4.Distribucions relacionades amb la Normal

Estadística
Grau en Matemàtiques

Josep A. Sanchez
Dept. Estadística i I.O.(UPC)

Distribucions en el mostreig

Suposem donada una mostra aleatoria simple (m.a.s) de mida n d'una variable aleatòria amb distribució Normal:

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$$

Tot estadístic que podem calcular a partir de la mostra és un cert funcional dels valors obtinguts. Com els valors de la mostra són variables aleatòries, l'estadístic també ho serà:

$$T(\underset{\sim}{X}) = T(X_1, X_2, \ldots, X_n)$$

Exemples:

- Mitjana mostral: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Variància mostral: $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

Quan s'estudien les distribucions d'alguns d'aquests estadístics, apareixen noves distribucions lligades al seu comportament:

- **Distribució** χ^2 (Chi-quadrat)
- **Distribució t-Student**
- **Distribució F-Snedecor**

Distribució χ^2

Sigui $X \sim N(\mu, \sigma^2)$, llavors estandarditzant la variable X tindrem:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

La distribució χ^2_n apareix amb la distribució de la suma de n variables aleatòries estandarditzades independents al quadrat:

Si $Z_i \sim N(0, 1)$ indep. $\Rightarrow \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$

El seu suport són els reals positius.

Distribució χ^2

Definició: X té distribució χ^2 amb ν graus de llibertat ($X \sim \chi^2_\nu$) si la seva funció de densitat és:

$$f(x;\nu) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}} x^{\nu/2-1} e^{-x/2} \quad x \in (0, +\infty)$$

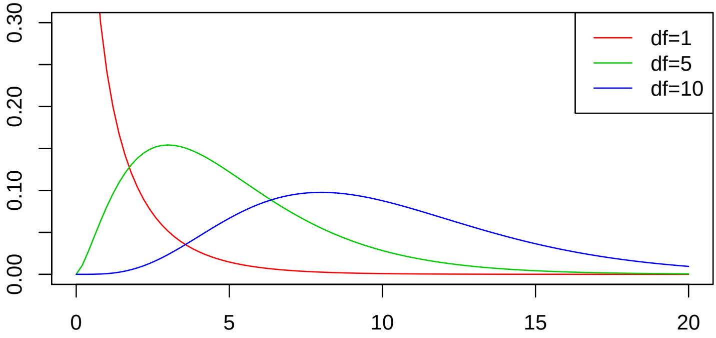
- Es pot comprobar que és un cas particular de la distribució gamma:
$$X \sim \chi^2_\nu \Leftrightarrow X \sim \Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right)$$
- $E(X) = \nu$ i $V(X) = 2\nu$

Definició alternativa: Si $Z_i \sim N(0,1)$, llavors

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$$

Distribució χ^2

```
curve(dchisq(x,df=1), col=2, xlim=c(0,20), ylim=c(0,0.3), xlab="",ylab="")
curve(dchisq(x,df=5), col=3, add=T)
curve(dchisq(x,df=10), col=4, add=T)
legend("topright", lty=1, col=2:4, legend=c("df=1","df=5","df=10"))
```



Distribució de l'estadístic variància mostral (S^2):

- Si coneixem μ, i per tant l'estadístic és $\hat{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}$

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} = \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_n^2$$

- Si no coneixem μ, fent servir l'estadístic $\bar{X}$ com a estimació de μ

$$\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Distribució t-Student

Si $X \sim N(\mu, \sigma^2)$ podem deduir facilment la distribució per a la mitjana mostral:

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu$$

$$V[\bar{X}] = V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

(en el segon cas, hem fet servir la propietat d'independència de la mostra, i per tant $cov(X_i, X_j) = 0$)

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \qquad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Distribució t-Student

- Si la mostra prové de la distribució Normal, llavors la distribució indicada és exacta (les combinacions lineals de variables Normals tenen distribució Normal)

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- Si la mostra no és Normal, el Teorema Central del Límit estableix que la distribució de l'estadístic mitjana mostral ($\bar{X}$) tendeix en llei a la distribució Normal quan n tendeix a infinit (aproximació assimptòtica)

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{L}} N(0,1)$$

Distribució t-Student

En el resultat anterior és necesari conèixer els dos paràmetres per tal que la distribució estigui determinada. Si volem fer inferència del paràmetre μ, el més freqüent és que el paràmetre σ^2 sigui desconegut.

