Statistiska begrepp och metoder som används i Successivprincipen

Generellt har statistiska procedurer antingen varit överförenklade eller opraktiska för projektteamen. Resultatet blir inte trovärdigt i något av fallen. Successivprincipens tillvägagångssätt ger mer rätt och är framförallt praktiskt genomförbart:

- Vi får en bra uppfattning av normalfallet med relaterade möjligheter och risker genom att använda kvalitativa scenarioanalysen
- Sedan kvantifierar vi normalfallet genom att använda trippel estimat för att inkludera osäkerheterna i de olika aktiviteterna/grupperna vilka kan hanteras oberoende av varandra eftersom scenariot för normalfallet är tämligen specifikt.
- Därefter räknar vi med de korrelerande faktorerna kopplade till “den verkliga världen” via kvantifiering av generella villkor som är applicerbara på modellen och åter igen använder vi trippel estimat.

Detta tillvägagångssätt som är en naturlig algoritm är tillräckligt enkel för att kunna förklaras för workshopdeltagarna så att alla är med på hur den kvantitativa metoden fungerar och vad resultaten betyder.

Metoden använder en god approximation för en variabel där ett viktat medelvärde och en standard avvikelse beräknas enligt följande:

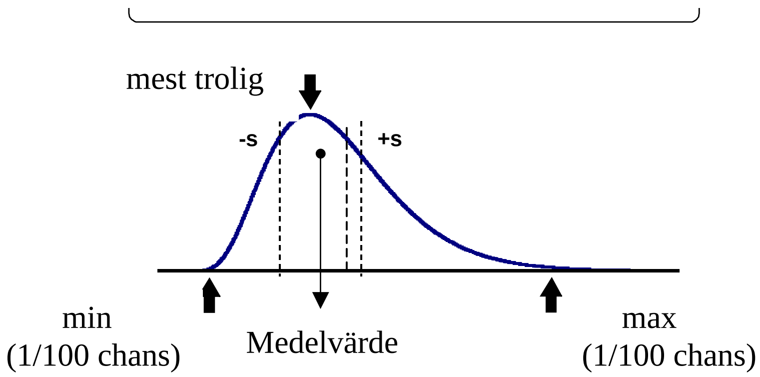
$$\text{Medelvärde} = \frac{[\text{min} + (3 \text{ x mest trolig}) + \text{max}]}{5}$$

egentligen 2,9 och ej 3

egentligen 4,9 och ej 5

$$\text{Standardavvikelse} = \frac{[\text{max} - \text{min}]}{5}$$

egentligen 4,65 och ej 5



De här algoritmerna relaterar till Erlang K5 fördelning vilket har visat sig vara ett perfekt val för just projektsituationer. Fastän medelvärdet för de nästan alla variabler i modellen verkar bli lite fel så har forskning i mer än 20 år visat att det kumulativa medelvärdet och den relaterade S-kurvan blivit en excellent förutsägelse av det verkliga utfallet, se sidan 5 och 6.

Successivprincipen baseras på Bayes utvecklat teorem, se sidan 2, där räkneregler för sannolikheten redovisas. Vidare använder Successivprincipen sannolikhetsfördelning kallad normalfördelning, se sidan 5. Varför just dessa teorem och fördelningar? Steen Lichtenbergs doktorsavhandling resulterade just i detta 1974. Sedan dess har metoden använts i hundratals (1000-tals?) projekt runt om i världen.

Beskrivande statistik

En grupps (population) fördelning kan beskrivas med flera olika typer av värden.
Medelvärden … *detta är ett mått på läge*

- aritmetiska medelvärden … *en grupps aritmetiska medelvärde visar gruppens läge och beräknas som summan av alla termer i gruppen dividerat med antalet individer i gruppen … det är detta man oftast avser när man enbart pratar om medelvärdet*
- viktat medelvärde … *du räknar ut medelvärdet av flera medelvärden*
- geometriskt medelvärde … *när du vill beräkna genomsnittlig avkastning*
- medianvärde … *det mittersta värdet i en grupp*

