SME0822 Análise Multivariada e Aprendizado Não-Supervisionado

Aula 4c: **Inferência sobre a média**

Prof. Cibele Russo

cibele@icmc.usp.br

http://www.icmc.usp.br/~cibele

Baseado em Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. Prentice Hall.

Estimação de parâmetros

Sejam $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ vetores que representam uma amostra aleatória de uma distribuição $N_p(\underset{\sim}{\mu}, \Sigma)$. A função densidade de probabilidade conjunta de $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ é dada por

$$
\begin{aligned}
f(\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n) &= \prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{(\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu})}{2}\right\} \\
&= \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} \exp\sum_{j=1}^{n}\left\{-\frac{(\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu})}{2}\right\}.
\end{aligned}
$$

Estimação de parâmetros

Resultado

Seja $A_{k\times k}$ *uma matriz simétrica e* $\underset{\sim}{x}_{k\times 1}$ *um vetor*

- $\underset{\sim}{x}^\top A\underset{\sim}{x} = tr(\underset{\sim}{x}^\top A\underset{\sim}{x}) = tr(A\underset{\sim}{x}\underset{\sim}{x}^\top)$
- $tr(A) = \sum_{i=1}^{k} \lambda_i$, *com* λ_i *autovalores de* A *para* $i = 1, \ldots, k$.

Estimação de parâmetros

Utilizando o resultado anterior, temos que

$$\sum_{j=1}^{n}\left\{(\underset{\sim}{x}_j-\underset{\sim}{\mu})^\top\Sigma^{-1}(\underset{\sim}{x}_j-\underset{\sim}{\mu})\right\}=$$

$$\sum_{j=1}^{n}\operatorname{tr}\left\{(\underset{\sim}{x}_j-\underset{\sim}{\mu})^\top\Sigma^{-1}(\underset{\sim}{x}_j-\underset{\sim}{\mu})\right\}=$$

$$\sum_{j=1}^{n}\operatorname{tr}\left\{\Sigma^{-1}(\underset{\sim}{x}_j-\underset{\sim}{\mu})(\underset{\sim}{x}_j-\underset{\sim}{\mu})^\top\right\}=$$

$$\operatorname{tr}\left\{\Sigma^{-1}\sum_{j=1}^{n}(\underset{\sim}{x}_j-\underset{\sim}{\mu})(\underset{\sim}{x}_j-\underset{\sim}{\mu})^\top\right\}=$$

Estimação de parâmetros

Utilizando o resultado anterior, temos que

$$\sum_{j=1}^{n} \left\{ (\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}) \right\} =$$

$$\sum_{j=1}^{n} \text{tr} \left\{ (\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}) \right\} =$$

$$\sum_{j=1}^{n} \text{tr} \left\{ \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu})(\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \right\} =$$

$$\text{tr} \left\{ \Sigma^{-1} \sum_{j=1}^{n} (\underset{\sim}{x}_j - \underset{\sim}{\mu})(\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \right\} =$$

Estimação de parâmetros

Utilizando o resultado anterior, temos que

$$\sum_{j=1}^{n} \left\{ (\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}) \right\} =$$

$$\sum_{j=1}^{n} \operatorname{tr} \left\{ (\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}) \right\} =$$

$$\sum_{j=1}^{n} \operatorname{tr} \left\{ \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu})(\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \right\} =$$

$$\operatorname{tr} \left\{ \Sigma^{-1} \sum_{j=1}^{n} (\underset{\sim}{x}_j - \underset{\sim}{\mu})(\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \right\} =$$

Estimação de parâmetros

Utilizando o resultado anterior, temos que

$$\sum_{j=1}^{n} \left\{ (\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}) \right\} =$$

$$\sum_{j=1}^{n} \text{tr} \left\{ (\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu}) \right\} =$$

$$\sum_{j=1}^{n} \text{tr} \left\{ \Sigma^{-1} (\underset{\sim}{x}_j - \underset{\sim}{\mu})(\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \right\} =$$

$$\text{tr} \left\{ \Sigma^{-1} \sum_{j=1}^{n} (\underset{\sim}{x}_j - \underset{\sim}{\mu})(\underset{\sim}{x}_j - \underset{\sim}{\mu})^\top \right\} =$$

Estimação de parâmetros

$$\mathrm{tr}\left\{\Sigma^{-1}\sum_{j=1}^{n}(\underset{\sim}{x}_j-\underset{\sim}{\mu})(\underset{\sim}{x}_j-\underset{\sim}{\mu})^\top\right\}=$$

$$\mathrm{tr}\left\{\Sigma^{-1}\sum_{j=1}^{n}(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}}+\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}}+\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})^\top\right\}=$$

(após alguns cálculos - exercício)

$$\mathrm{tr}\left\{\Sigma^{-1}\left[\sum_{j=1}^{n}(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}})(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}})^\top+n(\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})(\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})^\top\right]\right\}.$$

