Modelització i Inferència

Grau en Matemàtica Computacional i Analítica de Dades (MATCAD) UAB

Tema 1: Introducció a l'Estadística

1.1. Primers conceptes
1.2. Els estadístics més usuals
1.3. Les distribucions mostrals dels estadístics més usuals
1.4. Distribucions mostrals asimptòtiques: mitjana i proporció
1.5. Els estadístics d'ordre i la seva distribució

Definició 1. *Una* **població** *és una v.a. X (fent la identificació).*

Definició 2. *Una* **mostra aleatòria** *de mida n de X és un conjunt de n v.a.'s $X_1, X_2, \ldots, X_n$ tals que:*

1. *$X_1, X_2, \ldots, X_n$ són v.a. independents.*
2. *Per a tot $A \subset \mathbb{R}$ es compleix que*

$$P(X_i \in A) = P(X \in A)$$

per a tot $i = 1, \ldots, n$ (és a dir, les variables X_i són còpies de la variable X, amb la mateixa distribució de probabilitat).

Dit d'una altra manera, una mostra d'una v.a. X és un conjunt de n variables aleatòries independents idènticament distribuïdes (v.a.i.i.d), amb la mateixa distribució d'X.

Recordeu que $X_1, \ldots, X_n$ són independents si els esdeveniments de la forma

$$\{X_1 \leq x_1\}, \ldots, \{X_n \leq x_n\}$$

són independents, per a tot $x_1, \ldots, x_n \in \mathbb{R}$.

Per exemple, si tenim només dues variables ($n = 2$), X_1 i X_2 són independents si

$$P(X_1 \leq x_1,\, X_2 \leq x_2) = P(X_1 \leq x_1)\, P(X_2 \leq x_2) \quad \text{per a tot } x_1,\, x_2 \in \mathbb{R}\,.$$

Si tenim $n = 3$ variables, X_1, X_2 i X_3 són independents si per a tot x_1, x_2 i x_3 de $\mathbb{R}$,

$$\begin{aligned}
P(X_1 \leq x_1,\, X_2 \leq x_2) &= P(X_1 \leq x_1)\, P(X_2 \leq x_2) \\
P(X_1 \leq x_1,\, X_3 \leq x_3) &= P(X_1 \leq x_1)\, P(X_3 \leq x_3) \\
P(X_2 \leq x_2,\, X_3 \leq x_3) &= P(X_2 \leq x_2)\, P(X_3 \leq x_3) \\
P(X_1 \leq x_1,\, X_2 \leq x_2,\, X_3 \leq x_3) &= P(X_1 \leq x_1)\, P(X_2 \leq x_2)\, P(X_3 \leq x_3)
\end{aligned}$$

Si les variables són discretes, la definició d'independència és equivalent a que els esdeveniments

$$\{X_1 = x_1\}, \ldots, \{X_n = x_n\} \quad \text{siguin independents}$$

per a tot x_1 àtom d'X_1, …, x_n àtom d'X_n (pels valors que no són àtoms s'obté que la probabilitat de la intersecció és producte de probabilitats òbviament, ja que dóna $0 = 0$).

Comentaris:

- Mostra aleatòria de mida n d'una població finita sense reemplaçament o reposició agafada a l'atzar (tots els individus tenen la mateixa probabilitat de ser seleccionats).

$X_1, \ldots, X_n$ són els resultats de la variable X per a cada element seleccionat. Aleshores, **no** formen una mostra aleatòria perquè **no** són independents.

Exemple: Població amb 18 estudiants, dels quals:
3 estan a primer curs, 4 a segon, 5 a tercer i 6 a quart.
$X =$ "curs", és una variable discreta que pren els valors $\{1, 2, 3, 4\}$.
Funció de probabilitat de X:

$$P(X=1) = \frac{3}{18}, \quad P(X=2) = \frac{4}{18}, \quad P(X=3) = \frac{5}{18}, \quad P(X=4) = \frac{6}{18}.$$

Si prenem una mostra de dos estudiants **sense reemplaçament**, sigui
$X_1 =$ curs del primer estudiant seleccionat, i
$X_2 =$ curs del segon estudiant seleccionat.
Aleshores, X_1, X_2 **no** és una mostra aleatòria de mida $n = 2$ de la v.a. X, ja que no són v.a. independents.

En canvi sí que és cert que X_1 i X_2 tenen la mateixa distribució que X.

1) X_1 té la mateixa distribució que X.

És evident ja que estem seleccionant un dels 18 estudiants a l'atzar.

2) X_2 té la mateixa distribució que X (no és evident).

Per exemple, per la Fórmula de la Probabilitat Total:

$$\begin{aligned} P(X_2=1) &= P(X_2=1/X_1=1)\,P(X_1=1)+P(X_2=1/X_1\neq 1)\,P(X_1\neq 1) \\ &= \frac{2}{17}\times\frac{3}{18}+\frac{3}{17}\times\frac{15}{18}=\frac{3\,(2+15)}{17\times 18}=\frac{3\times 17}{17\times 18}=\frac{3}{18}=P(X=1). \end{aligned}$$

(la resta es fa anàlogament)

3) X_1 i X_2 **no** són independents (intuïtivament és clar!). En efecte,

$$P(X_1=1,X_2=1)=P(X_2=1/X_1=1)\,P(X_1=1)=\frac{2}{17}\times\frac{3}{18}=\frac{1}{51}$$

no dóna el mateix que

$$P(X_1=1)\,P(X_2=1)=\frac{3}{18}\times\frac{3}{18}=\frac{9}{18^2}=\frac{1}{36}$$

- Mostra aleatòria de mida n d'una població finita amb reemplaçament o reposició agafada a l'atzar (tots els individus tenen la mateixa probabilitat de ser seleccionats).

En aquest cas **sí** que és cert que les variables $X_1, \ldots, X_n$ formen una mostra aleatòria de X.

Cas habitual:

la mida de la mostra, n, és petita comparada amb la mida de la població. Encara que el mostreig es fa sense reemplaçament, hi ha poca dependència entre les variables de la mostra i es treballa com si fossin independents (con si es fes amb reemplaçament).

Exercici: Repetiu l'exemple anterior amb 1800 estudiants,
dels quals 300 estan a primer curs, 400 a segon, 500 a tercer i 600 a quart.

X_1 i X_2 continuen sent dependents, però ara el seu grau de dependència ha minvat, ja que la diferència entre

$$P(X_1 = 1, X_2 = 1) \quad \text{i} \quad P(X_1 = 1)\, P(X_2 = 1)\,,$$

per exemple, és molt menor que abans.

- Mostra aleatòria de mida n per a representar els resultats d'un experiment.

Fem n determinacions diferents del valor d'una v.a. X en un experiment.

X pot ser la quantitat de greix en una porció de 100 grams de formatge de certa marca, o podria ser si el formatge està fet amb llet de vaca o d'ovella.

L'experiment pot consistir en trobar els valors de X en n items similars (porcions de 100 grams del formatge d'aquesta marca), o bé en respetir n vegades l'observació del valor de X en el mateix item (la mateixa porció).

Com que els resultats de l'experiment són intrínsecament aleatoris, s'intenta que les n observacions de la variables X es facin sempre en les mateixes condicions.

D'aquesta manera, es pot considerar que les n observacions corresponen a una mostra de mida n de X, $X_1, \ldots, X_n$.

