Statistical and Mathematical Methods DS-GA 1002
, 2019

Recitation Session 11 Solutions

Problem

1. *Convergence* A fair dice is casted independently 600 times. Give an approximation (by CLT) for the probability that the number of 6s falls between 95 and 110.

   SOLUTION

   If the outcome of cast i is y_i. Define an indicator I_i which equals 1 when $y_i = 6$. Then we are interested in:

$$\mathbb{P}(95 < \sum_{i=1}^{600} I_i < 110)$$

   By the CLT, $\frac{1}{600}\sum_{i=1}^{600} I_i$ behaves like a normal distribution with mean $\mathbb{E}(I_i) = \frac{1}{6}$ and variance $\sigma_I^2/n = \frac{5}{36}/600$, and therefore $\sum_{i=1}^{600} I_i$ follows $\mathcal{N}(100, 250/3)$ Then we can rewrite the probability as:

$$\begin{aligned}
\mathbb{P}(95 < \sum_{i=1}^{600} I_i < 110) &= \mathbb{P}(\frac{95-100}{\sqrt{250/3}} < \frac{\sum_{i=1}^{600} I_i - 100}{\sqrt{250/3}} < \frac{110-100}{\sqrt{250/3}}) \\
&= \mathbb{P}(\frac{95-100}{\sqrt{250/3}} < Z < \frac{110-100}{\sqrt{250/3}}) \\
&= F_Z(\frac{110-100}{\sqrt{250/3}}) - F_Z(\frac{95-100}{\sqrt{250/3}}) \\
&= 0.5715
\end{aligned}$$

2. *Empirical covariance* Suppose $\{(X_i, Y_i)\}_{i=1}^n$ are n independent samples from a two-variable pdf $f_{X,Y}(x, y)$ where

$$\mu_X = \mathbb{E}(X), \quad \mu_Y = \mathbb{E}(Y)$$

   and

$$\sigma_X^2 = \mathrm{Var}(X), \quad \sigma_Y^2 = \mathrm{Var}(Y), \quad \sigma_{XY} = \mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))].$$

   Consider the empirical covariance between $\{X_i\}$ and $\{Y_i\}$:

$$\widehat{\sigma}_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n).$$

   Prove that

$$\widehat{\sigma}_{XY} = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i Y_i - n\overline{X}_n\overline{Y}_n\right)$$

SOLUTION

3. *PCA* Let $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ be a (sample) covariance matrix with an eigenvalue/eigenvector decomposition as

$$\mathbf{\Sigma} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_d] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix} [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_d]^\top$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, $\boldsymbol{u}_i \in \mathbb{R}^{d\times 1}$, and $\boldsymbol{u}_i \perp \boldsymbol{u}_j$, i.e., $\boldsymbol{u}_i^\top \boldsymbol{u}_j = 0$ for $i \neq j$.

a. Prove

$$\lambda_1 = \max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^\top \mathbf{\Sigma} \boldsymbol{v}, \quad \boldsymbol{u}_1 = \operatorname{argmax}_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^\top \mathbf{\Sigma} \boldsymbol{v}$$

(Hint: Rewrite $\mathbf{\Sigma}$ into $\mathbf{\Sigma} = \sum_{i=1}^d \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$ and use the fact that $\sum_{i=1}^d \boldsymbol{u}_i \boldsymbol{u}_i^\top = \boldsymbol{I}_d$.)

b. Prove that

$$\lambda_2 = \max_{\|\boldsymbol{v}\|=1, \boldsymbol{v}\perp \boldsymbol{u}_1} \boldsymbol{v}^\top \mathbf{\Sigma} \boldsymbol{v}, \quad \boldsymbol{u}_2 = \operatorname{argmax}_{\|\boldsymbol{v}\|=1, \boldsymbol{v}\perp \boldsymbol{u}_1} \boldsymbol{v}^\top \mathbf{\Sigma} \boldsymbol{v}$$

where $\|\boldsymbol{v}\|^2 = \boldsymbol{v}^\top \boldsymbol{v} = \sum_{i=1}^d v_i^2 = 1$.

SOLUTION

a. Since, $\|v\|^2 = 1$, $\|v\|^2 = \sum_{i=1}^d \alpha_i^2 \|u_i\|^2 = \sum_{i=1}^d \alpha_i^2 = 1$

$$\begin{aligned}
\boldsymbol{v}^\top \mathbf{\Sigma} \boldsymbol{v} &= \boldsymbol{v}^\top USU\top \boldsymbol{v} \\
&= \sum_{i=1}^d \lambda_i \alpha_i^2 \boldsymbol{u}_i^\top \boldsymbol{u}_i \boldsymbol{u}_i^\top \boldsymbol{u}_i \\
&= \sum_{i=1}^d \lambda_i \alpha_i^2 \|\boldsymbol{u}_i\|^2 \\
&\leq \lambda_1 \sum_{i=1}^d \alpha_i^2 \|\boldsymbol{u}_i\|^2 \\
&= \lambda_1 \cdot 1 \\
&= \lambda_1
\end{aligned}$$

Thus, $\lambda_1 = \max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^\top \mathbf{\Sigma} \boldsymbol{v}$
Since $\boldsymbol{u}_1^\top \mathbf{\Sigma} \boldsymbol{u}_1 = \lambda_1$, $\boldsymbol{u}_1 = \operatorname{argmax}_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^\top \mathbf{\Sigma} \boldsymbol{v}$

b. $u_1, u_2, ..., u_d$ is an orthonormal basis of $\mathbb{R}^{d\times d}$, and v is orthogonal with u_1, we can write v as: $v = 0u_1 + \alpha_2 u_2 + ... + \alpha_d u_d$ for α_i in $\mathbb{R}$. In addition, $\|\boldsymbol{u}_i\|^2 = 1$.
Since, $\|\boldsymbol{v}\|^2 = 1$, $\|\boldsymbol{v}\|^2 = \sum_{i=1}^d \alpha_i^2 \|\boldsymbol{u}_i\|^2 = \sum_{i=1}^d \alpha_i^2 = 1$

$$
\begin{aligned}
\boldsymbol{v}^\top \boldsymbol{\Sigma} \boldsymbol{v} &= \boldsymbol{v}^\top U S U \top \boldsymbol{v} \\
&= \sum_{i=1}^{d} \lambda_i \alpha_i^2 \boldsymbol{u}_i^\top \boldsymbol{u}_i \boldsymbol{u}_i^\top \boldsymbol{u}_i \\
&= \sum_{i=1}^{d} \lambda_i \alpha_i^2 \|\boldsymbol{u}_i\|^2 \\
&\leq \lambda_1 \sum_{i=1}^{d} \alpha_i^2 \|\boldsymbol{u}_i\|^2 \\
&= \lambda_2 \cdot 1 \\
&= \lambda_2
\end{aligned}
$$

Thus, $\lambda_2 = \max_{\|\boldsymbol{v}\|=1, \boldsymbol{v} \perp \boldsymbol{u}_1} \boldsymbol{v}^\top \boldsymbol{\Sigma} \boldsymbol{v}$
Since $\boldsymbol{u}_2^\top \boldsymbol{\Sigma} \boldsymbol{u}_2 = \lambda_2$, $\boldsymbol{u}_2 = \operatorname{argmax}_{\|\boldsymbol{v}\|=1, \boldsymbol{v} \perp \boldsymbol{u}_1} \boldsymbol{v}^\top \boldsymbol{\Sigma} \boldsymbol{v}$