SDS7102: Linear Models and Extensions

Multivariate Normal Distributions

Qiang Sun, Ph.D. <qiang.sun@mbzuai.ac.ae>

These slides are prepared by Eric Moulines.

August 19, 2025

MBZUAI

Scalar normal random variable

Definition

A random variable Y has the normal distribution with mean μ and variance σ^2, denoted $Y \sim N\left(\mu, \sigma^2\right)$ whose density is given by

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

We say that Y is standard normal if $\mu = 0$ and $\sigma = 1$.

The moment generating function (mgf) for the standard normal is

$$\begin{aligned} m_z(t) \equiv E\left[e^{tZ}\right] &= \int_{-\infty}^{\infty} e^{tz} f(z) dz = \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} \exp\left\{tz - z^2/2\right\} dz \\ &= \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} \exp\left\{-(z-t)^2/2 + t^2/2\right\} dz = \exp\left\{t^2/2\right\} \end{aligned}$$

Standard multivariate distribution

Definition

Let $\mathbf{Z}$ be a $p \times 1$ vector with each component $Z_i, i = 1, \ldots, p$ independently distributed with $Z_i \sim N(0, 1)$. Then $\mathbf{Z}$ has the standard multivariate normal distribution, denoted $\mathbf{Z} \sim N_p\left(0, \mathbf{I}_p\right)$, in p dimensions. The joint density of the standard multivariate normal can be written then as

$$p_{\mathbf{Z}}(\mathbf{z}) = (2\pi)^{-p/2} \exp\left\{-\sum_{i=1}^{p} z_i^2/2\right\}$$

Moment generating function of a random vector

Definition

The moment generating function of a multivariate random variable $\mathbf{X}$ is given by

$$m_{\mathbf{X}}(\mathbf{t}) = E\left\{e^{\mathbf{t}^T\mathbf{X}}\right\}$$

provided this expectation exists in a rectangle that includes the origin. More precisely, there exists $h_i > 0, i = 1, \ldots, p$, so that the expectation exists for all $\mathbf{t}$ such that $-h_i < t_i < h_i, i = 1, \ldots, p$

Key property of MGF I

Theorem

If moment generating functions for two random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ exist, then the cdf's for $\mathbf{X}_1$ and $\mathbf{X}_2$ are identical iff the MGF's are identical in an open rectangle that includes the origin.

Key property of MGF II

Theorem

Assume the random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p$ *each have MGFs* $m_{\mathbf{X}_j}(\mathbf{t}_j)$, $j = 1, \ldots, p$, *and that* $\mathbf{X} = \left(\mathbf{X}_1^T, \mathbf{X}_2^T, \ldots, \mathbf{X}_p^T\right)^T$ *has MGF* $m_{\mathbf{X}}(\mathbf{t})$, *where* $\mathbf{t}$ *is partitioned similarly. Then* $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p$ *are mutually independent iff*

$$m_{\mathbf{X}}(\mathbf{t}) = m_{\mathbf{X}_1}(\mathbf{t}_1) \times m_{\mathbf{X}_2}(\mathbf{t}_2) \times \ldots \times m_{\mathbf{X}_p}(\mathbf{t}_p)$$

for all $\mathbf{t}$ *in an open rectangle that includes the origin.*

MGF for a standard MVN distribution

The MGF for the standard multivariate normal distribution $\mathbf{Z} \sim N_p(0, \mathbf{I}_p)$ is:

$$
\begin{aligned}
m_{\mathbf{z}}(\mathbf{t}) &= E\left\{\exp\left(\mathbf{t}^T\mathbf{Z}\right)\right\} = E\left\{\exp\left(\sum_{i=1}^{p} t_i Z_i\right)\right\} = \prod_{i=1}^{p} m_{z_i}(t_i) \\
&= \exp\left\{\sum_{i=1}^{p} t_i^2/2\right\} = \exp\left\{\mathbf{t}^T\mathbf{t}/2\right\}
\end{aligned}
$$

From this the moment generating function for $\mathbf{X} = \mu + \mathbf{AZ}$ can be constructed:

$$
\begin{aligned}
m_{\mathbf{X}}(\mathbf{t}) &= E\left[e^{\mathbf{t}^T\mathbf{X}}\right] = E\left[e^{\mathbf{t}^T\mu + \mathbf{t}^T\mathbf{AZ}}\right] = e^{\mathbf{t}^T\mu} \times m_z\left(\mathbf{A}^T\mathbf{t}\right) \\
&= \exp\left\{\mathbf{t}^T\mu + \mathbf{t}^T\mathbf{A}\mathbf{A}^T\mathbf{t}/2\right\}
\end{aligned}
$$

which is a function of just μ and $\mathbf{AA}^T$.

