Stats 206 - 10/5/15

Multivariate normal distribution

- Practical: commonly used approximation, often justified by CLT, may be bad for discrete data, data on a bounded subset of $\mathbb{R}^p$, skew data.
- In theory: asymptotics (CLT again), often a simple or basic case, eg. results that usually hold only asymptotically may hold in finite normal samples, uncorrelated <=> independent (this class), best (mean-square) predictor is linear (next class)

Defn #1

Given iid $N(0,1)$ $z_1, \dots, z_p$,

$$f_z(z) = (2\pi)^{-1/2} \exp\left(-\frac{z^2}{2}\right), \qquad z \sim N(0,1),$$

the vector $\underset{p\times 1}{Z} = (z_1, \dots, z_p)$ is mv normal with mean $\underset{p\times 1}{0}$, covariance I_p,

$$f_Z(z) = (2\pi)^{-p/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{p} z_i^2\right) \qquad Z \sim N(\underset{p\times 1}{0}, \underset{p\times p}{I_p}).$$

The vector $\underset{p\times 1}{X}$ is mv normal with mean μ, covariance $\underset{p\times p}{\Sigma}$, $X \sim N_p(\mu, \Sigma)$, means

$$X = \mu + \Sigma^{1/2} Z, \qquad Z \sim N_p(0, I)$$

($\Sigma^{1/2}\Sigma^{1/2} = \Sigma$, unique for $\Sigma > 0$ — more later)

If Σ is rank p, there is a density in $\mathbb{R}^p$:

$$f_X(x) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

From HW#1, $\mathbb{E}X = \mu$, $\mathrm{Cov}\, X = \Sigma$

Comparing these densities

- chief similarity is negative quadratic in the exponential (gaussian tails)
- σ^2 corresponds to Σ. Where does $|\Sigma|$ come from? (Jacobian of $Z \mapsto \mu + \Sigma^{1/2} Z$, greater volume <=> diluted density)

- Writing $N_p(0,I)$ density as $f_Z(z) = (2\pi)^{-p/2} \exp\left(-\frac{\|z\|^2}{2}\right)$, squared Euclidean distance in $N_p(0,I)$ case corresponds to ——— in $N_p(\mu,\Sigma)$ case.

<u>Ex.</u> $p=2$, $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \cdot & \sigma_{22}^2 \end{pmatrix}\right)$

$$f_X(x) = (2\pi)^{-1}(\sigma_{11}^2\sigma_{22}^2 - \sigma_{12}^2)^{-1/2} \exp\Big(-\tfrac{1}{2} \underbrace{(x_1-\mu_1, x_2-\mu_2)\begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \cdot & \sigma_{22}^2 \end{pmatrix}^{-1}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}}\Big)$$

requires $|\Sigma|>0 \iff \left(1-\left(\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}\right)^2\right)>0 \iff \rho \neq \pm 1$

$$= \frac{1}{\sigma_{11}^2\sigma_{22}^2 - \sigma_{12}^2}(x_1-\mu_1, x_2-\mu_2)\begin{pmatrix} \sigma_{22}^2 & -\sigma_{12} \\ & \sigma_{11}^2 \end{pmatrix}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

$$= \frac{\sigma_{22}^2(x_1-\mu_1)^2 + \sigma_{11}^2(x_2-\mu_2)^2 - 2\sigma_{12}(x_1-\mu_1)(x_2-\mu_2)}{\sigma_{11}^2\sigma_{22}^2 - \sigma_{12}^2}$$

$$= \frac{1}{1-(\sigma_{12}/\sigma_{11}\sigma_{22})^2}\left(\left(\frac{x_1-\mu_1}{\sigma_{11}}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_{22}}\right)^2 - 2\frac{\sigma_{12}}{\sigma_{11}^2\sigma_{22}^2}(x_1-\mu_1)(x_2-\mu_2)\right)$$

$$f_X(x) = \frac{1}{2\pi\sigma_{11}\sigma_{22}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_{11}}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_{22}}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right)\right)$$

with $\rho := \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}} = \mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}}$

(Already getting messy in the $p=2$ case)

- We can play with the parameters at: haben.shinyapps.io/stats206
- R tools
  - MASS::mvrnorm to generate mv normal data — older routine but common
  - mvtnorm::rmvnorm, ::dmvnorm, etc — modern alternative, interface is more consistent with other R pseudorandom routines.

