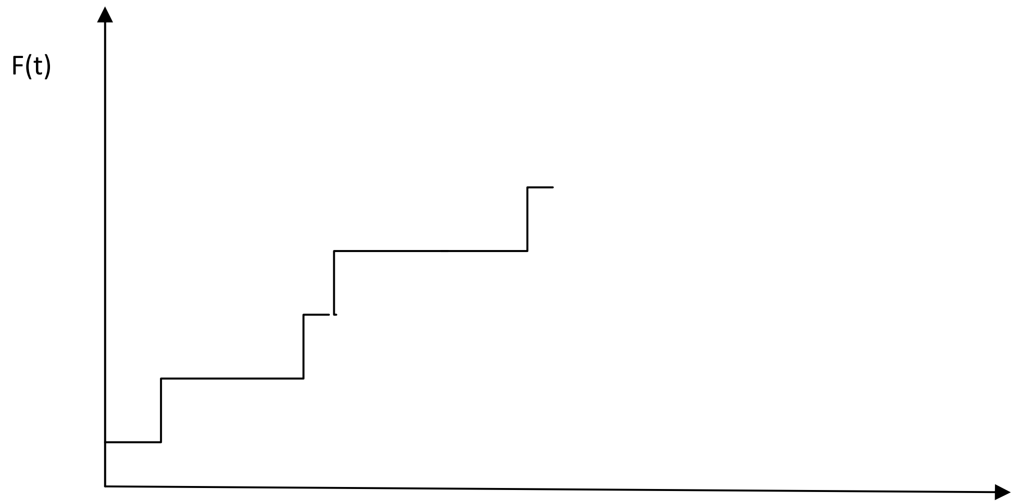
You start from time zero, and you can call the demand arrival. You can tally the total events that entered by time t.

Definition: [N(t), t>0] is a counting process (c.p) if N(t)= # of events by time t.

Property:

(i) N(0)=0

(ii) N(t) will increase in time



Time between events could be independent but generally not. For example in the diffusion process time between purchases is history dependent, and that is also counting process but is very complicated to analyze (There is a paper on this from professor Niu on his website).

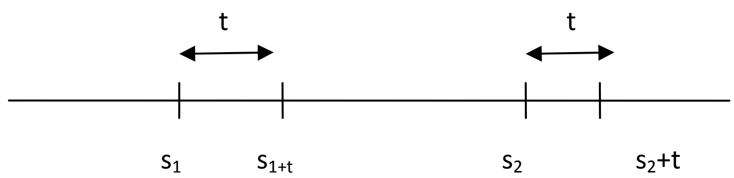
Possible properties of counting processes:

Telephone calls. # telephone calls between 10 -11 p.m. and 3-5 in the afternoon. One property we are going to consider is that if you look at number of calls in those intervals. The question is that is it reasonable to assume that these two are independent?

Independent increment is that we take two intervals and we assume that the number of calls in those intervals are independent. Increment of the first interval and increment of second interval. In some application it could be rejected and in some cases it cannot be rejected that these two are independent (probably the statistics will tell you whether they are dependent or independent).

Independent increment properties $N(t_1)-N(s_1)= N_1 \perp N(t_2)-N(s_2)= N_2$ for all disjoint (s_1,t_1) & (s_2,t_2)

The second property is ***Stationary increment*** property:



$$N(s_1+t)-N(s_1)\overset{st}{=}N(s_2+t)-N(s_2)$$

These are two possible properties that we assumed.

To test this assumption in the specific context you need to use the data.

Definitions of P.P. in Poisson process:

The intent is look at Poisson process from three different angles, and after that we will have fourth definition which is more theoretical, but we are not going to prove it since it is difficult to prove.

Def. 1. A counting process is a Poisson process at rate λ if:

(i) N(0)=0

(ii) The process has stationary and independent increments.

(iii) P{N(t)=n}= $e^{-\lambda t}\frac{(\lambda t)^n}{n!}, n\geq 0$

Over each interval the events are distributed with Poisson distribution. To calculate between duration s_1 and t_1 you move it to zero, and to calculate two disjoint interval you move both to 0 and multiply them, and if intervals have overlap you split it into three intervals and then multiply them to get the distribution. (The main property that is used here is the stationary incremental property to move to zero and just calculate the Marginal distribution). As a result for the process you are going to find distribution for all disjoint snapshots.

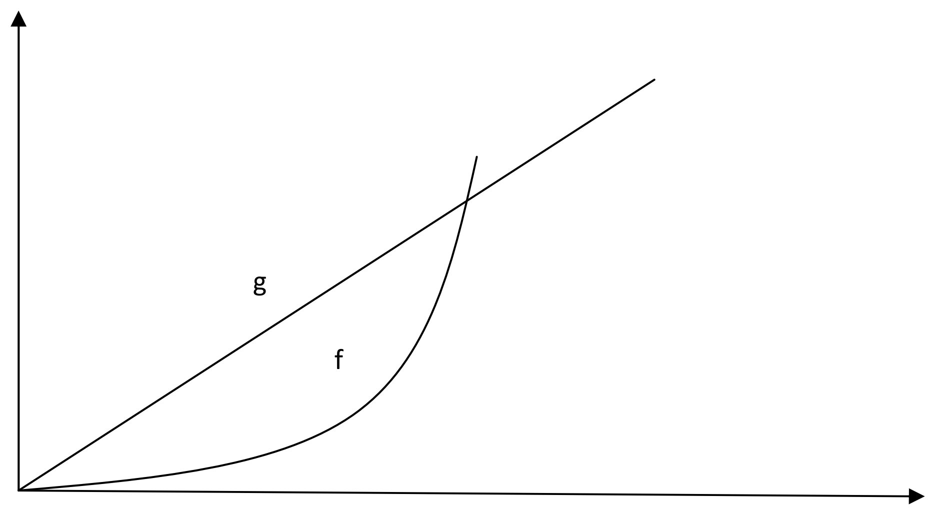
Dependence implies that the intervals should be disjoint: Cov $[X_1+X_2, X_2+X_3]$=Var$[X_2]$ when these three are independent.

Defintion 2: A counting process is a Poisson process at rate λ if:

(i) N(0)=0: The same as previous def.

(ii) The process has stationary and independent increments: The same as previous def.

Notation: Consider functions $f(x)=x^2$ & $g(x)=x$



$\lim_{x \to 0} \frac{f(x)}{g(x)} = 0$: f(x)~o(g(x)): little "o" of x (means of little order)

Big O means replace 0 with constant (c) and they would be of similar order

(iii) P{N(h)=1}= λh +o(h): probability of having one event [just λh is strong assumption and is greedy, so you will put a correction term and we want it to be irrelevant, so we will put lower order function of h]

(iv) P{N(h)=0}=1- λh +o(h)

Adding iii, and iv = 1+o(h) : (since o(h) only captures the property of this function)

1-P{N(h)=0 or 1}=P{N(h)>=2}=o(h)

Since the sum of all three is one, we are not violating the probability principle. Sum of 0, 1, and more than one is still one.

As a result either you have one or no event in the small interval. This will make it a Bernoulli, and number of events is depending on the length of the interval.

Third definition describes time between events, and inter events time and we assume that this inter event time is i.i.d. and we will establish the relation between λ and this enter event time.

The fourth definition is like the first definition, but the last assumption is that increment comes one at a time. That's minimal assumption but it implies hard to prove it.

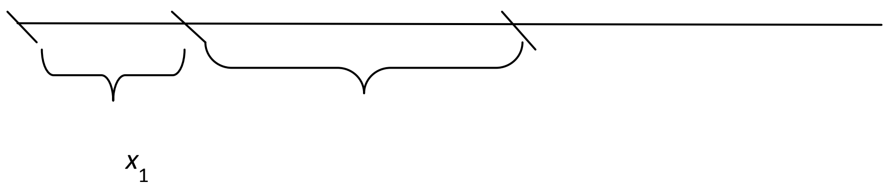
Equivalence of definitions:

Def 1. (iii) $P[N(t)=k]=e^{-\lambda t}(\lambda t)^k/k!, k\geq 0$

Def 2. (iii) $P[N(h)=1]=\lambda h+o(h)$

(iv) $P[N(h)=0]=1-\lambda h+o(h)$

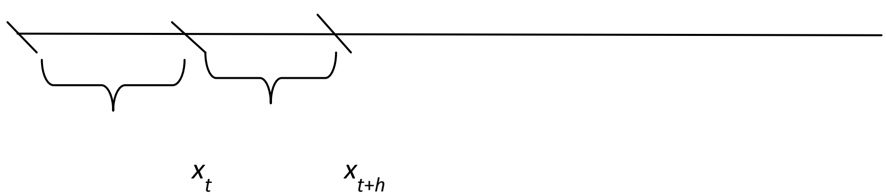
Def 3. The inter-event times are i.i.d. $exp(\lambda)$



We are going to show that they are the same:

(2)=> (1)

$P_0(t)=P\{N(t)=0\}=?$ means the probability of having no events



$p_0(t+h)=p\{N(t)=0, N(t+h)-N(t)=0\}$

Means: if I do not have any event in whole t+h then I should not have any event in either interval

They are two disjoint intervals, and distributions are independent.