MGF of a MVN distribution

- The moment generating function for $\mathbf{X} = \mu + \mathbf{AZ}$ can be constructed:

$$\begin{aligned} m_{\mathbf{X}}(\mathbf{t}) &= E\left[e^{\mathbf{t}^T\mathbf{X}}\right] = E\left[e^{\mathbf{t}^T\mu+\mathbf{t}^T\mathbf{AZ}}\right] \\ &= e^{\mathbf{t}^T\mu} \times m_z\left(\mathbf{A}^T\mathbf{t}\right) = \exp\left\{\mathbf{t}^T\mu + \mathbf{t}^T\mathbf{A}\mathbf{A}^T\mathbf{t}/2\right\} \end{aligned}$$

  which is a function of μ and $\mathbf{A}\mathbf{A}^T$.
- We know that $E[\mathbf{X}] = \mu$ and $\mathrm{Cov}(\mathbf{X}) = \mathbf{A}\mathbf{A}^T$.
- The multivariate normal distribution is characterized by its mean vector and covariance matrix.

Multivariate normal distribution

Definition

The p-dimensional vector $\mathbf{X}$ has the multivariate normal distribution with mean μ and covariance matrix $\mathbf{V}$, denoted by $\mathbf{X} \sim N_p(\mu, \mathbf{V})$, if and only if its moment generating function takes the form

$$m_{\mathbf{X}}(\mathbf{t}) = \exp\left\{\mathbf{t}^T\mu + \mathbf{t}^T\mathbf{V}\mathbf{t}/2\right\}$$

- An important point to be emphasized here is that the covariance matrix may be singular, leading to the singular multivariate normal distribution.
- In this singular normal distribution, the probability mass lies in a subspace, and the dimension of the subspace-the rank of the covariance matrix - will be important

How to sample an MVN(μ, V)

- for any nonnegative definite matrix $\mathbf{V}$, we can find a matrix $\mathbf{A}$ such that $\mathbf{V} = \mathbf{A}\mathbf{A}^T$.
- Hence, $\mathbf{Y} = \mu + \mathbf{A}\mathbf{Z}$ where $\mathbf{Z}$ is standard MVN is MVV$(\mu, \mathbf{V})$. The choice of the square root $\mathbf{A}$ does not matter.

Multivariate normal distribution

Theorem

The p-dimensional vector $\mathbf{X}$ is multivariate normal if and only if for any p-dimensional vector $\mathbf{a}$, $\mathbf{a}^\top \mathbf{X}$ is a scalar normal random variable.

Elementary properties 1

Theorem

If $\mathbf{X} \sim N_p(\mu, \mathbf{V})$ *and* $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X}$ *where* $\mathbf{a}$ *is* $q \times 1$, *and* $\mathbf{B}$ *is* $q \times p$, *then* $\mathbf{Y} \sim N_q\left(\mathbf{a} + \mathbf{B}\mu, \mathbf{B}\mathbf{V}\mathbf{B}^T\right)$.

Corollary

If $\mathbf{X}$ *is multivariate normal, then the joint distribution of any subset is multivariate normal.*

Elementary properties 2

Theorem

If $\mathbf{X} \sim N_p(\mu, \mathbf{V})$ *and* $\mathbf{V}$ *is nonsingular, then*

(a) *there exists a nonsingular matrix* $\mathbf{A}$ *such that* $\mathbf{V} = \mathbf{A}\mathbf{A}^T$,

(b) $\mathbf{A}^{-1}(\mathbf{X} - \mu) \sim N_p\left(\mathbf{0}, \mathbf{I}_p\right)$, *and*

(c) *the pdf is* $(2\pi)^{-p/2}|\mathbf{V}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{V}^{-1}(\mathbf{x} - \mu)\right\}$.

Decorrelation and independence

Theorem

Let $\mathbf{X} \sim N_p(\mu, \mathbf{V})$. Consider the following partition:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_m \end{bmatrix} \begin{array}{l} p_1 \\ p_2 \\ \vdots \\ p_m \end{array}, \; \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix}, \; \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \dots & \mathbf{V}_{1m} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \dots & \mathbf{V}_{2m} \\ \vdots & \vdots & & \vdots \\ \mathbf{V}_{m1} & \mathbf{V}_{m2} & \dots & \mathbf{V}_{mm} \end{bmatrix},$$

then $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$ are jointly independent iff $\mathbf{V}_{ij} = \mathbf{0}$ for all $i \neq j$.

Elementary property 3

Theorem

Let $\mathbf{X} \sim N_p(\mu, \mathbf{V})$, *and* $\mathbf{Y}_1 = \mathbf{a}_1 + \mathbf{B}_1\mathbf{X}, \mathbf{Y}_2 = \mathbf{a}_2 + \mathbf{B}_2\mathbf{X}$, *then* $\mathbf{Y}_1$ *and* $\mathbf{Y}_2$ *are independent iff* $\mathbf{B}_1 V \mathbf{B}_2^T = \mathbf{0}$.

