Variables aleatòries

Ana Escudero
Alícia Miralles
Alícia Vila

PID_00193846

Universitat Oberta de Catalunya

Índex

Introducció

Molt sovint és necessari relacionar el resultat d'una experiència amb un nombre. Imagineu, per exemple, que volem avaluar el senyal de sortida d'un circuit electrònic o saber quin és el temps de servei en què es processen peticions d'usuari que arriben a un servidor. Una primera aproximació a aquests problemes seria considerar que els valors que estem buscant són deterministes i que, per tant, es poden definir perfectament amb uns paràmetres que ens permeten obtenir valors exactes al llarg del temps. En la pràctica, però, sabem que hi ha molts factors que fan que la resposta dels sistemes de telecomunicació tingui una certa variabilitat. Així, per exemple, en el cas del circuit electrònic que acabem d'esmentar hauríem de tenir en compte la presència de soroll i altres interferències que fan que el senyal de sortida no sigui exactament l'esperat. També hem de tenir en compte que els dispositius electrònics no són ideals i que poden introduir errors. Les variables aleatòries ens permeten tenir en compte aquesta variabilitat i modelitzar els diferents resultats que obtenim en cada experiència, de manera que podem preveure quin serà el comportament del nostre sistema amb una certa probabilitat.

En l'apartat 1 d'aquest mòdul introduïm formalment el concepte de variable aleatòria. Veurem que hi ha dos tipus bàsics de variable aleatòria: la variable aleatòria discreta (apartat 2), que ens dóna un conjunt numerable de resultats possibles, i la variable aleatòria contínua (apartat 3), que ens dóna com a resultat qualsevol nombre d'un interval definit dins dels nombres reals. Veurem com les podem estudiar i mostrarem els casos de variables aleatòries que apareixen més habitualment. Treballarem, en particular, amb les distribucions més importants que estan relacionades amb les telecomunicacions. Definirem els conceptes de valor mitjà, $E(X)$ i variància, $Var(X)$, d'una variable aleatòria X. En l'apartat 4 veurem el teorema central del límit, que ens permet relacionar variables aleatòries contínues i discretes.

Objectius

Els objectius d'aquest mòdul són el següents:

1. Entendre què és una variable aleatòria i diferenciar-ne dos tipus: les discretes i les contínues.
2. Estudiar quatre tipus de variables aleatòries discretes: la distribució de Bernoulli, la distribució binomial, la distribució geomètrica i la distribució de Poisson.
3. Estudiar tres tipus de variables aleatòries contínues: la distribució uniforme, la distribució exponencial i la distribució normal o de Gauss.
4. Entendre els conceptes de valor mitjà i variància i caracteritzar les distribucions aleatòries amb aquests dos paràmetres.
5. Entendre els conceptes de funció de distribució i funció de densitat i caracteritzar les distribucions aleatòries amb aquestes funcions.
6. Saber triar el tipus de distribució aleatòria més adient per a modelitzar un fenomen determinat.
7. Estudiar el teorema central del límit i les seves aplicacions.

1. Concepte de variable aleatòria

A partir d'una experiència aleatòria es pot definir l'espai mostral, Ω, (omega), com el conjunt de tots els resultats possibles associats a aquesta experiència. Una variable aleatòria, X, assigna un nombre a cada un d'aquests resultats. Vegem-ne alguns exemples.

Vegeu també

L'espai mostral es defineix al mòdul "Introducció a la probabilitat".

Exemple 1.1

Considerem l'experiència de llençar una moneda. L'espai mostral és $\Omega = \{\text{cara,creu}\}$. Observeu que l'espai mostral Ω inclou tots els resultats possibles d'aquest experiment. Podem assignar a cada un d'aquests resultats els valors 0 o 1, segons el resultat de l'experiència sigui cara o creu. Escrivim, doncs, que $X(\text{cara}) = 0$ i $X(\text{creu}) = 1$. La variable X pot prendre els valors $\{0,1\}$.

Observació

La variable X de l'exemple 1.1 pren els valors $\{0,1\}$. És una variable aleatòria discreta.
La variable X de l'exemple 1.2 pren els valors a [0,2]. És una variable aleatòria contínua.

Exemple 1.2

Suposem que un aparell elèctric emet un senyal aleatori cada segon. Aquest senyal aleatori s'expressa en mil·livolts (mV) i pren valors dins de l'interval [0,2]. En aquest cas l'espai mostral està format per valors numèrics. Podem definir la variable aleatòria com l'aplicació identitat. A cada resultat de l'experiència li assigna el mateix valor. La variable X pot prendre un valor qualsevol de l'interval [0,2].

Definició 1.1. Una **variable aleatòria**, X, és una funció que assigna un nombre a cada element de l'espai mostral.

En els exemples 1.1 i 1.2 veiem la diferència entre una variable aleatòria discreta i una de contínua. En el primer cas, tenim un nombre determinat de resultats possibles, podem obtenir o bé cara o bé creu. En el segon cas el nostre aparell elèctric pot emetre un valor qualsevol dins de l'interval [0,2].

Conjunt discret

Un conjunt discret és aquell que està format per un nombre finit d'elements o bé per un nombre infinit d'elements que són numerables (és a dir, que es poden ordenar de manera que hi ha un primer element, un segon element, etc.). Per exemple, els conjunts de nombres naturals i enters ($\mathbb{N}$ i $\mathbb{Z}$) són discrets. El conjunt de nombres reals, $\mathbb{R}$, no és un conjunt discret.

Definició 1.2. Una **variable aleatòria discreta** (v.a.d.) pren valors d'un conjunt finit $\{a_1,a_1,\cdots,a_n\}$ o bé numerable infinit $\{a_1,a_2,\cdots\}$.

Definició 1.3. Una **variable aleatòria contínua** (v.a.c.) pot prendre valors en conjunts no numerables, com per exemple en un interval de $\mathbb{R}$ o en tot $\mathbb{R}$.

Això fa que el tractament matemàtic de les variables aleatòries discretes i contínues sigui molt diferent.

2. Variable aleatòria discreta

En aquest apartat definirem les variables aleatòries discretes, veurem les distribucions més importants i calcularem per a cadascuna dos paràmetres: el valor mitjà i la variància.

En l'apartat 1 hem vist que una variable aleatòria, X, ens dóna un valor numèric per al resultat d'una experiència. Aquest valor numèric es troba dins l'espai mostral Ω. Una variable aleatòria és discreta quan l'espai mostral Ω està format per un conjunt finit o un conjunt numerable infinit de valors. Per tant, X només pot prendre un valor especificat dins de l'espai mostral.

De manera natural la probabilitat que tenim definida en l'espai Ω es trasllada als valors que pren X.

En l'exemple 1.1, en què l'experiment és llençar una moneda a l'aire, escrivim $P(X=0) = P(\text{cara}) = \frac{1}{2}$ i $P(X=1) = P(\text{creu}) = \frac{1}{2}$.

Com que el resultat de la variable aleatòria X varia amb cada repetició de l'experiment, no podem definir el valor de X, però sí que podem descriure la probabilitat per a cadascun dels resultats possibles de X, és a dir, la probabilitat que X prengui un valor determinat. Això es descriu com $P(X = a_i)$, en què a_i és un valor possible de X.

Observació

Escrivim $X = 0$ quan volem indicar que X pren el valor 0 i $X = 1$ quan X pren el valor 1. En l'exemple 1.1 associem el valor 0 a treure cara i el valor 1 a treure creu:

$$P(X=0) = \frac{1}{2}$$

$$P(X=1) = \frac{1}{2}$$

Definició 2.1. S'anomena **funció de probabilitat** el conjunt de valors $P(X=a_i)$. Aquesta funció assigna una probabilitat a cada valor possible de X.

Per exemple, en el cas de l'experiència de llançar una moneda a l'aire, hem definit $\Omega = \{cara, creu\}$ i hem assignat els valors de 0 i 1 a aquest espai mostral. En aquest cas, la funció de probabilitat ens diu que la probabilitat que surti cara, $P(X = 0)$, és igual a $\frac{1}{2}$, i la probabilitat que surti creu, $P(X = 1)$, és igual a $\frac{1}{2}$.

La funció de probabilitat d'una variable aleatòria acompleix les propietats següents:

- La probabilitat $P(X=a_i)$ és un valor que està sempre entre 0 i 1, és a dir, $0 \leq P(X=a_i) \leq 1$.
- La suma de totes les probabilitats ha de ser 1, ja que els esdeveniments $X = a_i, \forall i$ formen una partició de l'espai mostral. És a dir, $\sum_i P(X=a_i) = 1$.

Definició 2.2. Diem que una **variable aleatòria discreta** que pren els valors $\{a_1,a_2,\cdots,a_n\}$, és **uniforme**, quan $P(X=a_i) = \frac{1}{n} \quad \forall i \in \{1,2,\cdots n\}$. És a dir, tots els elements a_i tenen la mateixa probabilitat assignada

Llançar una moneda a l'aire o un dau són exemples de variables aleatòries discretes uniformes. En el cas de la moneda la probabilitat d'obtenir cara o creu és $\frac{1}{2}$. En el cas del dau, la probabilitat d'obtenir qualsevol dels resultats és $\frac{1}{6}$.

2.1. Variables aleatòries discretes més importants

2.1.1. Variable aleatòria de Bernoulli: *B*(*p*)

Aquest és el tipus més senzill de variable aleatòria discreta i s'utilitza per a representar experiències en què només podem tenir dos resultats possibles.

Partim d'una experiència aleatòria i distingim entre els resultats $\Omega = \{A,A^c\}$. El resultat A s'anomena *èxit* i definim $X(A) = 1$. El resultat A^c s'anomena *no-èxit* i definim $X(A^c) = 0$. La variable aleatòria pren només els dos valors $\{0,1\}$. Només cal donar la probabilitat assignada a un d'aquests valors i la distribució queda definida completament. Si $P(A) = p$ llavors $P(X=1) = p$ i $P(X=0) = 1-p$.

A i A^c

Els conjunts A i A^c són complementaris. Recordeu el que havíem vist en la proposició 2.1 del mòdul "Introducció a la probabilitat": si $P(A) = p$, la probabilitat del conjunt complementari, $P(A^c)$, és $1-p$.

Diem que X és una **variable aleatòria de Bernoulli** amb probabilitat d'èxit p quan aquesta variable pot prendre els valors $X = 1$ (èxit) amb probabilitat p i $X = 0$ (no-èxit) amb probabilitat $(1-p)$. S'escriu: $X \sim B(p)$, en què p indica la probabilitat d'èxit i $p \in [0,1]$. Direm que s'ha produït èxit quan el resultat obtingut estigui dins del conjunt A, és a dir, si passa A, i no-èxit quan el resultat obtingut estigui en el conjunt A^c (complementari de A), és a dir, si no succeeix A.

La variable de l'exemple 1.1 d'aquest mòdul segueix una distribució $B(\frac{1}{2})$. Tornem al cas de la moneda. Si definim èxit, A, que surti cara, i el no-èxit, A^c, que surti creu, la nostra variable aleatòria, X, segueix una distribució de Bernoulli $B(\frac{1}{2})$. Noteu que també podríem haver definit A com a creu i A^c com a cara.

Exemple 2.1

En comunicacions binàries X pot indicar l'error en la transmissió d'un bit. L'espai mostral de l'experiència és determinat per $\Omega = \{\text{error},\text{no error}\}$ i la variable aleatòria pren els valors $X = 1$ si hi ha error i $X = 0$ si no n'hi ha. $P(X=1) = P(\text{error})$ i $P(X=0) = P(\text{no error})$. Fixeu-vos que en aquest exemple hem definit èxit, A, com la presència d'error en la transmissió.

Un altre exemple similar d'aplicació de la distribució de Bernoulli en les telecomunicacions es dóna en els sistemes de radar. Podem definir èxit, A (és a dir $X = 1$), quan el radar detecta la presència d'un objecte i A^c ($X = 0$) quan el radar no detecta cap objecte.

Exemple 2.2

En llançar un dau ens fixem en la màxima puntuació possible i definim $A = \{\textit{surt un } 6\}$, llavors $A^c = \{\textit{no surt un } 6\}$. Definim $X(A) = 1$ i $X(A^c) = 0$. La variable X segueix una distribució $B(\frac{1}{6})$, ja que la probabilitat del que hem definit com a èxit és $\frac{1}{6}$.

Observació

En l'exemple 2.2:

$$P(X=0) = \frac{5}{6}$$

$$P(X=1) = \frac{1}{6}$$

2.1.2. Variable aleatòria binomial: *Bin*(*n*,*p*)

La **distribució binomial**, $Bin(n,p)$, es dóna quan repetim n cops i de manera independent una experiència $B(p)$ de Bernoulli. A cada resultat, la variable aleatòria X li assigna el nombre d'èxits que han sortit. Així X pren els valors $\{0,1,2,\cdots,n\}$.