Estimação de parâmetros

$$\operatorname{tr}\left\{\Sigma^{-1}\sum_{j=1}^{n}(\underset{\sim}{x}_j-\underset{\sim}{\mu})(\underset{\sim}{x}_j-\underset{\sim}{\mu})^\top\right\}=$$

$$\operatorname{tr}\left\{\Sigma^{-1}\sum_{j=1}^{n}(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}}+\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}}+\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})^\top\right\}=$$

(após alguns cálculos - exercício)

$$\operatorname{tr}\left\{\Sigma^{-1}\left[\sum_{j=1}^{n}(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}})(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}})^\top+n(\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})(\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})^\top\right]\right\}.$$

Estimação de parâmetros

$$\text{tr}\left\{\Sigma^{-1}\sum_{j=1}^{n}(\underset{\sim}{x}_j-\underset{\sim}{\mu})(\underset{\sim}{x}_j-\underset{\sim}{\mu})^\top\right\}=$$

$$\text{tr}\left\{\Sigma^{-1}\sum_{j=1}^{n}(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}}+\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}}+\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})^\top\right\}=$$

(após alguns cálculos - exercício)

$$\text{tr}\left\{\Sigma^{-1}\left[\sum_{j=1}^{n}(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}})(\underset{\sim}{x}_j-\overline{\underset{\sim}{x}})^\top+n(\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})(\overline{\underset{\sim}{x}}-\underset{\sim}{\mu})^\top\right]\right\}.$$

Estimação de parâmetros

Reescrevemos então a função densidade de probabilidade agora como função de verossimilhança:

$$L(\underset{\sim}{\mu}, \Sigma | \underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \times$$
$$\exp\left\{ -\text{tr}\left\{ \Sigma^{-1} \left[\sum_{j=1}^{n} (\underset{\sim}{x}_j - \overline{\underset{\sim}{x}})(\underset{\sim}{x}_j - \overline{\underset{\sim}{x}})^\top + n(\overline{\underset{\sim}{x}} - \underset{\sim}{\mu})(\overline{\underset{\sim}{x}} - \underset{\sim}{\mu})^\top \right] \right\} / 2 \right\}$$

Estimação de parâmetros

e então, obtemos o logaritmo da verossimilhança:

$$\log\ L(\underset{\sim}{\mu}, \Sigma | \underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma|$$
$$-\mathrm{tr}\left\{\Sigma^{-1}\left[\sum_{j=1}^{n}(\underset{\sim}{x}_j - \overline{\underset{\sim}{x}})(\underset{\sim}{x}_j - \overline{\underset{\sim}{x}})^\top + n(\overline{\underset{\sim}{x}} - \underset{\sim}{\mu})(\overline{\underset{\sim}{x}} - \underset{\sim}{\mu})^\top\right]\right\}/2.$$

Estimação de parâmetros

Resultado

Os estimadores de máxima verossimilhança de $\underset{\sim}{\mu}$ *e* Σ *são dados por*

$$\hat{\underset{\sim}{\mu}} = \overline{X} \quad e \quad \hat{\Sigma}_{MV} = \frac{1}{n}\sum_{j=1}^{n}(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}})(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}})^\top.$$

As estimativas de máxima verossimilhança (após observar a amostra) de $\underset{\sim}{\mu}$ *e* Σ *são dados por*

$$\hat{\underset{\sim}{\mu}} = \overline{x} \quad e \quad \hat{\Sigma}_{MV} = \frac{1}{n}\sum_{j=1}^{n}(\underset{\sim}{x}_j - \overline{\underset{\sim}{x}})(\underset{\sim}{x}_j - \overline{\underset{\sim}{x}})^\top$$

Prova em Johnson (2007, p. 172).

Estimação de parâmetros

Note que, como já mostramos, $\hat{\Sigma}_{MV}$ é viesado para estimar Σ.

Assim, em muitas aplicações consideramos o estimador não viesado para Σ:

$$\hat{\Sigma} = S = \frac{1}{n-1}\sum_{j=1}^{n}(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}})(\underset{\sim}{X}_j - \overline{\underset{\sim}{X}})^\top$$

Distribuição amostral de $\underset{\sim}{\overline{X}}$ e S

Seja $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ uma amostra aleatória de uma distribuição $N_p(\underset{\sim}{\mu}, \Sigma)$ e

$$\hat{\underset{\sim}{\mu}} = \underset{\sim}{\overline{X}} \quad \text{e} \quad \hat{\Sigma} = S = \frac{1}{n-1}\sum_{j=1}^{n}(\underset{\sim}{X}_j - \underset{\sim}{\overline{X}})(\underset{\sim}{X}_j - \underset{\sim}{\overline{X}})^\top.$$

Temos o seguinte

Resultado

1. $\underset{\sim}{\overline{X}} \sim N_p\left(\underset{\sim}{\mu}, \dfrac{\Sigma}{n}\right)$.
2. $(n-1)S \sim Wishart(n-1)$.
3. $\underset{\sim}{\overline{X}}$ *e* S *são independentes.*