Ex If $X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \sim N_p(\mu, \Sigma)$ then a permutation of X, say $\tilde{X} = \begin{pmatrix} X_{\pi(1)} \\ \vdots \\ X_{\pi(p)} \end{pmatrix}$ is also mv normal. Use a permutation matrix Γ (single 1 in each row and column, 0 elsewhere) and the definition,

$$X = \mu + \Sigma^{1/2} Z \Rightarrow \tilde{X} = \Gamma X = \Gamma\mu + \Gamma \Sigma^{1/2} Z, \qquad Z \sim N_p(0, I).$$

Properties

- Level sets. $X \sim N_p(0, I)$, $f_X(x) = (2\pi)^{-p/2} \exp(-\frac{1}{2}\frac{\|z\|^2}{2})$ is a function only of the length $\|z\|^2$, so X is spherically symmetric. Alternatively look at ΓZ for a rotation matrix Γ (cf HWH1) and find $\Gamma Z \sim Z$.

  If $X \sim N_p(\mu, \Sigma)$, $f_X(x) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu))$ is a function only of the Mahalanobis (JW: "statistical") distance to μ. What are the level sets $\{x \in \mathbb{R}^p : (x-\mu)^T \Sigma^{-1}(x-\mu) = \text{constant}\}$?

Ex $p=2$. $\frac{(x_1-\mu_1)^2}{\sigma_{11}^2(1-\rho^2)} + \frac{(x_2-\mu_2)^2}{\sigma_{22}^2(1-\rho^2)} - \frac{2\rho}{1-\rho^2}\left(\frac{x_1-\mu_1}{\sigma_{11}}\right)\left(\frac{x_2-\mu_2}{\sigma_{22}}\right) = \text{constant}$ defines an ellipse centered at (μ_1, μ_2) rotated through some angle determined by the cross-term. To study the general case...

Review of eigenvectors / eigenvalues

- Given square $\underset{p\times p}{A}$, $\underset{p\times 1}{e} \neq 0$ is an eigenvector with eigenvalue λ means $Ae = \lambda e$
- These tell us the subspaces of $\mathbb{R}^p$ (A's domain/range) fixed by A.
- $Ae = \lambda e$ iff $(A - \lambda I)e = 0$ iff $\det(A - \lambda I) = 0$. This is degree p in λ so there are λ eigenvalues – maybe repeated, maybe complex.

- If $\lambda_1, \ldots, \lambda_p$ and $e_1, \ldots, e_p$ are eigenvalues and eigenvectors of A, the definition has matrix form

$$AE = E\Lambda \qquad \begin{pmatrix} | & & | \\ a_1 & \cdots & a_p \\ | & & | \end{pmatrix} \begin{pmatrix} | & & | \\ e_1 & \cdots & e_p \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ e_1 & \cdots & e_p \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix}$$ and if

$e_1, \ldots, e_p$ are independent, A can be written $A = E\Lambda E^{-1}$. But you don't generally have independent eigenvectors. Also they could be complex (eg rotation). The situation is simpler for (real) symmetric A, $A = A^T$.

$A = A^T$ implies

i) $\exists$ p orthogonal eigenvectors (not necessarily unique, eg. identity)

ii) all eigenvalues, eigenvectors are real

The eigenvalue equation $AQ = Q\Lambda$ can the be written

$$A = Q\Lambda Q^T \qquad \begin{pmatrix} | & & | \\ a_1 & \cdots & a_p \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ q_1 & \cdots & q_p \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} \begin{pmatrix} - & q_1 & - \\ & \vdots & \\ - & q_p & - \end{pmatrix}$$ ("spectral theorem")

Equivalently using 1-dimensional projection matrices,

$$A = \sum_{i=1}^{p} \lambda_i q_i q_i^T$$

Think of A's action as scaling by λ_i in direction q_i, $i = 1, \ldots, p$.