La distribució de probabilitats (probabilitat que té cada un dels valors que pren la variable X) és determinada per:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{amb} \quad k \in \{0,1,2,\cdots,n\}.$$

Diem que X és una variable aleatòria binomial i escrivim $X \sim Bin(n,p)$, en què n és el nombre de cops que repetim l'experiència de Bernoulli $B(p)$, de probabilitat d'èxit p.

Recordeu que $\binom{n}{k}$ (es llegeix *n* *sobre k*) es calcula com $\frac{n!}{k!(n-k)!}$.

En l'exemple 2.3 X segueix una distribució $Bin(10,0,1)$, en què 10 és el nombre de cops que es repeteix l'experiment i 0,1 la probabilitat d'èxit de cada experiment.

Exemple 2.3

Una persona, emissor, ha d'enviar un missatge de 10 elements, triats del conjunt $\{0,1\}$ i ordenats. Un missatge d'aquest tipus podria ser la paraula 0011111101 (formada amb 10 bits). Suposem que cada cop que la persona tria un bit per a formar la paraula, la probabilitat que sigui un 0 és 0,1 i, per tant, la que sigui un 1 és 0,9. En aquest exemple considerarem que es dóna la condició d'èxit, A, quan es transmet un zero i la condició de no-èxit, A^c, quan es transmet un u. Amb aquesta idea ens vénen al cap tota una sèrie de preguntes, com per exemple:

1) Quina és la probabilitat que l'emissor enviï exactament la paraula 0011111101?

2) Quina és la probabilitat que l'emissor enviï una paraula que tingui exactament tres zeros?

3) Quina és la probabilitat que l'emissor enviï una paraula que tingui exactament k zeros?

4) Quina és la probabilitat que l'emissor enviï una paraula que tingui com a màxim tres zeros?

Les qüestions anteriors les podem resoldre aplicant el que heu après en el tema anterior. S'obtenen els resultats següents:

1) $(0{,}1)^3(0{,}9)^7 = 0{,}00048$

La paraula per enviar té tres zeros i, per tant, la probabilitat que en 10 experiències obtinguem 3 èxits és: $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$ En aquest cas ens demanen una combinació concreta de totes les possibles que podrien incloure 3 zeros, i per tant, no tenim en compte el terme $\binom{n}{k}$ i $P(X=3) = 0{,}1^3 0{,}7^{10-3}$.

2) El resultat anterior ens dóna la probabilitat per a una paraula que té 3 zeros i 7 uns. Vam veure en l'apartat anterior que el nombre de paraules que es poden formar amb 3 zeros i 7 uns és $\binom{10}{3}$. Llavors la resposta és $\binom{10}{3}(0{,}1)^3(0{,}9)^7 = 0{,}0574$.

En aquest cas ens demanen la probabilitat d'una seqüència amb 3 zeros. Aquests zeros poden estar en qualsevol posició, i per tant, hem de considerar totes les combinacions possibles de paraules de 10 bits que poden contenir aquests zeros.

3) És clar que una paraula de mida 10 pot tenir entre 0 i 10 zeros. Així, $0 \leq k \leq 10$. Fent el mateix raonament que en l'apartat anterior tenim $\binom{10}{k}(0{,}1)^k(0{,}9)^{10-k}$. Observeu que k és el nombre de zeros que pot contenir la paraula, no el nombre d'experiències, que per a aquest cas és 10.

4) Ara hem de tenir en compte aquells casos en què el nombre de zeros sigui més petit o igual que 3. De l'expressió anterior sumem els casos en què k pren els valors 0, 1, 2 i 3. Obtenim:

$$\sum_{k=0}^{3}\binom{10}{k}(0{,}1)^k\cdot(0{,}9)^{10-k} = \binom{10}{0}(0{,}1)^0(0{,}9)^{10-0} + \binom{10}{1}(0{,}1)^1(0{,}9)^{10-1}$$

$$+\binom{10}{2}(0{,}1)^2(0{,}9)^{10-2} + \binom{10}{3}(0{,}1)^3(0{,}9)^{10-3} =$$

$$0{,}3487 + 0{,}3874 + 0{,}1937 + 0{,}0574 = 0{,}9872$$

Si a la variable X li assignem el nombre de zeros que té cada paraula, els apartats anteriors els podem escriure utilitzant la X:

2) $P(X=3) = \binom{10}{3}(0{,}1)^3(0{,}9)^7$

3) $P(X=n) = \binom{10}{n}(0{,}1)^n(0{,}9)^{10-n}$

4) $P(X \leq n) = \sum_{k=0}^{n}\binom{10}{k}(0{,}1)^k(0{,}9)^{10-k}$

Exemple 2.4

Enviem una paraula de n bits en què cada bit pot portar error o no, independentment dels altres. La variable $X \sim Bin(n,p)$ pren el valor del nombre de bits erronis que hi ha en la paraula, i per tant, els valors possibles són $\{0,1,2,\cdots,n\}$. p és la probabilitat que un bit sigui erroni*.

*** Recordeu que hem definit l'èxit *A* com el fet de transmetre un bit erroni.**

2.1.3. Variable aleatòria geomètrica: *Geom(p)*

La **distribució geomètrica**, $Geom(p)$, es dóna quan repetim, de manera independent, una experiència $B(p)$, fins a obtenir el primer èxit. X compta el nombre de cops que cal fer l'experiència per a obtenir el primer èxit. Per tant, X pren els valors $\{1,2,3,\cdots\}$. La seva distribució de probabilitats és determinada per:

$$P(X=k) = (1-p)^{k-1}p,$$

en què p és la probabilitat d'èxit i k el nombre d'intents que necessitem fins obtenir l'èxit. Diem que X és una variable aleatòria geomètrica amb probabilitat d'èxit p i escrivim $X \sim Geom(p)$.

En l'exemple que veurem a continuació, X segueix una distribució $Geom(0{,}2)$. Fixeu-vos que l'expressió de la distribució geomètrica és similar a l'expressió de la distribució binomial però sense el terme $\binom{n}{k}$, ja que en aquest cas estem fixant la seqüència de resultats com a A^c, A^c, A^c, ... A.

Exemple 2.5

Per a enviar missatges per Internet, els missatges es divideixen en paquets i després s'envien per la xarxa. Si la xarxa està congestionada els paquets es poden perdre. Suposem que en una xarxa molt congestionada la probabilitat de perdre un paquet és 0,8. Això significa que el paquet no es perd en una transmissió amb una probabilitat de 0,2. El paquet es transmet repetidament fins que el receptor el rep. Ens fem les preguntes següents:

1) Quina és la probabilitat que el paquet hagi de ser enviat almenys tres cops?

2) Quina és la probabilitat que hàgim d'enviar el paquet com a màxim 5 cops perquè el receptor el rebi?

Si la variable aleatòria X compta el nombre de vegades que cal enviar un paquet, pren valors del conjunt $\{1,2,3,\cdots\}$ (X és una variable discreta infinita). Podem escriure:

1) $P(X \geq 3) = 1 - P(X \leq 2) = 1 - (P(X{=}1) + P(X{=}2)) = 1 - (0{,}2 + 0{,}8 \cdot 0{,}2) = 0{,}64$

2) $P(X \leq 5) = \sum_{k=1}^{5} 0{,}2(0{,}8)^{k-1} = 0{,}2\left(1 + 0{,}8 + 0{,}8^2 + 0{,}8^3 + 0{,}8^4\right) = 0{,}67232$

2.1.4. Variable aleatòria de Poisson: *Poiss*(α)

La variable aleatòria de Poisson s'utilitza per a modelitzar alguns fenòmens com els següents:

- El nombre d'accidents en un encreuament donat i per a un interval de temps fixat.
- El nombre de trucades que arriben a una centraleta en un cert interval de temps.
- El nombre de peticions que arriben a un servidor en un cert interval de temps.
- El nombre d'electrons o forats que travessen una barrera de potencial.
- El nombre de defectes de fabricació d'un producte d'unes dimensions determinades.
- Teoria de cues en xarxes de comunicacions de veu i dades.

En aquest tipus de distribucions la variable aleatòria X pren els valors $\{0,1,2,3,\cdots\}$ i ens diu que el nombre mitjà d'èxits en un cert interval (temps, àrea, etc.) és α.

La distribució de probabilitats per a una distribució discreta de Poisson és:

$$P(X=k) = \frac{\alpha^k}{k!}e^{-\alpha}.$$

Diem que X és una variable aleatòria de Poisson de paràmetre α, i escrivim $X \sim Poiss(\alpha)$.

Exemple 2.6

Sabem que a un servidor arriben de mitjana 5 peticions per segon. Quina és la probabilitat que en un segon no hi arribi cap petició? Quina és la probabilitat que en un segon hi arribi una o més peticions?

La dada que ens dóna l'enunciat és precisament el paràmetre de la distribució de Poisson; així, $\alpha = 5$. Ara ja podem donar les respostes. En el primer cas hem de calcular la probabilitat que el nombre d'arribades, k, sigui zero; per tant:

$$P(X=0) = \frac{5^0}{0!}e^{-5} = e^{-5} = 0{,}0067.$$

En aquest segon cas hem de calcular la probabilitat que arribi una o més peticions al servidor:

$$P(X \geq 1) = 1 - P(X=0) = 1 - 0{,}0067 = 0{,}9933.$$

2.2. Paràmetres: valor mitjà i variància

Fins ara en aquest apartat hem definit què és una variable aleatòria discreta i hem vist algunes de les distribucions més importants: la distribució de Bernoulli, la binomial, la geomètrica i la de Poisson. En aquest subapartat veurem dos paràmetres molt utilitzats que ens permetran de manera molt global avaluar i comparar les diferents variables aleatòries. Aquests paràmetres són el **valor mitjà** i la **variància**.

Definició 2.3. Sigui X una variable aleatòria discreta que pren els valors següents: $\{a_1,a_2,\cdots,a_n\}$ (per al cas que X prengui infinits valors es fa de manera semblant, però en lloc d'una suma finita tenim una sèrie numèrica).

El **valor mitjà, esperança** o **moment d'ordre 1** de X:

$$E(X) = \sum_{i=1}^{n} a_i P(X=a_i)$$

El **moment d'ordre 2**:

$$E(X^2) = \sum_{i=1}^{n} a_i^2 P(X=a_i)$$

El **moment d'ordre** n:

$$E(X^n) = \sum_{i=1}^{n} a_i^n P(X=a_i)$$

La **variància**:

$$Var(X) = E((X-E(X))^2) = \sum_{i=1}^{n} (a_i - E(X))^2 P(X=a_i) = E(X^2) - E(X)^2$$

La **desviació típica**:

$$\sigma(X) = \sqrt{Var(X)}$$

Intuïtivament podem entendre l'esperança d'una variable aleatòria com el valor que esperem obtenir en fer un experiment, i la variància com la mesura en què ens podem allunyar d'aquest valor que esperem obtenir.

Observació

La desviació típica es representa amb la lletra σ, que es llegeix "sigma".

Variància i desviació típica

Noteu que la variància és una mitjana de les diferències al quadrat, mentre que la desviació típica és el mateix paràmetre però donat en les mateixes unitats que la variable aleatòria. La relació entre elles és: $\sigma(X) = \sqrt{Var(X)}$.

Exemple 2.7

Siguin X_1 i X_2 dues variables aleatòries equiprobables que prenen els valors {4,5,6} i {0,5,10}, respectivament. Aquests valors podrien ser les tres notes obtingudes en una determinada assignatura per dos alumnes diferents. Suposant que totes tres notes tenen el mateix pes, podem estimar l'esperança i variància de cada una de les variables:

$$E(X_1) = 4\frac{1}{3} + 5\frac{1}{3} + 6\frac{1}{3} = 5$$

$$E(X_2) = 0\frac{1}{3} + 5\frac{1}{3} + 10\frac{1}{3} = 5$$

$$Var(X_1) = (4-5)^2\frac{1}{3} + (5-5)^2\frac{1}{3} + (6-5)^2\frac{1}{3} = 0{,}66 \quad \text{i} \quad \sigma_{X_1} = 0{,}82$$

$$Var(X_2) = (0-5)^2\frac{1}{3} + (5-5)^2\frac{1}{3} + (10-5)^2\frac{1}{3} = 16.67 \quad \text{i} \quad \sigma_{X_2} = 4{,}08$$

El que podem dir és que els dos alumnes tenen la mateixa nota de mitjana $E(X_1) = E(X_2)$, però el segon alumne presenta més dispersió en les seves notes, ja que $\sigma_{X_2} > \sigma_{X_1}$ (és a dir, el segon alumne és menys regular).

