Machine Learning, STOR 565

Random Vectors and the Multivariate Normal

Andrew Nobel

September, 2021

Review of the Univariate Case

Variance and Covariance

Recall: The variance of a random variable X is

$$\mathrm{Var}(X) \;=\; \mathbb{E}(X - \mathbb{E}X)^2 \;=\; \mathbb{E}X^2 - (\mathbb{E}X)^2$$

and the covariance of random variables X, Y is

$$\mathrm{Cov}(X,Y) \;=\; \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \;=\; \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$$

Basic Properties

- $\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X,Y)$
- $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$
- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y)$

Univariate Normal

Recall: Given $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ the $\mathcal{N}(\mu, \sigma^2)$ distribution has the (bell-shaped) density function

$$f(x) \;=\; \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad -\infty < x < \infty$$

- $\mu \in \mathbb{R}$ and $\sigma > 0$ called *parameters*; fully determine f
- *standard normal* is special case $\mu = 0$ and $\sigma^2 = 1$

Notation: If X has density f above, write $X \sim \mathcal{N}(\mu, \sigma^2)$

Univariate Normal

Basic Properties: If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

- $\mathbb{E}X = \mu$ and $\text{Var}(X) = \sigma^2$
- $X \stackrel{\text{d}}{=} \sigma Z + \mu$ where $Z \sim \mathcal{N}(0, 1)$
- $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

Fact: If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y \sim \mathcal{N}(\eta, \tau^2)$ are independent then

$$X + Y \sim \mathcal{N}(\mu + \eta, \sigma^2 + \tau^2)$$

Random Vectors

Random Vectors

Definition: A d-dimensional *random vector* is a vector of d random variables

$$\mathbf{X} = (X_1, \cdots, X_d)^t \in \mathbb{R}^d$$

The *expected value* of X is

$$\mathbb{E}(\mathbf{X}) = (\mathbb{E}X_1, \cdots, \mathbb{E}X_d)^t \in \mathbb{R}^d$$

Basic Properties: Let $a \in \mathbb{R}$, $\mathbf{v} \in \mathbb{R}^d$, and $\mathbf{A} \in \mathbb{R}^{k \times d}$ be non-random

- $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}(\mathbf{X}) + \mathbb{E}(\mathbf{Y})$
- $\mathbb{E}(a\mathbf{X} + \mathbf{v}) = a\,\mathbb{E}(\mathbf{X}) + \mathbf{v}$
- $\mathbf{AX} \in \mathbb{R}^k$ is a random vector and $\mathbb{E}(\mathbf{AX}) = \mathbf{A}\,\mathbb{E}(\mathbf{X})$

Variance Matrix of a Random Vector

Definition: The *covariance matrix* of a d-dimensional random vector $\mathbf{X}$ is

$$\mathrm{Var}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))^t] \in \mathbb{R}^{d \times d}$$

Basic Properties: Let $\mathbf{v} \in \mathbb{R}^d$ and $\mathbf{A} \in \mathbb{R}^{k \times d}$ be non-random

- $\mathrm{Var}(\mathbf{X})$ is symmetric and non-negative definite
- $\mathrm{Var}(\mathbf{X})_{ij} = \mathrm{Cov}(X_i, X_j)$
- $\mathrm{Var}(\mathbf{X} + \mathbf{v}) = \mathrm{Var}(\mathbf{X})$
- $\mathrm{Var}(\mathbf{A}\mathbf{X}) = \mathbf{A}\,\mathrm{Var}(\mathbf{X})\,\mathbf{A}^t$ (a $k \times k$ matrix)

The Multivariate Normal

Multivariate Normal

Definition: A random vector $\mathbf{X} \in \mathbb{R}^d$ is *multinormal* if for each $v \in \mathbb{R}^d$ the random variable $\langle \mathbf{X}, v \rangle$ is univariate normal

Fact: If $\mathbf{X} = (X_1, \ldots, X_d)^t$ is multinormal then $X_1, \ldots, X_d$ are univariate normal. However, the converse is *not* true.