Observació:

Distingirem entre les variables aleatòries que defineixen la mostra, $X_1, \ldots, X_n$, i les realitzacions d'aquestes variables, que denotarem per $x_1, \ldots, x_n$, que són nombres reals.

Recordeu que els **paràmetres** són característiques numèriques poblacionals que solen ser desconegudes.

Els paràmetres més habituals son:

• Si X quantitativa (numèrica):

mitjana poblacional, $\mu = E(X)$, també denominada *valor esperat* o *esperança*.

variància poblacional, $\sigma^2 = \mathrm{Var(X)}$.

desviació típica poblacional, $\sigma = \sqrt{\mathrm{Var(X)}}$.

• Si X qualitativa dicotòmica (binària):

proporció poblacional de presència de l'atribut, p.

• Si X general:

paràmetre qualsevol de la llei o distribució de probabilitat de X, θ.

1.2. Els estadístics més usuals

Definició 3. *Donada una mostra aleatòria $X_1, \ldots, X_n$ de X, un* **estadístic** *és una funció d'aquestes variables, i potser també de constants conegudes.*

Exemples:

La mitjana mostral $\quad \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$

La variància mostral (corregida)

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

La variància mostral sense corregir

$$S'^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

Observació: Clarament S'^2 i S^2 s'obtenen fàcilment una a partir de l'altra:

$$S'^2 = \frac{n-1}{n} S^2 \quad \text{i} \quad S^2 = \frac{n}{n-1} S'^2.$$

En el cas (estrany) que coneguéssim la mitjana poblacional μ de X, podem considerar també aquesta altra "espècie" de variància mostral que anomenarem quasivariància mostral:

$$\widetilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2,$$

Definició 4. *Un* **estimador** *és un estadístic que es fa servir per a estimar un determinat paràmetre.*

Notació:

Un estadístic que s'usa per a estimar el paràmetre θ serà denotat per $\hat{\theta}$. Així per exemple si volem usar la mitjana mostral per a estimar la mitjana poblacional μ d'una certa variable aleatòria X, com és habitual, escriurem

$$\hat{\mu} = \overline{X}.$$

Si usem la variància mostral per a estimar la variància poblacional, σ^2, escriurem

$$\widehat{\sigma^2} = S^2.$$

Distingirem entre l'estimador, que és una variable aleatòria (lletra majúscula), i l'estimació, que és un valor concret, la seva realització (en minúscula). L'estimació s'obté amb la mateixa expressió de l'estimador, però per comptes de fer servir les variables de la mostra $X_1, \ldots, X_n$, fent servir les seves realitzacions $x_1, \ldots, x_n$.

- Estimador de μ: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (v.a) Estimació: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (nombre real)

- Estimador de σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ (v.a.)

  Estimació: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ (nombre positiu).

Exemple: EL TALLER MECÀNIC

Un taller de cotxes cobra 40, 45 i 50 euros per la posada a punt dels cotxes segons tinguin 2 o 3 cilindres en línia, 4 cilindres en línia, o bé 5 o 6 cilindres en línia, respectivament. El 20% de les posades a punt corresponen a cotxes de 2 o 3 cilindres, el 30% a cotxes de 4 cilindres i el 50% restant a cotxes amb 5 o 6.

Sigui $X =$ "ingressos per una posada a punt al taller". Clarament, X pren els valors x amb probabilitats $p_X(x)$ que anotem en la taula següent:

x	$p_X(x)$
40	0.20
45	0.30
50	0.50

Podem calcular la mitjana poblacional d'aquesta variable:

$$\mu_X = E(X) = 40 \times 0.20 + 45 \times 0.30 + 50 \times 0.50 = 46.5$$

i la seva variància poblacional:

$$\begin{aligned}\sigma_X^2 &= E(X^2) - E(X)^2 \\ &= \left(40^2 \times 0.20 + 45^2 \times 0.30 + 50^2 \times 0.50\right) - 46.5^2 = 15.25\end{aligned}$$

Exemple (continuació):

Sigui X_1, X_2 una mostra aleatòria de mida $n = 2$ de X.

Siguin x_1, x_2 les possibles realitzacions (taula), amb les corresponents mitjanes $\overline{x}$.

x_1	x_2	$P(X_1 = x_1, X_2 = x_2)$	$\overline{x} = \frac{x_1+x_2}{2}$
40	40	$0.04 = 0.2 \times 0.2$	40
40	45	$0.06 = 0.2 \times 0.3$	42.5
40	50	$0.10 = 0.2 \times 0.5$	45
45	40	$0.06 = 0.3 \times 0.2$	42.5
45	45	$0.09 = 0.3 \times 0.3$	45
45	50	$0.15 = 0.3 \times 0.5$	47.5
50	40	$0.10 = 0.5 \times 0.2$	45
50	45	$0.15 = 0.5 \times 0.3$	47.5
50	50	$0.25 = 0.5 \times 0.5$	50

De la taula deduïm que la v.a. $\overline{X} = \frac{X_1+X_2}{2}$, pren els següents valors, i amb les següents probabilitats:

$\overline{x}$	$P(\overline{X} = \overline{x})$
40	0.04
42.5	0.12
45	0.29
47.5	0.30
50	0.25

Exemple (continuació):

Per exemple, el valor 45 es pot obtenir de tres maneres diferents i, per tant, la probabilitat d'obtenir aquest valor s'obtindrà sumant les tres probabilitats corresponents de la taula anterior,

$$P(X_1 = 40,\, X_2 = 50) + P(X_1 = 45,\, X_2 = 45) + P(X_1 = 50,\, X_2 = 40) \\ = 0.1 + 0.09 + 0.1 = 0.29\,.$$

Tenim que:

1) La distribució de $\overline{X}$ és diferent de la de X.

2) Calculem l'esperança d'aquesta nova variable:

$$\mu_{\overline{X}} = E(\overline{X}) = 40 \times 0.04 + 42.5 \times 0.12 + 45 \times 0.29 + 47.5 \times 0.20 + 50 \times 0.25 = 46.5.$$

L'esperança, en canvi, coincideix amb la de la v.a. X.

Això és un fet absolutament general, com veurem a la secció següent.

3) Si calculem la seva variància, tenim:

$$\sigma^2_{\overline{X}} = E(\overline{X}^2) - E(\overline{X})^2 \\ = \left(40^2 \times 0.04 + 42.5^2 \times 0.12 + 45^2 \times 0.29 + 47.5^2 \times 0.30 + 50^2 \times 0.25\right) \\ - 46.5^2 = 7.625$$

i observem que és la variància de X dividida per $n = 2$ (i, per tant, és menor).