Therefore on one hand we have: $p_0(t+h)=p\{N(t)=0, N(t+h)-N(t)=0\}=p_0(t)(1-\lambda h+o(h))$

On the other hand: $$\frac{\partial p_0(t)}{\partial t}=\frac{p_0(t+h)-p_0(t)}{h}=\frac{p_0(t)(1-\lambda t+o(h))-p_0(t)}{h}=-\lambda h p_0(t)$$

This applies when p→0 we will have: $p_0(t)=c^{\lambda t}$

we have $p_0(0)=0$ therefore c=1

We are comparing the probability in the interval of [t,t+h] this is called forward approach. You can do the same to compare from (0,h) and then [h,t+h] this is called Backward approach

Try backward approach as an exercise.

$$p_k[t+h] = P[N(t)=k, N(t+h)-N(t)=0] + P[N(t)=k-1, N(t+h)-N(t)=1] + \sum_{i=2}^{k} P[N(t)=k-i, N(t+h)-N(t)=i]$$

We would also have:

$$\sum_{i=2}^{k} P[N(t)=k-i, N(t+h)-N(t)=i] = o(h)$$

By closing the eye on the first event of joint event then it would become higher then we can derive the inequality mentioned (generally think about what makes the life easier). Therefore we will have:

$$p_k[t+h] = P[N(t)=k, N(t+h)-N(t)=0] + P[N(t)=k-1, N(t+h)-N(t)=1] + o(h)$$

Therefore we have:

$$P_k[t+h] = P_k(t)(1-\lambda h + o(h)) + P_{k-1}(\lambda h + o(h))$$

$$\frac{P_1[t+h]-P_1(t)}{h} = P_1(t)\frac{(\lambda h+o(h))}{h} + P_0(t)\frac{(\lambda h+o(h))}{h} + \frac{o(h)}{h} \Rightarrow \frac{\partial p_1(t)}{t} = \lambda p_0[t] - \lambda p_1[t]$$

Multiply two sides to $e^{\lambda}t$

$$\mathrm{d}\ e^{\lambda}tP_1(t)/dt = \lambda \Rightarrow e^{\lambda}tP_1(t) = \lambda t + c$$

C=0 since $P_1(0)=0$

$P_1(t) = e^{-\lambda t}\frac{\lambda t^1}{1!}$ complete the proof by induction

This showed (1)=> (2)

$$\mathrm{PN(h)=0=e}^{-\lambda h} = 1 - \frac{\lambda h}{1!} + \frac{(\lambda h)^2}{2!} + \dots = 1 - \frac{\lambda h}{1!} + \dots$$

$$\mathrm{P\{N(h)=0\}=e}^{-\lambda h} = 1 - \frac{\lambda h}{1!} + \frac{(\lambda h)^2}{2!} + \dots = 1 - \frac{\lambda h}{1!} + o(h)$$

P{N(h)=1}=$(\lambda t)e^{-\lambda h} = \lambda t - \frac{(\lambda h)^2}{1!} + \frac{(\lambda h)^3}{2!} + ... = \lambda h + o(h)$

Since $\frac{\lambda t^2}{1!} - \frac{\lambda t^3}{2!} + ... = o(h)$

$\lim_{h \to 0}(\frac{f(h)}{1}) \to 0 : o(h)$

o(1)+o(h)=o(1)

(1) => (3)

X_i be the i^{th} inter-event time

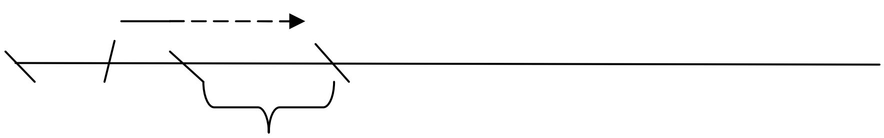
P{$x_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t} \Rightarrow X_1\tilde{}Exp(\lambda)$

$$P\{X_1 + X_2 > t\} = P\{N(t) = 0or1\} = P\{N(t) = 0\} + P\{N(t) = 1\} = e^{-\lambda t} + \lambda t e^{-\lambda t} \Rightarrow \frac{\partial P\{X_1+X_2>t\}}{\partial t} = -\lambda^2 t e^{-\lambda t}: Erlang(2, \lambda)$$

Take the derivative as HW.

Erlang(2,λ)

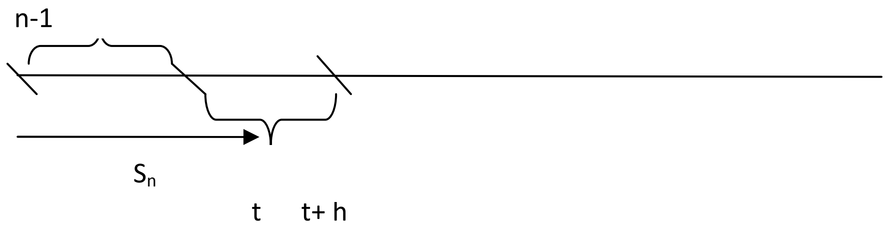
(3)=> (1)



$$P\{N(t)=\lambda\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!} = P\{N(t) \geq k\} - P\{N(t) \leq k+1\} = P\{\sum_{i=1}^{k} X_i \leq t\} + P\{\sum_{i=1}^{k+1} X_i \leq t\} = \int_0^t e^{-\lambda y}\frac{(\lambda y)^{k-1}}{(k-1)!}\lambda dy - \int_0^t e^{-\lambda y}\frac{(\lambda y)^k}{k!}\lambda dy = e^{\lambda y}\frac{(\lambda t)^k}{k!} + \int_0^t e^{-\lambda y}\frac{(\lambda y)^k}{k!}\lambda dy - \int_0^t e^{-\lambda y}\frac{(\lambda y)^k}{k!}\lambda dy$$

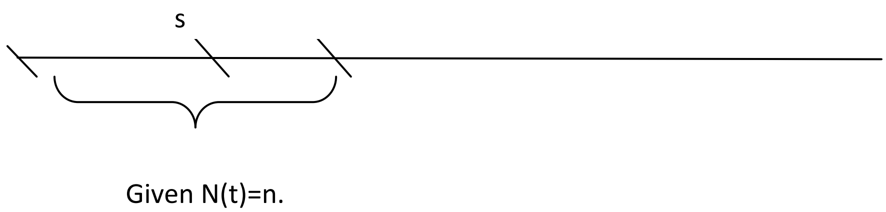
do the algebra yourself as homework

Comment on the Erlang density. Suppose you are interested in $P\{X_1 + .. + X_n = S_n \in (t, t+h)\}$



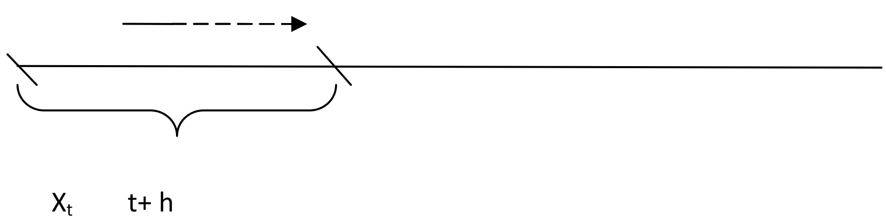
$$P\{X_1 + .. + X_n = S_n \in (t, t+h)\} \approx e^{-\lambda t}\frac{(\lambda t)^{n-1}[\lambda h + o(h)]}{n-1} \to e^{\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}$$

Conditional arrival Times:



$$\mathrm{P}\{\mathrm{X}_1 \le s | N(t) = 1\} = \frac{X_1 \le s, N=1}{N(t)=1} = \frac{e^{-\lambda s}\lambda s e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} = \frac{s}{t}, 0 \le s \le t$$

It is uniform since everything is symmetric so equally likely every event could happen



$$\mathrm{P}\{\mathrm{S}_i = \sum_{i \in (t_i, t_i+\epsilon_i)}^{i=1,2,\dots,n} X_i | N(t) = n\} = \frac{e^{-\lambda t_1}e^{-\lambda[t_1+\epsilon_1-t_1]}\lambda\epsilon_1 e^{-\lambda[t_2-\epsilon_1-t_2]}\dots}{\frac{e^{\lambda t}(\lambda t)^n}{n!}} = n!(\frac{1}{t})^n$$

Order statistics:

Let $Y_1,\dots,Y_n$ be n i.i.d random variables with distribution F(.)