Spridning … *anger hur väl samlad gruppen är*

- varians … *beskriver hur värdena i en grupp varierar, den är lika med kvadratavvikelsen från medelvärdet*
- standardavvikelse … *när du vill veta hur stor den genomsnittliga avvikelsen från medelvärdet i en grupp är*

Skevhet … *är fördelningen symmetrisk eller asymmetrisk? … en helt symmetrisk fördelning, där avvikelserna från medelvärdet är lika stora ovanför som under medelvärdet har ingen skevhet.*

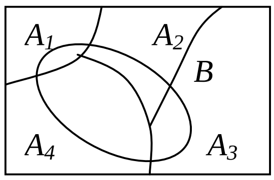
Sannolikhet

Vi gör experiment som kan upprepas många gånger. Vi bestämmer alla möjliga utfall av experiment som kan inträffa. Sedan bestämmer vi vilka fall som vi anser vara gynnsamma utfall, i detta fall benämnt A. Sannolikheten, P, för att vi skall få ett gynnsamt utfall blir då
P = antal gynnsamma utfall / totalt antal möjliga utfall

Räkneregler

- Snittsannolikhet … *sannolikheten att A och B inträffar, då A och B är oberoende av varandra* $P(A \cap B) = P(A) \times P(B)$
- Betingad sannolikhet … *sannolikheten att A inträffar när B har inträffat*
  $P(A \mid B) = P(A \cap B) / P(B)$
- Bayes teorem … *betingad sannolikhet enligt Thomas Bayes*
  $P(A \mid B) = [P(B \mid A) \times P(A)] / P(B)$
- Bayes utvecklat teorem för betingad sannolikhet, mer generell …

$$P(A_i \mid B) = [P(A_i)\, P(B \mid A_i)] \,/ \sum_{i=1}^{n} P(A_i)P(B \mid A_i)$$



vi kan också teckna den totala sannolikhetsekvationen för händelsen B

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B \mid A_i)$$

- Oberoende variabler … *två händelser, A och B, sägs vara oberoende om och endast om* $P(A \mid B) = P(A)$

- Additionsregeln … *om två händelser är ömsesidigt oberoende och den ena händelsen har N_1 utfall och den andra händelsen har N_2 utfall så kan hela försöket utfalla på $N_1 + N_2$ sätt.*
- Multiplikationsregeln … *om en händelse kan utfalla på N_1 sätt och den därpå följande händelsen kan utfalla på N_2 sätt så kan sekvensen av de två händelserna utfalla på N_1 x N_2 sätt. Mer generellt kan n händelser utfalla på N_1 x N_2 x … N_n olika sätt om N betecknar antalet möjliga utfall för en viss händelse.*

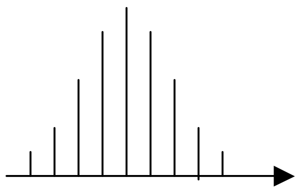
Väntevärden

- Stokastisk variabel … *resultatet i ett slumpmässigt försök kallar vi för en stokastisk variabel (ett stort X) - slumpvariabel. X kan sägas stå som en symbol för resultatet innan försöket är utfört.*
  Diskret fördelning *… en variabel som är diskret fördelad kan bara anta vissa värden inom ett visst intervall. Detta intervall kan vara hur stort eller litet som helst.*
  Kontinuerlig fördelning *… en variabel som är kontinuerligt fördelad kan anta alla värden inom ett intervall.*
- Sannolikhetsfördelningar … *en sannolikhetsfördelning är en sorterad lista på alla tänkbara värden som en stokastisk variabel kan anta tillsammans med respektive sannolikhet. $P(x) = P(X = x)$ där X visar att du undersöker alla tänkbara värden x.*
- Kumulativ fördelning … *den kumulativa sannolikhetsfördelningen ger dig sannolikheten för att utfallet, X, ska bli mer eller mindre än x. $F(x) = P(X \leq x)$*
- Väntevärde … *Det förväntade värdet, väntevärdet [E(X) eller μ], eller medelvärdet, av en stokastisk, slumpmässig variabel är lika med summan av produkterna av alla möjliga värden av variabeln och dess motsvarande sannolikhet. Väntevärdet är det medelvärde som utfallet kommer att få om du gör oändligt många försök.*
  Väntevärde för en diskret fördelad stokastisk variabel benämns E(X)
  $E(X) = \mu = \sum x\, P(x)$ där μ=väntevärdet, x=varje möjlig värde, P(X)=sannolikheten för respektive värde.
- Varians … *är det förväntade värdet av kvadraten på avvikelsen mellan en stokastisk variabel och dess väntevärde. $V(X) = \sigma^2 = E(X- \mu)^2$ där μ=väntevärdet*
- Standardavvikelse … *den definieras som den positiva roten av variansen.*
  σ = roten av $E(X- \mu)^2$ eller roten av $E(X^2)- \mu^2$
  (i Successivprincipen uttrycks osäkerheten som en standardavvikelse)