Si en el resultat anterior substituim el paràmetre σ^2 per a l'estimador S^2, quina distribució té el següent estadístic?

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Distribució t-Student

Figure 1: William Gosset (1876-1937)

Distribució t-Student

Definició: X té distribució t-Student amb ν graus de llibertat ($X \sim t_\nu$) si la seva funció de densitat és:

$$f(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

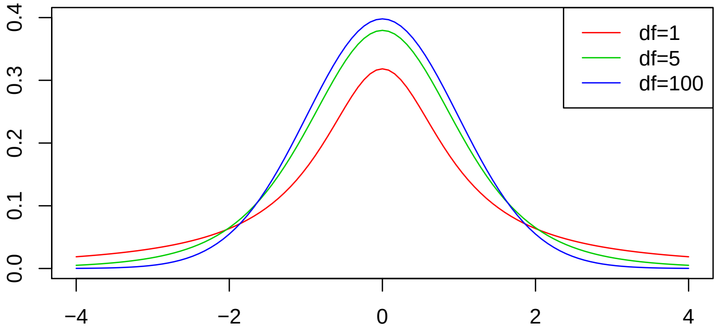
- La distribució t_1 també és coneguda com a distribució de Cauchy.
- $t_n \xrightarrow{n\to\infty} N(0,1)$
- $E(X) = 0 \quad (\nu > 1) \qquad \text{i} \qquad V(X) = \frac{\nu}{\nu-2} \quad (\nu > 2)$

Definició alternativa: Si $Z \sim N(0,1)$ and $U \sim \chi^2_\nu$ independents,

$$\frac{Z}{\sqrt{U/\nu}} \sim t_\nu$$

Distribució t-Student

```
curve(dt(x,df=1), col=2, xlim=c(-4,4), ylim=c(0,0.4), xlab="",ylab="")
curve(dt(x,df=5), col=3, add=T)
curve(dt(x,df=100), col=4, add=T)
legend("topright", lty=1, col=2:4, legend=c("df=1","df=5","df=100"))
```



Distribució t-Student

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \frac{1}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}$$

- $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- $U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

i per tant,

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Distribució F-Snedecor

George W. Snedecor (1882 - 1974)

Ronald A. Fisher (1890 - 1962)

Distribució F-Snedecor

Definició: X té distribució F-(Fisher-)Snedecor amb ν_1 graus de llibertat en el numerador i ν_2 en el numerador ($X \sim F_{\nu_1,\nu_2}$) si la seva funció de densitat és:

$$f(x;\nu) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{x^{(\nu_1-2)/2}}{(1+\frac{\nu_1}{\nu_2}x)^{(\nu_1+\nu_2)/2}} \qquad x \geq 0$$

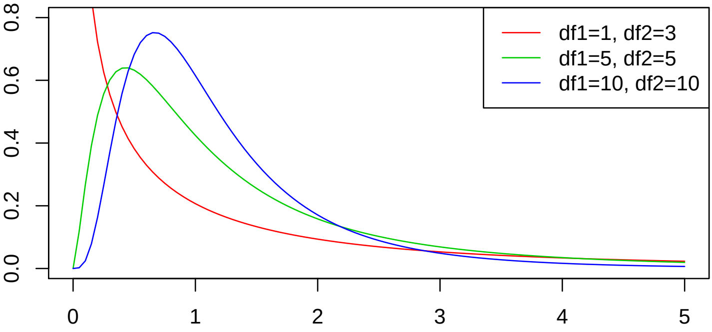
- $E(X) = \frac{\nu_2}{\nu_2-2} \quad (\nu_2 > 2)$ i $V(X) = 2(\frac{\nu_2}{\nu_2-2})^2 \frac{\nu_1+\nu_2-2}{\nu_1(\nu_2-4)} \quad (\nu_2 > 4)$
- Si $X \sim F_{n,m}$ llavors $\frac{1}{X} \sim F_{m,n}$
- Si $T \sim t_n$ llavors $T^2 \sim F_{1,n}$

