TEMA 7: MODELS DE DISTRIBUCIÓ PER A VARIABLES ALEATÒRIES

1.1 VARIABLES ALEATÒRIES DISCRETES

Quan una variable aleatòria pren un nombre finit de valors o infinit, però numerable, direm que és una variable aleatòria discreta.

Per a una variable aleatòria escriurem $P(X=x_i)=f(x_i)$ per i=1, 2, 3, etc.

La funció "*f*" o "p" rep el nom de ***funció probabilitat*** o ***funció quatia*** i compleix les següents propietats:

- $f(x_i)\geq 0$
- $\sum_i f(x_i)=1$

Amb el símbol *F(x)* representarem l'anomenada ***funció de distribució***, definida per:

$$\boxed{F(x)=P(X\leq x)=\sum_{x_i\leq x} f(x_i)}$$

La funció de distribució dóna la probabilitat que un valor observat sigui menor o igual que el nombre real "x".

Per tant, *F(x)* serà sempre una funció no decreixent. Les propietats de la funció distribució són:

- $0\leq F(x)\leq 1 \quad \forall x\in\Re$
- $x_1<x_2 \;\Rightarrow\; F(x_1)\leq F(x_2)\,,\quad \forall x_1,x_2\in\Re$
- $P(a<X<b)=F(b)-F(a)$

- **ESPERANÇA MATEMÀTICA**

La primera característica d'una distribució que es pot quantificar és l'anomenada ***característica de tendència central*** o ***valor central***. L'esperança matemàtica seria el valor teòric al qual tendiríem si repetíssim l'experiment moltes vegades i calculéssim la mitjana de tots els resultats; ***és el valor mitjà esperat*** μ .

V.A. Discreta

$$\boxed{E(x)=\sum_{i=1}^{\infty} x_i f(x_i)}$$

- **VARIANÇA**

La ***variança*** o variancia es representa per σ^2. A la seva arrel quadrada, σ, se l'anomena ***desviació típica***. ***Ambdues donen una idea de la dispersió de la variable aleatòria***. Quan més gran és la variança, més llunyans i "variats" seran els valors que pot prendre la variable respecte la seva mitjana o esperança. La variança es calcula de la següent manera:

$$\boxed{\sigma^2=E\left[(x-\mu)^2\right]=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx=\ldots=E(x^2)-E(x)^2}$$

1.1.1 VARIABLES ALEATÒRIES DISCRETES CONEGUDES

1. La distribució binomial B(n, p)

Definició

Una prova de Bernouilli és un ***experiment que només té dos resultats possibles***, anomenats habitualment ***èxit o fracàs***, com per exemple el llançament d'una moneda. Si es realitza una successió de 'n' proves de Bernouilli independents tindrem l'anomenada distribució de Bernouilli o distribució binomial.

La distribució de Bernouilli permet doncs calcular la probabilitat d'obtenir x èxits (amb probabilitat d'èxit p) en un total de n proves.

Així, si: $P(E) = p$ Probabilitat d'èxit

$P(F) = 1 - p = q$ Probabilitat de fracàs

x: número d'èxits;

n: número de proves

Direm que la probabilitat de x èxits en aquestes n proves és:

$$\boxed{P(x) = \binom{n}{x} p^x q^{n-x}} \quad \text{amb} \quad \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

- **La moda de la distribució binomial**

La moda d'una variable aleatòria discreta és el valor de l'espai de mostres que té una probabilitat més alta.

En general, direm que la moda serà aproximadament: $\boxed{x \approx np}$

- **Esperança matemàtica i variança de la binomial**

Si apliquem la fórmula discreta de l'esperança matemàtica a aquesta distribució, obtenim:

$$\boxed{E(X) = np} \qquad \boxed{\sigma^2 = E(X^2) - E(X)^2 = npq}$$

EXERCICI 1

La població d'un país te una tasa d'atur del 15% de la población activa. Si es seleccionen 9 persones d'aquets població. Quina será la probabilitat de que 4 d'aquests 9 estuguin a l'atur?