Chi-square distribution

Definition

Let $\mathbf{Z} \sim N_p\left(\mathbf{0}, \mathbf{I}_p\right)$, then $\mathbf{U} = \mathbf{Z}^T\mathbf{Z} = \sum_{i=1}^{p} \mathbf{Z}_i^2$ has the chi-square distribution with p degrees of freedom, denoted by $U \sim \chi_p^2$.

MGF of a chi-square distribution

The moment generating function for U can be computed directly from the normal distribution as

$$\begin{aligned} m_U(t) = E\left[e^{tU}\right] &= E\left[\exp\left\{t\sum_{i=1}^{p} Z_i^2\right\}\right] \\ &= \prod_{i=1}^{p}\int_{-\infty}^{\infty}(2\pi)^{-\frac{1}{2}}\exp\left\{tz_i^2 - \frac{1}{2}z_i^2\right\}dz_i = (1-2t)^{-\frac{p}{2}} \end{aligned}$$

since

$$\begin{aligned} \int_{-\infty}^{\infty}(2\pi)^{-\frac{1}{2}}\exp\left\{tz^2 - \frac{1}{2}z^2\right\}dz &= \int_{-\infty}^{\infty}(2\pi)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(1-2t)z^2\right\}dz \\ &= (1-2t)^{-\frac{1}{2}}. \end{aligned}$$

Density of central chi-square distribution

The density for $U \sim \chi_p^2$ is given by

$$p_U(u) = \frac{u^{(p-2)/2} e^{-u/2}}{\Gamma(p/2) 2^{p/2}}$$

for $u > 0$, and zero otherwise. Obtaining the MGF from the density we have

$$\begin{aligned} m_U(t) = \int_0^\infty e^{tu} p_U(u) du &= \int_0^\infty \frac{u^{(p-2)/2} e^{-u\left(\frac{1}{2}-t\right)}}{\Gamma(p/2) 2^{p/2}} du \\ &= \frac{\Gamma(p/2)\left(\frac{1}{2}-t\right)^{-p/2}}{\Gamma(p/2) 2^{p/2}} = (1-2t)^{-p/2} \end{aligned}$$

Non-central chi-square distribution

Definition

Let $J \sim \text{Poisson}(\phi)$, and $(U \mid J = j) \sim \chi^2_{p+2j}$, then unconditionally, U has the noncentral chi-square distribution with noncentrality parameter ϕ, denoted by $U \sim \chi^2_p(\phi)$.

Using the characterization above, the density of the noncentral χ^2 can be written as a Poisson-weighted mixture:

$$p_U(u) = \sum_{j=0}^{\infty} \left[\frac{e^{-\phi}\phi^j}{j!} \right] \times \frac{u^{(p+2j-2)/2} e^{-u/2}}{\Gamma\left(\frac{p+2j}{2}\right) 2^{j+p/2}}$$

for $u > 0$ and zero otherwise.

Property 1

Theorem

If $U \sim \chi_p^2(\phi)$, *then its MGF is* $m_U(t) = (1-2t)^{-p/2}\exp\{2\phi t/(1-2t)\}$.

Property 1

Theorem

If $U \sim \chi^2_p(\phi)$*, then its MGF is* $m_U(t) = (1-2t)^{-p/2}\exp\{2\phi t/(1-2t)\}$*.*

Proof.

Taking the conditional route rather than directly using the density and employing Result 5.8, we have

$$
\begin{aligned}
E\left[e^{tU}\right] = E\left[E\left[e^{tU} \mid J=j\right]\right] &= E\left[(1-2t)^{-(p+2J)/2}\right] \\
&= \sum_{j=0}^{\infty}(1-2t)^{-(p+2j)/2}\phi^j e^{-\phi}/j! \\
&= (1-2t)^{-p/2}e^{-\phi}\sum_{j=0}^{\infty}[\phi/(1-2t)]^j/j! \\
&= (1-2t)^{-p/2}e^{-\phi}e^{\phi/(1-2t)}
\end{aligned}
$$

□

Property 2

Theorem

If $U_1, U_2, \ldots, U_m$ *are jointly independent, and* $U_i \sim \chi^2_{p_i}(\phi_i)$, *then* $U = \sum_{i=1}^m U_i \sim \chi^2_p(\phi)$ *where* $p = \sum_{i=1}^m p_i$ *and* $\phi = \sum_{i=1}^m \phi_i$.

Property 2

Theorem

If $U_1, U_2, \ldots, U_m$ *are jointly independent, and* $U_i \sim \chi^2_{p_i}(\phi_i)$*, then* $U = \sum_{i=1}^m U_i \sim \chi^2_p(\phi)$ *where* $p = \sum_{i=1}^m p_i$ *and* $\phi = \sum_{i=1}^m \phi_i$*.*

Proof.