Figura 1. Notes obtingudes pels dos estudiants

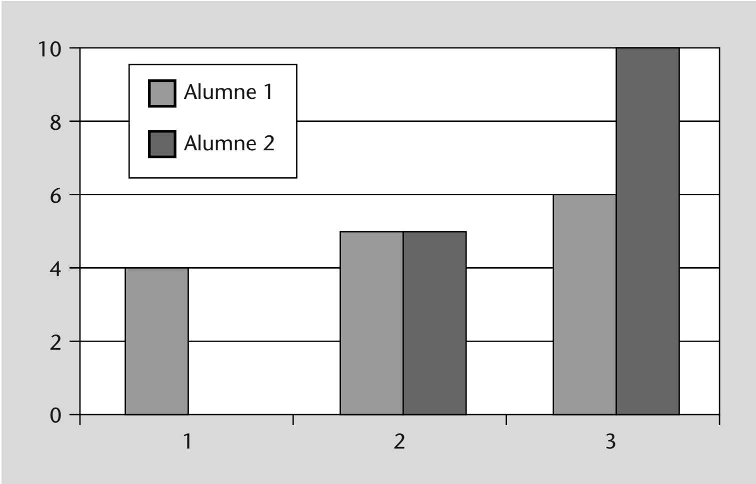


Figura 1

Distribució de les notes obtingudes pels dos estudiants. A l'eix horitzontal es representa el número de prova, i a l'eix vertical la nota obtinguda. Els dos estudiants obtenen la mateixa nota mitjana, però observeu que el primer és més regular (tendeix a treure notes en un interval més petit) que el segon, que obté resultats més dispersos.

Proposició 2.2. Per a cadascuna de les distribucions vistes, s'obtenen els valors de l'esperança i la variància de la taula següent.

Distribucions de variables aleatòries discretes

$X \sim$	k	$P(X=k)$	$E(X)$	$Var(X)$
$B(p)$	$\{0,1\}$	$P(X=1) = p$ $P(X=0) = 1-p$	p	$p(1-p)$
$Bin(n,p)$	$\{0,1,2,\ldots,n\}$	$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
$Geom(p)$	$\{1,2,3,\ldots\}$	$P(X=k) = (1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poiss(\alpha)$	$\{0,1,2,\ldots\}$	$P(X=k) = \frac{\alpha^k}{k!} e^{-\alpha}$	α	α

Comprovem els resultats de la taula anterior, que corresponen a $X \sim B(p)$:

- El valor mitjà, $E(X) = 1p + 0(1-p) = p$
- El moment d'ordre 2, $E(X^2) = 1^2 p + 0^2(1-p) = p$
- La variància, $Var(X) = p - p^2 = p(1-p)$

2.3. Funció de distribució

Una manera de donar el valor de les probabilitats acumulades és a partir de la funció de distribució.

Definició 2.4. La **funció de distribució** d'una variable aleatòria X es defineix com

$$F_X(x) = P(X \leq x), \quad \forall x \in \mathbb{R}. \tag{1}$$

És a dir, la funció de distribució ens dóna la probabilitat que la nostra variable aleatòria X prengui un valor igual o més petit que un valor x determinat. Vegem-ne un exemple.

Exemple 2.8

Per al cas de $X \sim B(p)$, la funció de distribució $F_X(x)$ presenta una discontinuïtat de salt en $X = 0$ i en $X = 1$:

$$F_X(x) = \begin{cases} 0 \quad \text{si} \quad x < 0 \\ 1 - p \quad \text{si} \quad 0 \leq x < 1 \\ 1 \quad \text{si} \quad 1 \leq x \end{cases}$$

Considerem aquí que zero és no-èxit.

Figura 2. Funció de distribució de $X \sim B(p)$

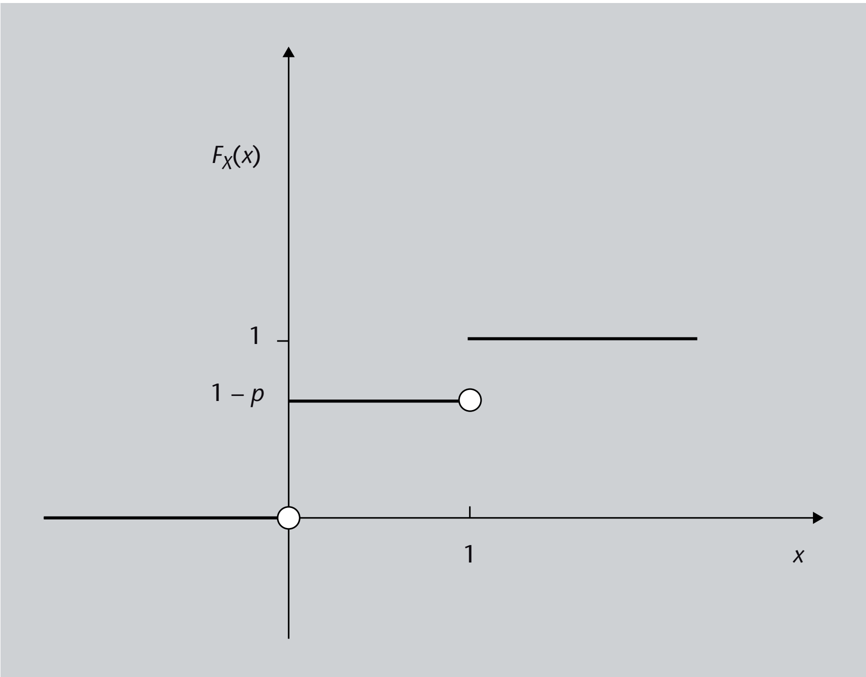


Figura 2

Recordeu del subapartat 2.1.1 d'aquest mòdul que l'espai mostral de la distribució de Bernoulli es definia com $\Omega = \{A^c, A\} = \{0,1\}$. En la figura s'ha definit $1 - p$ com la probabilitat de no-èxit, $P(X=0)$, i p com la probabilitat d'èxit $P(X=1)$.

Fixeu-vos en els intervals definits en la figura 2. El resultat de l'experiment aleatori ens ha de donar no-èxit o èxit, és a dir, 0 o 1. Per tant, la probabilitat d'obtenir un nombre més petit que 0 com a resultat és nul·la. En el segon interval de la gràfica tenim la probabilitat d'obtenir un zero (no-èxit) i aquesta probabilitat és $1 - p$. Per a la darrera part de la gràfica ($x \geq 1$) estem considerant la probabilitat acumulada d'obtenir o bé 0 o bé 1. Com que sabem segur que obtindrem un dels dos resultats, la funció de distribució val 1 a partir d'aquest punt. En els cas de variables aleatòries discretes, la funció de distribució és esglaonada. La funció experimenta un salt en cada nombre real que correspongui a un valor que pren X.

Funció de distribució

La funció de distribució ens dóna la probabilitat que una variable aleatòria, X, tingui un valor més petit o igual que una x determinada. Noteu que per aquesta raó també s'anomena **probabilitat acumulada**, és a dir, probabilitat de tots els valors fins a x.

Vegem ara un segon exemple: la funció de distribució per la variable aleatòria binomial.

Exemple 2.9

En el cas $X \sim Bin(4,1/2)$, $F_X(x)$ presenta una discontinuïtat de salt en els valors del conjunt $\{0,1,2,3,4\}$. Tenint en compte que $P(X{=}k) = \binom{4}{k}(\frac{1}{2})^k(\frac{1}{2})^{4-k}$ i que

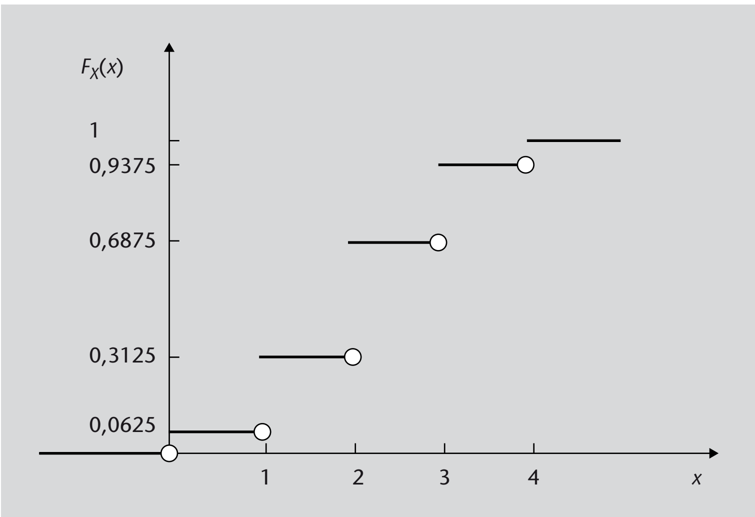
$$F_X(x) = P(X \leq k) = \sum_{i=0}^{k} \binom{4}{i}(\frac{1}{2})^i(\frac{1}{2})^{4-i} = \sum_{i=0}^{k} \binom{4}{i}(\frac{1}{2})^4,$$

obtenim els valors més significatius de la funció de distribució:

$$\begin{aligned}
F_X(0) &= 0{,}0625 \\
F_X(1) &= 0{,}0625 + 0{,}250 = 0{,}3125 \\
F_X(2) &= 0{,}0625 + 0{,}250 + 0{,}375 = 0{,}6875 \\
F_X(3) &= 0{,}0625 + 0{,}2500 + 0{,}3750 + 0{,}2500 = 0{,}9375 \\
F_X(4) &= 0{,}0625 + 0{,}2500 + 0{,}3750 + 0{,}2500 + 0{,}0625 = 1
\end{aligned}$$

$$F_X(x) = \begin{cases} 0 \quad \text{si} \quad x < 0 \\ 0{,}0625 \quad \text{si} \quad 0 \leq x < 1 \\ 0{,}3125 \quad \text{si} \quad 1 \leq x < 2 \\ 0{,}6875 \quad \text{si} \quad 2 \leq x < 3 \\ 0{,}9375 \quad \text{si} \quad 3 \leq x < 4 \\ 1 \quad \text{si} \quad 4 \leq x \end{cases}$$

Figura 3. Funció de distribució de $X \sim Bin(4, \frac{1}{2})$



Fixeu-vos en la figura. La probabilitat d'obtenir un nombre d'èxits més petit que zero és nul·la perquè si fem l'experiment de Bernoulli 4 vegades obtindrem o bé 0 èxits o bé 1, 2, 3 o 4, però en cap cas un valor negatiu. Per tant, $F(x < 0) = 0$.

La probabilitat d'obtenir 0 èxits és la següent:

$$P(X{=}0) = \binom{4}{0}(\frac{1}{2})^0(\frac{1}{2})^{4-0} = \frac{1}{16} = 0{,}0625. \quad (2)$$

Distribució binomial

Recordeu, com hem vist en el subapartat 2.1.2 d'aquest mòdul, que la distribució binomial consisteix a repetir n vegades un experiment de Bernoulli. En aquest cas, la nostra variable aleatòria comptabilitza el nombre d'èxits que obtenim en fer l'experiment n cops.

Figura 3

Noteu com la funció de distribució és un valor que va acumulant la probabilitat d'obtenir un cert nombre d'èxits igual o més petit que x.

Per tant, en $x = 0$ tenim un salt de la funció de distribució, $F(x)$, que passa de valer zero a tenir el valor de 0,0625.

Ara calculem la probabilitat de tenir un èxit en els quatre experiments. Aquesta probabilitat està determinada per l'expressió següent:

$$P(X=0) = \binom{4}{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{4-1} = \frac{1}{16} = 0{,}25. \qquad (3)$$

Així doncs, la funció de distribució, $F(x)$, en $x = 1$ és la probabilitat acumulada d'obtenir zero èxits o un èxit, és a dir, $F(1) = 0{,}0625 + 0{,}25 = 0{,}3125$.

Si fem els càlculs per a la resta de valors de x, en què x és el nombre d'èxits per als quatre experiments, obtenim els valors que es mostren a la figura 3.

