Lecture 4

Multivariate normal distribution and multivariate CLT.

We start with several simple observations. If $X = (x_1, \ldots, x_k)^T$ is a $k \times 1$ random vector then its expectation is

$$\mathbb{E}X = (\mathbb{E}x_1, \ldots, \mathbb{E}x_k)^T$$

and its covariance matrix is

$$\mathrm{Cov}(X) = \mathbb{E}(X - \mathbb{E}X)(X - \mathbb{E}X)^T.$$

Notice that a covariance matrix is always symmetric

$$\mathrm{Cov}(X)^T = \mathrm{Cov}(X)$$

and nonnegative definite, i.e. for any $k \times 1$ vector a,

$$a^T \mathrm{Cov}(X) a = \mathbb{E}a^T (X - \mathbb{E}X)(X - \mathbb{E}X)^T a^T = \mathbb{E}|a^T(X - \mathbb{E}X)|^2 \geq 0.$$

We will often use that for any vector X its squared length can be written as $|X|^2 = X^T X$. If we multiply a random $k \times 1$ vector X by a $n \times k$ matrix A then the covariance of $Y = AX$ is a $n \times n$ matrix

$$\mathrm{Cov}(Y) = \mathbb{E}A(X - \mathbb{E}X)(X - \mathbb{E}X)^T A^T = A\mathrm{Cov}(X)A^T.$$

Multivariate normal distribution. Let us consider a $k \times 1$ vector $g = (g_1, \ldots, g_k)^T$ of i.i.d. standard normal random variables. The covariance of g is, obviously, a $k \times k$ identity matrix, $\mathrm{Cov}(g) = I$. Given a $n \times k$ matrix A, the covariance of Ag is a $n \times n$ matrix

$$\Sigma := \mathrm{Cov}(Ag) = AIA^T = AA^T.$$

Definition. *The distribution of a vector Ag is called a (multivariate) normal distribution with covariance Σ and is denoted $N(0, \Sigma)$.*

One can also shift this disrtibution, the distribution of $Ag + a$ is called a normal distribution with mean a and covariance Σ and is denoted $N(a, \Sigma)$. There is one potential problem with the above definition - we assume that the distribution depends only on covariance matrix Σ and does not depend on the construction, i.e. the choice of g and a matrix A. For example, if we take a $m \times 1$ vector g' of i.i.d. standard normal random variables and a $n \times m$ matrix B then the covariance of Bg' is a $n \times n$ matrix

$$\mathrm{Cov}(Bg') = BB^T.$$

It is possible that $\Sigma = AA^T = BB^T$ so both constructions should give a normal distribution $N(0, \Sigma)$. This is, indeed, true - the distribution of Ag and Bg' is the same, so the definition of normal distribution $N(0, \Sigma)$ does not depend on the construction. It is not very difficult to prove that Ag and Bg' have the same distribution, but we will only show the simplest case.

Invertible case. Suppose that A and B are both square $n \times n$ invertible matrices. In this case, vectors Ag and Bg' have density which we will now compute. Since the density of g is

$$\prod_{i \leq n} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2} x_i^2\Big) = \Big(\frac{1}{\sqrt{2\pi}}\Big)^n \exp\Big(-\frac{1}{2}|x|^2\Big),$$

for any set $\Omega \in \mathbb{R}^n$ we can write

$$\mathbb{P}(Ag \in \Omega) = \mathbb{P}(g \in A^{-1}\Omega) = \int_{A^{-1}\Omega} \Big(\frac{1}{\sqrt{2\pi}}\Big)^n \exp\Big(-\frac{1}{2}|x|^2\Big) dx.$$

Let us now make the change of variables $y = Ax$ or $x = A^{-1}y$. Then

$$\mathbb{P}(Ag \in \Omega) = \int_{\Omega} \Big(\frac{1}{\sqrt{2\pi}}\Big)^n \exp\Big(-\frac{1}{2}|A^{-1}y|^2\Big) \frac{1}{|\det(A)|} dy.$$

But since

$$\det(\Sigma) = \det(AA^T) = \det(A)\det(A^T) = \det(A)^2$$

we have $|\det(A)| = \sqrt{\det(\Sigma)}$. Also

$$|A^{-1}y|^2 = (A^{-1}y)^T(A^{-1}y) = y^T(A^T)^{-1}A^{-1}y = y^T(AA^T)^{-1}y = y^T\Sigma^{-1}y.$$

Therefore, we get

$$\mathbb{P}(Ag \in \Omega) = \int_{\Omega} \Big(\frac{1}{\sqrt{2\pi}}\Big)^n \frac{1}{\sqrt{\det(\Sigma)}} \exp\Big(-\frac{1}{2} y^T \Sigma^{-1} y\Big) dy.$$

This means that a vector Ag has the density

$$\frac{1}{\sqrt{\det(\Sigma)}} \exp\Big(-\frac{1}{2} y^T \Sigma^{-1} y\Big)$$

which depends only on Σ and not on A. This means that Ag and Bg' have the same distributions.