Sannolikhetsfördelningar

Diskreta fördelningar

Med diskreta fördelningar menas sannolikhetsfördelningar av diskreta stokastiska variabler. För varje värd på den horisontella axeln, finns en motsvarande sannolikhet på den vertikala axeln.



- Binomialfördelning … *du använder binomialfördelningen då du har ett försök eller en händelse som uppfyller följande villkor:*
  - Varje delförsök kan ha två olika resultat, två utfall.
  - Delförsöken ska vara oberoende av varandra. Utfallet av nästa försök ska inte vara påverkat av det nyss gjorda för-söket.
  - Sannolikheten för att ett delförsök lyckas, benämnt p, ska vara lika stor för alla delförsök.

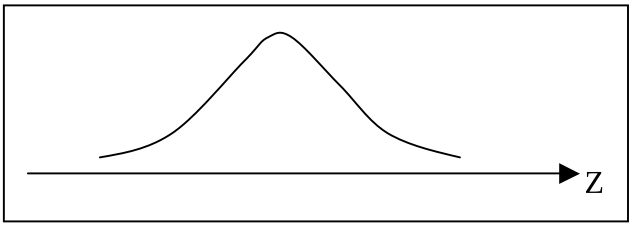
Binomialfördelningen ger alltså sannolikheten att lyckas exakt x gånger vid n försök med sannolikheten p.

- Geometrisk fördelning … *du använder geometrisk fördelning då du har ett försök som uppfyller samma villkor som för binomialfördelningen (se ovan). Geometrisk fördelning ger sannolikheten för att ett försök nummer x skall bli första lyckade.*
- Poissonfördelning … *används vid studier av olika slags trafikintensiteter, med vissa villkor, samt som approximation till binomialfördelning. Binomialfördelning övergår till Poissonfördelning när sannolikheten att någonting händer är liten och antalet försök är stort $\sigma^2=\mu$.*
  - Händelserna observeras på en kontinuerlig skala. Exempelvis under en viss tid, i en given volym eller på en specifik sträcka.
  - Händelserna skall vara oberoende av varandra.
  - Händelserna uppkommer slumpmässigt men med en viss frekvens som du betecknar med lambda, λ. $\lambda \geq 0$
  Du vill undersöka hur vanlig en händelse är på en viss del av en kontinuerlig skala samt hur stor sannolikhet det är att du finner en viss mängd av denna händelse vid en undersökning.

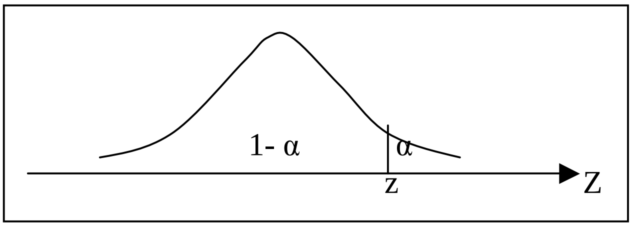
Kontinuerliga fördelningar

I en kontinuerlig sannolikhetsfördelning finns oändligt många utfall.