La funció de distribució verifica les propietats següents:

1) $0 \leq F_X(x) \leq 1$

2) $\lim_{x\to\infty} F_X(x) = 1$

3) $\lim_{x\to-\infty} F_X(x) = 0$

4) $F_X(x)$ és creixent, és a dir, si $a < b$ llavors $F_X(a) \leq F_X(b)$

5) $P(a < X \leq b) = F_X(b) - F_X(a)$

Les quatre primeres propietats de la funció de distribució, $F(X)$, s'observen fàcilment en la figura 3. Els valors de les probabilitats acumulades els obtenim directament de la funció de distribució. En donem alguns valors com a exemple:

$$\begin{aligned}
P(X \leq 0{,}5) &= F_X(0{,}5) = 0{,}0625 \\
P(X \leq 1{,}7) &= F_X(1{,}7) = 0{,}3125 \\
P(X \leq 2{,}4) &= F_X(2{,}4) = 0{,}6875 \\
P(1{,}7 < X \leq 2{,}4) &= F_X(2{,}4) - F_X(1{,}7) = 0{,}375
\end{aligned}$$

3. Variable aleatòria contínua

En l'apartat 2 d'aquest mòdul hem vist què és una variable aleatòria discreta i hem estudiat quatre de les distribucions més utilitzades: la distribució de Bernoulli, la binomial, la geomètrica i la de Poisson. Hem vist també els paràmetres valor mitjà i variància i finalment hem vist què és la funció de distribució d'una variable aleatòria discreta. L'estructura d'aquest apartat és molt similar a la de l'anterior. Aquí veurem els conceptes anteriors aplicats al cas de les variables aleatòries contínues.

En l'exemple 1.2 de l'apartat 1 quan hem definit què és una variable aleatòria, i hem vist que la variable aleatòria contínua X pot prendre un valor qualsevol de l'interval [0,2]. En aquest cas, si suposem que cap dels valors dins [0,2] té preferència, podríem trobar els resultats següents de manera intuïtiva:

1) Quina és la probabilitat que el senyal emès es trobi entre 0 i 1 mV? És a dir, $P(0 \leq X \leq 1)$? Tot ens fa pensar que és $\frac{1}{2}$, ja que estem jugant amb la meitat de possibilitats.

2) Quina és la probabilitat que el senyal emès es trobi entre 3 i 4 mV? És a dir, $P(3 \leq X \leq 4)$? Com que sabem que això no succeirà mai perquè el generador només ens dóna un senyal en l'interval [0,2], diem que és 0.

3) Quina és la probabilitat que el senyal emès sigui exactament d'1 mV? Aquest cas ens caracteritza les distribucions de variables aleatòries contínues. Diem que $P(X=1) = 0$. En una distribució de variable aleatòria contínua, la probabilitat en qualsevol punt x, és zero.

Observació

Noteu que per a les variables aleatòries contínues la probabilitat d'obtenir exactament un valor determinat és zero, ja que en aquest cas la variable aleatòria, en ser contínua, pot prendre infinits valors. O el que és el mateix: si X és una variable aleatòria contínua, llavors $\forall x \in \mathbb{R}$ se satisfà que $P(X=x) = 0$.

3.1. Funció de distribució i funció de densitat

La funció de distribució es defineix de la mateixa manera que per a una variable aleatòria discreta, tal com ho havíem definit en l'apartat 2.

Definició 3.1. La **funció de distribució d'una variable aleatòria** X es defineix com

$$F_X(x) = P(X \leq x), \quad \forall x \in \mathbb{R}. \tag{4}$$

Observeu les similituds en la definició de **funció de distribució** per a variables aleatòries discretes i contínues.

La funció de distribució $F_X(x)$ verifica les propietats següents:

1) $0 \leq F_X(x) \leq 1$

2) $\lim_{x \to \infty} F_X(x) = 1$

3) $\lim_{x \to -\infty} F_X(x) = 0$

4) $F_X(x)$ és creixent, és a dir, si $a < b$ llavors $F_X(a) \leq F_X(b)$

5) $P(a < X \leq b) = F_X(b) - F_X(a)$

Fixeu-vos que són exactament les mateixes propietats que havíem vist per al cas de les variables aleatòries discretes, però aplicades, en aquest cas, a les variables contínues.

En el cas de les variables aleatòries contínues, $F_X(x)$ és contínua en tot $\mathbb{R}$ i derivable en $\mathbb{R}$ (tret potser d'un nombre finit de punts). Això ens permet definir la funció de densitat.

Funcions contínues i derivables

Diem que una funció $f(x)$ és **contínua** si a mesura que ens anem desplaçant per l'eix de la variable independent, x, no es produeixen salts o canvis bruscos. Intuïtivament són funcions que podríem dibuixar sobre un paper sense aixecar el llapis. Diem que una funció és **derivable** o **diferenciable** en un punt si existeix la seva derivada en aquell punt. Recordeu que totes les funcions derivables són contínues.

Definició 3.2. Si X és una variable aleatòria contínua amb funció de distribució $F_X(x)$, la **funció de densitat** es defineix com

$$f_X(x) = \frac{dF_X(x)}{dx}, \quad \forall x \in \mathbb{R}. \tag{5}$$

És a dir, la funció de densitat, $f_X(x)$, és la derivada en funció de x (variable independent que ens diu quins valors pot prendre la nostra variable aleatòria X) de la funció de distribució, $F_X(x)$. Recordeu que la funció de distribució ens donava la probabilitat acumulada a mesura que anàvem considerant els valors possibles de la variable aleatòria. La funció de densitat, $f_X(x)$, ens dóna una idea de com varia la funció de distribució d'una variable aleatòria. I inversament, la funció de distribució és la integral sobre x de la funció de distribució.

Vegem-ne un exemple. Com podeu veure en la figura 4, l'àrea per sota de la corba de $f(x)$ correspon a un punt de $F(x)$.

Figura 4. Funció de densitat $f(x)$ i funció de distribució $F(x)$

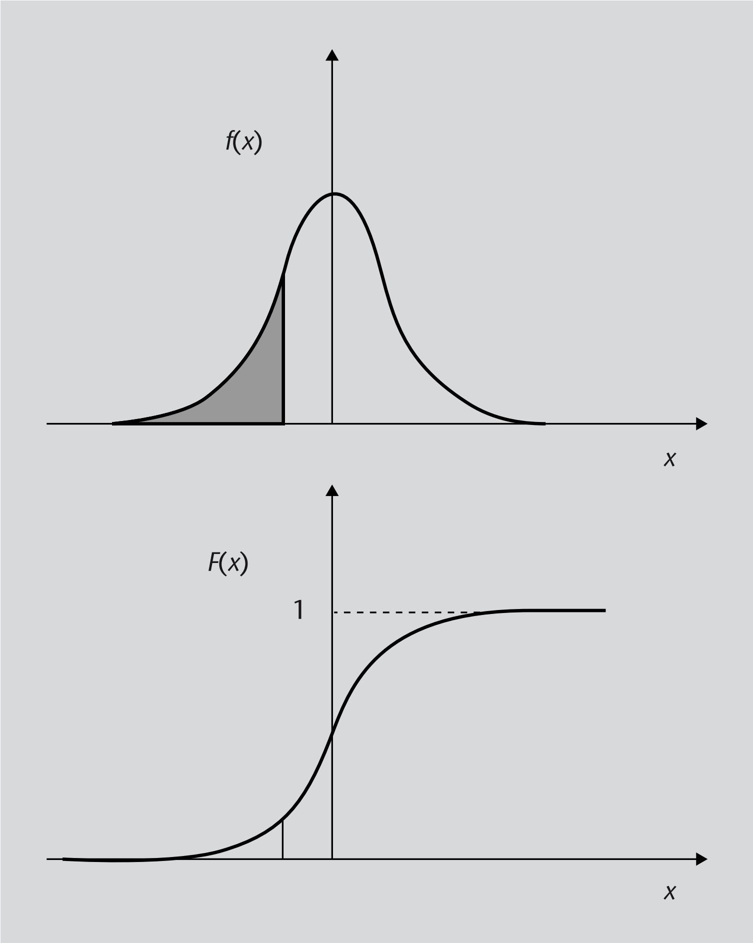


Figura 4

En la figura podeu observar la relació entre la funció de distribució, $F_X(x)$, i la funció de densitat, $f_X(x)$.

A continuació vegem quines relacions hi ha entre la **funció de distribució**, $F_X(x)$, i la **funció de densitat**, $f_X(x)$, d'una variable aleatòria contínua.

Proposició 3.1. Si X és una variable aleatòria contínua amb **funció de distribució** $F_X(x)$ i **funció de densitat** $f_X(x)$, llavors:

1) $f_X(x) \geq 0$ (això és clar si observem l'equació (5) de la definició 3.2 i pensem que $F_X(x)$ és creixent) ja que una funció creixent sempre té pendent positiu.

2) $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,dx$ (és a dir, la probabilitat entre dos punts a i b l'obtenim integrant la funció de densitat entre aquests dos punts, és a dir, l'àrea per sota de la corba de la funció de densitat).

3) $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ (és a dir, la probabilitat acumulada també la podem pensar com una àrea per sota de la funció de densitat $f_X(x)$).

4) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$. És a dir, l'àrea total per sota de la corba $f_X(x)$ és 1.

Observació

Si $X \sim$ és una variable aleatòria contínua, llavors $P(a \leq X \leq b) = P(a < X < b)$.

Arribats a aquest punt ens podríem preguntar: per què és necessari definir la funció de densitat d'una variable aleatòria contínua si ja tenim el concepte de funció de distribució, com havíem vist amb les variables aleatòries discretes? La resposta és que per al cas de les variables contínues no sempre és possible expressar la funció de distribució d'una manera senzilla i tancada. A més, moltes de les propietats d'aquestes variables es veuen més clarament quan utilitzem la funció de densitat en comptes de la funció de distribució.

3.2. Variables aleatòries contínues més importants

Com acabem de veure en el subapartat anterior, quan treballem amb variables aleatòries contínues, aquestes es poden caracteritzar amb la seva funció de densitat. Vegem-ne a continuació les més importants.

3.2.1. Variable aleatòria uniforme: $X \sim U(a,b)$

La variable X pot prendre un valor qualsevol de l'interval (a,b) i de manera uniforme. En aquest cas diem que X segueix una **distribució uniforme** en (a,b). Això ho indiquem amb la funció de densitat:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si} \quad a < x < b \\ 0 & \text{altrament} \end{cases}$$

La funció de distribució serà, doncs:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} 0 & \text{si} \quad x < a \\ \frac{1}{b-a}(x-a) & \text{si} \quad a \leq x < b \\ 1 & \text{si} \quad b \leq x \end{cases}$$

En la figura 5 podeu veure un exemple de variable uniforme i les seves funcions de densitat i de distribució. Com podeu veure, l'àrea indicada per sota de $f(x)$ correspon a un punt de $F(x)$.

Figura 5. Funció de densitat $f(x)$ i funció de distribució $F(x)$ de la variable aleatòria $X \sim U(a,b)$

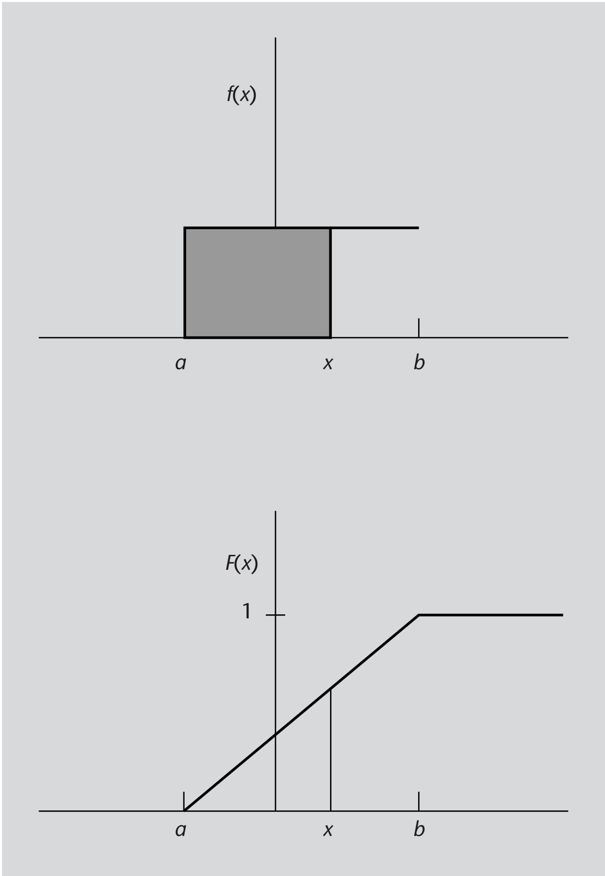


Figura 5

Funció de densitat $f(x)$ i funció de distribució $F(x)$ d'una variable aleatòria uniforme $X \sim U(a,b)$. Fixeu-vos com la probabilitat acumulada en el punt x, $F(x)$, correspon a l'àrea sota la funció de densitat $f(x)$.

L'exemple 1.2 segueix una distribució $X \sim U(0,2)$. Així, la probabilitat que el senyal emès es trobi entre 0 i 1 mV ens la dóna l'àrea per sota de la corba de la funció de densitat, que en aquest cas correspon a l'àrea d'un rectangle de base 1 i altura $\frac{1}{2}$. És a dir, $P(0 \leq X \leq 1) = \frac{1}{2}$.