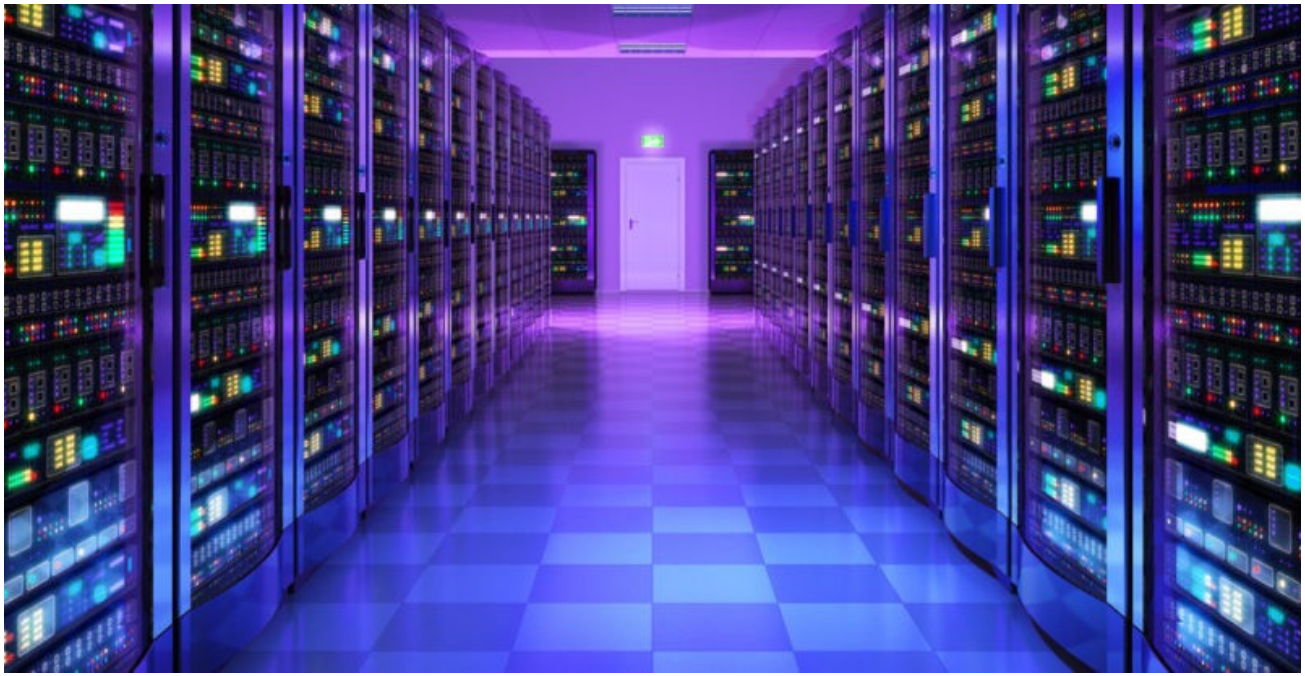
Això també és un fet absolutament general, com veurem a la secció següent.

Exemple: LA VELOCITAT DELS SERVIDORS

Esteu treballant en un projecte per a una empresa tecnològica que vol analitzar els temps de resposta dels seus servidors per optimitzar el rendiment i millorar l'experiència dels usuaris.

Recolliu els temps de resposta (en mil·lisegons, ms.) durant un període de temps:

120,135,150,125,140,130,145,160,155,130,135,140,150,145,160



Exemple
(continuació):

Aquests valors es poden considerar la realització d'una mostra de mida $n = 15$ d'una variable X amb la que identifiquem la població, que és el temps de resposta dels servidors de l'empresa.

A partir de la (realització d'aquesta) mostra, trobarem una estimació de

(la mitjana) $\mu = E(X)$, (la variància) $\sigma^2 = Var(X)$ i (la desviació) $\sigma = \sqrt{Var(X)}$,

que són paràmetres desconeguts.

- Estimació de μ: $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{2120}{15} = 141.33$ ms.
- Estimació de σ^2:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\,\overline{x}^2 = \frac{301750}{14} - \frac{15}{14}\,141.33^2 = 151.67 \text{ ms.}^2$$

- Estimació de σ: $s = \sqrt{s^2} = \sqrt{151.67} = 12.3153$ ms.

Exemple
(continuació):

Tenim una estimació dels temps promig que triguen els servidors de l'empresa:

$$\overline{x} = 141.33 \text{ ms.}$$

i també de les seves variància i desviació:

$$s^2 = 151.67 \text{ ms.}^2 \quad \text{i} \quad s = 12.3153 \text{ ms.}$$

Però sabem que aquests valors depenen de la (realització de la) mostra que hem obtingut, $x_1, \ldots, x_n$.

Si tornessim a agafar una mostra, obtindríem una realització diferent i les estimacions també serien diferents!! Però, com de diferents?

El necessitem per a poder contestar aquesta pregunta, i per tant poder treure conclusions i prendre decisions a partir de les estimacions dels paràmetres de la població que hem obtingut, és:

conèixer la distribució de les v.a. estimadors tals que les seves realitzacions són les estimacions obtingudes ($\overline{X}$, S^2, S).

D'aquesta manera, podrem saber com de diferents de les estimacions obtingudes serien les altres estimacions que podríem obtenir, i d'aquesta manera, fins a quin punt ens podem "refiar" d'elles.

Exemple: ÉS EFECTIVA LA CAMPANYA?

Una empresa de comerç electrònic vol avaluar l'efectivitat d'una nova campanya de màrqueting i està interessada en la proporció de clients que fan una compra després de rebre la campanya.

Dades: Sabem que després d'enviar la campanya a 1000 clients, 300 fan una compra i 700 no la fan. Què podem dir a partir d'aquesta informació?

Exemple
(continuació):

En aquest exemple, identifiquem la població dicotòmica dels clients que reben la campanya de màrqueting amb una v.a. $X \sim B(p)$, on p és la proporció (teòrica) de clientes que compren d'entre els que reben la campanya (als quals assignem, per tant, el valor 1).

El 0 correspon als clients que no compren d'entre els que reben la campanya.

Si considerem una mostra de mida $n = 1000$ de la població: $X_1, \ldots, X_n$, les dades que ens donen corresponen a una realització de la mostra, $x_1, \ldots, x_n$. Totes les x_i són 0 o 1, i ens diuen que hi ha exactament 300 que són 1, i 700 que són 0.

A partir de la (realització d'aquesta) mostra, trobem una estimació de la proporció teòrica de 1, p, que és un paràmetre desconegut:

$$\hat{\hat{p}} = \frac{300}{1000} = 0.3$$

Fixeu-vos que l'estimació $\hat{\hat{p}}$ coincideix amb $\overline{x}$, i també que $p = E(X)$, que habitualment denotem per μ.

El corresponent estimador (v.a.) del qual n'és una realització serà, per tant, $\overline{X}$, i ho denotem per $\hat{p}$.

Paràmetre: $(\mu =)\ p$, estimador: $\hat{p} = \overline{X}$, estimació: $\hat{\hat{p}} = \overline{x}$

Exemple
(continuació):

Hem obtingut una estimació de la proporció dels clients que compren d'entre els que reben la campanya de màrqueting, que és una mesura de la efectivitat de la campanya:

$$\hat{\hat{p}} = 0.3$$

Però, com amb l'exemple anterior, sabem que aquest valor varia amb la (realització de la) mostra que tinguem, $x_1, \ldots, x_n$.

Si tornessim a agafar una mostra, obtindríem una realització diferent i l'estimació també seria diferent!! Com abans, ens preguntem: com de diferent?

I com abans, el necessitem per a poder contestar aquesta pregunta, i treure conclusions a partir de l'estimació del paràmetre p de la població que hem obtingut, és:

conèixer la distribució de l'estimador $\hat{p}$, que és la v.a. tal que la seva realització ha estat l'estimació obtinguda $\hat{\hat{p}} = 0.3$

Com abans així podrem saber com de diferents de l'estimació obtinguda serien les altres estimacions que podríem obtenir, i fins a quin punt ens podem "refiar" d'ella per a poder prendre decisions sobre l'efectivitat (o no) de la campanya de màrqueting.

