Distributii si numere aleatoare

Variabile aleatoare

Variabilele aleatoare pot fi considerate ca fiind funcții de la un domeniu oarecare către mulțimea numerelor reale. Denumirea de "aleatoare" se datorează faptului că valoarea funcției este cunoscută doar DUPĂ realizarea unui experiment. Probabilitatea ca o variabilă aleatoare X să ia o valoare mai mică decât un număr real x este funcția de distribuție cumulată. Aceasta se scrie F(x)=P(x<=X), pentru orice x real. Proprietățile acestei funcții sunt:

- dacă x<=y atunci F(x)<=F(y).
- F(-∞)=0, F(+∞)=1.
- F(x) este strict continuă.

Dacă a şi b sunt două numere reale a.î. a<b, atunci P(a<X<=b) probabilitatea ca X să fie în intervalul (a, b] este dată de diferența F(b) - F(a).

Variabilele aleatoare pot fi continue sau discrete. Corespunzător, funcția de distribuție cumulată este continuă sau discretă.

În cazul variabilelor continue se defineşte funcția de densitate de probabilitate f(x) în felul următor: f(x) = dF(x)/dx. Astfel se ajunge la concluzia că probabilitatea ca o variabilă aleatoare să fie într-un interval este integrala funcției de densitate de probabilitate pe acel interval.

Pentru variabilele discrete se defineşte similar funcția de probabilitate de masă.

Distribuția uniformă

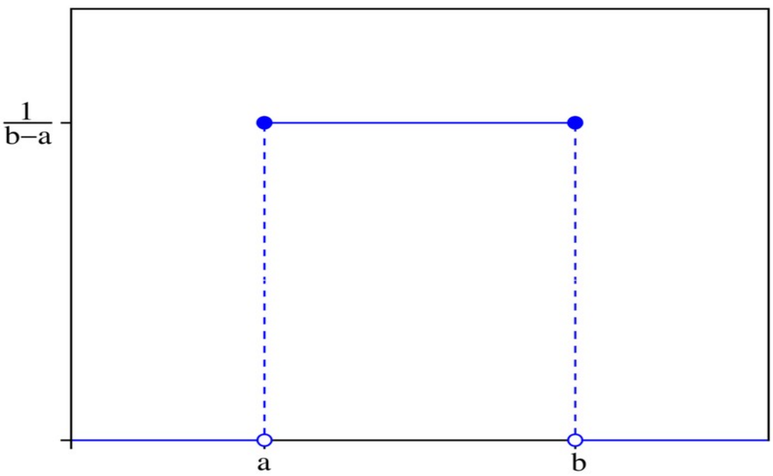
O variabilă aleatoare are o distribuție uniformă dacă poate lua echiprobabil orice valoare într-un interval (variabila este uniform distribuită în interval).

Daca a şi b sunt capetele intervalului, atunci densitatea de probabilitate va fi 1/(b-a) pe intervalul (a, b) şi 0 în rest.

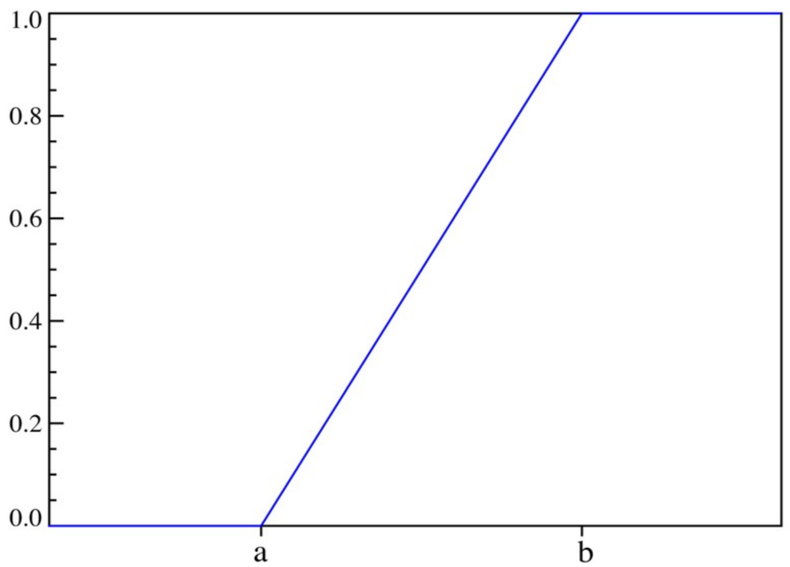
În cazul discret, în care variabila poate lua n valori, funcția de probabilitate este 1/n pe intervalul (a, b) şi 0 în rest.

