Theoretical Statistics, STOR 655

Random Vectors and the Multivariate Normal

Andrew Nobel

January 2022

Stein’s Lemma

Stein's Lemma (Gaussian Integration by Parts)

Stein's Lemma: Let $Z \sim \mathcal{N}(0,1)$ and let $f : \mathbb{R} \to \mathbb{R}$ have derivative f'. If $\mathbb{E}|f'(Z)|$ is finite then

$$\mathbb{E}(Zf(Z)) = \mathbb{E}f'(Z)$$

Idea of proof

- Show that $\mathbb{E}|f'(Z)|$ finite implies $\mathbb{E}|f(Z)|$ and $\mathbb{E}|Zf(Z)|$ finite.
- If f is zero outside a finite interval, use integration by parts.
- For general f use a truncation/approximation argument.

Corollary: If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $\mathbb{E}|f'(Z)| < \infty$ then $\mathbb{E}((X-\mu)f(X)) = \sigma^2 \mathbb{E}f'(X)$.

Application: Moments of the Normal Distibution

Let $X \sim \mathcal{N}(0, \sigma^2)$. We know $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = \sigma^2$. What about higher moments?

Fact: If $X \sim \mathcal{N}(0, \sigma^2)$ then $\mathbb{E}X^k = 0$ when k odd and for all $k \geq 1$

$$\mathbb{E}X^{2k} = \sigma^{2k} \prod_{l=1}^{k} (2l - 1)$$

Random Vectors

Random Vectors

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A d-dimensional *random vector* is a Borel-measurable function $X : \Omega \to \mathbb{R}^d$. Write

$$X = (X_1, \cdots, X_d)^t$$

where $X_i : \Omega \to \mathbb{R}$ is the i'th component of X. Note

1. X is a random vector iff each component X_i is a random variable.
2. If $A \in \mathbb{R}^{k \times d}$ then $Y = AX$ is a k-dimensional random vector

Distribution of a Random Vector

Definition: The *distribution* of X is the probability measure on $\mathbb{R}^d$ defined by

$$P(A) = \mathbb{P}(X \in A) \text{ for Borel } A \subseteq \mathbb{R}^d$$

- X is *continuous* if there is a function $f : \mathbb{R}^d \to [0, \infty)$ such that

  $$P(A) \; = \; \int_A f(x)\, dx$$

  Say f is the probability density function (pdf) of P, write $X \sim f$

- X is *discrete* if there is a function $p : \mathbb{R}^d \to [0, 1]$ such that

  $$P(A) \; = \; \sum_{x \in A} p(x)$$

  Say p is the probability mass function (pmf) of P, write $X \sim p$

Expectation of a Random Vector

Definition: Let $X = (X_1, \ldots, X_d)^t$ be a random vector. If $\mathbb{E}|X_i|$ is finite for each i, the *expected value* of X is given by

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

Basic Properties

1. If $v \in \mathbb{R}^k$ and $A \in \mathbb{R}^{k \times d}$ are non-random, $\mathbb{E}(AX + v) = A\,\mathbb{E}(X) + v$
2. If $Y \in \mathbb{R}^d$ is defined on the same probability space as X then $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$.

Note: Entry-wise definition of expectation extends to random matrices

Variance Matrix of a Random Vector

Definition: Let $X = (X_1, \ldots, X_d)^t$ be a random vector. If $\mathbb{E}X_i^2$ is finite for each i, the *variance matrix* of $\mathbf{X}$ is given by

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

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

1. $\mathrm{Var}(X)$ is symmetric and non-negative definite
2. $\mathrm{Var}(X) = \mathbb{E}(XX^t) - \mathbb{E}(X)\mathbb{E}(X)^t$
3. $\mathrm{Var}(X)_{ij} = \mathrm{Cov}(X_i, X_j)$
4. $\mathrm{Var}(X + v) = \mathrm{Var}(X)$
5. $\mathrm{Var}(AX) = A\,\mathrm{Var}(X)\,A^t$ (a $k \times k$ matrix)

Covariance Matrix of Two Random Vectors

Definition: Let $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^l$ be random vectors with $\mathbb{E}X_i^2$, $\mathbb{E}Y_j^2$ finite. The covariance matrix of X, Y is given by

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

Note that $\mathrm{Cov}(X, Y)$ is a $k \times l$ matrix.