- $A = Q\Lambda Q^T$ makes transparent inverse $Q\Lambda^{-1}Q$ and square root $Q\Lambda^{1/2}Q^T$

- If additionally A is positive definite,

$$A \text{ p.d.} \iff x^T A x > 0 \quad \text{f.a.} \quad x \in \mathbb{R}^p,\ x \neq 0$$

then also

i) $\lambda_i > 0$, $i = 1, \ldots, p$

ii) $A^{1/2}$ is unique (This makes the notation $N_p(\mu, \Sigma)$ for $X = \mu + \Sigma^{1/2} Z$ well-defined for full-rank Σ. Another fact: $\text{rank}(AA^T) = \text{rank}(A)$ so $\text{rank}(\Sigma^{1/2}) = \text{rank}\,\Sigma$.)

- Back to level curves of $(x-\mu)^T \Sigma^{-1}(x-\mu)$. If $\text{rank}\,\Sigma = p$, $\Sigma = Q\Lambda Q^T$ is pd,

$$(x-\mu)^T \Sigma^{-1}(x-\mu) = c \iff (x-\mu)^T Q \Lambda^{-1} Q^T (x-\mu) = c$$
$$\iff y^T \Lambda^{-1} y = c \iff \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} = c$$

giving a p-dimensional ellipse with axes given by the eigenvectors of Σ, $q_1, \ldots, q_p$, and axis lengths $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_p}$, the eigenvalues. (Visualize bivariate case using Rshiny app given earlier.)

- uncorrelated $\iff$ independent

Given $X \sim N_p(\mu, \Sigma)$ written $\begin{matrix} q\times 1 \\ (p-q)\times 1 \end{matrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ & \Sigma_{22} \end{pmatrix} \right)$,

X_1 and X_2 uncorrelated means $\underset{q\times(p-q)}{\Sigma_{12}} = 0$, so Σ is block diagonal.

Block diagonal matrices are easy to work with. The density of X is

$$f_X(x) = (2\pi)^{-p/2} \begin{vmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{vmatrix}^{-1/2} \exp\left(-\frac{1}{2} (x_1-\mu_1, x_2-\mu_2) \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} \right)$$

$$= (2\pi)^{-p/2} |\Sigma_{11}|^{-1/2} |\Sigma_{22}|^{-1/2} \exp\left(-\frac{1}{2} \left((x_1-\mu_1)^T \Sigma_{11}^{-1} (x_1-\mu_1) + (x_2-\mu_2)^T \Sigma_{22}^{-1} (x_2-\mu_2) \right) \right)$$

ie the density factors.

- Holds for any uncorrelated subsets of X using the permutation example.
- Note for this equivalence to hold, (X_1, X_2) must be multivariate normal. Standard example: $X \sim N(0,1)$, flip a coin, take $Y = X$ if heads and $Y = -X$ if tails. Then X, Y are marginally normal, uncorrelated but not independent.
- Note in the last example $X + Y = 0$ w.p. $1/2$, so this linear combination is not normal. This leads to another definition of mv normal.

Definition #2 $(X_1, \ldots, X_p)$ is mv normal if $\sum_{i=1}^{p} c_i X_i$ is normal f.a. $c \in \mathbb{R}^p$.

- Linear transformations. Given $\underset{q \times p}{A}$, $X \sim N_p(\mu, \Sigma)$, then $AX \sim N_q(A\mu, A\Sigma A^T)$.

Use definition #2: given $c \in \mathbb{R}^p$, X normal $\Rightarrow$ $c^T AX = (c^T A)X$ is normal.

- Partitions. $\begin{array}{r} {\scriptstyle q \times 1} \\ {\scriptstyle (p-q) \times 1} \end{array} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \cdot & \Sigma_{22} \end{pmatrix} \right)$, taking $A := (I_q \quad 0_{q \times (p-q)})$,

$X_1 \sim N_q(\mu_1, \Sigma_{11})$. Nothing special about first q coordinates - permutation example.

- MGF/characteristic function/Fourier transform - another way to characterize the mv normal distribution. They would shorten many of the preceding results, but maybe not as illuminating.

SUMMARY

- 2 characterizations: transformation $X = \mu + \Sigma^{1/2} Z$, and normality preserved under linear combinations
- level sets
- permutations, transformations, lower dimensional projections all preserve normality
- uncorrelated $\Leftrightarrow$ independent
- aside on eigenvalues/eigenvectors, spectral representation