Distribuição amostral de $\overline{X}$ e S

Obs: A distribuição Wishart é uma generalização da distribuição Gama, e é definida como a soma de produtos de normais multivariadas independentes de média $\underset{\sim}{0}$ e variância Σ:

Em outras palavras, seja

$$W = \sum_{j=1}^{n} \underset{\sim}{Z}_j \underset{\sim}{Z}_j^\top \text{ com } \underset{\sim}{Z}_j \overset{i.i.d}{\sim} N(\underset{\sim}{0}, \Sigma),$$

Então

$$W \sim Wishart(\Sigma, n).$$

A distribuição assintótica de $\overline{\underset{\sim}{X}}$

Teorema do Limite Central

Seja $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ uma amostra aleatória de uma distribuição qualquer p-variada com $\mathrm{E}(\underset{\sim}{X}_i) = \underset{\sim}{\mu}$ e $\mathrm{Var}(\underset{\sim}{X}_i) = \Sigma$, para $i = 1, \ldots, n$ e Σ positiva definida.

Então, para n suficientemente grande e $n >> p$, temos

$$\sqrt{n}(\overline{\underset{\sim}{X}} - \underset{\sim}{\mu}) \sim N_p(\underset{\sim}{0}, \Sigma)$$

e ainda

$$n(\overline{\underset{\sim}{X}} - \underset{\sim}{\mu})^\top S^{-1}(\overline{\underset{\sim}{X}} - \underset{\sim}{\mu}) \sim \chi^2_p.$$

Testes de hipóteses para a média

Seja $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ uma amostra aleatória de uma distribuição normal p-variada com vetor de médias $\underset{\sim}{\mu}$ e matriz de variâncias e covariâncias Σ. Sejam $\overline{\underset{\sim}{X}}$ e S o vetor de médias amostrais e a matriz de variâncias e covariâncias amostrais.

Queremos avaliar se

$$\begin{array}{l} H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0 \text{ contra} \\ H_1 : \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0, \end{array} \quad .$$

Testes de hipóteses para a média

Relembramos o resultado anterior

Resultado

1. $\underset{\sim}{\overline{X}} \sim N_p\left(\underset{\sim}{\mu}, \frac{\Sigma}{n}\right)$.
2. $(n-1)S \sim Wishart(n-1)$.
3. $\underset{\sim}{\overline{X}}$ *e* S *são independentes.*

Além disso, sob H_0,

$$T^2 = \sqrt{n}(\underset{\sim}{\overline{X}} - \underset{\sim}{\mu}_0)^\top \left(\frac{(n-1)S}{n-1}^{-1}\right) \sqrt{n}(\underset{\sim}{\overline{X}} - \underset{\sim}{\mu}_0) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

Testes de hipóteses para a média

A quantidade

$$T^2 = n(\underset{\sim}{\overline{X}} - \underset{\sim}{\mu}_0)^\top S^{-1}(\underset{\sim}{\overline{X}} - \underset{\sim}{\mu}_0) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

é conhecida como a **Estatística T^2 de Hotelling**.

Testes de hipóteses para a média

Assim, rejeitamos H_0 a um nível de significância α se

$$T^2_{obs} = n(\overline{\underset{\sim}{X}} - \underset{\sim 0}{\mu})^\top S^{-1}(\overline{\underset{\sim}{X}} - \underset{\sim 0}{\mu}) > \frac{(n-1)p}{n-p}\; q_{F_{p,n-p,\alpha}}$$

em que $q_{F_{p,n-p,\alpha}}$ é o quantil $\alpha-$superior de uma distribuição $F_{p,n-p}$.

Propriedade de T^2

Seja $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ uma amostra aleatória de uma distribuição normal p-variada com vetor de médias $\underset{\sim}{\mu}_X$ e matriz de variâncias e covariâncias Σ_X.

Seja $\underset{\sim}{Y}_i = C\underset{\sim}{X}_i + \underset{\sim}{d}$, com C uma matriz não singular fixa e $\underset{\sim}{d}$ um vetor fixo.

Temos que $\mathrm{E}(\underset{\sim}{Y}_i) = C\underset{\sim}{\mu}_X + \underset{\sim}{d}$, $\mathrm{Var}(\underset{\sim}{Y}_i) = C\Sigma_X C^\top$.

Além disso, $\overline{\underset{\sim}{Y}} = C\overline{\underset{\sim}{X}} + \underset{\sim}{d}$ e $S_Y = CS_x C^\top$ (exercício).

Propriedade de T^2

É possível mostrar que a estatística T^2 para avaliar

$$\begin{array}{l} H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0 \text{ contra} \\ H_1 : \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0, \end{array} .$$

é equivalente à estatística T^2 para avaliar

$$\begin{array}{l} H_0 : C\underset{\sim}{\mu} + \underset{\sim}{d} = C\underset{\sim}{\mu}_0 + \underset{\sim}{d} \text{ contra} \\ H_1 : C\underset{\sim}{\mu} + \underset{\sim}{d} \neq C\underset{\sim}{\mu}_0 + \underset{\sim}{d}, \end{array} .$$