Valoarea medie a unei astfel de variabile este (a+b)/2.

Densitatea de probabilitate:



Funcția de distribuție cumulată:

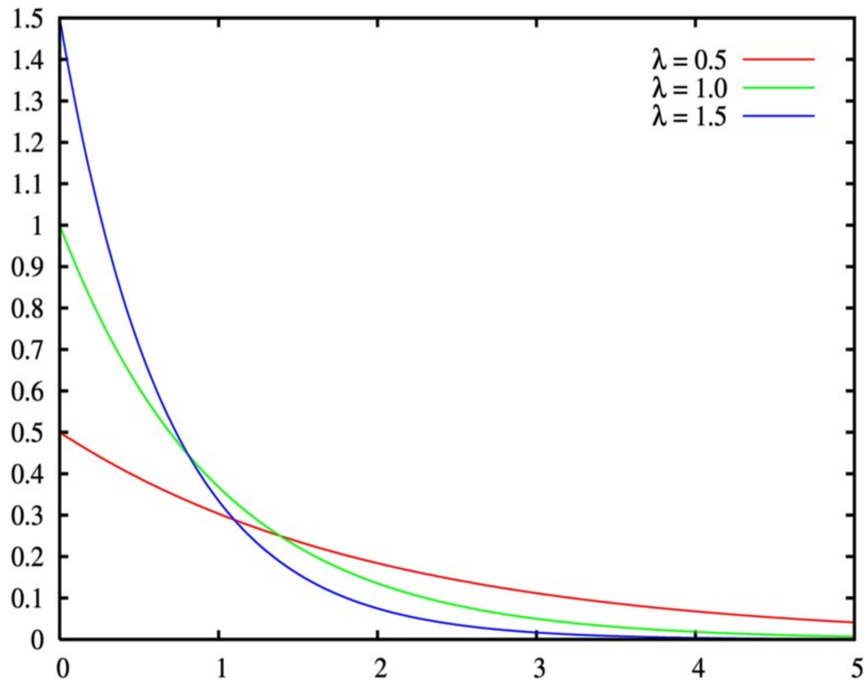


Distribuția exponențială

O variabilă continuă X are distribuție exponențială de parametru λ>0 dacă are funcția de probabilitate de forma F(x) = 1 - e ^(-λ*x), x>=0. Prin derivare se obține densitatea de probabilitate f(x) = λ*e^(-λ*x), x>=0. Se calculează valoarea medie a acestei variabile ca fiind 1/λ.

Cea mai importantă proprietate a acestei distribuții este lipsa de memorie. Daca interpretăm x ca fiind timpul, atunci faptul că activitatea care respectă aceasta distribuție a fost în desfăşurare la un moment dat nu influențează viitorul activității. Altfel spus nu contează că activitatea a început mai demult, ea se comportă ca şi cum ar începe ACUM. Aceasta este proprietatea Markov.

Densitatea de probabilitate:



Distribuția geometrică

Este echivalentul discret al distribuției exponențiale. Avem f(x) = P(x=X) = p*q^(x-1) unde p+q=1 şi x=1, 2, ... Variabila discretă poate fi considerată ca index într-o listă posibil infinită de experimente independente, unde p este probabilitatea de succes, iar q=1-p este probabilitatea de eşec. Funcția de probabilitate este F(X) = 1 - q^x. Media acestei ditribuții este 1/p.