Define $Y_{[1]}$ =the smallest of $Y_1,\dots,Y_n$

$Y_{[2]}$ =2nd smallest of $Y_1,\ldots,Y_n$

..

Sample could be bids coming for auction, or device life time.

Consider $(Y_{[1]},\ldots,Y_{[n]})$

Question is what is the joint density of $(Y_{[1]},\ldots,Y_{[n]})$?

$$f_{Y_{[1]},\ldots,Y_{[n]}}(Y_1,\ldots,Y_n) = n!\prod_{i=1}^{N} f(y_i)$$

We can look it in this form that you generate random variables in the form of i.i.d and then you rearrange them in descending order and then present them to user.

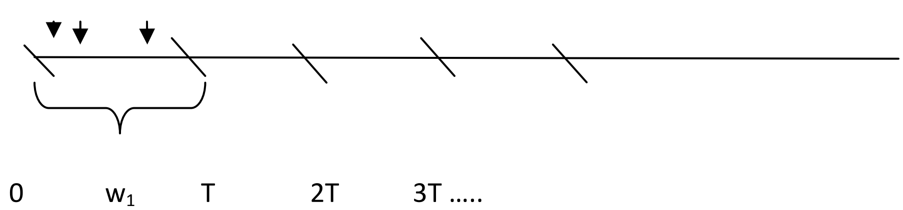
$P(S_1,\ldots,S_n \mid N(t)=n) \sim (Y_{[1]},\ldots,Y_{[n]})$ where the Y_i's are i.i.d. uniform (0,t)

When we arrange them in descending order they will become independent. if in a condition the permutation is not important then this can help you.

Examples:

1. Train departure

   Suppose you are at the train station and



Passengers arrive according to a poison process at rate λ

When train arrives all people will be picked up

$W = \sum_{i=1}^{N(t)} w_i$ waitings

E(w)=?

Brutal approach: $E(W) = E[\sum_{i=1}^{\infty} w_i] = \sum_{i=1}^{\infty} E[w_i = (T - \sum_{j=1}^{i} X_j)^+] = \ldots$

$x^+ = \max(x,0)$

$x^- = -\min(x,0)$

$x = x^+ - x^-$

One way is to $\begin{array}{l} E(x) = \int_0^\infty P[X > t]dt \\ \sum_0^\infty p\{X > n\} \end{array}$ Do the rest yourself but it would be brutal

Clever approach:

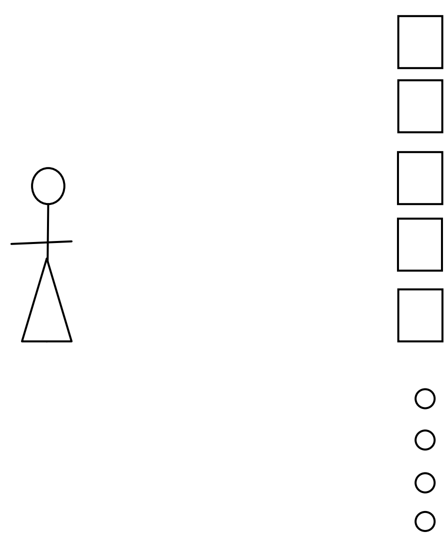
$$E(w) = E[E(W \mid N(T))] = E[W \mid N(T)] = E[\sum_{i=1}^{N(T)} W_i \mid N(T)] = E[\sum_{i=1}^{N(T)} W'_{[i]}], W'_i, i = 1,.., N(T): i.i.d. Unif(0,T)$$

$$= E[\sum_{i=1}^{N(T)} W'_{[i]}] = N(T)T/2$$

$$Ans. = E[N(T)T/2] = \frac{\lambda T^2}{2}$$

Example:

2) M/G/∞ Queue. M: nature of the arrival (Poisson process: M means exponential), G: nature of the service (General), ∞: infinite number of servers



M/G/2/5: means the queue has the capacity of 2 [door will be closed after for example] and 5 is the number of servers

This model would be useful for the number of cars that are parked in the parking lot

$N(t)$ =# of customers in the system at time t

M: Poisson at rate λ

G: service distribution

Question is P[N(t)=k]=?

A(t) number of arrivals at time t

Number of people in the system will be calculated as follows:

$$\sum_{n=k}^{\infty} P\{N(t)=k \mid A(t)=n\}\{A(t)=n\} = \binom{n}{k}[(p(t)]^{k}[1-p(t)]^{n-k}$$

Calculate departure between t and t+h, when arrival has been done in s<t+h

Following reasoning shows that if the arrival is position then then departure would be exponential:

Whether you depart or not is G(s-t), so $\overline{G}(t-y)$ will talk about whether it is still in the system after the entrance, also dy/t would be the probability of entrance at any point applying the uniform distribution, so it would be t

$$P(t) = \int_{0}^{t} \overline{G}(t-y)dy/t$$

$$Ns.N(t) \sim Poisson(\lambda t.p(t))$$

As a result number of client in the system that are not yet served comply the following:

$$\lambda t \int_{0}^{t} \overline{G}(t-y)dy/t = \lambda t \int_{0}^{t} \overline{G}(y)dy$$

Try the calculation that two subcounts of departed and not departed are independent and both Poisson.

Probability Course of Professor Niu @ UTD: Poisson process

Meisam Hejazinia

11/7/2012

Solution to the problems of posterior chapters of midterm is posted. M/G/∞ $N(t) = \#$ in the system of time t A(t)-N(t)= Number that arrived and departed in [0,t] N(t)=Poisson (tP(t)) where

$P(t) = \int_0^t \bar{G(y)} \frac{d\theta}{t}$
$D(t) = A(t) - N(t)$ ~Poisson $(\lambda t(1 - P(t)))$
Furthermore, $N(t) \perp D(t)$

Comment.
1. $E[N(t)] = \lambda P(t) = \lambda \int_0^t (\bar{G}(t-y))dy$
Interpretation for E[N(t)]
$lim_{n\to n} \sum_{k=1}^{n} [\frac{kt}{n} - \frac{(k-1)t}{n}] \bar{G}(t - \frac{k-1}{n} t)$

Let N ~Poisson (λ)
$I_1, I_2, ...$ be i.i.d. Beroulli (p)
Define $N_1 \equiv \sum_{i=1}^{N(t)} I_i \sim Poisson \lambda p$
$N_2 \equiv \sum_{i=1}^{N(t)} (1 - I_i) \sim Poisson \lambda (1-p)$

Proof of $N_1 \perp N_2$:
$P\{N_1 = m, N_2 = n\} = P\{N_1 = m, N_2 = n | N = m+n\} P\{N = m+n\} = \frac{(m+n)!}{m!n!} p^m (1-p)^n e^{-\lambda} \frac{\lambda^{m+n}}{(m+n)!} = e^{-\lambda p} \frac{(\lambda p)^m}{m!} e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^n}{n!}$

You can substitute Bernoulli(p) with Multinomial type probablities $P_1, P_2, ..., P_m$

In this event we had not time dependence, but in the following we will have it.

Decomposition of Poisson process
Let $\{N(t), t \geq 0\}$ be a Poisson process at rate λ. Each event be classified as type 1 with Probability p and are type 2 with probability of (1-p), independently. This yields to counting processes:

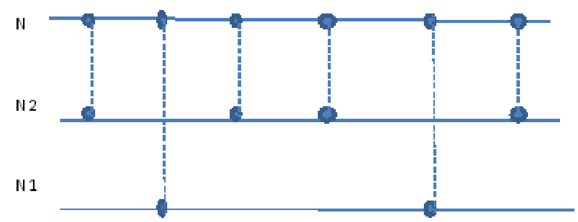


Figure 1: Decomopossition of Poisson Process

$\{N_1(t), t \geq 0\} \{N_2(t), t \geq 0\}$

Claim. $\{N_i(t), i \geq 0\}, i = 1, 2$ are indpendent poisson process with parameter λp and $\lambda(1-p)$.

$\sum_{i=1}^{K:geometric} X_i$ where X_i is i.i.d. $Exp(\lambda)$

Based on what would be the result of our tossing the coin we will put the point on the first or second line, and since each of these poisson events were independent of each other, our event would be independent as well.

The above concept was called Thinning process. By doing the Bernouli trial we made it thinner.

If we substitute p with $p(y)$ if an event occurs at time y. This is called " Time-dependent thining ".