1.3. Les distribucions mostrals dels estadístics més usuals

Sigui $X_1, \ldots, X_n$ una mostra de mida n d'una variable aleatòria X tal que

$$E(X) = \mu \quad \text{i} \quad \text{Var}(X) = \sigma^2 .$$

Definició 5. *Donat un estadístic funció de la mostra $X_1, \ldots, X_n$, que és una variable aleatòria, la seva distribució (o llei) s'anomena distribució mostral de l'estadístic.*

• Propietats de la llei de la mitjana mostral $\overline{X}$

La mitjana mostral és una variable aleatòria que té una distribució que depèn de la distribució de la variable X. En general:

Proposició 1:

$$\mu_{\overline{X}} = E(\overline{X}) = \mu, \quad \sigma^2_{\overline{X}} = \text{Var}(\overline{X}) = \frac{\sigma^2}{n} \quad \left(\text{i, per tant, } \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\right)$$

Demostració:

Provem primer que $E(\overline{X}) = \mu$. En efecte,

$$E(\overline{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}\, n\, \mu = \mu,$$

on hem usat primer les propietats de linealitat de l'esperança i després que totes les X_i tenen la mateixa esperança que la v.a. X.

No fem servir la seva independència per a demostrar això!

Vegen ara que

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}\,.$$

En efecte,

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n},$$

on hem usat

1) primer que si a és una constant i Y una variable aleatòria, aleshores

$$\boxed{\mathrm{Var}(a\,Y) = a^2\,\mathrm{Var}(Y)}$$

2) Després hem utilitzat, ara **sí**, que *per a variables aleatòries independents*, la variància de la suma és la suma de les variàncies

(recordeu que si les variables no són independents, això no té perquè ser cert).

3) Finalment, hem usat que totes les variables X_i tenen la mateixa variància que X, és a dir, σ^2 . □

• Cas particular: $X \sim B(p)$, $p \in (0,\, 1)$ (Població dicotòmica).

En aquest cas,

$$n\,\overline{X} = \sum_{i=1}^{n} X_i \sim B(n,\, p)$$

Per la Proposició 1:

$$\mu_{\overline{X}} = E(\overline{X}) = \mu = E(X) = p \quad \Longrightarrow E(n\,\overline{X}) = n\,E(\overline{X}) = n\,p$$

que correspon exactament a l'esperança de la binomial.

Anàlogament, per la mateixa Proposició 1:

$$\sigma^2_{\overline{X}} = Var(\overline{X}) = \frac{\sigma^2}{n} = \frac{Var(X)}{n} = \frac{p\,(1-p)}{n}$$

$$\Longrightarrow Var(n\,\overline{X}) = n^2\,Var(\overline{X}) = n^2\,\frac{p\,(1-p)}{n} = n\,p\,(1-p)$$

que és la variància de la binomial. I, per tant,

$$\sigma_{n\,\overline{X}} = \sqrt{n\,p\,(1-p)}\,.$$

• Cas particular: $X \sim N(\mu,\, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ (Població Normal).

En aquest cas es pot assegura que la distribució mostral de l'estadístic $\overline{X}$ també és normal.

Proposició 2: *Si* $X \sim N(\mu, \sigma^2)$, *aleshores*

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad o,\ equivalentment, \quad Z = \frac{\overline{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0,\, 1)$$

Demostració:

La Proposició 1 ens diu que $E(\overline{X}) = \mu$ i $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$.

Que la distribució de $\overline{X}$ sigui també normal és conseqüència del fet que les combinacions lineals de variables aleatòries independents amb distribució normal també tenen distribució normal.

Això és un resultat de la **Teoria de la Probabilitat** que es demostra calculant la funció generatriu de moments que, recordeu, caracteritza la distribució.

Recordeu que la funció generatriu de moments d'una variable $X \sim N(\mu,\, \sigma^2)$ és

$$\varphi_X(t) = E\left(e^{t\,X}\right) = \int_{-\infty}^{+\infty} e^{t\,x}\, \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\, e^{-(x-\mu)^2/2\,\sigma^2}\, dx = \cdots = e^{\mu\,t + \frac{1}{2}\,\sigma^2\,t^2}\,.$$

Veurem que la funció generatriu de moments de la variable aleatòria $\overline{X}$ és igual a la funció generatriu de moments d'una normal amb mitjana μ i variància σ^2/n en t, que és:

$$e^{\mu\,t + \frac{1}{2}\,\frac{\sigma^2}{n}\,t^2}\,.$$

En efecte,

$$\varphi_{\overline{X}}(t) = E(e^{t\,\overline{X}}) = E\left(e^{\frac{t}{n}\sum_{i=1}^{n} X_i}\right) = E\left(\prod_{i=1}^{n} e^{\frac{t}{n}X_i}\right) = \prod_{i=1}^{n} E\left(e^{\frac{t}{n}X_i}\right) = \prod_{i=1}^{n} \varphi_{X_i}(t/n),$$

on hem fet servir que per la independència, l'esperança del producte és producte d'esperances.

Aleshores, fent servir que coneixem la f.g.m. de X_i, que és la mateixa que la de $X \sim N(\mu,\, \sigma^2)$, tindrem que

$$\varphi_{\overline{X}}(t) = \prod_{i=1}^{n} \varphi_X(t/n) = (\,\varphi_X(t/n)\,)^n = e^{n\left(\mu\,\frac{t}{n} + \frac{1}{2}\,\sigma^2\,(t/n)^2\right)} = e^{\mu\,t + \frac{1}{2}\,\frac{\sigma^2}{n}\,t^2} = \varphi_{N(\mu,\,\frac{\sigma^2}{n})}(t)$$

i, per tant, $\overline{X} \sim N(\mu,\, \frac{\sigma^2}{n})$. $\square$

• Propietats de la llei de la variància mostral (corregida) S^2, sense corregir S'^2 i quasivariància $\widetilde{S}^2$

En general, tenim que les seves esperances són:

Proposició 3:

$$E(\widetilde{S}^2) = \sigma^2,\; E(S^2) = \sigma^2,\; E(S'^2) = \frac{n-1}{n}\,\sigma^2$$

Demostració:

Provarem la propietat primera i la tercera. La segona serà conseqüència immediata de la tercera, degut a que $S^2 = \frac{n}{n-1}\, S'^{\,2}$.

Comprovem la primera, suposant μ coneguda:

$$E(\widetilde{S}^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) = \frac{1}{n}\sum_{i=1}^{n}E((X_i-\mu)^2) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(X) = \frac{1}{n}\, n\,\sigma^2 = \sigma^2.$$

Per a comprovar la tercera usarem el fet que per a qualsevol variable Y es té que

$$\mathrm{Var}(Y) = E(Y^2) - E(Y)^2, \quad \text{d'on tenim} \quad E(Y^2) = \mathrm{Var}(Y) + E(Y)^2. \qquad (1)$$

També usarem la següent expressió de la variància mostral:

$$S'^{\,2} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 \;-\; \overline{X}^2.$$

Llavors,

$$E(S'^{\,2}) = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 \;-\; \overline{X}^2\right) = \frac{1}{n}\sum_{i=1}^{n}E(X_i^2) \;-\; E(\overline{X}^2).$$

Aplicant la igualtat (1) a les variables $Y = X_i$ i a $Y = \overline{X}$, obtenim

$$\begin{aligned} E(S'^2) &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Var}(X_i) + E(X_i)^2\right) - \left(\mathrm{Var}(\overline{X}) + E(\overline{X})^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2+\mu^2\right) - \left(\frac{\sigma^2}{n}+\mu^2\right) = \frac{1}{n}\,n\,(\sigma^2+\mu^2) - \frac{\sigma^2}{n} - \mu^2 \\ &= \sigma^2\left(1-\frac{1}{n}\right) = \frac{n-1}{n}\,\sigma^2. \quad \square \end{aligned}$$

Com en el cas de la mitjana mostral, és important a les aplicacions conèixer les distribucions dels estimadors de la variància poblacional en el cas que la variable X tingui distribució Normal.

