UNIVERSITAT DE BARCELONA

Distribucions multivariants discretes

Diplomatura d'Estadística

Estadística Matemàtica I

Jordi Ocaña Rebull

Departament d'Estadística

Divisió de Ciències Experimentals i Matemàtiques

UNIVERSITAT DE BARCELONA

Distribucions multivariants discretes

- Distribució multinomial
  - Definició
  - Propietats
- Repàs de la distribució hipergeomètrica
- Distribució multihipergeomètrica
  - Definició
  - Propietats

Distribució multinomial
definició

$\mathbf{X}=\left(X_1,\ldots,X_k\right)'$ té distribució multinomial de paràmetres

n i $\mathbf{p}=\left(p_1,\ldots,p_k\right)', \mathbf{X}\sim M(n,\mathbf{p})$, amb n enter positiu i $p_i\geq 0, \sum_{i=1}^{k} p_i=1$,

sii la seva funció de densitat conjunta és :

$$f(\mathbf{x})=P\left[\mathbf{X}=\mathbf{x}\right]=P\left[X_1=x_1,\ldots,X_k=x_k\right]=\frac{n!}{x_1!\,x_2!\cdots x_k!}p_1^{x_1}\ldots p_K^{x_K}$$

per x_i enters no negatius tals que $\sum_{i=1}^{k} x_i=n$ (i 0 en cas contrari).

Distribució multinomial
propietats

- Les marginals són binomials

$$X_i \sim b(n, p_i)$$

- El vector de mitjanes és:

$$E(\mathbf{X}) = \begin{pmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{pmatrix}$$

Distribució multinomial
propietats

- La matriu de covariàncies té la forma:

$$\Sigma = \mathrm{cov}(\mathbf{X}) = \begin{pmatrix} np_1q_1 & -np_1p_2 & \cdots & -np_1p_k \\ -np_2p_1 & np_2q_2 & \cdots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \cdots & np_kq_k \end{pmatrix}$$

$$q_i = 1 - p_i$$

Distribució multinomial
propietats

- I la de correlacions:

$$\mathrm{P} = \mathrm{corr}(\mathbf{X}) = \begin{pmatrix} 1 & -\sqrt{\dfrac{p_1 p_2}{q_1 q_2}} & \cdots & -\sqrt{\dfrac{p_1 p_K}{q_1 q_K}} \\ -\sqrt{\dfrac{p_2 p_1}{q_2 q_1}} & 1 & \cdots & -\sqrt{\dfrac{p_2 p_K}{q_2 q_K}} \\ \vdots & \vdots & \ddots & \vdots \\ -\sqrt{\dfrac{p_K p_1}{q_K q_1}} & -\sqrt{\dfrac{p_K p_2}{q_K q_2}} & \cdots & 1 \end{pmatrix}$$

Distribució hipergeomètrica
definició (repàs)

- Direm que una v.a. discreta té distribució hipergeomètrica

$$X \sim H(N,n,M) = H(N,n,p) \quad p = \frac{M}{N}$$

- si la seva densitat és

$$f(x) = P[X = x] = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$màx\{0, n-N+M\} \leq x \leq min\{n, M\}$$

Distribució hipergeomètrica
propietats (repàs)

- Esperança $E(X) = np$
- Variància $\operatorname{var}(X) = npq\left(\frac{N-n}{N-1}\right)$ $\quad q = 1 - p = \frac{N-M}{N}$
- Sota certes condicions convergeix en llei a la binomial.

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \cong \binom{n}{x} p^x q^{n-x}$$

- A la pràctica podem fer l'aproximació si

$$N \geq 50 \quad \frac{n}{N} < 0{,}1$$

Distribució multihipergeomètrica
definició

- Direm que un vector aleatori discret té distribució multihipergeomètrica

$$\mathbf{X} \sim MH(N;n;\mathbf{M}) = MH(N;n;\mathbf{p})$$

$$\mathbf{M} = (M_1,\ldots,M_k) \qquad \mathbf{p} = (p_1,\ldots,p_k)$$

$$N \geq n, \quad M_i \geq 0, \text{enters amb, } \sum_{i=1}^{k} M_i = N, \quad p_i = \frac{M_i}{N}$$

- si la seva densitat és:

$$f(\mathbf{x}) = P[\mathbf{X} = \mathbf{x}] = \frac{\binom{M_1}{x_1}\binom{M_2}{x_2}\cdots\binom{M_k}{x_k}}{\binom{N}{n}} \qquad x_i \geq 0, \text{ enters, } \sum_{i=1}^{k} x_i = n$$

Distribució multihipergeomètrica
propietats. I

- Les marginals univariants són hipergeomètriques.

$$X_i \sim H(N, n, M_i)$$

- Per tant, el vector de mitjanes té la mateixa forma que el de la multinomial

$$E(\mathbf{X}) = \begin{pmatrix} np_1 \\ \vdots \\ np_k \end{pmatrix} \qquad p_i = \frac{M_i}{N}$$

Distribució multihipergeomètrica

propietats. II

- La covariància entre dues components qualsevol és

$$\mathrm{cov}(X_i, X_j) = -np_i p_j \frac{N-n}{N-1} \quad i \neq j$$

- Per tant, la matriu de covariàncies és

$$\mathrm{cov}(\mathbf{X}) = \frac{N-n}{N-1} \begin{pmatrix} np_1q_1 & -np_1p_2 & \cdots & -np_1p_k \\ -np_2p_1 & np_2q_2 & \cdots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \cdots & np_kq_k \end{pmatrix}$$

$$q_i = 1 - p_i$$