Superposition of Poisson process

Let $\{N_i(t), t \geq 0\}, i = 1, 2$ be two independent poisson process with rate λ_i Define $N(t) = N_1(t) + N_2(t)$
Claim. $N(t), t \geq 0$ is a poisson process at rate $\lambda_1 + \lambda_2$.

We proved this around 6 sessions ago, so the proof is up to you as homework.

Nonhomomogeneous poisson process

Def 1.
A Conditioning Poission process $\{N(t), t \geq 0\}$ is a non homogeneous Poisson Process with intensity funcntion $\{\lambda(t), t \geq 0\}$ if

- $N(0) = 0$
- the process with indpendenet increments
- $p\{N(t) - N(s)\} = e^{-[m(t)-m(s)]} \frac{[m(t)-m(s)]^k}{k^1}, k = 0;$

Let $m(t) = \int_0^t \lambda(y)dy$, the mean value function
Def. 2

- same as previous def. 1
- same as previous def. 1
- $P\{N(t+h) - N(h) = 1\} = \lambda(t)h + o(h)$
- $P\{N(t+h) - N(t) = 0\} = 1 - \lambda(t)h + o(h)$

Note. The interevent times are not i.i.d.

Conditional arrival Time

Poisson Case:

$(S_1, ..., S_n | N(t) = n) \sim (\mathrm{y}_{[1]}, ..., y_{[n]})$ where the y_i's are i.i.d. Unif (0,t).
for the intervals that you bring each distribution then you will have $\frac{\lambda(u)h}{m(t)}$

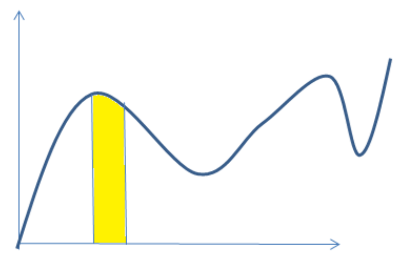


Figure 2: Non Homogeneous Poisson Process

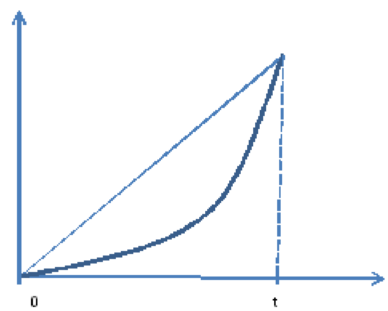


Figure 3: Non Homogeneous Poisson Process

Nonhomogeneous case
$(S_1, ..., S_n | N(t) = n) \sim (\mathrm{y}_{[1]}, ..., y_{[n]})$ where the y_i's are i.i.d. with density $\frac{\lambda(u)h}{m(t)}, 0 \leq t \leq t$

Proof same as before.

As a homeowork also show that $X \sim F$, then show that if $X^* = F^{-1}(0)$ then $X^* \sim X$.

Simulation of a Nonhomogeneous Poisson Process

Compund Poisson Process

Let $X_1, X_2, ...$ be i.i.d. Let $[N(t), t \geq 0]$ be a poisson process.
Define $Y(t) = \sum_{i=1}^{N(t)} X_i, t \geq 0$.
Define $\{Y(t), t \geq 0\}$ is a compound poisson process.

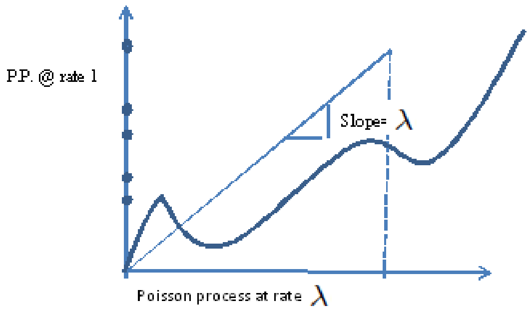


Figure 4: Random Variable Creation

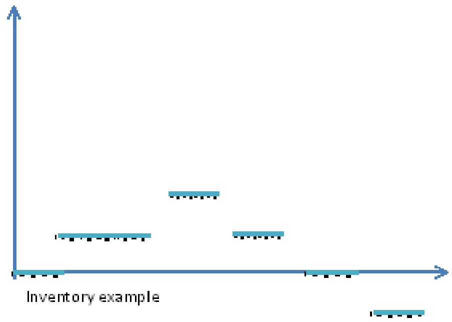


Figure 5: Inventory Problem

This definition could be used for the problem that some people come to your website and each of them spend random amount of money, or your ordering for inventory.

Birth & Death Process

A Poisson process is pure birth process.

Each time event occures you can go up one time, or you go down one time. In contrast, in Contineous Time Process which would be the next level, you would have jumps to other states that could be complicated.

Example. Inventory

- Units are produced according to Poisson process at rate λ
- Demand also occur according to Poisson process at rate μ. Demands occur one at a time.

Let I(t)=inventory level at time t.
An example could be the taxi in the airport. Also backlog could exist means the event go negative and the demand comes and there would be no supply.

Time intervals are going to be random. The intervals will have the distribution of μ.

Once you are in the random state (i) then there would be duration which would be random and is called "Sojourn in state" i ~Exp($\lambda_i \mu_i$) that would be Min (X,Y) where X ~Exp(λ_i), Y ~Exp(μ_i) and going up would have the probability of $\frac{\lambda_i}{\lambda_i+\mu_i}$ and the probability of going down would be in the form of $\frac{\mu_i}{\lambda_i+\mu_i}$.

Another Example: M/M/1 Queue

The entrance is exponentially distributed with parameter λ, and the service time would be exponentially distributed with parameter μ.

Departure rate would be dependent on the state that you are in, and if there would 5 people there the departure rate would be 5λ, but for this example the arrival time is with the same rate which is μ, but in other examples it may be different based on the state for example in the bank when you see people standing in the line in the bank, so you leave.

In the birth and death process we may be interested in knowing $P\{N(t) = k\} =?$ you may go to the calculation based on $P_1(t) = 1 - \lambda t + o(h)$ and then calculate the differential equations, and these equations would be state based.

You should try problems for chapter 5.

Probability Course of Professor Nui @ UTD: Markov Chain

Meisam Hejazinia

11/14/2012

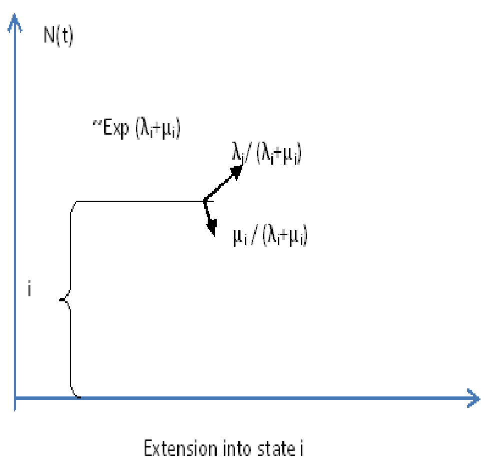


Figure 1: Birth and Death Process

Birth and Death Process

State space: 0,1,2,...
λ_i = Birth rate in state i
μ_i = Death rate in state i
Birth and Death Process is shown in Figure 1.

$\{N(t), t \geq 0\}$ have i called a Birth and Death process.

This is an example of Markov process.

Let $P_n(t) \equiv P\{N(t) = n\}, t \geq 0$

Forward Approach (Figure 2 is showing this approach):

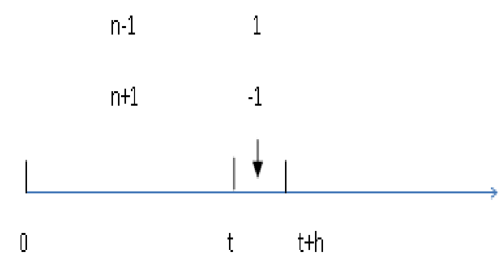


Figure 2: Birth and Death interpretation Markov Chain

In problem set of chapter five you will be asked to show that whether Death wins or Birth wins will not affect the exponential inter arrival distribution.

$P_n(t+h) = P\{N(t) = n+1\}[(\lambda_{n+1}+\mu_{n+1})h + o(h)]\frac{\mu_{n+1}}{\lambda_{n+1}+\mu_{n+1}} + P\{N(t) = n\}[1-(\lambda_n+\mu_n)h+o(h)] + P\{N(t) = n-1\}[((\lambda_{n-1}+\mu_{n-1})h+o(h)]\frac{\lambda_{n-1}}{\lambda_{n-1}+\mu_{n-1}} + o(h)$

o(h) is correction factor in the above equations. The formulation of above is standard and is used very often.

