Дифференцирование следа матрицы по матрице[1]

1. След матрицы $\mathbf{A}$, определяемый как сумма элементов этой матрицы, находящихся на главной диагонали, обозначается символами $\mathrm{sp}[\mathbf{A}]$ или $\mathrm{tr}[\mathbf{A}]$. Имеют место следующие свойства следа матрицы:

$$\mathrm{sp}[\mathbf{ABC}] = \mathrm{sp}[\mathbf{CAB}] = \mathrm{sp}[\mathbf{BCA}];$$

$$\mathrm{sp}[\mathbf{A}] = \mathrm{sp}[\mathbf{A}^T];\ \mathrm{sp}[\mathbf{A} + \mathbf{B}] = \mathrm{sp}[\mathbf{A}] + \mathrm{sp}[\mathbf{B}].$$

При $\mathbf{A} = \mathbf{XY}^T$, $\mathrm{sp}[\mathbf{A}] = \mathbf{X}^T\mathbf{Y}$ ($\mathbf{X}$ и $\mathbf{Y}$ — векторы-столбцы одной размерности).

2. Основные операции дифференцирования следа матрицы по матрице имеют вид ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{X}$ — матрицы согласованной размерности):

$$\textit{а)}\ \frac{\partial}{\partial \mathbf{X}}\mathrm{sp}[\mathbf{X}] = \mathbf{E};\quad \textit{б)}\ \frac{\partial}{\partial \mathbf{X}}\mathrm{sp}[\mathbf{AX}] = \mathbf{A}^T;\quad \textit{в)}\ \frac{\partial}{\partial \mathbf{X}}\mathrm{sp}[\mathbf{AX}^T] = \mathbf{A};$$

$$\textit{г)}\ \frac{\partial}{\partial \mathbf{X}}\mathrm{sp}[\mathbf{AXB}] = \mathbf{A}^T\mathbf{B}^T;\quad \textit{д)}\ \frac{\partial}{\partial \mathbf{X}}\mathrm{sp}[\mathbf{AX}^T\mathbf{B}] = \mathbf{BA};$$

$$\textit{е)}\ \frac{\partial}{\partial \mathbf{X}}\mathrm{sp}[\mathbf{AXBX}] = \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T + \mathbf{B}^T\mathbf{X}^T\mathbf{A}^T;\quad \textit{ж)}\ \frac{\partial}{\partial \mathbf{X}}\mathrm{sp}[\mathbf{AXBX}^T] = \mathbf{A}^T\mathbf{XB}^T + \mathbf{AXB}.$$

---

[1] Ивановский Р.И. Теория вероятностей и математической статистики. Основы, прикладные аспекты с примерами и задачами в среде Mathcad. — СПб.: БХВ, 2008. –528 с., Приложение 15, стр. 513.