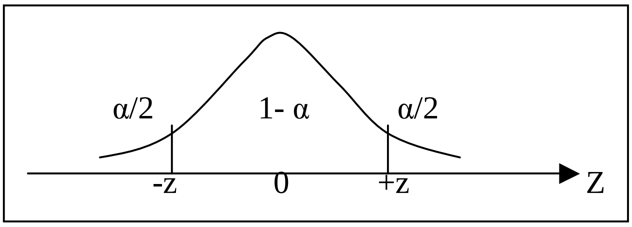
- Kontinuerlig fördelning … *ett utfall för en kontinuerlig variabel kan anta alla värden inom ett intervall som kan beskrivs av en täthetsfunktion, f(x)*



Sannolikheten för hela utfallsrummet, P(-∞ ≤ Z ≤ +∞) är 1 vilket är arean under kurvan för alla värden på Z.



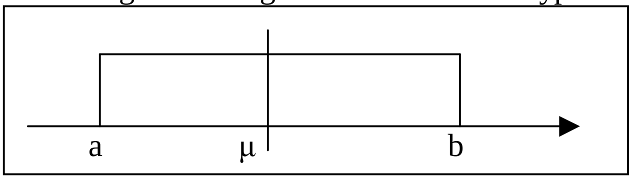
α betecknar sannolikheten för att utfallet ska ligga över en viss punkt z Sannolikheten att utfallet ska hamna under denna punkt är 1- α



En symmetrisk fördelning, jämt fördelade respektive sidor om medelvärdet, används ibland α/2 för att beteckna svansen, sannolikheten att utfallet kommer att ligga utanför aktuellt intervall, konfidensintervall.

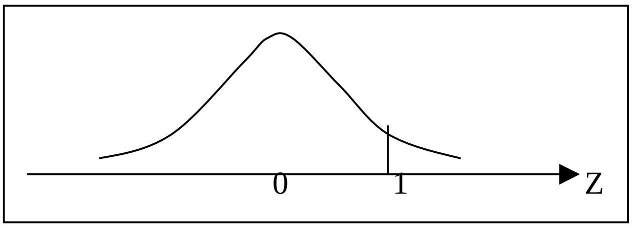
Du söker sannolikheten för att ett utfall ska hamna runt noll med intervallet ±z. Du vet att inom detta intervall är sannolikheten 1- α. Du finner värdet på z där den kumulativa sannolikheten är 1- (α/2) för utfall.

- Likformig fördelning … den enklaste typen av kontinuerlig fördelning.

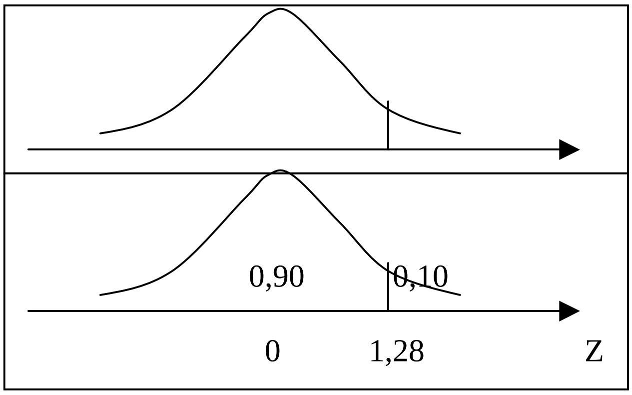


Sannolikheten för utfallet är helt jämnt fördelat kring medelvärdet, μ, mellan två punkter a och b, sådana att μ ligger i mitten och där a < b.

- Normalfördelning *(även kallad Gaussian fördelning) den absolut viktigaste fördelningen. Kurvan är klockformad och fördelningen är symmetrisk kring medelvärdet, μ, och standardavvikelse, σ.*