Notation: If $\mathbf{X} \in \mathbb{R}^d$ is multinormal with $\mathbb{E}(\mathbf{X}) = \mu$ and $\mathrm{Var}(\mathbf{X}) = \Sigma$ write

$$\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma)$$

Write $\mathbf{X} \sim \mathcal{N}_d$ if $\mathbf{X} \in \mathbb{R}^d$ is multinormal, mean variance unspecified

Standard Multinormal

Example: Let $\mathbf{Z} = (Z_1, \ldots, Z_d)^t$ where $Z_1, \ldots, Z_d$ are iid $\mathcal{N}(0, 1)$. Then

$$\mathbb{E}(\mathbf{Z}) \;=\; \mathbf{0} \;\; \text{and} \;\; \mathrm{Var}(\mathbf{Z}) \;=\; \mathbf{I}_d$$

Moreover, $\mathbf{Z}$ is multinormal. Thus we have $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.

Terminology: $\mathbf{Z}$ is called the standard d-dimensional multinormal

Singular Multinormal

Example: Let Z be iid $\mathcal{N}(0,1)$ and define $Y = (Z,Z)^t$. Then

$$\mathbb{E}(Y) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \mathrm{Var}(Y) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Moreover, Y is multinormal. Thus we can write

$$Y \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right)$$

Basic Properties of Multivariate Normal

Fact: Suppose that $\mathbf{X} = (X_1, \ldots, X_d)^t \sim \mathcal{N}_d(\mu, \Sigma)$

- If $\mathbf{A} \in \mathbb{R}^{k \times d}$ and $\mathbf{v} \in \mathbb{R}^k$ then $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{v} \sim \mathcal{N}_k(\mathbf{A}\mu + \mathbf{v}, \mathbf{A}\Sigma\mathbf{A}^t)$
- $X_i \perp\!\!\!\perp X_j$ iff $\mathrm{Cov}(X_i, X_j) = 0$
- If $\mathbf{Y} \sim \mathcal{N}_d(\mu', \Sigma')$ is independent of $\mathbf{X}$ then

$$\mathbf{X} + \mathbf{Y} \ \sim \ \mathcal{N}_d(\mu + \mu', \Sigma + \Sigma')$$

Multivariate Normal Representation Theorem

Theorem: If $\mathbf{X}$ is multinormal with mean μ and variance Σ then

$$\mathbf{X} \overset{\mathsf{d}}{=} \Sigma^{1/2}\mathbf{Z} + \mu$$

- $\overset{\mathsf{d}}{=}$ means equality in distribution
- $\Sigma^{1/2} \geq 0$ is such that $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$
- $\mathbf{Z}$ is a standard multinormal, has iid $\mathcal{N}(0, 1)$ components

Upshot: Any multinormal random vector can be expressed as an affine transformation of a standard multinormal

Multivariate Normal Density

Note: Density of $\mathcal{N}(\mu, \sigma^2)$ can be written in the form

$$g(v) \;=\; \frac{1}{(2\pi)^{1/2}\,\sigma}\exp\left\{-\frac{1}{2}(v-\mu)(\sigma^2)^{-1}(v-\mu)\right\}$$

Fact: If $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma)$ with $\Sigma > 0$ then $\mathbf{X}$ has density

$$f(x) \;=\; \frac{1}{(2\pi)^{d/2}\det(\Sigma)^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu)^t\,\Sigma^{-1}\,(x-\mu)\right\}$$

Proof: Note $\Sigma > 0$ implies $\det(\Sigma) > 0$ and Σ^{-1} exists. Applying change of variables theorem to representation $\mathbf{X} \overset{d}{=} \Sigma^{1/2}\mathbf{Z} + \mu$ gives density $f(x)$.

Density of Standard Multinormal

Example: Standard multinormal vector $\mathbf{Z} \sim \mathcal{N}_d(0, I)$ has density

$$f(z) \;=\; \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2} z^t z\right\} \;=\; \prod_{i=1}^{d} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{z_i^2}{2}\right\}$$

Note: Here $z = (z_1, \ldots, z_d)^t$. Product form follows as components of $\mathbf{Z}$ are independent standard normals.

Bivariate Normal Density

Ex: Random vector $(X, Y)^t \sim \mathcal{N}_2$ with $\mathrm{Corr}(X, Y) = \rho$ has joint density

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\}$$

- Here $\mu_X = \mathbb{E}X$, $\mu_Y = \mathbb{E}Y$, $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$
- Density is defined only if $-1 < \rho < 1$
- X and Y are independent if and only if $\rho = 0$