$P_n(t+h) = p_{n+1}(t)\mu_{n+1}h + p_n(t)[1-(\lambda_n+\mu_n)h] + P_{n-1}(t)\lambda_{n-1}h + o(h)$

$P_n(t+h) - P_n(t) = P_m(t)\mu_{n+1}\mu_{n+1}h + P_{n-1}(t)\lambda_{n-1}h - P_n(t)(\lambda_n+\mu_n)h + o(h)$
$\frac{P_n(t+h)-P_n(t)}{h} = \frac{P_m(t)\mu_{n+1}\mu_{n+1}h + P_{n-1}(t)\lambda_{n-1}h - P_n(t)(\lambda_n+\mu_n)h + o(h)}{h}$

$\frac{d}{dt}P_n(t) = P_{n+1}\mu_{n+1} + P_{n-1}\lambda_{n-1} - P_h(t)(\lambda_n +$

$\mu_n), n \geq 0$

This can be solved (next semester)

Assumption. $\lim_{t\to\infty} P_n(t) = P_n \Rightarrow \frac{d}{dt}P_n(t) \to 0$

This yields:

$P_n(\lambda_n + \mu_n) = P_{n+1}\mu_{n+1} + P_{n-1}\lambda_{n-1}, n \geq 0$: This is called Balance Equation

$P_n(\lambda_n + \mu_n)$: Rate of Leaving state n

$P_{n+1}\mu_{n+1}$: Rate of Entering state n from state n+1

$P_{n-1}\lambda_{n-1}$: Rate of Entering state n from n-1

Intuition: Think of P_n as

$P_n = lim_{T\to\infty} \frac{1}{T}\int_0^T 1_{\{N(y)=n\}}dy$

(This is called the time average)

The expectation of the indicator variable would become the probability.

Rate interpretation is shown in the Figure 3.

Transition rate out of state n:

# of transition into state in $[0, T]$:

$lim_{T\to\infty}\frac{\text{\# of transitions out of n in [0,T]}}{T}$
$\approx \frac{P_n T(\lambda_n+\mu_n)}{t} = P_n(\lambda_n + \mu_n)$

$P_{n+1}\mu_{n+1} = P_{n+1}(\lambda_{n+1} + \mu_{n+1})\frac{\mu_{n+1}}{\lambda_{n+1}+\mu_{n+1}}$

$P_{n-1}\lambda_{n-1} = P_{n-1}(\lambda_{n-=1} + \mu_{n-1})\frac{\lambda_{n-1}}{\lambda_{n-1}+\mu_{n-1}}$

Solution:

$n = 0 : \lambda_0 P_0 = \mu_1 P_1 \to P_1 = \frac{\lambda_0}{\mu_1}P_0$
$n = 1 : (\lambda_1 + \mu_1)P_1 = \lambda_0 P_0 + \mu_2 P_2 \to P_2 = \frac{\lambda_1}{\mu_2}P_1 = \frac{\lambda_1\lambda_0}{\mu_2\mu_1}P_0$
$n = 2 : (\lambda_2 + \mu_2)P_2 = \lambda_1 P_1 + \lambda_3 P_3$

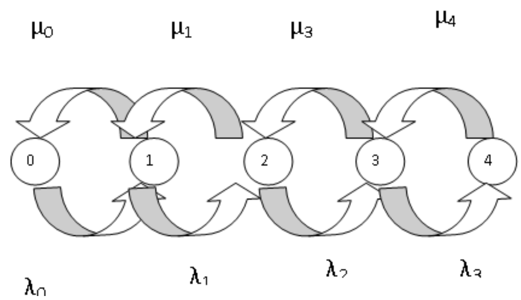


Figure 3: Birth and Death Transition Diagram

$\vdots$
$P_n = \frac{\lambda_{n-1}\cdots\lambda_0}{\mu_n\cdots\mu_1}P_0, n \geq 0$
From normalization, we require:
$1 = \sum_{i=0}^{\infty} P_0(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_1\lambda_0}{\mu_2\mu_1} + \cdots)$

Hence,

$P_n = \frac{\lambda_{n-1}\cdots\lambda_0}{\mu_n\cdots}(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_1\lambda_0}{\mu_2\mu_1} + \cdots)^{-1}, n \geq 0$

Special case: M/M/1 Queue.

λ = arrived rate

μ = service rate

Let $N(t)$ = # of customers in the system at time t.

$\{N(t), t \geq 0\}$ is a Birth and Death process with:

$\lambda_n = \lambda \forall n \geq 0$
$$\mu_n = \begin{cases} \lambda_n = \lambda & \forall n \geq 1 \\ 0 & \text{for } n = 0 \end{cases}$$
Hence, $\frac{\lambda_n}{\mu_{n+1}} = \frac{\lambda}{\mu} = \rho \forall n \geq 0$
Therefore,
$P_0 = (1 + \rho + \rho^2 + \cdots)^{-1}$
$= (\frac{1}{1-\rho})^{-1} = 1 - \rho$ provided that $\rho < 1$
and $\rho_n = \rho^n(1-\rho) = \rho\rho^{n-1}(1-\rho), n \geq 0$

Other characteristics:

L = average # in the system (in steady state)

$= \sum_{n=0}^{\infty} nP_n = \rho \sum_{n=0}^{\infty} n\rho^{n-1}(1-\rho) = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$

W = Average waiting time in the system

$= \sum_{n=0}^{\infty}(n+1)\frac{1}{\mu} = \frac{1}{\mu}(\sum_{n=0}^{\infty}(nP_n + 1)$

$= \frac{1}{\mu}(\frac{\lambda}{\mu-\lambda} + 1) = \frac{1}{\mu-\lambda}$

$P_n = P_n'$ Poisson Arrivals see Time Averages, PASTA

Little's Formula

$L = \lambda W$

Intuition: Impose a cost structure: collect 1 dollar pen unit time per unit time per customer (in the system)

Total amount collected in $[0, T] \approx$
$LT =$
λTW

More generally we have

$H = \lambda G$

H: The average "cost"
λ : Arrival Rate
G: Customer Average "Cost"

Example. Collect 1 dollar per unit time whenever the server is busy. Figure 4 shows this example. Here by taking the view of the customer on average you will spend $\frac{1}{\mu}$, so you must pay the amount of $\frac{1}{\mu}$ dollar. As a result time average of having a busy server would be $\lambda.\frac{1}{\mu}$ which would be equal to $1 - P_0 = \rho$ without calculation here you can see we get it. All you need is the average service time, so if it is generally distributed you just plug that in.

Generally for any system even a prison that a person comes in with rate λ and stay in the system around $\frac{1}{\mu}$ you just multiply them and it shows how many hours the prison has been busy.

Rule Example.PNG

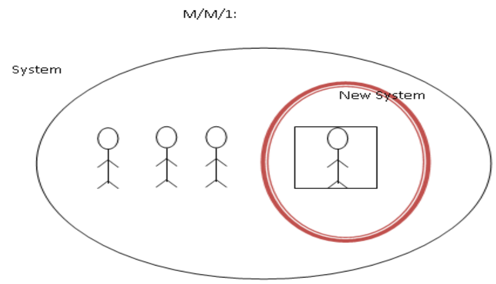


Figure 4: Little Rule Example

Probability Course of Professor Nui @ UTD: Markov Chain

Meisam Hejazinia

11/28/2012

M/M/1-Variations

M/M/1/N

N: is the capacity.

Truncate the states after state N. Therfore: $\lambda_N = 0$

M/M/K

k servers are in parallel.

It means the departure rate would be $k\mu$, so for this case: $\lambda_i = \lambda$

$$\mu_i = \begin{cases} i\mu & 1 \leq i \leq k \\ k\mu & i > k \end{cases}$$

M/M/k/k: Erlang Loss model

It is like the case that there are k servers/ telephone line would be availbale, and for random time the line would be occupied. The loss is when a call arrives, and all lines are busy. It means you make sure noone is in the queue, when you arrive either a server is free and you get service, or your call will be discarded.

The solution does not require second "M", mean service time. It means if you have general distribution, you can simply replace it with exponential variables. This assumption is called insensitivity.

An Extension- M/M/1 with Batch Arrival

The markov chain, the probability of transition only depends on the current state, and not previous states. You can also have contineous time states.

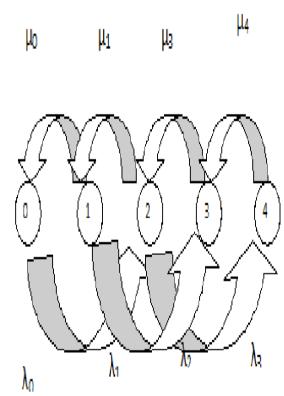


Figure 1: M/M/1/BatchArrival

For simplicity, assume that Batch sie = k, deterministically. Figure 1, shows this extension.