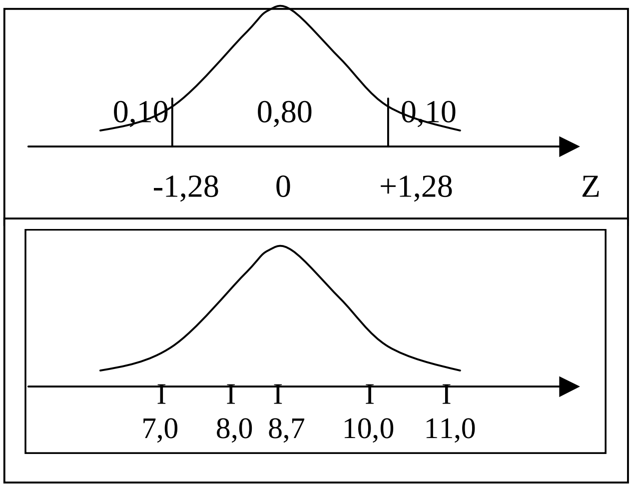


Här är medelvärdet = 0 och standardavvikelse = 1.



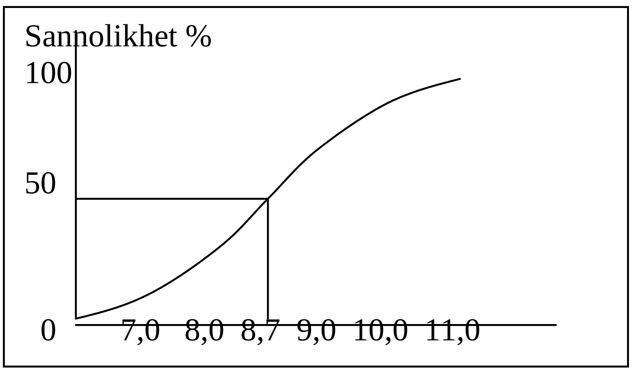
I den vanliga normalfördelningskurvan studerar du sannolikheten för utfallet X. I en standard normalfördelningskurva studerar du sannolikheten för utfallet Z.

Sannolikheten att utfallet ska bli mindre än, eller lika med, Z är 0,9 eller 90%.

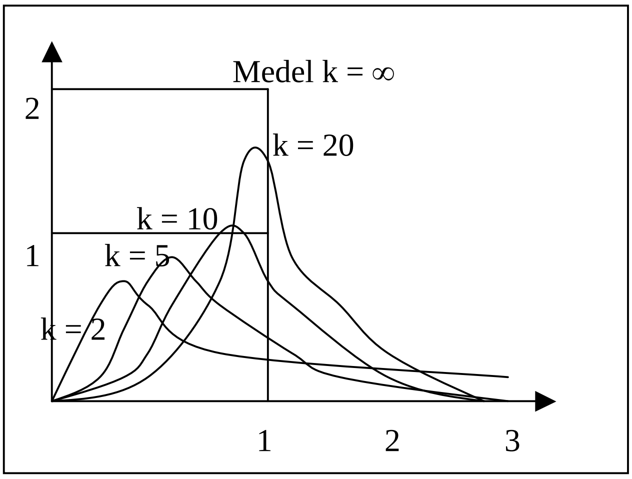


Symmetri i normalfördelningskurvan: normalfördelningskurvan är symmetrisk, den är exakt likadan på vardera sidan om medelvärdet, eller nollpunkten för standard normalfördelning.

Här en typisk frekvens funktion …

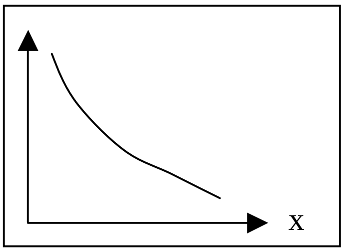


… och här en motsvarande kumulativ S-kurva.



Här en förklaring på hur en snedfördelad Erlang K5 fördelning närmar sig en normalfördelning när k närmar sig oändlighet. Alltså när vi upprepar försöket ett mycket stort antal gånger.