Definició alternativa: Si $U \sim \chi^2_m$ and $V \sim \chi^2_n$ independents,

$$\frac{U/m}{V/n} \sim F_{m,n}$$

Distribució F-Snedecor

```
curve(df(x,df1=1,df2=3), col=2, xlim=c(0,5), ylim=c(0,0.8  ), xlab="",ylab="")
curve(df(x,df1=5,df2=5), col=3, add=T)
curve(df(x,df1=10,df2=10), col=4, add=T)
legend("topright", lty=1, col=2:4, legend=c("df1=1, df2=3","df1=5, df2=5","df1=10, df2=10"))
```



Muestras de la distribución Normal

Si $X_1, \dots X_n$ m.a.s i $X_i \sim N(\mu, \sigma^2)$:

- $\bar{X}$ i S^2 són v. a. independents
- $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Muestras de la distribución Normal

Teorema de Fisher:

Sigui $X_1, \ldots X_n$ m.a.s. i $X_i \sim N(\mu, \sigma^2)$, llavors,
$\bar{X}$ i $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ són independents

Demostració (preliminars):

- La f.g.m. d'una variable aleatòria X és $M_X(t) = E(e^{tX})$
- La f.g.m. d'un vector $\underset{\sim}{X}$ és $M_{\mathbf{X}}(t_1, \ldots t_k) = E(e^{\sum_{i=1}^{k} t_i X_i})$
- Per tant, $M_{\mathbf{X}}(1, 0, \ldots, 0) = M_{X_1}(t)$
- $X_1, \ldots, X_n$ són independents si i només si, $M_{\mathbf{X}}(t_1, \ldots, t_n) = \prod_{i=1}^{n} M_{X_i}(t_i)$
- La f.g.m. d'una variable $N(\mu, \sigma^2)$ és $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

$$
\begin{aligned}
M_{\left(\overline{X},\mathbf{X}-\overline{X}\right)}(t,\mathbf{t}) &= E\left(e^{\left\{t\overline{X}+t_1(X_1-\overline{X})+t_2(X_2-\overline{X})+\ldots+t_n(X_n-\overline{X})\right\}}\right)\\
&= E\left(\exp\left\{\sum_{i=1}^n t_iX_i - \left(\sum_{i=1}^n t_i - t\right)\overline{X}\right\}\right)\\
&= E\left(\exp\left\{\sum_{i=1}^n X_i\left(t_i - \frac{\sum_{i=1}^n t_i - t}{n}\right)\right\}\right)\\
&= E\left[\prod_{i=1}^n \exp\left\{\frac{X_i(nt_i-n\overline{t}+t)}{n}\right\}\right] \text{ with } \left(\overline{t} = \frac{\sum_{i=1}^n t_i}{n}\right)\\
&= \prod_{i=1}^n E\left(\exp\left\{\frac{X_i[t+n(t_i-\overline{t})]}{n}\right\}\right)\\
&= \prod_{i=1}^n \exp\left\{\frac{\mu[t+n(t_i-\overline{t})]}{n} + \frac{\sigma^2}{2}\frac{1}{n^2}\left[t+n\left(t_i-\overline{t}\right)\right]^2\right\}\\
&= \exp\left\{\frac{\mu}{n}\left[nt+n\sum_{i=1}^n\left(t_i-\overline{t}\right)\right]\right\}\exp\left\{\frac{\sigma^2}{2n^2}\sum_{i=1}^n\left[t+n\left(t_i-\overline{t}\right)\right]^2\right\}\\
&= \exp(\mu t)\exp\left\{\frac{\sigma^2}{2n^2}\left(nt^2+n^2\sum_{i=1}^n\left(t_i-\overline{t}\right)^2\right)\right\}, \text{ given: } \left(\sum_{i=1}^n\left(t_i-\overline{t}\right)=0\right)\\
&= \exp\left(\mu t+\frac{\sigma^2}{2n}t^2\right)\exp\left\{\frac{\sigma^2}{2}\left(\sum_{i=1}^n\left(t_i-\overline{t}\right)^2\right)\right\}\\
&= M_{\overline{X}}(t)\; M_{\left(X_1-\overline{X},X_2-\overline{X},\ldots,X_n-\overline{X}\right)}(t_1,t_2,\ldots,t_n)\\
&= M(t,0,\ldots,0)\; M(0,t_1,t_2,\ldots,t_n)
\end{aligned}
$$