Properties of the Covariance Matrix

1. $\operatorname{Cov}(X,Y) = \mathbb{E}XY^t - (\mathbb{E}X)(\mathbb{E}Y)^t$
2. $\operatorname{Cov}(X,Y)_{i,j} = \operatorname{Cov}(X_i, Y_j)$
3. If X and Y are independent then $\operatorname{Cov}(X,Y) = 0$
4. $\operatorname{Var}(X) = \operatorname{Cov}(X,X)$
5. If A, B are non-random matrices $\operatorname{Cov}(AX, BY) = A\operatorname{Cov}(X,Y)B$
6. If X, Y are of the same dimension then
$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + \operatorname{Cov}(X,Y) + \operatorname{Cov}(Y,X)$$
7. In general, $\operatorname{Cov}(X,Y) \neq \operatorname{Cov}(Y,X)$

The Multivariate Normal

Multivariate Normal

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

Note: A constant $c \in \mathbb{R}$ is regarded as $\mathcal{N}(c, 0)$

Fact: If $X = (X_1, \ldots, X_d)^t$ is multinormal then components $X_1, \ldots, X_d$ are univariate normal. In particular, $\mathbb{E}(X)$ and $\mathrm{Var}(X)$ are well defined.

Note: Converse is *not* true.

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

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

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

Standard Multinormal

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

$$\mathbb{E}(Z) = 0 \text{ and } \mathrm{Var}(Z) = I_d$$

Moreover, Z is multinormal. Thus $Z \sim \mathcal{N}_d\,(0, I_d)$.

Terminology: Call Z the standard d-dimensional multinormal

Singular Multinormal

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

$$Y \sim \mathcal{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right)$$

Basic Properties of Multivariate Normal

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

a. If $A \in \mathbb{R}^{k \times d}$ and $u \in \mathbb{R}^k$ then $Y = AX + u \sim \mathcal{N}_k(A\mu + u, A\Sigma A^t)$

b. Components $X_i \perp\!\!\!\perp X_j$ iff $\mathrm{Cov}(X_i, X_j) = 0$

c. If $Y \sim \mathcal{N}_d(\mu', \Sigma')$ is independent of $\mathbf{X}$ then

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

d. If $1 \leq i_1 \leq \cdots \leq i_r$ then $Y = (X_{i_1}, \ldots, X_{i_r}) \sim \mathcal{N}_r$

Cramer-Wold Theorem

Definition: Write $X \stackrel{\mathrm{d}}{=} Y$ if X, Y have the same distribution

Theorem: Let X, Y be d-dimensional random vectors. Then $X \stackrel{\mathrm{d}}{=} Y$ if and only if $\langle X, u \rangle \stackrel{\mathrm{d}}{=} \langle Y, u \rangle$ for each $u \in \mathbb{R}^d$

Proof: Characteristic functions

Upshot: The distribution of a random vector is fully determined by the distributions of its one-dimensional projections

Multivariate Normal Representation Theorem

Theorem: If X is multinormal with mean μ and variance Σ then

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

- Matrix $\Sigma^{1/2} \geq 0$ is such that $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$
- Z is a standard multinormal with iid $\mathcal{N}(0, 1)$ components

Corollary

1. The distribution of multinormal random vector is fully determined by its mean and variance
2. If $X \sim \mathcal{N}_d(\mu, \Sigma)$ with $\Sigma > 0$ then $(X - \mu)^t \Sigma^{-1} (X - \mu) \sim \chi^2_d$

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

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

Density of Standard Multinormal

Example: Standard multinormal vector $Z \sim \mathcal{N}_d(0, I)$ 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 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 $\rho = 0$

Independence of Multinormals

Definition: Random vectors $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^l$ are jointly multinormal if

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_{k+l}$$

Fact: If X, Y are jointly multinormal then $X \perp\!\!\!\perp Y$ iff $\mathrm{Cov}(X, Y) = 0$.

Cor: If $X \sim \mathcal{N}_d(\mu, \Sigma)$ then $AX \perp\!\!\!\perp BX$ if and only if $A\Sigma B^t = 0$.