Proposició 4. *Si* $X \sim N(\mu, \sigma^2)$, *aleshores*

$$\frac{n\tilde{S}^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2_n$$

on χ^2_n *denota la distribució khi-quadrat, amb paràmetre (graus de llibertat)* n.

(es fa servir quan μ és coneguda)

Demostració:

$$\frac{n\widetilde{S}^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 .$$

Com que cada variable $\frac{X_i-\mu}{\sigma}$ té distribució $N(0,1)$ i són independents entre elles, la demostració és evident a partir de la definició de χ^2_n de l'Apèndix. □

L'estimador $\widetilde{S}^2$ no és gaire utilitzat ja que habitualment desconeixem el valor de μ.

El següent resultat és un dels teoremes més importants de l'Estadística. Ens dóna la distribució de S^2 (equivalentment, de S'^2).

Teorema de Fisher.
Si $X_1,\ldots,X_n$ és una mostra aleatòria d'una variable aleatòria $X \sim N(\mu,\sigma^2)$, aleshores:

(a) $\overline{X}$ i S^2 són independents.

(b) A més,

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 \sim \chi^2_{n-1}.$$

(es fa servir quan μ és desconeguda)

Demostració (idea):

(a) La seva demostració supera el nivell d'aquest curs. **Important:** aquest resultat només és cert quan la variable X de la que procedeix la mostra té distribució normal. Si no és així, aquestes dues v.a. no són independents, en general.

(b) Idea: es pot comprovar fàcilment que es compleix la següent identitat:

$$\frac{n\,\tilde{S}^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \;=\; \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 + \frac{n}{\sigma^2}\,(\overline{X}-\mu)^2 \;=\; \frac{n-1}{\sigma^2}\,S^2 + \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2,$$

fent servir que

$$(X_i-\mu)^2 = (X_i-\overline{X}+\overline{X}-\mu)^2 = (X_i-\overline{X})^2 + (\overline{X}-\mu)^2 + 2\,(X_i-\overline{X})\,(\overline{X}-\mu)$$

i que $\sum_{i=1}^{n}(X_i-\overline{X}) = 0$.

Per la Proposició 4, el membre de l'esquerra d'aquesta igualtat, $\frac{n\,\tilde{S}^2}{\sigma^2}$, té distribució χ^2_n, i el segon sumand del membre dret, $\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$, és una $N(0,\,1)$ elevada al quadrat, és a dir, és una χ^2_1. D'altra banda, per **(a)**, els dos sumands del membre dret, $\frac{(n-1)S^2}{\sigma^2}$ i $\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$, són v. a. independents. Així, tenim dues v.a. independents, una d'elles χ^2 amb 1 grau de llibertat i l'altra que no sabem quina distribució té, que quan les sumem ens dóna una distribució χ^2 amb n graus de llibertat. Es pot demostrar (i és molt raonable!) que això implica que el sumand amb distribució desconeguda tingui distribució χ^2 amb $n-1$ graus de llibertat, cosa que acaba la demostració. $\square$

Fins ara hem suposat que σ era coneguda. Què en podem dir quan no és així?

Proposició 5. *Si* $X \sim N(\mu, \sigma^2)$, *aleshores:*

$$T = \frac{\overline{X} - \mu}{\left(\frac{S}{\sqrt{n}}\right)} \sim t_{n-1}$$

on $S = \sqrt{S^2}$ *(amb signe positiu) i* t_{n-1} *denota la distribució t d'Student amb paràmetre (graus de llibertat)* $n - 1$.

Demostració:

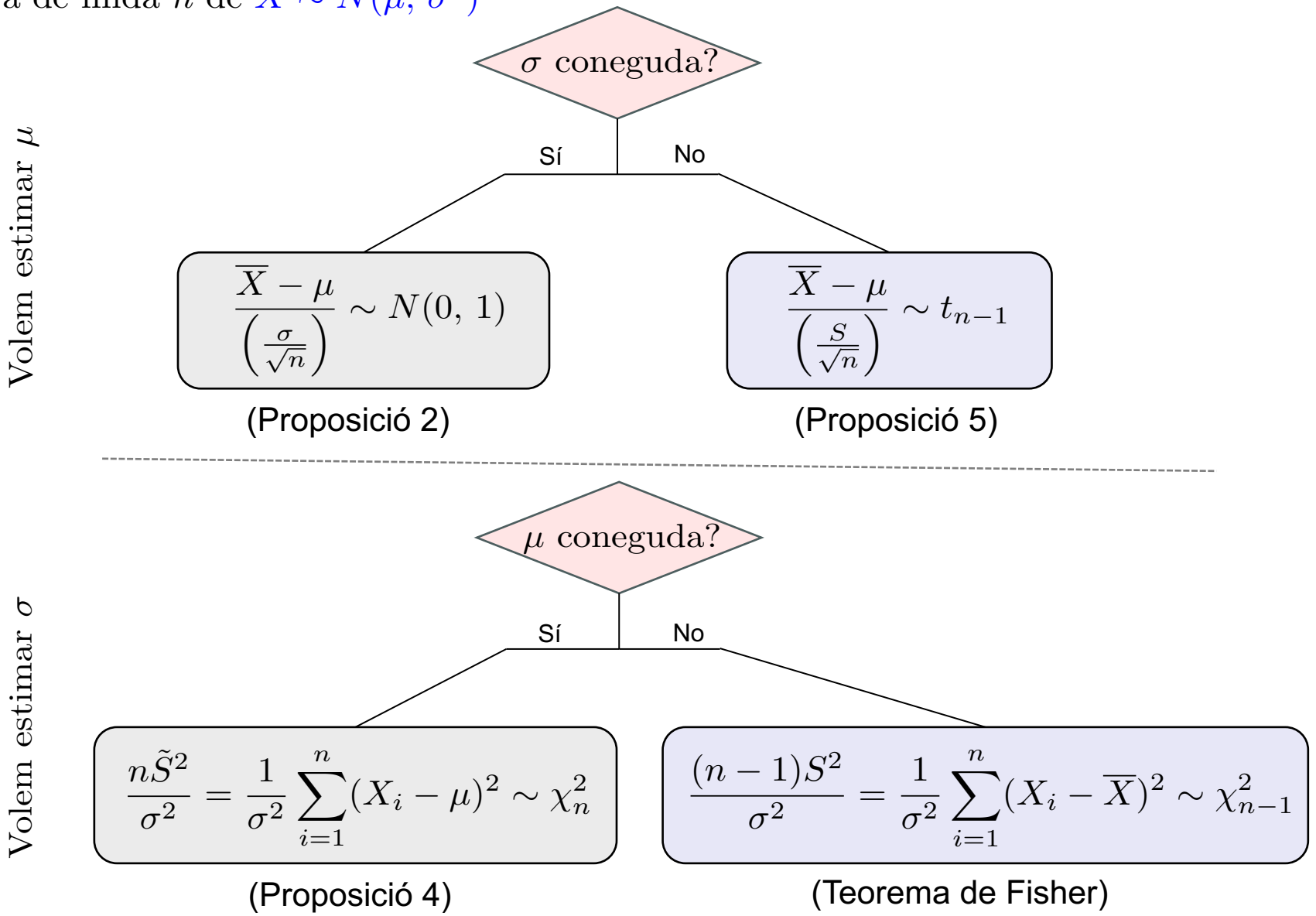
Podem escriure

$$\frac{\overline{X} - \mu}{\left(\frac{S}{\sqrt{n}}\right)} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)\,S^2/\sigma^2}{n-1}}}\,.$$

Així, hem expressat $\frac{\overline{X}-\mu}{\left(\frac{S}{\sqrt{n}}\right)}$ com el quocient entre una v.a. amb distribució $N(0,\,1)$ per la Proposició 2 i l'arrel quadrada d'una variable, $(n-1)\,S^2/\sigma^2$, que té distribució χ^2_{n-1} (ho sabem pel Teorema de Fisher) dividida pels seus graus de llibertat, essent aquestes dues variables independents (també pel Teorema de Fisher).