EXERCICI 2

Una associació promou entre els seus membres la conexión a internet. Un any després es tria a 20 dels seus membre de forma aleatòria i el nombre mig de socis que disposen de connexió és de 5. Quina és la probabilitat que algún dels socis triats tingui internet?

2. <u>La distribució de Poisson</u> P(λ)

El càlcul de la probabilitat per a una distribució binomial és complicat quan el nombre d'experiments és molt alt. Davant aquest problema, ***Poisson va aconseguir una aproximació de la fórmula de la probabilitat de distribucions binomials***, per a casos on:

- El nombre d'experiments n és molt alt
- La probabilitat d'èxit p és molt baixa
- L'esperança d'èxits és constant: $\lambda = np$

Substituint el valor de λ la fórmula de la distribució binomial, obtenim la següent aproximació, vàlida per a valors grans de n:

$$\boxed{P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}}$$

- **Esperança i variança de la distribució de Poisson**

Aplicant la fórmula de l'esperança i variança matemàtica obtenim:

$$\boxed{E(X) = \sum_0^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \lambda} \qquad \boxed{Var(X) = \lambda}$$

EXERCICI 3

En una adreça de correu electrònic es reben per terme mig 5 e-mails cada 30 minuts. Quina és la probabilitat que durant la próxima mitja hora es rebin al menys 2 e-mails?

EXERCICI 4

Una sucursal del BBVA concedeix per terme mig 2 crèdits per semana. Quina és la probabilitat que en 2 setmanes concedeixi al menys 3 crèdits?

3. La distribució geomètrica

Aquesta distribució té una definició molt semblant a la distribució binomial. De la mateixa manera que en la distribució binomial, cada prova pot tenir dos resultats: èxit o fracàs. La probabilitat d'èxit la designarem amb p i es manté constant en cada prova.

A la distribució geomètrica, la variable aleatòria X és el ***nombre de proves que es necessiten per tenir el primer èxit.*** (a la binomial o de Bernouilli el nombre n de proves està fixat a priori i estudiem el nombre d'èxits en aquestes n proves, representat per la variable aleatòria X).

A la geomètrica, X pren els valors possibles 1, 2, 3,..., k . X serà igual a k si les k-1 primeres proves produeixen fracàs i la k-èssima, èxit. Per tant,

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

Si apliquem la fórmula de l'esperança matemàtica, i la distribució discreta a aquesta distribució, obtenim:

$$\boxed{E(X)=\frac{1}{p}} \qquad \boxed{Var(X)=\frac{q}{p^2}}$$

Teorema

Si X és una variable aleatòria amb distribució geomètrica, es verifica per h i t enters qualsevols que:

$$\boxed{\mathrm{P(X>h+t/X>h)=P(X>t)}}$$

És a dir, podem observar que ***la distribució geomètrica no té memòria***.

1.2 VARIABLES ALEATÒRIES CONTÍNUES

Si la variable aleatòria *X* pot prendre tots els infinits valors possibles del conjunt dels reals, o bé tots els valors d'un interval (a,b), direm que es tracta d'una variable aleatòria contínua.

La probabilitat que la variable aleatòria *X* prengui un valor menor que *x* la designarem per *F(x)* i escriurem:

$$\boxed{P(X < x) = F(x)}$$

Aquesta funció *F(x)* l'anomenarem ***funció de distribució***. Les propietats d'aquesta funció distribució són les següents:

- $0 \le F(x) \le 1 \quad \forall x \in \Re$
- $F(x)$ és una funció no decreixent:
  $x_1 < x_2 \Rightarrow F(x_1) \le F(x_2), \quad \forall x_1, x_2 \in \Re$
- $P(a < X < b) = F(b) - F(a)$
- $\lim_{x \to -\infty} F(x) = 0 \; ; \; \lim_{x \to \infty} F(x) = 1 \; ; \; \frac{\partial F(x)}{\partial x} \ge 0 \; \forall x$

De la tercera propietat podem observar que la probabilitat que *X* prengui un valor determinat és igual a zero.