$\square$

It is not difficult to show that in a general case the distribution of Ag depends only on the covariance Σ, but we will omit this here. Many times in these lectures whenever we want to represent a normal distribution $N(0,\Sigma)$ constructively, we will find a matrix A (not necessarily square) such that $\Sigma = AA^T$ and use the fact that a vector Ag for i.i.d. vector g has normal distribution $N(0,\Sigma)$. One way to find such A is to take a matrix square-root of Σ. Since Σ is a symmetric nonnegative definite matrix, its eigenvalue decomposition is

$$\Sigma = QDQ^T$$

for an orthogonal matrix Q and a diagonal matrix D with eigenvalues $\lambda_1, \ldots, \lambda_n$ of Σ on the diagonal. In Matlab, '[Q,D]=eig(Sigma);' will produce this decomposition. Then if $D^{1/2}$ represents a diagonal matrix with $\lambda_i^{1/2}$ on the diagonal then one can take

$$A = QD^{1/2} \quad \text{or} \quad A = QD^{1/2}Q^T.$$

It is easy to check that in both cases $AA^T = QDQ^T = \Sigma$. In Matlab $QD^{1/2}Q^T$ is given by 'sqrtm(Sigma)'. Let us take, for example, a vector $X = QD^{1/2}g$ for i.i.d. standard normal vector g which by definition has normal distribution $N(0,\Sigma)$. If $q_1, \ldots, q_n$ are the column vectors of Q then

$$X = QD^{1/2}g = (\lambda_1^{1/2}g_1)q_1 + \ldots + (\lambda_n^{1/2}g_n)q_n.$$

Therefore, in the orthonormal coordinate basis $q_1, \ldots, q_n$ a random vector X has coordinates $\lambda_1^{1/2}g_1, \ldots, \lambda_n^{1/2}g_n$. These coordinates are independent with normal distributions with variances $\lambda_1, \ldots, \lambda_n$ correspondingly. When $\det\Sigma = 0$, i.e. Σ is not invertible, some of its eigenvalues will be zero, say, $\lambda_{k+1} = \ldots = \lambda_n = 0$. Then the random vector will be concentrated on the subspace spanned by vectors $q_1, \ldots, q_k$ but it will not have density on the entire space $\mathbb{R}^n$. On the subspace spanned by vectors $q_1, \ldots, q_k$ a vector X will have a density

$$f(x_1, \ldots, x_k) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\lambda_i}} \exp\Big(-\frac{x_i^2}{2\lambda_i}\Big).$$

Linear transformation of a normal random vector.

Suppose that Y is a $n \times 1$ random vector with normal distribution $N(0,\Sigma)$. Then given a $m \times n$ matrix M, a $m \times 1$ vector MY will also have normal distribution $N(0, M\Sigma M^T)$. To show this, find any matrix A and i.i.d. standard normal vector g such that Ag has normal distribution $N(0,\Sigma)$. Then, by definition, $M(Ag) = (MA)g$ also has normal distribution with covariance

$$(MA)(MA)^T = MAA^TM^T = M\Sigma M^T.$$

□

Orthogonal transformation of an i.i.d. standard normal sample.

Throughout the lectures we will often use the following simple fact. Consider a vector $X = (X_1, \ldots, X_n)^T$ of i.i.d. random variables with standard normal distribution $N(0,1)$. If V is an orthogonal $n \times n$ matrix then the vector $Y := VX$ also consists of i.i.d. random variables $Y_1, \ldots, Y_n$ with standard normal distribution. A matrix V is orthogonal when one of the following equivalent properties hold:

1. $V^{-1} = V^T$.
2. The rows of V form an orthonormal basis in $\mathbb{R}^n$.
3. The columns of V form an orthonormal basis in $\mathbb{R}^n$.
4. For any $x \in \mathbb{R}^n$ we have $|Vx| = |x|$, i.e. V preserves the lengths of vectors.

Below we will use that $|\det(V)| = 1$. Basically, orthogonal transformations represent linear transformations that preserve distances between points, such as rotations and reflections. The joint p.d.f of a vector X is given by

$$f(x) = f(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(\sqrt{2\pi})^n} e^{-|x|^2/2},$$

where $|x|^2 = x_1^2 + \ldots + x_n^2$. To find the p.d.f. of a vector $Y = VX$, which is an linear transformation of X, we can use the change of density formula from probability or the change of variables formula from calculus as follows. For any set $\Omega \subseteq \mathbb{R}^n$,

$$\begin{aligned} \mathbb{P}(Y \in \Omega) &= \mathbb{P}(VX \in \Omega) = \mathbb{P}(X \in V^{-1}\Omega) \\ &= \int_{V^{-1}\Omega} f(x)dx = \int_{\Omega} \frac{f(V^{-1}y)}{|\det(V)|} dy. \end{aligned}$$

where we made the change of variables $y = Vx$. We know that $|\det(V)| = 1$ and, since $|V^{-1}y| = |y|$, we have

$$f(V^{-1}y) = \frac{1}{(\sqrt{2\pi})^n} e^{-|V^{-1}y|^2/2} = \frac{1}{(\sqrt{2\pi})^n} e^{-|y|^2/2} = f(y).$$

Therefore, we finally get that

$$\mathbb{P}(Y \in \Omega) = \int_{\Omega} f(y)dy$$

which proves that a vector Y has the same joint p.d.f. as X.

$\square$

Multivariate CLT.

We will state a multivariate Central Limit Theorem without a proof. Suppose that $X = (x_1, \ldots, x_k)^T$ is a random vector with covariance Σ. We assumed that $\mathbb{E}x_i^2 < \infty$. If $X_1, X_2, \ldots$ is a sequence of i.i.d. copies of X then

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbb{E}X_i) \to^d N(0, \Sigma),$$

where convergence in distribution $\to^d$ means that for any set $\Omega \in \mathbb{R}^k$,

$$\lim_{n\to\infty} \mathbb{P}(S_n \in \Omega) = \mathbb{P}(Y \in \Omega)$$

for a random vector Y with normal distribution $N(0, \Sigma)$.

$\square$