Per tant, té distribució t de Student amb $n-1$ graus de llibertat. □

Mostra de mida n de $X \sim N(\mu, \sigma^2)$



Distribució asimptòtica de la mitjana

Sigui $X_1, \ldots, X_n$ una mostra de mida n d'una variable aleatòria X amb llei qualsevol tal que

$$E(X) = \mu \quad \text{i} \quad \text{Var}(X) = \sigma^2 \, .$$

Si denotem per $\overline{X}_n$ la mitjana de la mostra de mida n, és a dir, $\overline{X}_n = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i \, ,$

la Proposició 1 ens diu quina és la mitjana i la desviació de $\overline{X}_n$:

$$\mu_{\overline{X}_n} = \mu, \quad \sigma^2_{\overline{X}_n} = \frac{\sigma^2}{n} \quad \left(\text{i, per tant, } \sigma_{\overline{X}_n} = \frac{\sigma}{\sqrt{n}}\right)$$

A més, el Teorema Central del Límit (TCL) ens assegura que si n és prou gran, aleshores, $\overline{X}_n$ té distribució normal i, per tant,

$$\boxed{\overline{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)}$$

Equivalentment, la mitjana estandaritzada és aproximadament $N(0,\,1)$, és a dir,

$$\boxed{Z_n = \frac{\overline{X}_n - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \approx N(0,\,1)}$$

També, pel Teorema Central del Límit (TCL), es compleix que si n és prou gran i σ és desconeguda, aleshores,

$$\boxed{\frac{\overline{X}_n - \mu}{\left(\frac{S_n}{\sqrt{n}}\right)} \approx N(0,\,1)}$$

on S_n denota la desviació mostral de la mostra de mida n.

A la pràctica, per a la majoria de distribucions X (comentarem a part el cas de la distribució de Bernoulli) l'aproximació de la llei de la mitjana mostral per la normal es considera prou bona per a $n \geq 30$.

Mostra de mida n de X amb distribució qualsevol

$E(X) = \mu, \quad Var(X) = \sigma^2$

n qualsevol

$$\mu_{\overline{X}_n} = \mu, \quad \sigma^2_{\overline{X}_n} = \frac{\sigma^2}{n} \quad \left(\text{i, per tant, } \sigma_{\overline{X}_n} = \frac{\sigma}{\sqrt{n}}\right)$$

(Proposició 1)

Volem estimar μ

$n \geq 30$

σ coneguda

$$\frac{\overline{X}_n - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \approx N(0, 1)$$

σ desconeguda

$$\frac{\overline{X}_n - \mu}{\left(\frac{S_n}{\sqrt{n}}\right)} \approx N(0, 1)$$

(Teorema Central del Límit)

Distribució asimptòtica de la proporció mostral

Cas particular: $X \sim B(p)$ amb $p \in (0,\, 1)$ (població dicotòmica).

Aleshores, si n és prou gran, el TCL ens diu que:

$$\overline{X}_n \approx N\big(\mu, \frac{\sigma^2}{n}\big) \Longleftrightarrow \frac{\overline{X}_n - \mu}{\big(\frac{\sigma}{\sqrt{n}}\big)} \approx N(0,\, 1) \Longleftrightarrow \frac{Y_n - n\,p}{\sqrt{n\,p\,(1-p)}} \approx N(0,\, 1)$$

amb $Y_n = \sum_{i=1}^{n} X_i \sim B(n,\, p)$, $\mu = E(X) = p$ i $\sigma^2 = Var(X) = p\,(1-p)$.

Aquest resultat, que s'obté aplicant el TCL pel cas particular $X \sim B(p)$, ja s'havia demostrat abans i es coneix com a Teorema de DeMoivre-Laplace.

Aquest teorema ens diu que la distribució de la Binomial s'aproxima per la Normal, sota certes condicions.

D'altra banda, tenint en compte que $\overline{X}_n$ en aquest cas és, precisament, la proporció mostral, que denotem per $\hat{p}_n$, el que ens diu el Teorema de De Moivre-Laplace és que sota certes condicions, la distribució de la proporció mostral es pot aproximar per la Normal. Concretament:

$$\boxed{\frac{\hat{p}_n - p}{\sqrt{\frac{p\,(1-p)}{n}}} \approx N(0,\, 1)}$$

Sota quines condicions es pot fer servir l'aproximació?

A la pràctica, no es fa automàticament per a $n \geq 30$.

Quan p és molt petita la distribució de la binomial és molt asimètrica i és més adequat usar una aproximació per la distribució de Poisson per al nombre d'èxits:

$$\boxed{B(n,\, p) \approx Pois(\lambda) \quad \text{amb} \quad \lambda = n\,p \quad \text{si} \quad n \geq 10,\, p < 0.05}$$

o per al nombre de fracassos (si p propera a 1, cosa que implica que la probabilitat de fracàs és molt petita).

La regla pràctica diu que si n és gran i p no és ni molt gran ni molt petita, per exemple si es compleix:

$$\boxed{n\,p \geq 5 \quad \text{i} \quad n\,(1-p) \geq 5}$$

ja es pot aproximar per la Normal.

Si a més es compleix la condició més restrictiva $\boxed{n\,p\,(1-p) \geq 5}$ l'aproximació és força bona, és excel.lent si $\boxed{n\,p\,(1-p) \geq 18}$ i encara més si $\boxed{n\,p\,(1-p) \geq 20}$

De fet, quan m'es gran sigui el producte $n\,p\,(1-p)$, millor serà l'aproximació!

