Varianza muestral generalizada y normal multivar.

Wednesday, February 9, 2022 8:53 PM

Varianza muestral generalizada y normal multivar.

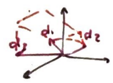
$S_n = \frac{1}{n} \sum_j (x_{ji} - \bar{x}_i)^2$

$S = \frac{1}{n-1} \sum_j (x_{ji} - \bar{x}_i)^2$

$S = [\quad]$ → matriz varianzas y covarianzas insesgada ↳ $1/n-1$

Varianza muestral generalizada $= |S| = \det(S)$.

↳ No solo como se comportan los datos de una var., sino una general de todas.



→ Asociado al volumen.
↳ Más disperos, más vol.

Cuando 1+ cols es lin dep → $|S| = 0$.
↳ Puedo asociar las vars. pues una "sobra". (reducir dims.)

$|S| = 0$ cuando y solo cuando al menos un vec. desviaº yace en el hiperplano formado por todas, combinaciones lins. de los vecs resto

↳ $n \leq p$: hay menos obs. que vars. → $|S| = 0$.

Varianza generalizada determinada por $|R|$ → corr.

La var muestral generalizada se ve afectada indebidam. por la variabilidad de las mediciones en una sola var.

Ej: Si S_{11} es muy grande (o chiki), geometricam. el vec. de desviaº correspondiente $d_i = (y_i - \bar{x}_i \mathbf{1})$ será muy largo (o muy corto) y afectará el volumen.

↳ Es útil escalar todos los vecs. de desviaº para q' tengan la misma long

↳ reemplazar x_{ji} por $\frac{(x_{ji} - \bar{x}_i)}{\sqrt{s_{ii}}}$

Así, la matriz de var. y cov. es R → la matriz de correlaº

$|R|$ → varianza muestral generalizada de las vars. estandarizadas.

↳ El ángulo dicta.

Varianza total de la muestra: $S_{11} + \ldots + S_{pp}$

Normal multivariada

PDF de una v.a con media μ y var σ^2 es:

$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)/2\sigma^2} \quad x \in \mathbb{R}.$

$X \sim N(\mu, \sigma^2)$ ↳ $N(\mu, \sigma^2)$

$\frac{(x-\mu)^2}{\sigma^2}$ → Dist. estad. entre x y μ.
$= (x-\mu)(\sigma^2)^{-1}(x-\mu)$

Generalizando $\bar{X}$ vec. de obs. de x (fila de $\mathbb{X}$).

Distancia estadística de x a μ:

$$(x-\mu)^T \Sigma^{-1} (x-\mu)$$

Entonces, generalizando se obtiene la PDF de la ~N multivar.

$$f(\vec{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \cdot e^{-(x-\mu)^T \Sigma^{-1}(x-\mu)/2}$$

$x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ con p: # vars.

☆ Normal multivar: $\mathbb{X} \sim N_p(\underbrace{\mu_i}_{\text{vec}}, \underbrace{\Sigma}_{\text{matriz}})$.

– Si $p=2$ y $S_{12}=0$: $(\mathrm{cov}(x_1,x_2)=0)$ ⟶ $x_1 x_2$ indep. pues las vars son normales.

Teo:

Si Σ def pos. $\exists\ \Sigma^{-1}$

Entonces, si e es un vec propio de Σ con val. propio λ_1 asociado, e es un vec propio de Σ^{-1} con val propio asociado $1/\lambda$.

Además, Σ^{-1} es def. pos.

Obj:

1) $(x-\mu)' \Sigma^{-1} (x-\mu) \leq \chi^2_p(\alpha)$ Chi² con proba $1-\alpha$ y p gl.

Proposiº

1º Si $\mathbb{X}$ vec. aleatorio normal multivariado

$\mathbb{X} \sim N_p(\mu, \Sigma)$ entonces

$\sim N(\mu, \Sigma)$ entonces,

Si $a \in \mathbb{R}^p$, $a'\chi = a_1 x_1 + \dots + a_p x_p$
$\sim N(a'\mu, a'\Sigma a)$.