- Students t-fördelning … *är en mycket vanlig sannolikhetsfördelning för små urval, där variablarna är normalfördelade, men standardavvikelsen är okänd. Du använder t-fördelning för att finna konfidensintervall och medelvärden. Kurvan ser ut som normalfördelningskurvan, fast mer utspridd.*
- Exponentialfördelning … *används för att beskriva sannolikhetsfördelning av kvarvarande tid. Det kan gälla olika typer av väntetid, tiden från start till någonting speciellt händer. Passar speciellt bra för beräkningar av livslängd för komponenter som inte åldras.*



- x^2-fördelning … *används bland annat för hypotesprövning, konfidensintervall samt för test av modeller.*
- Gamma-fördelning … *visar fördelningen för summan av n stycken oberoende slumpvaribler som är exponentialfördelade.*
- Beta-fördelning … *används bland annat för att uppskatta när en aktivitet är avslutad baserad på t.ex. PERT-diagram, med optimistiska och pessimistiska uppskattningar av tidsåtgång.*

Konfidensintervall, av latinets confidentia, tillförsikt, tilltro. En statistisk term som anger graden av osäkerhet för ett försök eller mätvärde. För varje konfidensintervall finns ett värde, **konfidensgrad**, till exempel 68%, 95% eller 99%. Konfidensgraden anger sannolikheten för att det sanna värdet för den uppmätta storheten ligger inom konfidensintervallet. Konfidensgraden för en bedömning som används vid beräkning av medelvärdet och standardavvikelsen skall vara 1% i Successivprincipen.

Det totala slutresultatet efter en analys med ett totalt medelvärde på M samt en standardavvikelse på S (= σ) ger en konfidensgrad på 68%. Alltså sannolikheten att utfallet hamnar inom konfidensintervallet M+/-S är 68%.

Monte Carlo simulering

En av de viktigaste numeriska metoderna för att beräkna flerdimensionella integraler som utvecklades under senare hälften av 1900-talet ar Monte Carlo-metoden. Den grundläggande idén med denna metod är att välja punkter på ett slumpmässigt sätt inom ett avgränsat område och sedan använda den vägda summan av funktionsvärdena i dessa punkter som en uppskattning för integralen. De första Monte Carlo-simuleringarna gjordes redan på 1950–talet då de första datorerna blev tillgängliga. Monte Carlo-metoden har tillämpningar bl.a. inom statistisk fysik och kvantmekanik.

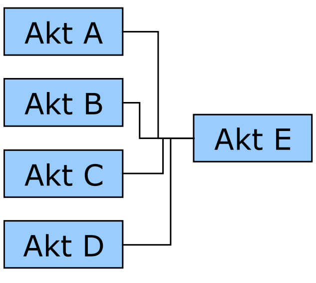
Vid en tidsanalys enligt Successivprincipen används Monte Carlosimulering för att beräkna kritiskt index. Metoden används för att beräkna sannolikheten för att en aktivitet är kritisk. Monte Carlo-metod är en brett använd klass av algoritmer som används för att simulera matematiska system. De skiljer sig från andra simuleringsmetoder genom att vara stokastiska, d.v.s ickedeterministiska i någon form - vanligtvis genom att förlita sig på slumptalsgeneratorer, till skillnad från deterministiska algoritmer.
En Monte Carloalgoritm är en numerisk Monte Carlometod som används för att finna lösningar till matematiska problem (som kan ha flera variabler) som inte kan lösas enkelt med andra numeriska metoder.

Monte Carlo-simulering, beräkningsmetod som baseras på slumptal. "Monte Carlo" kallas simuleringen därför att man spelar mycket roulett där och det kan ses som en simpel slumptalsgenerator.

När ett problem inte längre kan lösas analytiskt tillgrips ofta inom den tillämpade matematiken en numerisk metod. Inom statistiken använder man då istället någon form av simulering. Om den funktion (y) vars statistiska fördelning sökes, är en funktion av ett antal stokastiska variabler (x1, x2, ... xn), så att y = f(x1, x2, ...xn) kan ett stickprov på funktionen y erhållas genom följande metod, ofta benämnd Monte Carlo-metoden.

