

Math/Stat Review (prerequisites)

Stats 500, Fall 2017

Brian Thelen, University of Michigan
443 West Hall, bjthelen@umich.edu

Vectors and Matrices

- Vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

- Length

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{x^\top x}$$

- Inner product

$$x^\top y = x_1 y_1 + x_2 y_2 + \cdots x_n y_n$$

- Geometric interpretation:

$$\cos \theta = \frac{x^\top y}{\|x\| \cdot \|y\|}$$

- x and y are **orthogonal** if

Example:

- Vectors x_1, \dots, x_m are **linearly dependent** if there exist scalars a_1, \dots, a_m such that at least one $a_j \neq 0$ and

$$a_1 x_1 + \dots + a_m x_m = 0$$

Example:

Vector Spaces

- Vector space:
 - Basis vectors: x_1, \dots, x_m
 - Linear span:

$$\begin{aligned}\mathcal{X} &= \text{span}(x_1, \dots, x_m) \\ &= \{x : x = a_1x_1 + a_2x_2 + \dots + a_mx_m, a_j \in \mathbb{R}\}\end{aligned}$$

Examples:

Matrices

- Matrix

$$A_{n \times m} = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nm} \end{pmatrix}$$

Examples:

Matrix Multiplication

- Matrix multiplication

$$\begin{aligned} A_{n \times m} B_{m \times k} &= \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nm} \end{pmatrix} \cdot \begin{pmatrix} B_{11} & \cdots & B_{1k} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mk} \end{pmatrix} \\ &= \begin{pmatrix} \cdots & \cdots & \cdots \\ \cdots & \sum_{\ell=1}^m A_{i\ell} B_{\ell j} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}_{n \times k} \end{aligned}$$

Examples:

Matrix: Transpose and Inverse

- Matrix **transpose**

$$A^{\top} = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \ddots & \ddots & \vdots \\ A_{1m} & A_{2m} & \cdots & A_{nm} \end{pmatrix}_{m \times n}$$

$$(A_{n \times m} B_{m \times k})^{\top} = B^{\top} A^{\top}$$

- Matrix **inverse** (of a square matrix $A_{n \times n}$)
 - Defined by $A^{-1}A = AA^{-1} = I$
 - If the columns (rows) of A are **linearly independent**, then A is **invertible**; otherwise, A is **singular**, and **determinant** $\det(A) = 0$.
 $(AB)^{-1} =$
 $(A^{-1})^{\top} =$
 $(AB)^{-\top} =$

Derivatives of Functions

- **Derivatives:** $x = (x_1, \dots, x_m)$
 $y_{n \times 1} = f(x_{m \times 1})$; $D_{m \times n} = \partial y / \partial x$ is defined by

$$D_{ij} = \frac{\partial y_j}{\partial x_i}$$

Examples:

$$\partial(A_{n \times m} x_{m \times 1}) / \partial x =$$

$$\partial(x^\top B_{m \times m} x) / \partial x =$$

Eigendecomposition

- **Eigenvectors and eigenvalues** : there exists a vector $u \neq 0$

$$A_{n \times n} u_{n \times 1} = \lambda u_{n \times 1}$$

- There are exactly n eigenvalues and eigenvectors (not all eigenvalues are distinct and/or real)
- If A is **symmetric** ($A^\top = A$), all eigenvalues are
- A is **singular** iff at least one of the $\lambda = 0$
- A is **positive definite** ($x^\top A x > 0$ for all $x \neq 0$) \Leftrightarrow all eigenvalues $\lambda > 0$. Positive semi-definite: all $\lambda \geq 0$.
- $\det(A) = \lambda_1 \times \cdots \times \lambda_n$

Probability and Statistics

- **Continuous** random variables
- Probability density function (**p.d.f.**)

$$f(z) = \lim_{\Delta z \downarrow 0} \frac{Pr(z \leq Z \leq z + \Delta z)}{\Delta z}$$

Example: Normal r.v.

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-(z-\mu)^2/2\sigma^2}, \quad -\infty < z < \infty$$

Standard normal: $\mu = 0, \sigma = 1$ (**parameters**)

- Cumulative distribution function (**c.d.f.**)

$$F(z) = Pr(Z \leq z) = \int_{-\infty}^z f(z') dz'$$

$$f(z) = \frac{dF(z)}{dz}$$

- Implications

$$Pr(a \leq Z \leq b) =$$

$$\int_{-\infty}^{\infty} f(z) dz =$$

$$P(Z = z) =$$

Quantiles

- **Quantile** : the $100\alpha\%$ quantile q_α satisfies

$$\int_{-\infty}^{q_\alpha} f(z)dz = \alpha$$

Example: for standard normal, the 97.5% quantile is about
and 2.5% quantile is about

- **Median** : the 50% quantile
- For **symmetric** distributions, median = mean.

Probability and Statistics Ctd

- **Mean and variance**

$$E(Z) = \int_{-\infty}^{\infty} z f(z) dz$$

$$Var(Z) = \int_{-\infty}^{\infty} (z - E(Z))^2 f(z) dz$$

- **Covariance and correlation**

$$Cov(Z_1, Z_2) = E((Z_1 - E(Z_1))(Z_2 - E(Z_2)))$$

$$Cov(Z_1, Z_2) = Cov(Z_2, Z_1)$$

$$Cor(Z_1, Z_2) = \frac{Cov(Z_1, Z_2)}{\sqrt{Var(Z_1)}\sqrt{Var(Z_2)}}$$

If Z_1 and Z_2 are **independent** , $Cov(Z_1, Z_2) =$

Probability and Statistics Ctd

- Properties

$$E(aZ + b) = aE(Z) + b$$

$$Var(aZ + b) = a^2 Var(Z)$$

$$E(Z_1 + \cdots + Z_m) = E(Z_1) + \cdots + E(Z_m)$$

$$Var(Z_1 + \cdots + Z_m) = Var(Z_1) + \cdots + Var(Z_m) + 2 \sum_{j < j'} Cov(Z_j, Z_{j'})$$

- Vector of r.v.s: $Z = (Z_1, \dots, Z_m)$

$$E(Z) = (E(Z_1), \dots, E(Z_m))$$

$$Cov(Z) = \begin{pmatrix} Var(Z_1) & \cdots & \cdots & Cov(Z_1, Z_m) \\ Cov(Z_2, Z_1) & Var(Z_2) & \cdots & Cov(Z_2, Z_m) \\ \vdots & \ddots & \ddots & \vdots \\ Cov(Z_m, Z_1) & \cdots & \cdots & Var(Z_m) \end{pmatrix}$$

- If Z_j 's are independent, $Cov(Z) =$.
- General formulas for expected value/covariance of linear transformations

$$E(A_{n \times m} Z) =$$

$$Cov(A_{n \times m} Z) =$$

- Sample version: observe

	Z_1	Z_2	\cdots	Z_m
obs 1	z_{11}	z_{12}	\cdots	z_{1m}
obs 2	z_{21}	z_{22}	\cdots	z_{2m}
\vdots	\vdots	\vdots	\vdots	\vdots
obs n	z_{n1}	z_{n2}	\cdots	z_{nm}

Then

$$E_n(Z_j) = \bar{z}_j = \frac{1}{n}(z_{1j} + z_{2j} + \cdots + z_{nj})$$

$$Var_n(Z_j) = \frac{1}{n-1} \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2$$

$$Cov_n(Z_j, Z_{j'}) = \frac{1}{n-1} \sum_{i=1}^n (z_{ij} - \bar{z}_j)(z_{ij'} - \bar{z}_{j'})$$