Un altre exemple de sistema que utilitza les distribucions uniformes són els generadors de nombres enters aleatoris. Aquests dispositius generen nombres enters dins d'un interval determinat de manera uniforme, de tal manera que tots el nombres tenen la mateixa probabilitat de ser generats. Vegem-ne un exemple a continuació.

Exemple 3.1

Triem a l'atzar un nombre, X, en l'interval (0,5). La funció de densitat és:

$$f(x) = \begin{cases} \frac{1}{5} & \text{si} \quad x \in (0,5) \\ 0 & \text{altrament} \end{cases}$$

Calculem algunes probabilitats.

1) Probabilitat que el nombre sigui més petit que 3, $P(X < 3) = \int_0^3 \frac{1}{5}\,dx = \frac{3}{5}$.

2) Sabent que el nombre és més gran que 2, probabilitat que sigui més petit que 3,

$$P(X < 3/X > 2) = \frac{P(2 < X < 3)}{P(X > 2)} = \frac{1/5}{3/5} = \frac{1}{3}.$$

Vegeu també

Recordeu el subapartat 2.3 del mòdul "Introducció a la probabilitat", en què vam veure la probabilitat condicionada, és a dir, la probabilitat d'un succés sabent que s'ha produït un altre succés conegut.

3.2.2. Variable aleatòria exponencial: *Exp*(λ)

La distribució exponencial se sol utilitzar per a modelitzar experiències en què intervé un temps d'espera, com:

- Temps d'espera en una consulta sense cita prèvia.
- Temps d'espera en un servidor per a rebre resposta a una petició enviada.
- La vida d'un component electrònic.

La distribució de Poisson que hem vist en l'apartat de les variables aleatòries discretes està molt relacionada amb la distribució exponencial. Si un procés és de Poisson (succés aleatori en el temps), la variable temps, t, que passa fins que té lloc el primer succés, és exponencial. Cal destacar que el paràmetre de la variable de Poisson val $\alpha = \lambda T$, en què T és l'interval en què comptem els esdeveniments que succeeixen.

La distribució exponencial té per funció de densitat:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si} \quad x \geq 0 \\ 0 & \text{altrament} \end{cases}$$

Obtenim la funció de distribució integrant. Així,

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ 1 - e^{-\lambda x} & \text{si} \quad x \geq 0 \end{cases}$$

En la figura 6 veiem la representació de la funció de densitat, per a tres valors diferents de λ. (No s'ha representat l'eix negatiu d'abscisses en què la funció és 0.)

Figura 6. Funcions densitat de $X \sim Exp(1)$, $X \sim Exp(2)$ i $X \sim Exp(3)$

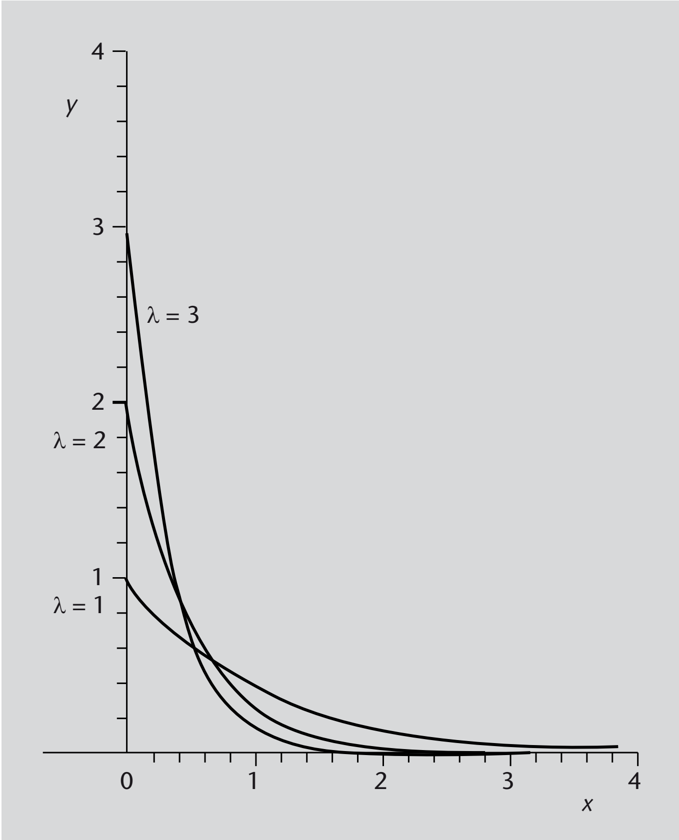


Figura 6

Representació de la funció de densitat d'una variable exponencial per als valors de λ 1, 2 i 3. Com més gran és λ (més trucades per unitat de temps, per exemple) és més probable que hàgim d'esperar poc temps fins que arribi una trucada.

Fixeu-vos que tal com λ creix la funció és menys suau. La causa d'això és que si arriben molts successos per unitat de temps, la major part de la probabilitat de tenir una arribada es concentra en valors de x petits. Per contra, si tenim pocs successos per unitat de temps (α o λ petits) la probabilitat d'arribada és més uniforme (més suau).

Exemple 3.2

Suposem que el temps, en hores, que es necessita per a arreglar un cert tipus d'avaria telefònica és una variable aleatòria, T, que segueix una llei exponencial de paràmetre $\lambda = 0{,}5$. En aquest cas tenim $f(t) = 0{,}5 \cdot e^{-0{,}5t}$ i $F(t) = 1 - e^{-0{,}5t}$ per a $x \geq 0$. Calculem algunes probabilitats:

1) Probabilitat que el temps de reparació passi de les 2 h. És a dir,

$$P(T > 2) = 1 - P(T < 2) = 1 - F(2) = e^{-1} = 0{,}368.$$

Vegeu també

Aquí tornem a utilitzar la noció de probabilitat condicionada del subapartat 2.3 del mòdul "Introducció a la probabilitat".

2) Sabent que el temps de reparació ja ha sobrepassat les 9 h, quina és la probabilitat que la reparació trigui almenys 10 h? En aquest cas es tracta de trobar una probabilitat condicionada, escrivim:

$$P(T > 10/T > 9) = \frac{P((T > 10) \cap (T > 9))}{P(T > 9)} = \frac{P((T > 10)}{P(T > 9)} = \frac{1 - F(10)}{1 - F(9)} = e^{-0{,}5 \cdot 1} = 0{,}606.$$

3.2.3. Variable aleatòria normal o de Gauss: $N(m,\sigma)$

És una de les distribucions de probabilitat més utilitzades. Molts fenòmens físics que afecten circuits i aparells de telecomunicacions es modelitzen utilitzant la distribució normal o de Gauss. També s'utilitza molt freqüentment per al control de qualitat estadístic de components electrònics. Depèn de dos paràmetres, m i σ, que veurem en el subapartat següent.

Paràmetres de la distribució normal

La distribució normal o de Gauss es caracteritza per dos paràmetres: el valor mitjà m (paràmetre de posició), i la desviació típica σ, paràmetre que ens mesura la dispersió de la variable aleatòria respecte de m.

Una particularitat que presenta aquesta distribució és que és la forma límit d'algunes distribucions discretes quan s'augmenta indefinidament el nombre de repeticions d'un experiment. Moltes variables aleatòries com pesos, alçades, talles, consums de gas, etc., segueixen una distribució normal perquè cadascuna és la suma d'un gran nombre de variables aleatòries independents. Així, l'alçada d'una persona és la suma de molts factors, hereditaris, alimentació, tipus de vida, etc.

Els errors, anomenats *aleatoris*, que es presenten en observacions astronòmiques, pesades d'una balança, el soroll generat en els aparells de telecomunicació, etc., i, en general, en la majoria de mesures amb algun aparell, són la suma d'un gran nombre d'errors elementals independents com corrents d'aire, vibracions, error d'apreciació, etc. Per això, els errors aleatoris segueixen una distribució normal.

La distribució normal té per funció de densitat:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad \forall x \in \mathbb{R}.$$

En aquest cas la funció de distribució no es pot trobar integrant analíticament de manera senzilla com ho hem fet abans. Per aquesta raó ens serà més útil treballar amb la funció de densitat.

Vegem alguns gràfics de la funció de densitat en variar els paràmetres m i σ.

Càlcul de probabilitats

Per a calcular probabilitats utilitzarem taules estadístiques o bé algun programari matemàtic de tipus Scilab, Excel, Wiris, Minitab, SPSS, R, etc.

En la figura 7 hem representat $N(0,1)$ i $N(2,1)$. El primer paràmetre de la funció $N(x,y)$ (0 i 2 en aquest exemple) fa referència al valor mitjà. El segon paràmetre (1 en els dos exemple) és la desviació estàndard. Fixeu-vos que com que totes dues variables aleatòries tenen la mateixa desviació típica, σ (sigma), la forma de la funció no ha variat. $N(2,1)$ està traslladada dues unitats a la dreta respecte de $N(0,1)$, ja que el valor mitjà és diferent en cada cas.

Figura 7. Funcions de densitat de $X \sim N(0,1)$ i $X \sim N(2,1)$

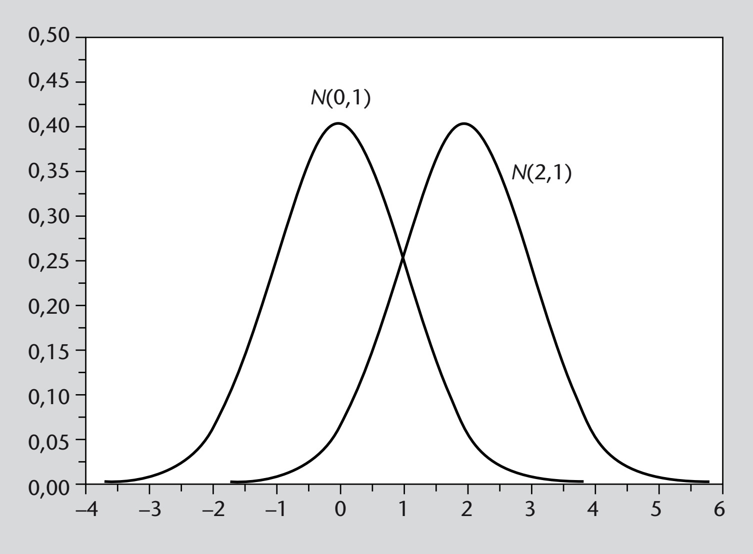
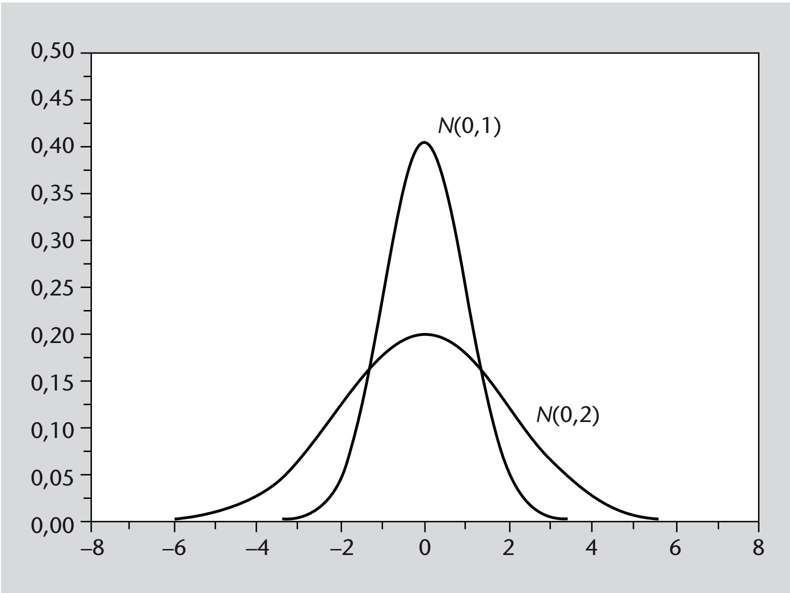


Figura 7

La corba de l'esquerra representa la distribució $N(0,1)$, centrada a 0 i amb una desviació típica de valor 1. La corba de la dreta representa la distribució $N(2,1)$, amb valor mitjà 2 i desviació estàndard 1.

En la figura 8 fixem el valor a $m = 0$ i modifiquem σ. Observem que per a un valor més petit que σ hi ha menys dispersió. Per a σ = 2 tenim més dispersió, i per tant la funció de densitat és menys punxeguda.