Det krävs då att frekvensfunktionen för x1, x2, ... xn är kända, så att deras fördelningsfunktioner också är kända. Simuleringen tillgår på följande sätt. Med hjälp av en slumpgenerator genereras ett slumptal, i intervallet 0-1000 för varje variabel. Detta divideras sedan med 1000, så att alla slumptal faller mellan 0 och 1. Slumptalen kallas här r1, r1, ... rn. Dessa betraktas sedan som värden på fördelningsfunktionen och med hjälp av fördelningsfunktionen bestäms tillhörande värden på variabeln, x1, x2, ... xn. Därigenom har en uppsättning slumpmässigt valda värden på variablerna x1, x2, ... xn bestämts och det första slumpmässiga värdet på funktionen f1(x1, x2, ... xn) kan beräkns. Sedan upprepas proceduren ett stort antal gånger och ett stickprov av y = f(x1, x2, ... xn) erhålls, till vilken en statistisk fördelning kan anpassas, eller alternativt kan medelvärde och standardavvikelse beräknas.

Stokastiskt tidstillägg.
I en tidsanalys beräknas den totala genomförandetiden för hela projektet. Vid en sådan beräkning tas hänsyn till om en aktivitet är beroende av många andra aktiviteter. T.ex:



Om aktivitet E är beroende av aktivitet A-D, så beräknas ett tidstillägg för aktivitet E som är en funktion av antal aktiviteter och dess kritiska index.

Planeringsmodeller

Gantt-scheman är enkla tidsdiagram som används för planering, schemaläggning och för att åskådliggöra ett projekt grafiskt.

I **nätverksmodeller** beskriver vi samband mellan olika aktiviteter; vilka som föregår vilka och vilka kan göras parallellt. CPM, Critical Path Method, inför tidsbegreppet i nätverket och ger oss möjlighet att räkna fram projektets sluttidpunkt eller kostnader. Vi kan också se vilka aktiviteter som är kritiska, kritisk linje, och på sätt avgöra vad som händer vid förseningar av enskilda aktiviteter. PERT-metoden, Program Evaluation and Review Technique, hjälper oss att bedöma sannolikheten för att projekt och delar av projekt blir färdiga i tid.

Till skillnad från CPM förutsätter inte PERT att aktivitetstiderna är fastställda med säkerhet.

PERT tillåter dig att ange ett intervall där tre tidpunkter används, mest trolig tid (m), optimistisk tid (a) och pessimistisk tid (b). Personer med god erfarenhet av liknande projekt är den bästa källan för rimliga tidsuppskattningar.
$t = (a + 4m + b)/6$
$v = [(b-a)/6]^2$

t = medelvärde v = varians m = mest trolig tid a = optimistisk tid b = pessimistisk tid

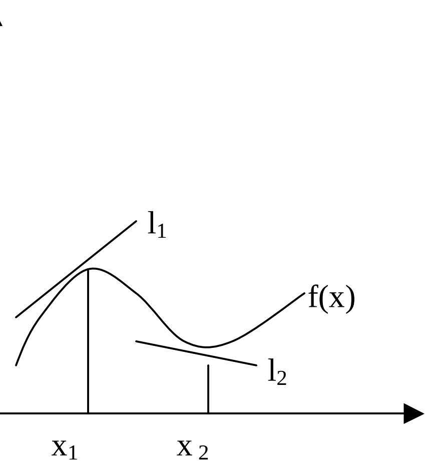
Derivata och Integraler

En funktions derivata betecknar lutningen på kurvan i en viss punkt. Derivatan av en funktion f(x) betecknas f´(x).

Integral betyder summa. Integralen av en funktion, f(x), mellan två värden, gränsen för integralen, är lika med summan av alla värden för funktionen.

f(x) är en matematisk funktion.
I punkterna x_1 och x_2 är tangenterna till funktionerna utritade, l_1 respektive l_2.
Derivatan till funktionen f(x), som är f´(x), anger lutningen på linjerna. Derivatan i punkten x_1 är således positiv och derivatan i punkt två är negativ.
Arean som täcks in av x-axeln, kurvan f(x), punkten x_1 samt punkten x_2 beräknas som integralen:

$$\int_{x_1}^{x_2} f(x)\,dx$$



Källor:
Proactive Management of Uncertainty using Successive Principle, Steen Lichtenberg, 2000
Ekonomi & Kalkyler, Akelius, 1993
Punktskattningsmetoden, G. Sällfors, 1990
Wikepedia
samt ett urplock från Internet t.ex. www.susning.nu