$$P(X = x) = 0$$

I, en conseqüència:

$$P(a \le x \le b) = P(a < x \le b) = P(a \le x < b) = P(a < x < b)$$

El fet que la probabilitat d'una variable contínua prengui un valor concret és 0 ens indica que no té cap sentit plantejar-se aquesta probabilitat. ***En casos de funcions contínues haurem d'estudiar la probabilitat que la variable caigui dins un determinat interval.***

- **Funció densitat de probabilitat**

S'anomena funció densitat de la variable aleatòria contínua *X* a la derivada de la funció de distribució:

$$\boxed{f(x) = F'(x)}$$

Recordem que la funció distribució (o de probabilitat acumulada) ens dóna la probabilitat que la variable aleatòria *X* prengui un valor igual o inferior que el nombre *x*.

A partir d'aquesta definició, podem plantejar la igualtat:

$$\boxed{P(a < X < b) = \int_a^b f(x)\,dx = F(b) - F(a)}$$

La funció densitat de probabilitat compleix:

- $f(x) \ge 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a \le x \le b) = \int_a^b f(x)dx$

És important entendre que el valor de la funció densitat en un punt no dóna la probabilitat que la variable aleatòria prengui aquell valor. La gràfica de la funció densitat dóna, únicament, idea de quines zones són més probables i quines ho són menys.

- **Funció distribució**

Obtenim la funció distribució o funció de probabilitat acumulada calculant:

$$\boxed{F(x) = P(X \leq x) = P(-\infty \leq X \leq x) = \int_{-\infty}^{x} f(t)dt}$$

- **ESPERANÇA MATEMÀTICA**

La primera característica d'una distribució que es pot quantificar és l'anomenada ***característica de tendència central*** o ***valor central***. L'esperança matemàtica seria el valor teòric al qual tendiríem si repetíssim l'experiment moltes vegades i calculéssim la mitjana de tots els resultats; ***és el valor mitjà esperat*** μ .

V.A. Contínua

$$\boxed{E(x) = \int_{a}^{b} x\, f(x)dx}$$

- **VARIÀNÇA**

La ***variança*** o variancia es representa per σ^2. A la seva arrel quadrada, σ , se l'anomena ***desviació típica***. ***Ambdues donen una idea de la dispersió de la variable aleatòria***. Quan més gran és la variança, més llunyans i "variats" seran els valors que pot prendre la variable respecte la seva mitjana o esperança. La variança es calcula de la següent manera:

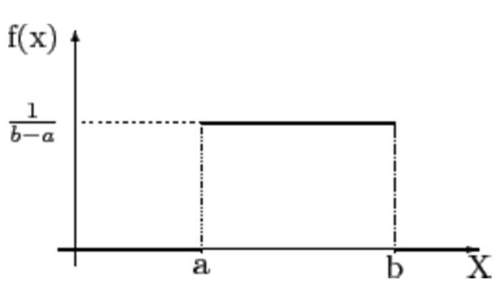
$$\boxed{\sigma^2 = E\left[(x-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \ldots = E(x^2) - E(x)^2}$$

1.2.1 VARIABLES ALEATÒRIES CONTINUES CONEGUDES

1. Distribució uniforme U(a, b)

Suposem que una variable aleatòria ve definida per la funció densitat següent:

$$f(x) = \begin{cases} 0 & x < a \\ \dfrac{1}{b-a} & a < x < b \\ 0 & x > b \end{cases}$$



En aquest cas, *els valors de la variable es troben distribuïts igualment sobre l'interval (a,b) i direm que la variable està distribuïda uniformement. És el cas d'equiprobabilitat. Qualsevol subinterval de valors d'aquest interval de valors (a,b) té les mateixes probabilitats de succeir.*

D'aquesta manera, la funció de distribució serà:

$$F(X) = P[X \leq x] = \int_a^x \frac{1}{b-a} dx = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x > b \end{cases}$$

I la seva esperança matemàtica i la seva mitjana seran:

$$\boxed{E(X) = \frac{b+a}{2}} \qquad \boxed{V(X) = \sigma^2 = \frac{(b-a)^2}{12}}$$

EXERCICI 5

Si X és una variable que segueix una distribució Unifore entre [0,5; 2].Quina és la probabilitat que X prengui un valor superior a 1,5?