At a transition time: P_{ij} = probability for the next state to be j, given that the current state is i.

$$P_{ij} = \begin{cases} \frac{\lambda}{\lambda+\mu} & for \text{j=i+k} \\ \mu\frac{}{\lambda+\mu} & for \text{j=i-1, i} \geq 1 \\ 0 & otherwise \end{cases}$$

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & \dots \\ \frac{\mu}{\lambda+\mu} & 0 & 0 & 0 & \frac{\lambda}{\lambda+\mu} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Balance equations

$n = 4: (\lambda + \mu)p_4 = \lambda p_1 + \mu p_5$

Solve these equations to get $\{p_i\}_{i=0}^{\infty}$

$L = \sum_{n=0}^{\infty} nP_n$

$W = \sum_{n=0}^{\infty} p'_n \frac{1}{\mu}(n + \frac{k+1}{2})$

The logic here is that n people are in the system, and you are part of a batch of k that arrived, and if you were the first you should just add to 1, yet, you could be any, so you take expectation of $\sum_{j=1}^{k} j\frac{1}{k}$ which would be $\frac{k+1}{2}$

As a result $W = \cdots = \frac{1}{\mu}\sum_{n=0}^{\infty} nP_n + \frac{1}{\mu}\frac{k+1}{2} = \frac{L}{\mu} + \frac{k+1}{2\mu}$, on the other hand we previously had $L = \Lambda W$, now here $\Lambda = \lambda k$. W can be obtained by solving above.

Server utilization is $\rho = \lambda k \frac{1}{\mu} < 1$

There would be fluctation, otherwise the system will explode.

$L_Q = L - \rho$ is the length of the queue.

$W_Q = W - \frac{1}{\mu}$

Figure 2 shows this M/M/1 with batch arrival analysis.

A Related Method - M/E_k/1

$E_k = Erlang(k, \mu)$

Mean $= k\frac{1}{\mu}$ since it is sum of k i.i.d. random variable. If you just monitor of people in the system, then it would not be the markov chain, since we don't know the remaining of time for the person, unless we know in which phase it is, if you know which phase the person is we can solve markovian model. The goal is to create enough informaitont o create markov chain. If you tell me the current phase, then the remining time revealing is not a problem.

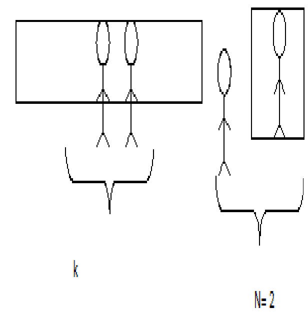


Figure 2: Batch Arrival

This is an example of an "Exponential Model", mean in the model every piece is defined in term of exponential variables, also called contineous markove chain (CTMV).

State Definition:

Mean you have to define state of the process carefully to have the result is that we need to include sufficient information in its definition so that the resulting process would be Markovian (memoryless).

In this example, for $N(t)$ previously it was the number of people in the system, now we change it to $N(t) = \#$ of phases in the entire system. The person out of system will have k pieces remain, and if you know $k - 5$ phases has passed, where k is for the person waiting in the queue, then this person has only remining 5 peices of exponential variable remains.

Only in case the person's getting service has finished all the phases he was to go through, then the other one could enter.

If we say there are 14 phases in the system and $k = 5$, then you will know that there are 3 people int he system, one has only finished one phase, and the other two are waiting.

$\{N(t), t \geq 0\}$ is a Countineous time Markov chain (CTMC).

You should look at the system carefully to come with the clever definition of the system so that you could not only create a CTMC, but also recover the value you were looking for. $\left\lceil \frac{N(t)}{k} \right\rceil$ would # of customers in the system.

The same transition diagram that we had on figure 1, would apply here. Let's try to work out average waiting time directly. $W_{M/E_k/1} = W_{M/M/1\text{Batch k}} + \frac{k-1}{2}\frac{1}{\mu}$, previously we had the random time within the batch completed, yet here we need to wait until everyone is done, so in comparison with previous calculation you need to add $\frac{k-1}{2}$ which is the number of people behind you.

Method of phases

We want to relax exponential assumptions, and generalize the model.

Consider $X = \sum_{i=1}^{k} X_i$ where X_i's are iid Exp($k\mu$). $E(x) = k\frac{1}{k\mu} = \frac{1}{\mu}$, so $Var = k.\frac{1}{(k\mu)^2} = \frac{1}{k}\frac{1}{\mu^2} \to^{k\to\infty} 0$

k is number of phases, and X is duration. Number of peices go up, and the rate is shrinking.

Any constant can be approximated, arbitrarily closely by the random variable.

Consider

$$Y = \begin{cases} c_1 & \text{with probability} p \\ c_2 & \text{with probability} 1 - p \end{cases}$$

From the above:

$C_1 \approx \sum_{i=1} k_1 X_{1i}$, where X_{1i}'s are i.i.d exp. (*)

$C_2 \approx \sum_{i=1} k_2 X_{2i}$, where X_{2i}'s are i.i.d exp. (**)

Following statement was discussed previously and is very important; never underestimate it.

$z_1 \sim Exp(\lambda_1)$, $z_2 \sim Exp(\lambda_2)$, and $\lambda_1 > \lambda_2$

$Z_2 = \sum_{i=1}^{N} z_{1i}$

$N \sim Geometric(\frac{\lambda_2}{\lambda_1})$

Note: The proof is based on conditioning, or using moment generating function.

Suppose X_{2i} have a smaller rate than X_{1i} Then. $X_{2i} = \sum_{l=1} N_i z_l$, where $z_l \sim X_{1i}$, substituting this back to (**) will result in sum of sum of random variables, and then we replace it back in (*), you will end up in $\sum_{i=1}^{N_2} z_l$, it is saying that you stop randomly at some point, and putting this together we will consider $y' = \sum_{l=1} NZ_l$ where N is random. It means you can use the constraint like this having c_1, c_2, you can approximate Y. It means that you could approximate anything contineous, and class of all distribution function, by sum of the smaller range you could regenerate the greater number, and using this you could approximate it, by sum of discrete approximates. This concept is called denseness, mean whatever the interval is, you can find rational numbers in.

Definition

Any random variable has the form $\sum_{i=1}^{N} Z_i$, where the Z_i's are i.i.d. exponential is said to be of phase type.

Claim. The class of phase-type distribution PH is dense in the set of all distributions.

Application. Consider $GI/GI/1$ Queue (GI means General Independent) , to generalize the distribution. Suppose GI's are complicated to mange, then you replace them with PH's. If we can approximate a given model with phase type distribution, then all the building block would be exponentially distributed, and then you can use balance equations.

Now lets focus on $E_k/E_l/1$, we don't know distribution until next distribution, and if you put k, and l, and put more information into state, expanding your definition, of what is the phase that you are in, and then you will know how many more phases you will need. You just define the state space that specializes what is the current state space, and with that definition you will have contineous time markov chain. This will still be messy and you will end up in large state space. This is called curse of dimensionality, and putting lots of information to create markove process, and there would be complicated structure. There is a field called computational probability that you will try to put in phases so that they would not be computationaly exponential.

Other Examples of CTMC

Barbar shop:

You have two chairs in this barbar shop.

There would be no waiting room. The first one haircut, and the second one does washing.

After entering you need to wait for the next chair to be emptied so that you get service there as well.

The first chair rate would be μ_1, and for the second one would be μ_2, and if you enter and see the first chair is taken you will go away.

This means that a person in chair one is either recieving service, or is being blocked.

You are independent you can not go as brother or sister.

Let's check different definition, first start with the stupid one, which is number of people in the system, and in this case we do not have enough information since the person in the first chair could be idle, for the markovian property you need to make sure that everything is clear for each state.

state definition

$(0,0)$: means the system is empty: it is makovian, since time until the next state would be exponential.

$(1,0)$: 1 in system at chair 1

$(0,1)$: 1 in the system at chair 2

$(1,1)$: 1 at each chair, both "active"

$(b,1)$: 1 at chair one and inactive, the other at chair 2

Next time we will continue with some examples, and then we will go over couple of pages (beggining) of chapter 7. This discussion that we had covered both chapter 6 and 8. Figure 3 shows this case.

The last session we will go over the homework problems.

Next session we will do couple of normalization, and then try to solve it.