Mostra de mida n de $X \sim B(p)$

$E(X) = \mu = p, \quad Var(X) = \sigma^2 = p\,(1-p)$

$\hat{p}_n = \overline{X}_n$

Volem estimar p

n qualsevol (usualment, petit)

$$n\,\hat{p}_n = \sum_{i=1}^{n} X_i \sim B(n,\, p)$$

(distribució exacta)

---

n gran (distribució asimptòtica)

$n\,p\,(1-p) \geq 18$

$$n\,\hat{p}_n = \sum_{i=1}^{n} X_i \approx N\Big(p,\, \frac{p\,(1-p)}{n}\Big)$$

(Teorema de DeMoivre-Laplace)

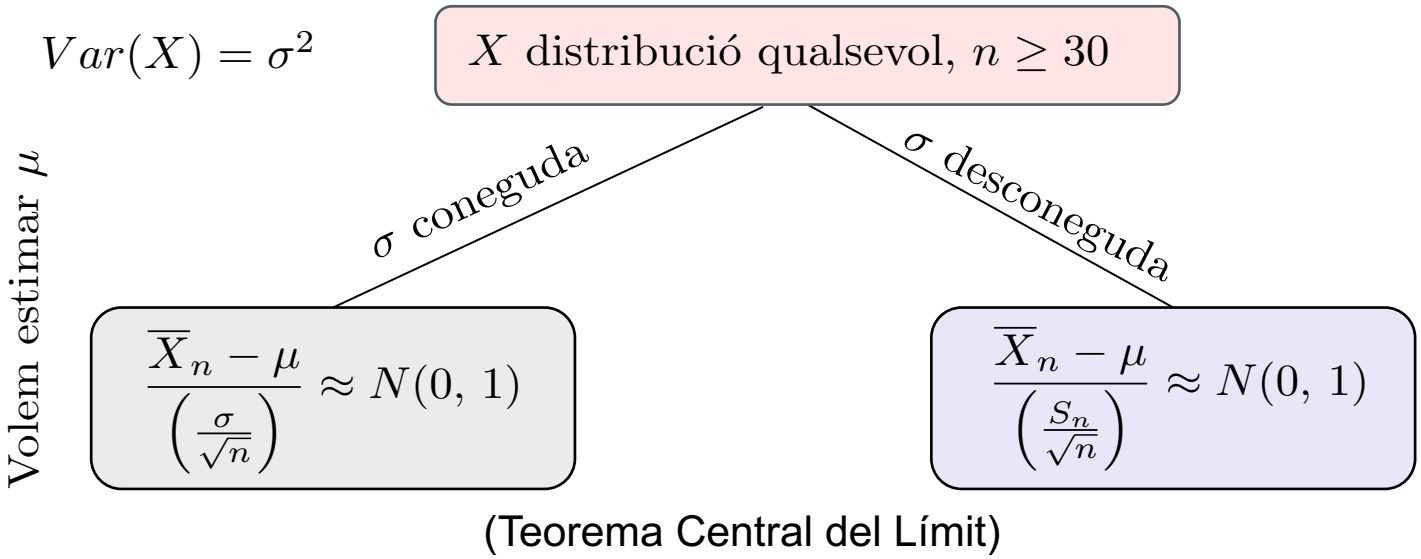
$n \geq 10,\, p < 0.05$

$$n\,\hat{p}_n = \sum_{i=1}^{n} X_i \approx Pois(\lambda),\, \lambda = n\,p$$

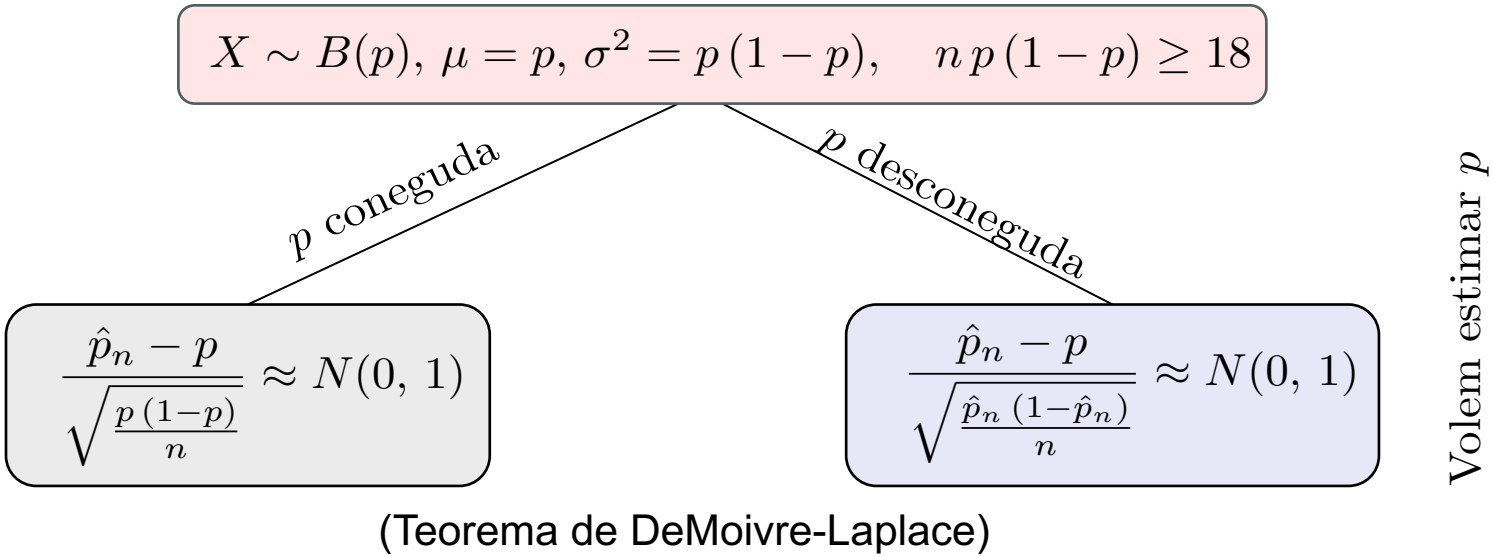
(Teorema de Poisson)

Mostra de mida n de X

$E(X) = \mu, \quad Var(X) = \sigma^2$



(Teorema Central del Límit)



(Teorema de DeMoivre-Laplace)

1.5. Els estadístics d'ordre i la seva distribució

Exemple introductori:

Suposeu que esteu treballant en una empresa que monitoritza la temperatura de les màquines que s'utilitzen en el procés de fabricació, mitjançant diversos sensors.

Recopileu dades de temperatura cada minut i es vol detectar quan una màquina està funcionant fora dels seus límits normals de temperatura, cosa que podria indicar un problema potencial.

Pas 1: Recollida i Ordenació de Dades

Obteniu una mostra de dades de temperatura d'una màquina específica durant un dia. Les temperatures (en graus Celsius) registrades són (n = 20 dades):

70, 72, 71, 68, 69, 70, 73, 75, 71, 69, 72, 74, 76, 73, 70, 68, 69, 71, 70, 72

Ordenem aquestes dades per obtenir els estadístics d'ordre:

68, 68, 69, 69, 70, 70, 70, 71, 71, 71, 72, 72, 72, 73, 73, 74, 75, 76

Pas 2: Càlcul dels Quartils i el Rang Interquartílic (IQR)

Utilitzem els estadístics d'ordre per calcular els quartils:

Primer quartil $Q_1 = X_{(n/4)} \approx X_{(5)} = 70$

Tercer quartil $Q_3 = X_{(3\,n/4)} = X_{(15)} = 73$

El rang interquartílic és $IQR = Q_3 - Q_1 = 73 - 70 = 3$

Pas 3: Detecció de valors anòmals

Un valor es considera "anòmal" o "sospitós" si es troba fora de l'interval:

$$[Q_1 - 1.5 \times IQR,\, Q_3 + 1.5 \times IQR] = [70 - 1.5 \times 3,\, 73 + 1.5 \times 3] = [66.5,\, 77.5]$$

Pas 4: Aplicació d'Algorismes de Aprenentatge Automàtic

Després de detectar i eliminar els valors anòmals, podem utilitzar les dades netes per entrenar un model de regressió o classificació que predigui la temperatura futura o classifiqui l'estat de la màquina (normal o anòmal).