Figura 8. Funcions de densitat de $X \sim N(0,1)$ i $X \sim N(0,2)$



Observeu que encara que canviï la forma de la funció, l'àrea total per sota de la corba és 1, la probabilitat total.

Figura 8

La funció $N(0,1)$ és una distribució gaussiana amb mitjana 0 i desviació estàndard 1. En la funció $N(0,2)$ hem augmentat la desviació estàndard, i com que tenim més dispersió, aquesta funció és més plana, menys punxeguda. Noteu, però, que l'àrea total sota les dues corbes ha de ser la mateixa i igual a 1.

Ja que no podem integrar analíticament $f(x)$, per a trobar probabilitats cal utilitzar taules o bé algun programari matemàtic o estadístic. En el nostre cas utilitzarem el programari Scilab*.

*** http://www.scilab.org**

Vegem-ne a continuació algun exemple. Si el que volem calcular és l'àrea indicada en la figura 9, $P(X < n)$, per a un cert valor de x, i $N(0,1)$, escrivim en l'Scilab:

```
y = integrate('1/sqrt(2*%pi)*exp(-t^2/2)','t',-10,n)
```

Figura 9. Funció de distribució de $X \sim N(0,1)$

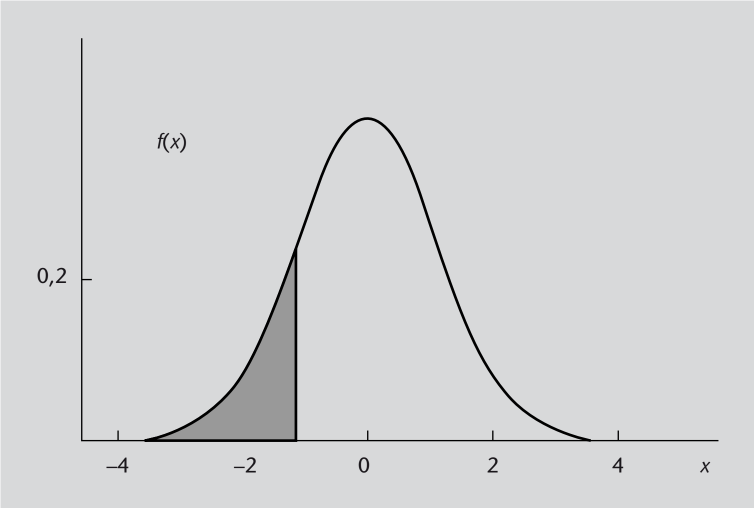


Figura 9

En aquest exemple s'ha calculat la probabilitat que $x < -1$. Per a fer-ho hem calculat l'àrea sota la distribució de Gauss des de $-\infty$ fins a $x = 1$. Utilitzant l'Scilab, com s'indica en aquest apartat, s'ha trobat que $P(x < -1) = 0{,}158655$.

En el cas particular $x = 0$, obtenim una àrea de 0,5. Proveu en aquest cas de fer-ho amb l'Scilab. Us sortirà:

```
y = integrate('1/sqrt(2*%pi)*exp(-t^2/2)','t',-10,0)

y = 0,5
```

Fixeu-vos que s'ha posat com a límit inferior –10. Proveu altres valors, i veureu que no cal considerar valors gaire negatius, ja que la funció és pràcticament zero.

3.3. Paràmetres: valor mitjà (esperança) i variància

En el subapartat anterior hem vist tres de les distribucions contínues més freqüents: la distribució uniforme, l'exponencial i la de Gauss. De manera semblant a com ho hem fet per al cas de les variables aleatòries discretes, en aquest subapartat veurem dos paràmetres que defineixen aquestes distribucions: el valor mitjà i la variància. La diferència fonamental entre les variables discretes i les contínues és que en aquest cas transformarem els sumatoris que havíem vist en el subapartat 2.2 en integrals.

Definició 3.3. Sigui X una variable aleatòria contínua.

El **valor mitjà, esperança** o **moment d'ordre 1** de X:

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx$$

El **moment d'ordre 2**:

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx$$

El **moment d'ordre *n***:

$$E(X^n) = \int_{-\infty}^{+\infty} x^n f(x)\,dx$$

La **variància**:

$$Var(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)\,dx = E(X^2) - E(X)^2$$

La **desviació típica**:

$$\sigma = \sqrt{Var(X)}$$

Proposició 3.2. Per a cada una de les distribucions vistes, s'obtenen els valors de l'esperança i la variància de la taula següent.

Distribucions de variables aleatòries contínues

$X \sim$	Funció de densitat	$E(X)$	$Var(X)$
$U(a,b)$	$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a,b) \\ 0 & \text{altrament} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$N(m,\sigma)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \forall x \in \mathbb{R}$	m	σ^2
$Exp(\lambda)$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{altrament} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Comprovem algun resultat per al cas de $X \sim Exp(\lambda)$

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xf(x)\,dx = \int_0^{+\infty} x\lambda e^{-\lambda x}\,dx = \left\{ \begin{array}{c} u = x \quad \text{i} \quad du = dx \\ dv = \lambda e^{-\lambda x} \quad \text{i} \quad v = -e^{-\lambda x} \end{array} \right\} = \\ &= \left\{-xe^{-\lambda x}\right\}_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}\,dx = \lim_{x\to+\infty}(-xe^{-\lambda x}) - \lim_{x\to 0}(-xe^{-\lambda x}) + \frac{1}{\lambda} = \\ &= 0 + 0 + \frac{1}{\lambda}. \end{aligned}$$

4. Teorema central del límit. Aplicació

En aquest mòdul hem vist les variables aleatòries discretes i contínues. Per a cada variable aleatòria n'hem calculat el valor mitjà i la variància. Aquests paràmetres ens donen una idea global del comportament de la variable aleatòria. També, per a cada una de les variables estudiades, hem vist la funció de distribució, que ens permet calcular la probabilitat acumulada donat un cert valor x i la funció de densitat, que ens diu com varia (és la derivada en funció de x) la funció de distribució.

En aquest darrer apartat d'aquest mòdul veurem com podem relacionar les variables aleatòries discretes vistes en l'apartat 2 amb la variable aleatòria contínua normal o de Gauss que hem vist en l'apartat 3. Tal com hem dit abans, la distribució normal és la forma límit d'algunes distribucions discretes, quan s'augmenta indefinidament el nombre de repeticions d'un experiment.

Recordeu l'exemple 2.8 que hem vist en el subapartat 2.2, en què teníem la distribució de notes de dos alumnes com a exemple de distribució discreta. Fixeu-vos que allà hem calculat el valor mitjà i la desviació típica, justament els paràmetres que defineixen una distribució normal. Què passaria si en la nostra distribució discreta augmentéssim el nombre de mostres i en comptes de 3 resultats prenguéssim moltes més notes? O si dibuixéssim la distribució de notes, no de 2, sinó de molts més alumnes? Això és el que ens permet el teorema central del límit: aproximar una distribució discreta a una normal quan augmentem el nombre de mostres i es donen una sèrie de condicions que veurem a continuació.

Enunciem el teorema central del límit que reflecteix aquest fet.

Observació

El teorema central del límit ens permet aproximar una variable aleatòria discreta a una distribució de Gauss quan repetim un nombre suficientment gran de vegades un experiment.

Proposició 4.1 (Teorema central del límit). Sigui (X_n) amb $n \geq 1$ una successió de variables aleatòries independents, que segueixen la mateixa llei de probabilitat, amb una esperança m i variància σ^2. Considerem la nova variable aleatòria definida per

$$Y_n = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - m}{\frac{\sigma}{\sqrt{n}}}. \tag{6}$$

Aquesta variable tal com l'hem definida té un valor mitjà igual a zero i una variància igual a 1. Es té que la variable Y_n convergeix cap a la distribució $N(0,1)$ quan n tendeix a infinit. La distribució $N(0,1)$ també es coneix com a **distribució normal estàndard**.

4.1. Aproximació de llei binomial a la normal

Si estem treballant amb una variable discreta $X \sim Bin(n,p)$, sota certes condicions la podrem aproximar a una variable $Y \sim N(np,\sqrt{np(1-p)})$, en què la variable Y té la mateixa esperança i variància que X. Fixeu-vos que en aquest cas el valor mitjà de la distribució normal, m, és igual a np, i la variància, σ, és $\sqrt{np(1-p)}$, en què n i p són els paràmetres de la distribució binomial.

Encara que en el teorema anterior es parla d'aproximació quan n tendeix a infinit, en la pràctica aquesta aproximació és vàlida quan es compleix $np > 5$ i $n(1-p) > 5$.

Cal tenir en compte que passem d'una distribució discreta que pren valors en $\{0,1,2,\cdots,n\}$ a una variable contínua que pren valors en tot $\mathbb{R}$. A més, en el cas de la llei binomial, la probabilitat en un punt és diferent de zero, mentre que no és així en el cas de la llei normal perquè és una distribució contínua.

Observació

Recordeu que $X \sim Bin(n,p)$ és el nombre de n experiències de Bernoulli $B(p)$ indexades.

Per aquestes raons, quan aproximem una distribució binomial a una normal cal fer una correcció de continuïtat de la manera següent:

Si $X \sim Bin(n,p)$ amb $np > 5$ i $n(1-p) > 5$ i volem calcular $P(a \leq X \leq b)$, considerem la variable $Y \sim N(np,\sqrt{np(1-p)})$ i calculem $P(a–0,5 < Y < b+0,5)$.

Observació

Fixeu-vos en els factors –0,5 i +0,5, que hem afegit als límits a i b a causa de la correcció de continuïtat.

Exemple 4.1

En un magatzem s'ha analitzat durant un any el percentatge de peces defectuoses i se n'ha detectat un 8%. És dir, podem considerar que la probabilitat que una peça sigui defectuosa és de 0,08. S'agafa una mostra de 100 peces i es defineix la variable aleatòria X com el nombre de peces defectuoses dins de la mostra de 100. La variable aleatòria X segueix una llei binomial $Bin(100,0,08)$, ja que repetirem l'experiment de prendre una peça 100 cops i la probabilitat d'èxit (aquí definim èxit com l'obtenció d'una peça defectuosa) és 0,08. Calculem la probabilitat que en les 100 peces n'hi hagi entre 10 i 20 de defectuoses.

Primer ho calculem sense fer l'aproximació. Com que $X \sim Bin(100,0,08)$,

$$P(10 \leq X \leq 20) = \sum_{k=10}^{20} \binom{100}{k} 0{,}08^k 0{,}92^{100-k} = 0{,}28.$$

Comprovem a continuació que podem fer l'aproximació de la distribució binomial a una normal:

- $np = 100 \cdot 0{,}08 = 8 > 5$
- $n(1-p)100 \cdot 0{,}92 = 92 > 5$

Un cop confirmat que podem fer l'aproximació prenem una distribució normal amb els paràmetres següents: $m = np = 8$ i $\sigma = \sqrt{np(1-p))} = 2{,}713$. Per tant, la nostra distribució normal és $Y \sim N(8;\ 2{,}713)$. La probabilitat que en 100 peces n'hi hagi entre 10 i 20 de defectuoses és:

$$P(10 \leq X \leq 20) = P(9{,}5 < Y < 20{,}5) = 0{,}29.$$

Recordeu que hem de corregir els límits de l'interval per considerar amb el factor 0,5. L'últim valor numèric s'ha obtingut amb l'ajut d'un programari matemàtic. Per exemple, amb l'Scilab caldria fer ús de les ordres següents:

```
--> y = integrate('1/(2.713 * sqrt(2*%pi))*exp(-(t-8)^2/(2*2.713^2))','t',9.5,20.5)
 y = 0.2901661
```

Resum

En l'apartat 1 hem vist què és una **variable aleatòria** i l'hem definida com una funció que assigna un nombre a cada element de l'espai mostral Ω. Per exemple, si llancem una moneda podem definir $X = 0$ si obtenim cara i $X = 1$ si obtenim creu.

També hem vist que hi ha dos tipus de variables aleatòries:

- Variables aleatòries discretes: els valors que pot prendre X es troben dins d'un conjunt finit o infinit numerable d'elements.

- Variables aleatòries contínues: X pot prendre qualsevol valor en conjunts no numerables.

L'apartat 2 l'hem dedicat a estudiar en detall algunes de les variables aleatòries discretes més importants. Hem definit què és la **funció de probabilitat** i hem vist les distribucions següents:

- Variable aleatòria discreta **uniforme**.
- Distribució de **Bernoulli**, $B(p)$, definida pel paràmetre p (probabilitat d'èxit).
- Distribució **binomial**, $Bin(n,p)$, que consisteix en un experiment de Bernoulli repetit n cops.
- Distribució **geomètrica**, $Geom(p)$, que es dóna quan repetim l'experiment de Bernoulli fins que obtenim el primer èxit.
- Distribució de **Poisson**, $Poiss(\alpha)$, caracteritzada pel paràmetre α (nombre mitjà d'èxits dins d'un cert interval).