EXERCICI 6

Una empresa d'autobusos sap que la demanda de bitllets per a un determinat trajecte és una varibale aleatòria amb una distribució uniforme amb un mínim de 57 i un màxim de 98. Si l'empresa utilitza per realitzar el trajecte un únic autocar de 80 places. Quina és la probabilitat que un dia qualsevol hi hagi persones que NO poden realitzar el viatge per falta de places?

2. <u>Distribució Normal</u> N(μ,σ)

La distribució normal és segurament la més important de totes les distribucions discretes o contínues. Una variable aleatòria pot modelitzar-se segons una distribució normal quan la seva variabilitat és deguda a que l'experiment depèn d'un gran número de causes, cap d'elles preponderant sobre les altres, de manera que els seus efectes es sumen.

- **Definició:**

La variable X segueix una distribució normal si la seva funció densitat és:

$$\boxed{f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \qquad -\infty < x < \infty$$

on μ és l'esperança de la variable i σ la seva desviació típica.

Que la variable X es distribueix segons una distribució normal ho designarem com:

$$X \approx N(\mu,\sigma)$$

La funció de densitat f(x) és una funció no integrable analíticament. Per treballar amb la normal, doncs, utilitzarem valors tabulats de la funció de distribució de la normal estàndard N(0,1) (de mitjana 0 i desviació tipus 1).

Per operar amb qualsevol distribució normal, per tant, ***serà necessari fer un canvi de variable, per convertir la variable normal en una variable normal estandarditzada N(0,1)*** (aquest procés rep el nom de tipificació).

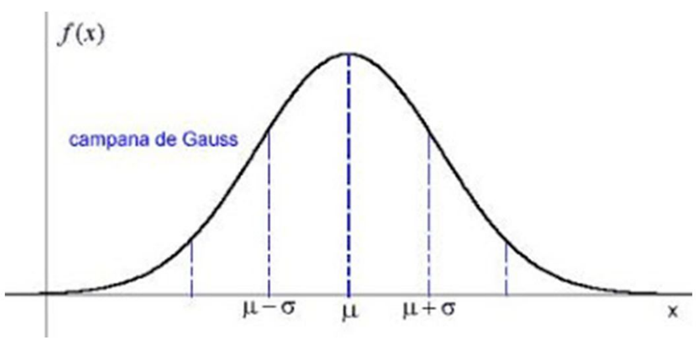
$$\boxed{Z=\frac{X-\mu}{\sigma}} \quad \text{on} \quad Z \approx N(0,1)$$

- **Forma i propietats de la distribució normal**

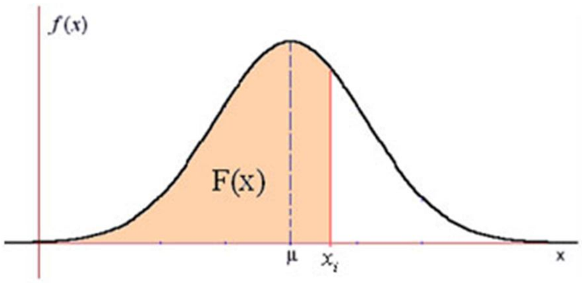
Les propietats d'aquesta funció seran les següents:

1. *f(x)>0* $\forall x$
2. Per $x=\mu$ la gràfica passa per un màxim.
3. La corba és simètrica respecte $x=\mu$
4. En les abscisses corresponents a $x=\mu\pm\sigma$ tenim els punts d'inflexió de la corba.
5. L'eix d'abscisses és una asímptota horitzontai.