Els estadístics d'ordre són els valors d'un conjunt de dades ordenats de menor a major. L'ús dels estadístics d'ordre és crucial en aquest exemple per a detectar valors anòmals de manera efectiva.

En l'aprenentatge automàtic, la qualitat de les dades d'entrenament és fonamental per obtenir models precisos. Els estadístics d'ordre ens ajuden a netejar les dades i a millorar la qualitat dels models.

Definició 6. *Donada una mostra de mida n de X, $X_1, \ldots, X_n$, els estadístics d'ordre són les variables aleatòries:*

$$
\begin{aligned}
X_{(1)} &= \min\{X_1, \ldots, X_n\} \quad (\textit{el mínim}) \\
X_{(2)} &= \min\{\, \{X_1, \ldots, X_n\} \backslash X_{(1)} \,\} \quad (\textit{el segon més petit}) \\
&\ldots \\
X_{(n)} &= \max\{X_1, \ldots, X_n\} \quad (\textit{el màxim})
\end{aligned}
$$

Exemples importants.

- La mediana és el valor que separa la meitat superior de la meitat inferior d'un conjunt de dades ordenades:

$$
Q_2 = \begin{cases} X_{((n+1)/2)} & \text{si } n \text{ és senar,} \\ \frac{1}{2}\left(X_{(n/2)} + X_{(n/2+1)}\right) & \text{si } n \text{ és parell.} \end{cases}
$$

- Els quartils divideixen les dades en quatre parts iguals.

  El primer quartil $Q_1 = X_{(n/4)}$ és el valor que separa el 25% inferior del 75% superior.

  El tercer quartil $Q_3 = X_{(3\,n/4)}$ separa el 75% inferior del 25% superior.

- El rang interquartílic (IQR) és la diferència $Q_3 - Q_1$.

  Aquest valor ens ajuda a entendre la dispersió de les dades centrals.

Distribució del màxim

Proposició 6. *Si X és una v.a. amb funció de distribució F_X, i $X_1, \ldots, X_n$ és una mostra de mida n de X, aleshores la funció de distribució de la v.a. màxim és, per a tot $t \in \mathbb{R}$:*

$$F_{X_{(n)}}(t) = \big(F_X(t)\big)^n$$

Demostració:

Usant tant la independència de les v.a. $X_1, \ldots, X_n$, com el fet que totes tenen la mateixa distribució que X, tenim que

$$F_{X_{(n)}}(t) = P(X_{(n)} \leq t) = P\big(\max(X_1, X_2, \ldots, X_n) \leq t\big)$$
$$= P(X_1 \leq t, X_2 \leq t, \ldots, X_n \leq t) = P(X_1 \leq t)\, P(X_2 \leq t) \cdots P(X_n \leq t) = (F_X(t))^n. \quad \square$$

Exemple: *Llei del màxim de $X \sim U(0, \theta)$ $(\theta > 0)$.*
La funció de densitat de X és: $f_X(x) = \frac{1}{\theta} I_{(0,\theta)}(x)$, i la funció de distribució de X és

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

Per la Proposició 6 tenim que

$$F_{X_{(n)}}(t) = (F_X(t))^n = \begin{cases} 0 & \text{si} \quad t \leq 0 \\ \frac{t^n}{\theta^n} & \text{si} \quad 0 \leq t \leq \theta \\ 1 & \text{si} \quad \theta \leq t\,. \end{cases}$$

I derivant, es té que la funció de densitat és

$$f_{X_{(n)}}(t) = \frac{n\, t^{n-1}}{\theta^n}\, I_{(0,\theta)}(t).$$

Distribució del mínim

Proposició 7. *Si X és una v.a. amb funció de distribució F_X, i $X_1, \ldots, X_n$ és una mostra de mida n de X, aleshores la funció de distribució de la v.a. mínim és, per a tot $t \in \mathbb{R}$:*

$$F_{X_{(1)}}(t) = 1 - \big(1 - F_X(t)\big)^n$$

Demostració:

$$\begin{aligned} F_{X_{(1)}}(t) &= P(X_{(1)} \leq t) = 1 - P(X_{(1)} > t) = 1 - P(\min\{X_1, X_2, \ldots, X_n\} > t) \\ &= 1 - P(X_1 > t, \ldots, X_n > t) = 1 - P(X_1 > t) \cdots P(X_n > t) = 1 - (1 - F_X(t))^n . \quad \square \end{aligned}$$

Exemple: *Llei del mínim de $X \sim U(\theta, 1)$ $(0 < \theta < 1)$.*
En aquest cas, $f_X(x) = \frac{1}{1-\theta}\, I_{(\theta,\,1)}(x)$ i

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\, dx = \begin{cases} 0 & \text{si} \quad t \leq \theta \\ \frac{t-\theta}{1-\theta} & \text{si} \quad \theta \leq t \leq 1 \\ 1 & \text{si} \quad t \geq 1\,. \end{cases}$$

Per tant, per la Proposició 7,

$$F_{X_{(1)}}(t) = 1 - (1 - F_X(t))^n = \begin{cases} 0 & \text{si} \quad t \leq \theta \\ 1 - \left(\frac{1-t}{1-\theta}\right)^n & \text{si} \quad \theta \leq t \leq 1 \\ 1 & \text{si} \quad t \geq 1\,. \end{cases}$$

I derivant, s'obté l'expressió de la funció de densitat:

$$f_{X_{(1)}}(t) = \frac{n\,(1-t)^{n-1}}{(1-\theta)^n}\, I_{(\theta,\,1)}(t).$$

Distribució de l'r-èssim estadístic d'ordre

Proposició 8. *Si X és una v.a. amb funció de distribució F_X, i $X_1, \ldots, X_n$ és una mostra de mida n de X, aleshores la funció de distribució de la v.a. r-èssim estadístic d'ordre és, per a tot $t \in \mathbb{R}$:*

$$F_{X_{(r)}}(t) = \sum_{j=r}^{n} \binom{n}{j} (F_X(t))^j \ (1 - F_X(t))^{n-j} .$$

Demostració:

Cadascuna de les variables X_i de la mostra pot satisfer que $X_i \leq t$ (èxit) o bé pot satisfer que $X_i > t$ (fracàs).
Podem pensar que per a cada X_i estem fent una prova que pot donar èxit si $X_i \leq t$ o fracàs si $X_i > t$, a més, com que les variables X_i són independents, les proves seran independents. Aleshores la variable

$$Y = \text{“nombre de variables entre les } X_i \text{ que satisfan } X_i \leq t\text{”}$$

té distribució binomial amb paràmetres n que és el nombre total de proves i $p = P(X_i \leq t) = F_X(t)$, que és la probabilitat d'èxit a cada prova.

Per acabar, només cal observar que l'esdeveniment $\{X_{(r)} \leq t\}$, que vol dir que la variable que ocuparia el lloc r quan les ordenem de menor a major pren un valor menor o igual que t, es pot expressar de manera equivalent com $\{Y \geq r\}$, és a dir, hi ha com a mínim r variables entre les X_i que han pres un valor menor o igual que t. Així, com que $Y \sim B(n, F_X(t))$, tenim

$$F_{X_{(r)}}(t) = P(Y \geq r) = \sum_{j=r}^{n} P(Y = j) = \sum_{j=r}^{n} \binom{n}{j} (F_X(t))^j \ (1 - F_X(t))^{n-j} . \quad \square$$