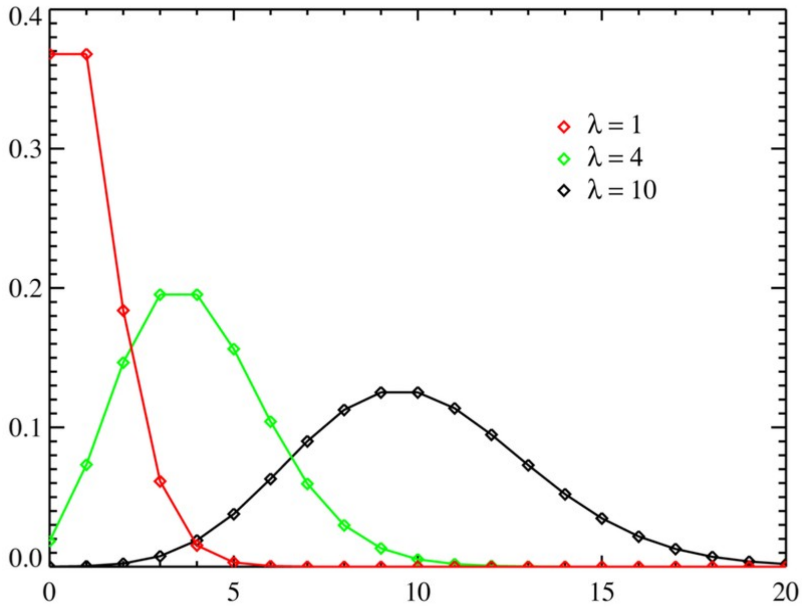
Distribuția binomială

O variabila X are o distribuție binomială de ordin n dacă ia valorile 0, 1, ..., n cu probabilitatea P(x=k) = (n/k)*(p^k)*(q^(n-k)), unde p+q=1. Următoarele două distribuții derivă din aceasta.

Distribuția Poisson

O variabilă care poate lua valorile 0, 1, 2, ..., are o distribuție Poisson de parametru λ dacă P(X=k) = (λ^k) * (e^(-λ)) / k!, k=0, 1, ... Media acestei distribuții este λ şi ea reprezintă de cele mai multe ori în simulări numărul mediu de sosiri în unitatea de timp.

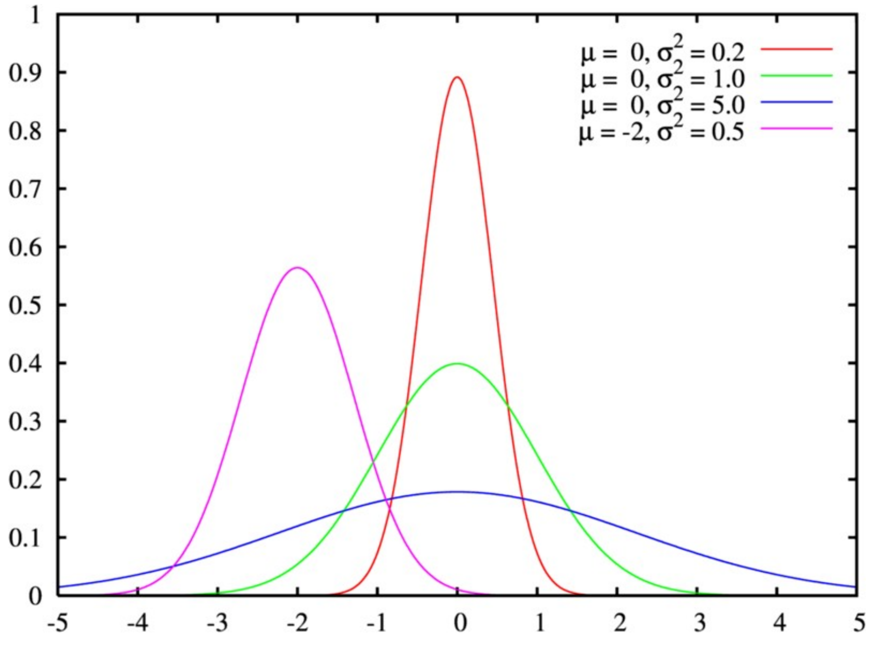
Densitatea de probabilitate:



Distribuția normala (Gauss)

Provine din distribuția binomială pentru n mare, când densitatea de probabilitate poate fi aproximată prin f(x) = K * e^(-(x-m)*(x-m)/(2*S*S)), unde K = 1/(S*K') si K'*K'=2*π. Media acestei distribuții este m, iar abaterea medie patratică S.