Gràficament, la funció densitat de probabilitat de la distribució normal presenta la següent forma:



Tal i com pot veure's gràficament, la funció de F(x) complirà que:



Es tracta d'una distribució simétrica i complirà:

$$F(-x) = 1 - F(x) \qquad \forall x$$

- <u>Ús de les taules per treballar amb la distribució normal</u>

Si X és una variable N(0,1), tenim:

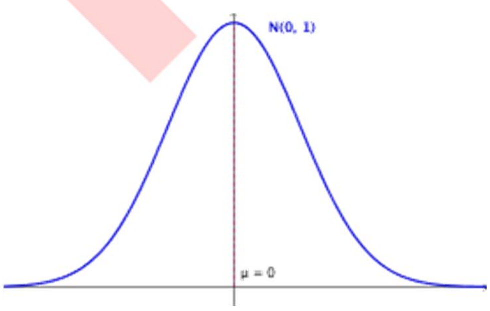
$$P(a \le x \le b) = \frac{1}{2\pi} \int_a^b e^{-\frac{x^2}{2}} dx$$

integral on no podem aplicar el teorema fonamental del càlcul, ja que no podem trobar una primitiva de $\int_a^b e^{-\frac{x^2}{2}} dx$.

És per això que la funció de distribució $F(x) = P(X \le x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ ha estat tabulada. Amb aquestes taules podrem calcular:

$$\boxed{P(a \le x \le b) = F(b) - F(a)}$$

***Si la variable X no és normal tipificada (o estandarditzada) sinó normal** $N(\mu, \sigma)$ **fem el canvi ja explicat i obtenim f(z) corresponent a** $z \approx N(0,1)$*



$$P(a \le x \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

És útil, per últim, conèixer els resultats següents:

$$P(-1 < z < 1) = 0{,}6826$$
$$P(-2 < z < 2) = 0{,}9544$$
$$P(-3 < z < 3) = 0{,}9974$$

EXERCICI 7

La temperatura d'una ciutat es distribueix segons una normal de valor esperat 0 graus i variança desconeguda. Quina és la probabilitat que la temperatura sigui superior a 0 graus?

EXERCICI 8

El pes del productes d'una fàbrica es distribueix segons una distribució Normal de valor esperat 1,7 Kg i desviació típica de 100 grams. El control de qualitat de la fàbrica rebutja aquells productes que tinguin unpes que difereixi en més de 300 grams del pes mitjà. En aquestes condicions, quina és la proporció de productes rebutjats?

EXERCICI 8

La fàbrica MINO té una màquina que envasa els seu producte en bosses de 200gr. El pes dels envasos, x, segueix una distribució Normal amb mitjana 198 gr. I desviació 10 gr. Quiné s el valor de “a” que verifica la següent relació:

$$p(a \leq x \leq 225) = 0,6519$$

3. Distribució exponencial Exp(α)

Una variable aleatòria té distribució exponencial de paràmetre α:

$$f(x) = \alpha e^{-\alpha x} \text{ si } x > 0$$

Esperança i variància:

$$E(X) = 1/\alpha$$
$$\sigma^2 = 1/\alpha^2$$

Propietat (la distribució exponencial no té memòria):

$$P(X < a + b | X > a) = P(X < b)$$

1.3 APROXIMACIONS A NORMAL

- Aproximació de la distribució Binomial per la Normal

Si X_n són variables aleatòries binomials $B(n, p)$ llavors la successió de variables:

$$E(X) = \mu = np$$

$$Var(x) = \sigma^2 = npq$$

té límit la variable normal tipificada Z quan $n \rightarrow \infty$.

$$z = \frac{X_n - np}{\sqrt{np(1-p)}}$$

- Aproximació de la distribució Poisson per la Normal

Si X_n són variables aleatòries binomials $P(\lambda)$ llavors la successió de variables:

$$E(X) = \mu = \lambda$$

$$Var(x) = \sigma^2 = \lambda^2$$

té límit la variable normal tipificada Z quan $n \rightarrow \infty$.

$$z = \frac{X_n - \lambda}{\lambda}$$

1.4 TEOREMA CENTRAL DEL LÍMIT

El teorema del límit central estableix que si se tenim n variables aleatorias, X_1, X_2, ..., X_n, independients i amb idèntica distribució de mitjana . μ i varianza σ^2 , a mesura que creix n, la suma (i la mitjana) d'aquestes variables tendeix a seguir una distribució Normal.

$$\mu_n = n\mu$$

$$\sigma_n^2 = n\sigma^2$$

$$\sigma_n = \sqrt{n}\sigma$$