Hem definit els moments d'ordre n d'una variable aleatòria discreta i, en particular, hem vist l'**esperança** o **valor mitjà**, la **variància** i la **desviació típica**. Definim el valor mitjà o esperança com s'indica a continuació:

$$E(X) = \sum_{i=1}^{n} a_i P(X=a_i). \tag{7}$$

El moment d'ordre 2 és:

$$E(X^2) = \sum_{i=1}^{n} a_i^2 P(X=a_i). \tag{8}$$

La variància i la desviació típica són:

$$Var(X) = E(X^2) - E(X)^2, \tag{9}$$

$$\sigma = \sqrt{Var(X)}.$$

Finalment, hem vist el concepte de **funció de distribució** com a funció de probabilitat acumulada que ens dóna la probabilitat que la variable aleatòria X sigui igual o més petita que un cert valor x:

$$F_X(x) = P(X \leq x), \forall x \in \mathbb{R}. \tag{10}$$

En la taula següent podeu veure els paràmetres més importants per a les variables aleatòries discretes que hem vist.

Distribucions de variables aleatòries discretes

$X \sim$	k	$P(X=k)$	$E(X)$	$Var(X)$
$B(p)$	$\{0,1\}$	$P(X=1) = p$ $P(X=0) = 1-p$	p	$p(1-p)$
$Bin(n,p)$	$\{0,1,2,\ldots,n\}$	$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
$Geom(p)$	$\{1,2,3,\ldots\}$	$P(X=k) = (1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poiss(\alpha)$	$\{0,1,2,\ldots\}$	$P(X=k) = \frac{\alpha^k}{k!} e^{-\alpha}$	α	α

Hem dedicat l'apartat 3 a l'estudi de les variables aleatòries contínues. Hem començat l'apartat definint la **funció de distribució**, que conceptualment és la mateixa que per al cas de les variables discretes. Per al cas de les variables contínues hem definit una nova funció, la **funció de densitat**, que és la derivada de la **funció de distribució**. A continuació hem vist les variables aleatòries contínues més importants:

- La distribució **uniforme**, $X \sim U(a,b)$, caracteritzada per l'interval (a,b), en què la **funció de densitat** és constant i val $\frac{1}{b-a}$ en aquest interval.
- La distribució **exponencial**, $Exp(\lambda)$, en què λ és la taxa de successos per unitat de temps.
- La distribució **normal** o de **Gauss**, $N(m,\sigma)$, caracteritzada pel valor mitjà i la desviació típica.

De la mateixa manera que hem fet en l'apartat 2 amb les variables aleatòries discretes, hem vist també en aquest apartat que per a les variables aleatòries contínues podem definir els **moments d'ordre** n, la **variància** i la **desviació típica**. El valor mitjà o esperança s'expressa com segueix:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx. \tag{11}$$

El moment d'ordre 2 és:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx. \tag{12}$$

La variància i la desviació estàndard són:

$$Var(X) = E(X^2) - E(X)^2, \tag{13}$$

$$\sigma = \sqrt{Var(X)}.$$

La funció de distribució i de densitat les definim com segueix:

$$F_X(x) = P(X \leq x), \tag{14}$$

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

En la taula següent podeu veure els paràmetres més importants per a les variables aleatòries discretes que hem vist.

Distribucions de variables aleatòries contínues

$X \sim$	Funció de densitat	$E(X)$	$Var(X)$
$U(a,b)$	$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a,b) \\ 0 & \text{altrament} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$N(m,\sigma)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \ \forall x \in \mathbb{R}$	m	σ^2
$Exp(\lambda)$	$f(x) = \begin{cases} \lambda\, e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{altrament} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Finalment, en l'apartat 4, hem vist el **teorema central del límit**. Aquest teorema ens permet aproximar una successió de variables aleatòries independents a una **distribució normal** sota unes certes condicions. En particular, hem vist com podem aplicar aquesta llei per a aproximar una distribució binomial a una normal.

Activitats

1. Una empresa de fabricació de microxips observa que el nombre de components electrònics que fallen abans de complir 100 h de funcionament segueix una distribució de Poisson. De mitjana, el nombre de microxips que fallen en aquest interval de temps és 8. Es demana:

a) Comproveu que la funció de probabilitat corresponent a la Poisson satisfà la condició que la suma de totes les probabilitats de valors possibles de la variable té un valor de 1, és a dir, que $1 = \sum_{k=0}^{+\infty} P(X=k)$.

Pista: podeu fer servir el programari matemàtic Wiris per a calcular el valor de $\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}$.

b) Quina és la probabilitat que falli un microxip al cap de 50 h de funcionament?

c) Quina és la probabilitat que no fallin més de dos microxips en 100 h?

d) Quina és la probabilitat que fallin almenys 10 microxips en 125 h?

2. Una font binària genera dígits 1 i 0 de manera aleatòria amb probabilitats 0,6 i 0,4, respectivament. Es demana:

a) Amb quina de les variables aleatòries vistes podríem modelitzar el comportament d'aquesta font?

b) Quina és la probabilitat que en una seqüència de 5 dígits surtin dos 1 i tres 0?

c) Quina és la probabilitat que en la seqüència de 5 dígits s'obtinguin almenys tres 1?

3. Suposeu que el temps(en segons) que triga un servidor de bases de dades a donar resposta a una consulta SQL és una variable aleatòria exponencial amb paràmetre $\lambda = 1/10$. Si el servidor rep una altra consulta SQL just abans de la vostra, es demana:

a) De quin tipus de variable aleatòria es tracta?

b) Quina és la probabilitat que el temps d'espera per a poder llançar una consulta sigui menys de 5 s.

c) Quina és la probabilitat que el temps d'espera per a poder llançar una consulta estigui entre 5 i 10 s.

Pista: recordeu que si $X \sim Exp(\lambda)$, llavors $f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$

4. En una xarxa de telecomunicacions, s'ha calculat que la probabilitat que un encaminador falli en una jornada d'activitat extrema és de 0,04. Si es considera un total de 2500 jornades d'activitat extrema, llavors:

a) De quin tipus de variable aleatòria es tracta?

b) Quina és la probabilitat que l'encaminador falli més de 120 cops en aquest període?

c) I que falli entre 100 i 120 cops (tots dos inclosos) en aquest període?

Pista: gràcies al teorema central del límit, és possible aproximar una binomial mitjançant una normal.

5. El nombre de consultes que un servidor de bases de dades processa en un interval de 10 s és una variable aleatòria de Poisson, X, amb paràmetre $\lambda = 5$ consultes per segon. Es demana:

a) Quina és la probabilitat que cap consulta sigui processada en un interval de 10 s?

b) Quina és la probabilitat que almenys 2 consultes siguin processades en un interval de 10 s?

6. Un emissor A transmet un missatge a un receptor B. Sigui p la probabilitat que B rebi correctament el missatge. Per a assegurar-se que el missatge serà rebut almenys un cop, A tornarà a enviar el missatge fins a un màxim de n intents. Suposant que les n transmissions són independents, es demana:

a) Identifiqueu quina és la distribució estadística associada a la variable aleatòria X = *nombre de missatges rebuts correctament per B en els n intents*.

b) Si $p = 0,7$ i $n = 3$, calculeu la probabilitat que B acabi rebent el missatge.

c) Si $p = 0,8$, quin és el valor mínim de n que fa que la probabilitat que el missatge es rebi sigui, com a mínim, de 0,95?

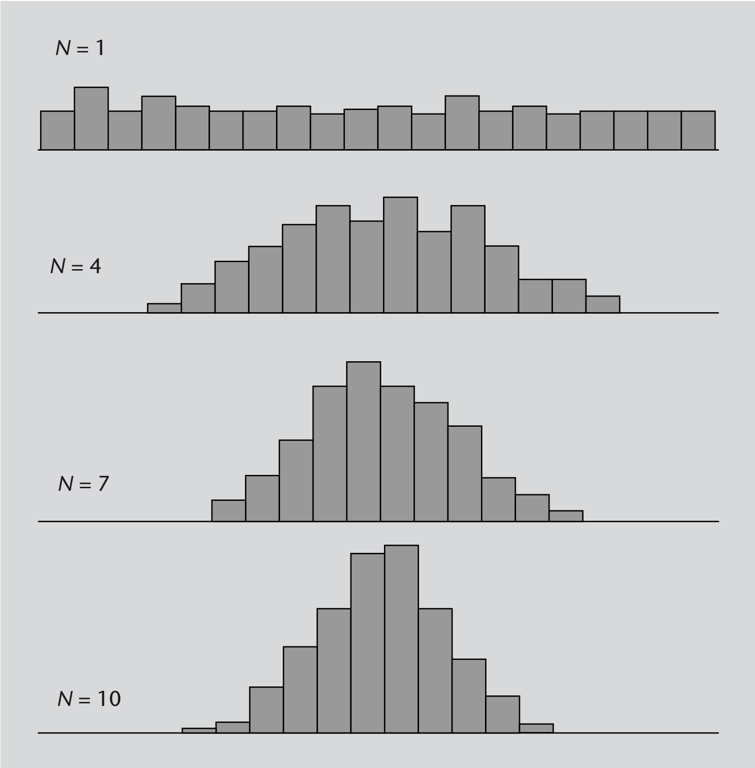
7. Suposeu que la temperatura T a què ha de treballar una sonda de mesura durant una missió espacial és una variable aleatòria gaussiana (distribució normal) amb mitjana 85 graus Fahrenheit i desviació estàndard de 10 graus Fahrenheit.

a) En un moment determinat, quina és la probabilitat que la temperatura estigui entre 75 i 95 °F?

b) I que estigui entre 65 i 105 °F?

c) Busqueu a Internet informació sobre la regla 68/95/99 i justifiqueu que els resultats anteriors són coherents amb això.

d) Feu ús del teorema central del límit per a explicar la gràfica següent.



8. Una centraleta telefònica rep 300 trucades per hora. La centraleta està dimensionada de tal manera que no es poden establir més de 12 connexions per minut. Amb aquestes dades ens demanen el següent:

a) Quina és la probabilitat que la centraleta quedi saturada en un minut determinat?

b) Quina és la probabilitat que es rebi una única trucada en un minut determinat?

9. D'una estació parteix un tren cada 20 min. Si arribem a l'estació en un moment qualsevol, ens demanen determinar el següent:

a) La funció de distribució de la variable aleatòria "temps d'espera".

b) La probabilitat que hàgim d'esperar a l'estació menys de 7 min.

c) L'esperança i variància de la variable aleatòria "temps d'espera".

d) La probabilitat que hàgim d'esperar exactament 12 min.

10. Un avió d'alt rendiment té una computadora central i dues més d'idèntiques preparades per si falla alguna de les altres. Durant una hora d'operació la probabilitat que falli la computadora principal o qualsevol de les altres és de 0,0005. Suposant que cada hora representa un experiment independent de la resta:

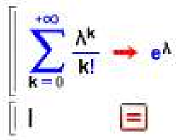
a) Quin és el temps mitjà que passa perquè fallin totes tres computadores?

b) Quina és la probabilitat que les 3 computadores fallin en un vol de 5 h?

Solucionari

1.

a) Es tracta de comprovar que $1 = \sum_{k=0}^{+\infty} P(X=k)$. Fent ús del Wiris, observem que:



i, per tant,

$$\sum_{k=0}^{+\infty} P(X=k) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

b) Sabem que el nombre mitjà de microxips que fallen abans de complir 100 h de funcionament és $\lambda = 8$. Si assumim certes condicions de regularitat podem dir que el nombre mitjà d'errades en un interval de 50 h serà $\lambda' = 4$. La probabilitat que falli un microxip en aquest interval de temps és: $P(X=k) = \frac{\lambda'^k}{k!} e^{-\lambda'}$ Substituint pels valors de l'enunciat: $P(X=1) = \frac{4^1}{1!} e^{-4} =$ 0,073.

c) En aquest cas l'interval per considerar són 100 h, el mateix que ens donen per enunciat. En aquest interval $\lambda = 8$. La probabilitat que fallin com a molt dos microxips és: $\sum_{k=0}^{2} \frac{\lambda^k}{k!} e^{-\lambda}$ $\sum_{k=0}^{2} \frac{8^k}{k!} e^{-8} = (\frac{8^0}{0!} + \frac{8^1}{1!} + \frac{8^2}{2!}) e^{-8} = 0{,}013$.

d) Considerant ara l'interval de 125 h podem dir que de mitjana el nombre de microxips que fallen és de $\lambda' = 10$. La probabilitat que fallin 10 o més microxips és igual a 1 menys la probabilitat que falli un nombre més petit que 10. És a dir:

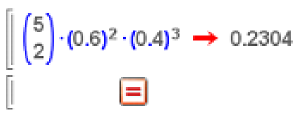
$P(X \geq 10) = 1 - (X \leq 9)$
$P(X \leq 9) = 1 - \sum_{k=0}^{9} \frac{\lambda^k}{k!} e^{-\lambda}$
$P(X \leq 9) = 1 - (\sum_{k=0}^{9} \frac{8^k}{k!}) e^{-8} = 0{,}283$

2.

a) Podem modelitzar el comportament d'aquest emissor amb una variable aleatòria discreta binomial, ja que cada bit es considera una variable aleatòria de Bernoulli i el nombre d'experiències serà el nombre de bits emesos.

b) Si denominem X la variable aleatòria que representa el nombre d'observacions 1 (èxits) que s'obtenen en generar la seqüència de 5 dígits, X segueix una distribució binomial de paràmetres $n = 5$ i $p = 0{,}6$, és a dir, $X \sim Bin(5;\ 0{,}6)$. Per tant, $P(X=2) = \binom{5}{2} 0{,}6^2 \cdot 0{,}4^3 = 0{,}23$.