2. Si $a'\chi \sim N(a'\mu, a'\Sigma a)$, $\forall a \in \mathbb{R}^p$
entonces, $\chi \sim N_p(\mu, \Sigma)$

3. $\chi \sim N_p(\mu, \Sigma)$
Sea $\underset{q \times p}{A} \in \mathbb{R}$ $\Rightarrow AX \sim N_q(A\mu, A\Sigma A')$

4. $\chi \sim N_p(\mu, \Sigma)$ $d \in \mathbb{R}^p$ $\rightarrow$ d es un vec. cualquiera.
$\chi + d \sim N_p(\mu + d, \Sigma)$

5. Todas las particiones de vecs. aleatorios normales, resultan en vecs. aleat. normales.

$\chi \sim N_p(\mu, \Sigma)$ con $\chi = \begin{pmatrix} \chi_{(1)} \\ \chi_{(2)} \end{pmatrix} \begin{matrix} \}q \\ \}p-q \end{matrix}$ $\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}$ $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

$\chi_{(1)} \sim N_q(\mu_{(1)}, \Sigma_{11})$
$\chi_{(2)} \sim N_{p-q}(\mu_{(2)}, \Sigma_{22})$

---

6. Si $\underset{q_1 \times 1}{\chi_{(1)}}$ y $\underset{q_2 \times 1}{\chi_{(2)}}$ normales multivariadas e indep.
$\mathrm{cov}(\chi_{(1)}, \chi_{(2)}) = 0 \rightarrow$ matriz 0 $q_1 \times q_2$

- Si $\chi = \begin{bmatrix} \chi_{(1)} \\ \chi_{(2)} \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$ con $\chi_{(1)}$ y $\chi_{(2)}$ indeps sii $\Sigma_{12} = 0$

cov 0 sí implica indep cuando hay normalidad

- $\chi_{(1)}$ y $\chi_{(2)}$ normales indep sii $\chi = \begin{bmatrix} \chi_{(1)} \\ \chi_{(2)} \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}, \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}\right)$

7. Sea $\chi \sim N_p(\mu, \Sigma)$, $|\Sigma| > 0$
entonces:
- $(\chi - \mu)^T (\Sigma^{-1})(\chi - \mu) \sim \chi^2_p$ $\longrightarrow$ la dist. de los vals a la media tiene $\sim \chi^2$ (chi 2).

La proba del evento: $\{(x-\mu)' \Sigma^{-1} (x - \mu) \leq \chi^2_p(\alpha)\}$ es $1 - \alpha$

Muestreo de la normal multivariada

Suponga q' $\chi_1, \dots, \chi_n$ son una m.a de una poblaº normal multivariada con media μ y cov. Σ

Densidad conjunta de $\chi_1, \dots, \chi_n$:

$$\prod_{j=1}^{n} \left(\frac{1}{(2\pi)^{p/2} \Sigma^{1/2}}\right) \exp\left(-\frac{1}{2}(x_j - \mu)' \Sigma^{-1} (x_j - \mu)\right)$$

fº de verosimilitud (de μ y Σ) dadas las obs $x_1, \dots, x_n$: $L(\mu, \Sigma)$

Con esto se pueden obtener estimadores de Σ y μ q' mejor explican los datos. - Método de máx-verosimilitud

Teo: sean $\chi_1, \dots, \chi_n$ una m.a de una poblaº normal con media μ y cov. Σ, entonces los MLE son:

$$\hat{\mu} = \bar{x} \qquad \hat{\Sigma} = \frac{1}{n}\left(\sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})'\right) = S_n = \frac{n-1}{n} S.$$

Obs: $\hat{\mu}$ y $\hat{\Sigma}$ son vecs aleatorios antes de tomar los datos.

$$L(\hat{\mu}, \hat{\Sigma}) = \frac{1}{(2\pi)^{np/2}} \cdot \exp\left(-\frac{np}{2} \cdot \frac{1}{|\Sigma|^{n/2}}\right) \qquad |\Sigma| \rightarrow \frac{n-1}{n} S$$

$cte(|S|)^{-n/2}$.

Su varianza generalizada determina el máx. de la fº de verosimilitud.