Densitatea de probabilitate:



Numere aleatoare

Motto: Un generator de numere aleatoare nu se alege aleator. (D. Knuth)

Cel mai important lucru de care trebuie ținut cont când se vorbeşte de generatoare de numere aleatoare este faptul că **numerele generate NU sunt aleatoare!!!** Deoarece se folosesc algoritmi determinişti, este mai corect sa vorbim de numere pseudo-aleatoare.

Folosirea unor algoritmi determinişti are şi o latură pozitivă, deoarece se permite reluarea exact a aceluiaşi experiment de mai multe ori. Însă de obicei valorile generate trebuie să fie diferite de la un experiment la altul. Pentru a rezolva acest paradox, majoritatea generatoarelor au nevoie de o valoare de start, numită "seed". Această valoare inițializează generatorul de numere aleatoare pentru a furniza un anumit şir de valori. Pentru a obține altă secvență de valori se foloseşte alt seed; folosirea aceluiaşi seed asigură reluarea aceleiaşi secvențe de numere aleatoare.

Un generator bun trebuie testat riguros, dar chiar şi în aceste condiții se pot găsi aplicații pentru care să nu fie utilizabil. De aceea, pentru a verifica consistența rezultatelor este bine să se foloseasca cel puțin două generatoare.

Pentru o folosire eficientă a unui generator, trebuie respectate câteva principii:

- Se inițializează totdeauna generatorul înaintea folosirii.
- Se folosesc seed-uri cât mai aleatoare (de exemplu 2731774 şi 10293082 par mai sigure decât 1 sau 4096 sau orice putere a lui 2).
- Două seed-uri consecutive pot produce secvențe corelate, deci nu se va folosi ca seed identificatorul unui proces dintr-un grup de procese.
- Generatorul de numere aleatoare implicit trebuie privit cu rezervă.
- Se folosesc cel puțin două generatoare, care să se verifice reciproc.

Tipuri de generatoare

Generatorul liniar congruențial (Linear Congruent Generator – LCM)

Formula matematică a sa este: X(n) = (a * X(n-1) + c) mod m, unde X(0) este seed, a este multiplicatorul, c este constanta. Daca m se alege bine, acest generator are multiple avantaje: este stabil, trece cu bine numeroase teste statistice, se calculează rapid, cere memorie puțină, teoria care sta la baza este bine înțeleasă.

Generator bazat pe deplasare (Shift Register Generator)

Are formula X(n) = X(n-1)T, unde X e un vector binar de m biți, iar T o matrice mXm de valori binare, aleasă astfel încât să corespundă unei shift-ări a biților în X. Dacă Y = X(n-1) xor (X(n-1) >> s), atunci X(n) = Y xor (Y << (m - s)). Avantaje: uşurință de implementare, viteză foarte mare de calcul, portabilitate, lucrul cu valori pe numar mare de biți, memorie puțină.

Generator Fibonnacci cu întârziere

Se numeşte aşa deoarece numărul aleator la pasul i se obține folosind două valori anterioare, oarecum asemănător şirului lui Fibonnacci. Formula este:

X(i) = (X(i-q) - X(i-p)) mod M

unde M=1 dacă se doresc numere reale în intervalul [0,1). Pentru a obține o perioada maximă de repetiție (2^p - 1)*2^(n-1) pentru întregi modulo 2^n sau pentru reali cu n biți la mantisa se folosesc următoarele valori:

P................................Q
9689........................471
4423.......................1393
2281........................715
1279........................418
607..........................273
521..........................168
127............................63

Avantaje: este rapid, elimină dezavantajele LCM, are o perioadă foarte mare. Este însă greu de găsit un seed potrivit şi cere multă memorie.

Generatorul Middle Square

Este un algoritm vechi, dar interesant. Metoda este următoarea:

Fie X(n-1) un număr de patru cifre ABCD. Se calculează numărul EFGHIJKL = (ABCD)*(ABCD).

Se alege X(n) ca fiind numărul de patru cifre GHIJ.

Algoritmul e rapid, cere puțină memorie, dar poate să se "agațe" la un moment dat, dacă se obține valoarea 0.

Pentru a obține un caracter şi mai aleator se pot folosi combinații de generatoare.