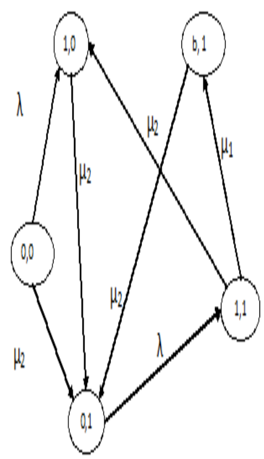


Figure 3: Barbar Shop

Probability Course of Professor Nui @ UTD: Renewal Theory

Meisam Hejazinia

12/05/2012

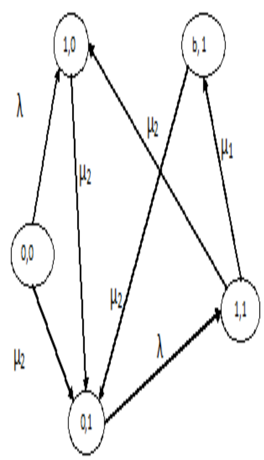


Figure 1: Barbar Shop

We continue the barber shop example.

We had two chairs in the barber shop, and customers continue arriving with poisson process, and each server served with rate μ_i, and there are two servers. When the person on the other chair finished then person on chair one would be able to move on.

Status would be in the form of $(0,0),(0,1),(1,0),(1,1),(b,1)$ shown in figure 1.

Put down the balance equations and solve for $p_{00}, p_{01}, p_{10}, p_{11}, p_{b1}$

Question 1: Loss probability=?

If someone sees the third, fourth, and fifth probabilities that one person would be in the first chair would be the answer so:

$A1 : p_{10} + p_{11} + p_{b1}$

Question 2: Entrance rate =?

(i) Calculating using L

$A_2 : \lambda[1 - p_{10} + p_{11} + p_{b1}]$

Question3: W(the expected waiting time) for all arriving customers=?

$A_3 : L = 0.p_{00} + 1.(p_{10} + p_{01}) + 2.(p_{11} + p_{b1})$
$W = \frac{L}{\lambda}$ This is not proper equation to be used here since it is for people who may not have enter as well. $w_e = \frac{L}{\lambda_e}$ for those who actually enter.

(ii) Conditioning on the state at the time of arrival, we have:

$W = 0.(p_{10}+p_{11}+p_{b1})+p_{00}(\frac{1}{\mu_1}+\frac{1}{\mu_2})+p_{01}[\frac{1}{\mu_1+\mu_2}+\frac{\mu_1}{\mu_1+\mu_2}(\frac{1}{\mu_2}+\frac{1}{\mu_2})+\frac{\mu_2}{\mu_1+\mu_2}(\frac{1}{\mu_1}+\frac{1}{\mu_2})]$

(iii) $A = E[max(x_1, x_2)] + min(x_1 + x_2) = \frac{1}{\mu_1} + \frac{1}{\mu_2} - \frac{1}{\mu_1+\mu_2} + \frac{1}{\mu_2}$
The logic behind this equation is that first we find the maximum time that it will take for the person to move to the chair, and then the time it would take to land from second chair to out of barbary shop.

(iv) $A = \frac{1}{\mu_1} + \frac{\mu_2}{\mu_1+\mu_2}.\frac{1}{\mu_2} + \frac{\mu_1}{\mu_1+\mu_2}(\frac{1}{\mu_2} + \frac{1}{\mu_2})$

Question 4 $W_e = \frac{p_{00}}{p_{00}+p_{01}}(\frac{1}{\mu_1}+\frac{1}{\mu_2})\frac{p_{01}}{p_{00}+p_{01}}.A$

Question 5: Server utilizations=?

A_s : Server 1: $p_{10}+p_{11}$
Server 2: $p_{01}+p_{11}+p_{b1}$
Q_s : Server occupancy?
Server 1: $p_{10}+p_{11}+p_{b1}$
Server 2: $p_{01}+p_{11}+p_{b1}$
Extension: Service at the second chair $\sim Erlang(2, \mu_2)$

States: $(0,0),(1,0),((0,1),(0,1'))$ $,((1,1),(1,1')),((b,1),(b,1'))$
We split each of the states into two since now the Erlang would have two phase, the normal one on the first state, and the second one in the form of prime.

2) Server Breakdown M/M/1

α: failure rate (exponentially distributed)

β: Repair

Everyone is lost when the server breaks down.

We define the states int he form of the following:

0_u: mean no one is in the system yet server is up

0_d: mean non one in the system and the server is down

$1,2,3$: states based on the number of people in the system

In the sequence of states: $(0_u,1,2,3)$ we would have λ to go up, and μ would go down.

Moving from all $0_u,1,2,3,\dots$ states to 0_d would be done with rate α, yet the transition from 0_d to 0_u is done by the rate of β.

Generally for mutlidimensional spaces you will just put in the Moment Generating funciton and multiply them to get the result.

Repairman problem

3 Machines [Customers]
2 Repair persons [Servers]

Uptime of the machines would be $\sim Exp(\lambda)$

Repair time of the repair person is $\sim exp(\mu)$

States: # of down machines

States are $0,1,2,3$ showing the number of machins down. The transition from state 0 to state 1 will have rate of 3λ, and the rest in the following form:

1 to 2: 2λ

2 to 3: λ

3 to 2: 2μ

2 to 1: 2μ

1 to 0: μ

Tandem Queue: M/ M/ 1 $\to$ M/ 1

There would be two queue and two server were connected int the serial form, and the rate of entery to the system is λ, and the servers rate would be in the form of μ_1 and μ_2

States would be in the form of (i,j) where the first one is the number in the first queue, and the seocnd is the number in the second queue

Transition diagram would be transition according to the following:

$(i,j) \to (i,j-1) : \mu_2$

$(i,j) \to (i+1,j) : \lambda$

$(i,j) \to (i-1, j+1) : \mu_1$

The solution $\{p_{ij}\}_{i=0,j=0}^{00,00}$ has the form:

Product form:

$p_{ij} = \rho_1^i(1-\rho_1)\rho_2^j(1-\rho_2)$: It is conjecture, and each part was calculated previously

$\rho_1 = \frac{\lambda}{\mu_1}$

$\rho_2 = \frac{\lambda}{\mu_2}$

We assume that $\lambda \geq \mu_1, \mu_2$

Also we need to use balance equations and substitute and verify that these are independent, and we are allowed to multiply each of the results based on what we previously derieved.

Interesting property of product form is that the output of the first process is the poisson process, and it is stronger.

Another extension is that let say we have finite number of people circulate in the system, in the form of the close system, you would go through this, and then yous how that this solution works.

Also in **Network** system, that in general form you can come out from any queue, and the customer can come out with some probability or go to one of the other servers, then you can put the factor for each state, and using this conjecture of the product form, then for each station you will find the solution and multiply together and the solution still works.
This whole process is reversible, mean forward or backward in time the stochastic behavior would be the same. The forward process is input, and if you reverse it the output would become stochastically equivalent. This is called network extension. You can change to Erlang, yet on that case product form will break down.

For chapter 6 only the first part is covered, and three last section is not covered.

For chapter 7, part of section 3, and 4 only covered.

Chapter 8 the first couple of section, 1,2,3, 4 is only covered, which is the application area for queuing model.

Removal Theory

Renewal process is relaxing the assumption of equal number of events or events function of time.

Let say that separate intervals are x_1, x_2, x_3, x_4, the x_i's are i.i.d. $\sim F(.)$, where F is arbitrary. Let $N(t) = \#$ of events by time t.

Definition: $\{N(t), t \geq 0\}$ is called a renewal process. Let $m(t) = E[N(t)]$

It would not necessarily exponential, yet it could not be normal, since it is not negative.

$\frac{N(t)}{t} \to \frac{1}{E(x_1)=\mu_f}$ with respect to 1.

Let have t between two sum of random variables then we will have:

$\frac{\sum_{i=1}^{N(t)} X_i}{N(t)} \leq \frac{t}{N(t)} \leq \frac{\sum_{i=1}^{N(t)+1} X_i}{N(t)+1} \frac{N(t)+1}{N(t)}$

$\Rightarrow \mu_f \leq \frac{t}{N(t)} < \mu_F$

$\frac{E[N(t)]}{t} \to \frac{1}{\mu_f}$

Renewal Reward process

To make life easier, let assume y_i = Reward recieve at the end of the i'th renewal interval

Assumption: $(X_i, Y_i), i \geq 1$, are i.i.d. random vector.

X_i, Y_i are not necessarily independent.

$R(t) = \sum_{i=1}^{N(t)} Y_i$

$\frac{R(t)}{t} \to \frac{E(Y_1)}{E(X_1)}$ with probability of 1

$\frac{R(t)}{t} = \frac{[R(t)]}{t}$

$\frac{N(t)}{t} \frac{\sum_{i=1}^{N(t)} Y_i}{N(t)} \leq \frac{R(t)}{t} < \frac{\sum_{i=1}^{N(t)+1} Y_i}{N(t)+1} \frac{N(t)+1}{t}$

In general case, the reward could accumulate contineously overtime, and they can be negative, but here we made simplifying assumption.

i.i.d. mean they have the same distribution, and the same mean, yet independent.