Amb el Wiris:



c) Ara es demana $P(X \geq 3) = 1 - P(X \leq 2)$, i

$$P(X \leq 2) = \sum_{k=0}^{2} \binom{5}{k} 0{,}6^k \cdot 0{,}4^{5-k} = 0{,}317.$$

En efecte:

$$\sum_{k=0}^{2} \binom{5}{k} \cdot (0.6)^{k} \cdot (0.4)^{5-k} \rightarrow 0.31744$$

Per tant, $P(X \geq 3) = 1 - 0{,}317 = 0{,}683$.

3. Segons es deriva de l'enunciat, $X \sim Exp(1/10)$; per tant:

$$f(X{=}x) = \begin{cases} \frac{1}{10}e^{-x/10} & \text{si } x \geq 0 \\ 0 \text{ si } x < 0 \end{cases}$$

a) En aquest cas es tracta d'una variable aleatòria contínua i exponencial. Fixeu-vos que el temps de donar resposta a una consulta SQL és un nombre real.

b) $P(X < 5) = \int_0^5 \frac{1}{10}e^{-x/10}dx = -\left[e^{-x/10}\right]_0^5 = 1 - \frac{1}{\sqrt{e}} = 0{,}393$.

En efecte:

$$\int_0^5 \frac{1}{10} e^{\frac{-x}{10}} dx \rightarrow \frac{e-\sqrt{e}}{e}$$

c) Anàlogament: $P(5 < X < 10) = \int_5^{10} \frac{1}{10}e^{-x/10}dx = \frac{1}{\sqrt{e}} - \frac{1}{e} \approx 0{,}239$.

En efecte:

$$\int_5^{10} \frac{1}{10} e^{\frac{-x}{10}} dx \rightarrow \frac{\sqrt{e}-1}{e}$$

4. Si representem per X = "nombre de cops que falla l'encaminador en les 2500 jornades", i entenem per "èxit" (en sentit estadístic) el fet que l'encaminador falli, llavors: $X \sim Bin(2500;\ 0{,}04)$.

Amb l'aproximació d'una binomial a una normal, es té que:

$$X \sim N(2500 \cdot 0{,}04, v(2500 * 0{,}04 * 0{,}96)) = N(100;\ 9{,}798)$$

a) En aquest cas podem definir com A, èxit, el fet que l'encaminador falla. Això passa amb una probabilitat de 0,04. El nombre de vegades que repetim l'experiment, segons l'enunciat, és de 2500. Per tant, podem considerar que la nostra variable aleatòria segueix una distribució binomial amb paràmetres $X \sim Bin(2500{,}0{,}04)$.

b) El que es demana és:

$$P(X > 120) = 1 - P(X \leq 120) = \{\text{correc. contin.}\} =$$
$$1 - P(0 < X < 120{,}5) = \{\text{Wiris}\} = 1 - 0{,}98179 = 0{,}01821$$

$$\int_0^{120.5} \frac{1}{9.798 \cdot \sqrt{2\pi}} \cdot e^{\frac{-(x-100)^2}{2 \cdot (9.798)^2}} dx \rightarrow 0.98179$$

c) Ara el que es demana és:

$$P(100 \leq X \leq 120) = \{\text{correc. contin.}\} = P(99.5 < X < 120{,}5) =$$
$$= \{\text{Wiris}\} = 0{,}50214$$

$$\int_{99.5}^{120.5} \frac{1}{9.798\cdot\sqrt{2\pi}}\cdot e^{\frac{-(x-100)^2}{2\cdot(9.798)^2}}\,dx \rightarrow 0.50214$$

5.

a) Per teoria, $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$ i, per tant, $P(X=0) = e^{-5}\frac{5^0}{0!} = 0{,}0067$.

b) Ara ens demanen $P(X \geq 2)$:

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(X=0) + P(X=1)] = 1 - 2e^{-1} \approx 0{,}264.$$

6.

a) Si representem per X la variable aleatòria que representa el nombre d'èxits (missatges rebuts) en les n proves, tenim que X segueix una distribució binomial de paràmetres n i p, és a dir, $X \sim Bin(n,p)$.

b) En aquest cas, $n = 3$ i $p = 0{,}7$, i per tant:

$$P(X \geq 1) = 1 - P(X=0) = 1 - \binom{3}{0}0{,}7^0 0{,}3^3 = 1 - 0{,}027 = 0{,}973.$$

Amb el Wiris:

$$\binom{3}{0}(0.7)^0(0.3)^3 \rightarrow 0.027$$

c) Ara n és desconegut i $p = 0{,}8$. Es demana trobar n tal que $P(X \geq 1) \geq 0{,}95$:

$$P(X \geq 1) = 1 - P(X=0) = 1 - \binom{n}{0}(0{,}8)^0(0{,}2)^n = 1 - (0{,}2)^n$$

Plantegem la inequació: $1 - (0{,}2)^n \geq 0{,}95 \rightarrow n \geq 2$, és a dir, agafarem $n = 2$.

En efecte:

$$1-\binom{2}{0}(0.8)^0(0.2)^2 \rightarrow 0.96$$
$$1-\binom{1}{0}(0.8)^0(0.2)^1 \rightarrow 0.8$$

7.

a) El que es demana és: $P(75 < X < 95)$ = {Wiris} = 0,68269 ≈ 0,69 (coherent amb la regla).

$$\int_{75}^{95} \frac{1}{10\cdot\sqrt{2\pi}}\cdot e^{\frac{-(x-85)^2}{2\cdot(10)^2}}\,dx \rightarrow 0.68269$$

b) Ara el que es demana és: $P(65 < X < 105)$ = {Wiris} = 0,9545 ≈ 0,95 (coherent amb la regla).

$$\int_{65}^{105} \frac{1}{10\cdot\sqrt{2\pi}}\cdot e^{\frac{-(x-85)^2}{2\cdot(10)^2}}\,dx \rightarrow 0.9545$$

c) Per exemple:

Figura 10. Distribució de les notes obtingudes pels dos estudiants. En l'eix horitzontal es representa el número de prova, i en l'eix vertical, la nota obtinguda.

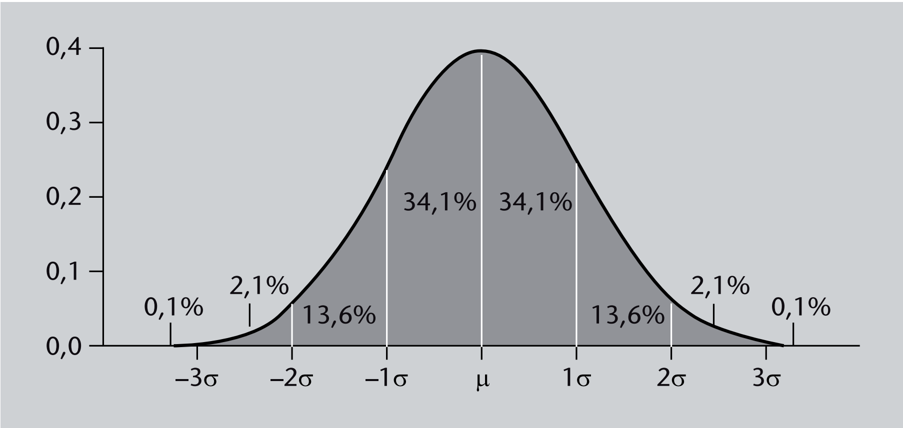


Figura 10

La part en blau fosc és menys d'una desviació estàndard des de la mitjana. Per a una distribució normal, això representa el 68% del conjunt (blau fosc). La part en blau menys fosc està situada fins a dues desviacions estàndard i representa el 95% del conjunt. La part en blau clar (fins a tres desviacions estàndard) representa el 99,7%.

d) La primera fila correspon a la distribució original de la variable X. La resta de files corresponen a les distribucions de les mitjanes mostrals per a diferents mides de la mostra (n = 4, n = 7 i n = 10). S'observa que, amb independència de com es distribueixi la variable X, les distribucions de les mitjanes mostrals es van aproximant a una distribució normal tal com augmenta la mida de la mostra. Aquest resultat és conegut com a teorema central del límit.

8. Per a resoldre aquest exercici tindrem en compte que podem modelitzar l'arribada de trucades a una centraleta mitjançant una distribució de Poisson, *Poiss*(α).

El paràmetre que defineix la distribució, α, coincideix amb l'esperança de la variable, com hem vist en el subapartat 2.2. L'enunciat ens diu que arriben 300 trucades en una hora; per tant, $\alpha = \frac{300}{60}$ = 5 trucades/minut.

La probabilitat que el nombre de trucades per minut sigui més gran que 12 i que, per tant, la central quedi saturada, és:

$$P(X > 12) = 1 - P(X \leq 12) = 1 - e^{-5}\sum_{k=0}^{12}\frac{5^k}{k!} = 0{,}002.$$

La probabilitat que es rebi una única trucada en un minut és:

$$P(X = 1) = e^{-5}\frac{5^1}{1!} = 0{,}0337.$$

9.

a) Si arribem a l'estació en qualsevol moment i tots els moments són equiprobables, podem assumir que la variable aleatòria "temps d'espera" segueix una distribució uniforme en l'interval (0, 20). La funció de distribució és la següent:

$$F(x) = P(X \leq x) = \begin{cases} 0 \text{ si } x \leq 0 \\ \frac{x}{20} \text{ si } 0 < x < 20 \\ 1 \text{ si } x \geq 20 \end{cases}$$

La funció de densitat $f(x)$ és $\frac{1}{20}$ dins de l'interval (0, 20).

b) La probabilitat que esperi almenys 7 min és la següent:

$$P(X < 7) = F(7) = \frac{7}{20} = 0{,}35.$$

c) L'esperança de la variable "temps d'espera" és:

$$E(X) = \frac{0 + 20}{2} = 10.$$

I la variància és:

$$var(X) = \frac{20^2}{12} = \frac{20^2}{12} = 33{,}33.$$

d) La probabilitat que el temps d'espera sigui exactament de 12 min és zero, ja que és una variable aleatòria contínua i, en aquest cas, la probabilitat d'un valor determinat és zero.

10.

a) Representem per X el nombre d'hores que passa fins que fallen els 3 sistemes. X_1, X_2 i X_3 són el nombre d'hores que passa fins que falla el primer, segons i tercer sistema, respectivament. Llavors $X = X_1 + X_2 + X_3$. Sabem que $p = 0{,}0005$ i que tenim 3 computadores independents. Per tant: $E(X) = 3/0{,}0005 = 6000$ h.

b) A continuació ens demanen la probabilitat que les 3 computadores fallin en un vol de 5 h, és a dir:

$$P(x \leq 5) = P(x = 3) + P(x = 4) + P(x = 5) = 0{,}0005^3 + \binom{3}{2} 0{,}0005^3 (0{,}9995) + \tag{15}$$

$$+ \binom{4}{2} 0{,}0005^3 (0{,}9995) = 1{,}249 \cdot 10^{-10}$$

Bibliografia

Gubner, J. A. (2006). *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press.

Hsu, H. (2010). *Schaum's Outline of Probability, Random Variables, and Random Processes*. McGraw-Hill.

Leon-Garcia, A. (2008). *Probability, Statistics, and Random Processes For Electrical Engineering*. Prentice Hall.

Miller, S.; Childers, D. (2004). *Probability and Random Processes: With Applications to Signal Processing and Communications*. Academic Press.

Yates, R. D.; Goodman, D. (2004). *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*. Wiley.