Obtaining the MGF for U we have

$$\begin{aligned}
m_U(t) = E\left[e^{t(\sum U_i)}\right] &= \prod_{i=1}^m m_{U_i}(t) \\
&= \prod_{i=1}^m \left[(1-2t)^{-p_i/2} \exp\left\{2t\phi_i/(1-2t)\right\}\right] \\
&= (1-2t)^{-p/2} \exp\{2t\phi/(1-2t)\}.
\end{aligned}$$

□

Property 3

Theorem

If $U \sim \chi_p^2(\phi)$, *then* $E(U) = p + 2\phi$ *and* $\mathrm{Var}(U) = 2p + 8\phi$.

Property 4

Theorem

If $X \sim N(\mu, 1)$*, then* $U = X^2 \sim \chi_1^2\left(\mu^2/2\right)$.

Property 4

Theorem

If $X \sim N(\mu, 1)$, *then* $U = X^2 \sim \chi_1^2\left(\mu^2/2\right)$.

Proof.

Finding the moment generating function for U, we have

$$
\begin{aligned}
m_U(t) =& E\left[e^{tX^2}\right] = \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} \exp\left\{tx^2 - (x-\mu)^2/2\right\} dx \\
=& \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[x^2 - 2x\mu + \mu^2 - 2tx^2\right]\right\} dx \\
=& \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} \exp\left\{-(1-2t)(x - \mu/(1-2t))^2/2\right\} dx \\
& \times \exp\left\{-\frac{1}{2}\left(\mu^2 - \mu^2/(1-2t)\right)\right\} \\
=& (1-2t)^{-\frac{1}{2}} \times \exp\left\{\left(\frac{1}{2}\mu^2\right) 2t/(1-2t)\right\}
\end{aligned}
$$

□

Theorem

If $\mathbf{X} \sim N_p(\mu, \mathbf{I}_p)$, *then* $W = \mathbf{X}^T\mathbf{X} = \sum_{i=1}^{p} X_i^2 \sim \chi_p^2\left(\frac{1}{2}\mu^T\mu\right)$.

Proof.

Since $W = \sum_{i=1}^{p} U_i$ where U_i are independent (since $V_{ij} = 0$ for $i \neq j$), and $U_i \sim \chi_{p_i}^2(\phi_i)$ where $p_i = 1, \phi_i = \frac{1}{2}\mu_i^2$, Property 2 provides the result, since $\sum_{i=1}^{p} \phi_i = \frac{1}{2}\mu^T\mu$. $\square$

Property IV

Theorem

Let $\mathbf{X} \sim N_p\left(\mu, \mathbf{I}_p\right)$ *and* $\mathbf{A}$ *be symmetric; then if* $\mathbf{A}$ *is idempotent with rank* s*, then* $\mathbf{X}^T \mathbf{A} \mathbf{X} \sim \chi_s^2\left(\phi = \frac{1}{2}\mu^T \mathbf{A} \mu\right)$*.*

Property V

Theorem

Let $\mathbf{X} \sim N_p(\mu, \mathbf{V})$ *and* $\mathbf{A}$ *be symmetric with ranks; if* $\mathbf{BVA} = \mathbf{0}$, *then* $\mathbf{BX}$ *and* $\mathbf{X}^T\mathbf{AX}$ *are independent. Here* $\mathbf{B}$ *is* $q \times p$.

Mean and variance of Gaussian sample

Suppose that $X_1, \cdots, X_n$ are i.i.d. Normal random variables with mean μ and variance σ^2 and define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

$\bar{X}$ and S^2 are called the sample mean and sample variance respectively. We know already that $\bar{X} \sim N\left(\mu, \sigma^2/n\right)$. The following results indicates that $\bar{X}$ is independent of S^2 and that the distribution of S^2 is related to a χ^2 with $n-1$ degrees of freedom.

Gosset theorem

Proposition

$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$ *and is independent of* $\bar{X} \sim N\left(\mu, \sigma^2\right)$.

Definition

Let $Z \sim N(0,1)$ and $V \sim \chi^2(n)$ be independent random variables. Define $T = Z/\sqrt{V/n}$; the random variable T is said to have Student's t distribution with n degrees of freedom. ($T \sim \mathcal{T}(n)$.)

p.d.f of a Student's t-distribution

Suppose that $Z \sim N(0,1)$ and $V \sim \chi^2(n)$ are independent random variables, and define $T = Z/\sqrt{V/n}$.

- determine the density of T

Student's t-distribution

Suppose that $X_1, \cdots, X_n$ are i.i.d. Normal random variables with mean μ and variance σ^2. Define the sample mean and variance of the X_i 's:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

Now define $T = \sqrt{n}(\bar{X} - \mu)/S$;

- Show that $T \sim \mathcal{T}(n-1)$