This means total reward would be equal to the number of rewards, over expected time of thw reward.

Applications

1. Train Dispatching.

Distpatch a train when number of passengers reaches K.
$C=$ waiting cost per unit time per customer.

$Q=$ what is the long-run average waiting time of customer? (K is given)
Cost would be equal to $= c.Area$

If we call it regenerated cycle until the dispatch hapens.

Long-run Average cost $= \frac{E[\text{Cost in a cycle}]}{E[\text{Cycle time}]}$

$E[\text{Cycle time}] = k\frac{1}{\lambda}$

$E[\text{Cost in a cycle}] = c[\frac{1}{\lambda} + \frac{2}{\lambda} + \ldots + \frac{k-1}{\lambda}]$

Work on the 7.8, and 7.26 from problems of the book, and be prapared for.

These problems, are hard, the concept is simple you calculate the reward during whole regeneration process, and devide it by cycle time.

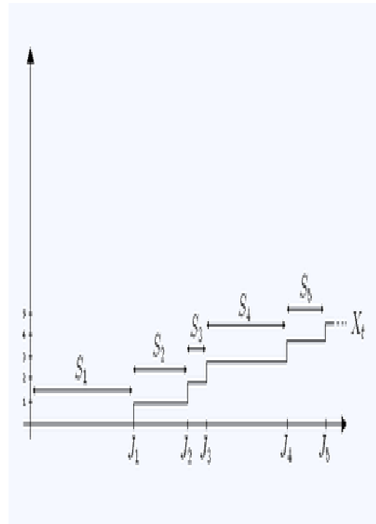


Figure 2: Renewal process

Car replacement

Life time $\sim F(.)$

Replacement cost $=$

c_1 if the car is still working
$c_1 + c_2$ if the car is dead

Age replacement policy: Replace the car whenever it reaches age T or dies, whichever occurs first.

Let's assume that T is fix number.

Long-run Average cost $= \frac{E[\text{Cost in a cycle}]}{E[\text{Cycle time}]}$

$E[\text{Cycle Time}] = E\min[T, X]$ where X is Life Time.

The hint is you just calculate the min by using tale probability.

$E[\text{Cost in a cycle}] = c_1 + c_2.px < T = c_1 + c_2.F(T)$

Next week we will do the homework problems.

Probability Course of Professor Nui @ UTD: Homework Solution

Meisam Hejazinia

12/12/2012

In renewal process time intervals are not exponential as poisson, yet could have general distribution.

Problems:

Chapter 5: 7,10,12(a), 71,77, 91, 80, 94(b), 18,45
Chapter 6: 11, 14, 5
Chapter 8: 29

Chapter 5 7. $p\{x_1 < x_2 | min(x_1, x_2) = t\} = \frac{p\{x_1<x_2, min(x_1,x_2)=5\}}{p\{min(x_1,x_2)=5\}} = \frac{p\{x_1<x_2, min[x_1,x_2]\in(t,t+1)\}}{p\{min(x_1,x_2)=5\}}$

Another way is $\frac{f_1(t)\bar{F_2}(t)}{f_1(t)\bar{F_2}(t)+f_2(t)\bar{F_1}(t)} = \frac{f_1(t)|\bar{F_1}(t)}{\frac{f(t)}{\bar{F_1}(t)}+\frac{f_2(t)}{\bar{F_2}(t)}} = \frac{r_1(t)}{r_1(t)+r_2(t)}$

10. a) $E(MX|M = X) = E[M^2]$
$P\{M > t, M = x\} = p\{t < x < Y\}$ where $M : X < Y$
$= \int_t^\infty e^{-\mu x}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda+\mu}\int_t^\infty(\lambda+\mu)e^{-(\lambda+\mu)x}dx = p\{M = X\}.P\{M > t\}$
This is important relation since most of the problems would become easier by it.

b) $E[MX|M = Y] = E[M(M + X')] = E(M^2) + E[MX'] = E[M^2] + E(M)E(X') = E(M^2) + E(M)E(X')$

c) $cov[X, M] = E(XM) - E(X)E(M)$
$E(XM) = E(XM|M = X)P\{M = X\} + E(XM|M = Y)P\{M = Y\}$

12. $PX_1 < X_2 < X_3 = \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}\frac{\lambda_2}{\lambda_2+\lambda_3}$

18. $X_{(1)} = Min[X_1, X_2]$ and $X_{(2)} = Max[X_1, X_2]$

a) $E[X_{(1)}] = \frac{1}{2\mu}$
b) $Var[X_{(1)}] = (\frac{1}{2\mu})^2$
c) $E[X_{(1)]=E[X_{(1)}+R]=\frac{1}{2}}$
$+1\frac{}{\mu}$ where R is residual, and both are independent.
d) $Var[X_{(2)}] = Var[X_{(1)}] + Var(R)$

71. $E[\sum_{i=1}^{N(t)} g(s_i)] = E[\sum_{i=1}^{N} g(u_i)]$
$N \sim poisson(\lambda t)$
$U_i \sim$ i.i.d. Uniform $(0, t)$.
$= E(N)E[g(u)] = \lambda t\int_0^t g(x)\frac{dx}{x} = \lambda int_0^t g(x)dx = \lambda\int_0^t g(x)dx = \int_0^t \lambda g(x)dx$
$Var[\sum_{i=1}^{N} g(u_i)] = Var[E[\sum_{i=1}^{N} g(u_i)|N)] + E[Var(\sum_{i=1}^{N} g(U_i)|N)]$

77. a) $p\{N = 1\} = \frac{\mu}{\lambda+\mu}$
b) $p\{N = 2\} = \frac{\lambda}{}$
$2\mu\lambda + 2\mu$
c) $p\{N = j\} = \frac{\lambda}{}$
$\lambda\lambda + 2\mu \ldots \frac{\lambda}{\lambda+(j-1)\mu}\frac{j\mu}{\lambda+j\mu}$
d) $= \sum_j p\{event|N = j\}p\{N = j\} = \sum_j \frac{1}{j}p\{N = j\}$
e) $= \sum_j E[Time|N = j]p\{N = j\} = (\frac{1}{\lambda+\mu} + \frac{1}{\lambda+2\mu} + \ldots)$
This is actually the sequence that you allow each arrival win and then last time allow the departure to win; you can see the helpfulness of memoryless property, else calculation would be total mess.

45. a) $Cov[T, N(t)] = E[TN(t)] - E(T)E[N(t)]$ where $E[N(t)] = \lambda E(T)$
$E[E(TN(T)|T)] = E(\lambda T^2) = \lambda E(T^2)$
b) $Var[N(T)] = E[N(T)^2] - (E[N(t)])^2$
$E[N(T)^2] = E[E[N(T)^2|T]] = E[\lambda T + (\lambda T)^2] =$

$\lambda E(T) + \lambda^2 E(T^2)$

80. We assume that distribution of interarrival are given in the standard nonhomogeneous poisson process. If we know x_1, x_2 we can find out x_3 and when we know these three we would be able to determine x_4, since sum of previous will tell us what is the rate. These processes are Markovian.

91) $pX_1 > \sum_{i=2}^{n} X_i = (\frac{1}{2})^{n-1}$

94. b) $p\{x > t\} = e^{-\lambda \pi t^2}$
$E(x) = \int_0^\infty p\{X > t\}dt = \int_0^\infty e^{-\lambda \pi t^2} dt$

Integrating tail for taking the expectation is very useful formula: $E(x) = int_0^\infty p\{x > tdt$, to derieve this you can use indicator variable.

Chapter 6

5. Two groups of people, part infected: i, and parts are not infected: N-i

The probability of all interaction would be $\frac{1}{C(N,2)}$, and you want infected and uninfected interaction would be in the form of $i(N-i)$, and you will get $\frac{i(N-i)}{C(N,2)}$, the probability would be p, so given current state i, then for the poisson we would have number of people which creates markov chain. and the rate would be then in the form of $\lambda p \frac{i(N-i)}{C(N,2)}$. The only difference between here and BASS model is that there is not self initiated here.

11. $p_{1j} = pX(t) = j = PT_1 + \ldots + T_{j-1} \leq t - pT_1 + \ldots + T_j \leq t = pT1 + \ldots + Tj - 1 \leq t, T_1 + \ldots + T_j > t$

(c) p$_{ij}(t)$ $\sim$ sum of i independent variable each $\sum p_{ij}$