18.175 Theory of Probability
Fall 2008

Section 10

Multivariate normal distributions and CLT.

Let $\mathbb{P}$ be a probability distribution on $\mathbb{R}^k$ and let

$$g(t) = \int e^{i(t,x)} d\mathbb{P}(x).$$

We proved that $\mathbb{P}^\sigma = \mathbb{P} * \mathcal{N}(0, \sigma^2 I)$ has density

$$p^\sigma(x) = (2\pi)^{-k} \int g(t) e^{-i(t,x) - \frac{1}{2}\sigma^2 |t|^2} dt.$$

Lemma 20 *(Fourier inversion formula) If $\int |g(t)| dt < \infty$ then $\mathbb{P}$ has density*

$$p(x) = (2\pi)^{-k} \int g(t) e^{-i(t,x)} dt.$$

Proof. Since

$$g(t) e^{-i(t,x) - \frac{1}{2}\sigma^2 |t|^2} \to g(t) e^{-i(t,x)}$$

pointwise as $\sigma \to 0$ and

$$\left| g(t) e^{-i(t,x) - \frac{1}{2}\sigma^2 |t|^2} \right| \leq |g(t)| \text{ - integrable,}$$

by dominated convergence theorem $p^\sigma(x) \to p(x)$. Since $\mathbb{P}^\sigma \to \mathbb{P}$ weakly, for any $f \in C_b(\mathbb{R}^k)$,

$$\int f(x) p^\sigma(x) dx \to \int f(x) d\mathbb{P}(x).$$

On the other hand, since

$$|p^\sigma(x)| \leq (2\pi)^{-k} \int |g(t)| dt < \infty,$$

by dominated convergence theorem, for any compactly supported $f \in C_b(\mathbb{R}^k)$,

$$\int f(x) p^\sigma(x) dx \to \int f(x) p(x) dx.$$

Therefore, for any such f,

$$\int f(x) d\mathbb{P}(x) = \int f(x) p(x) dx.$$

It is now a simple exercise to show that for any bounded open set U,

$$\int_U d\mathbb{P}(x) = \int_U p(x)dx.$$

This means that $\mathbb{P}$ restricted to bounded sets has density $p(x)$ and, hence, on entire $\mathbb{R}^k$.

□

For a random vector $X = (X_1, \ldots, X_k) \in \mathbb{R}^k$ we denote $\mathbb{E}X = (\mathbb{E}X_1, \ldots, \mathbb{E}X_k)$.

Theorem 24 *Consider a sequence $(X_i)_{i\geq 1}$ of i.i.d. random vectors on $\mathbb{R}^k$ such that $\mathbb{E}X_1 = 0, \mathbb{E}|X_1|^2 < \infty$. Then $\mathcal{L}\Big(\frac{S_n}{\sqrt{n}}\Big)$ converges weakly to distribution $\mathbb{P}$ which has characteristic function*

$$f_p(t) = e^{-\frac{1}{2}(Ct,t)} \quad \textit{where} \quad C = \mathrm{Cov}(X_1) = \Big(\mathbb{E}X_{1,i}X_{1,j}\Big)_{1\leq i,j\leq k}. \tag{10.0.1}$$

Proof. Consider any $t \in \mathbb{R}^k$. Then $Z_i = (t, X_i)$ is i.i.d. real-valued sequence and by the proof of the CLT on the real line,

$$\mathbb{E}e^{i\left(t,\frac{S_n}{\sqrt{n}}\right)} = \mathbb{E}e^{i\frac{1}{\sqrt{n}}\sum_i (t,X_i)} \to e^{-\frac{1}{2}\mathrm{Var}((t,X_i))} = e^{-\frac{1}{2}(Ct,t)}$$

as $n \to \infty$, since

$$\mathrm{Var}\Big(t_1 X_{1,1} + \cdots + t_k X_{1,k}\Big) = \sum_{i,j} t_i t_j \mathbb{E}X_{1,i}X_{1,j} = (Ct,t) = t^T Ct.$$

The sequence $\Big\{\mathcal{L}\Big(\frac{S_n}{\sqrt{n}}\Big)\Big\}$ is uniformly tight on $\mathbb{R}^k$ since

$$\mathbb{P}\Big(\Big|\frac{S_n}{\sqrt{n}}\Big| \geq M\Big) \leq \frac{1}{M^2}\mathbb{E}\Big|\frac{S_n}{\sqrt{n}}\Big|^2 = \frac{1}{nM^2}\mathbb{E}\left|(S_{n,1},\ldots,S_{n,k})\right|^2 = \frac{1}{nM^2}\sum_{i\leq k}\mathbb{E}S_{n,i}^2 = \frac{1}{M^2}\mathbb{E}|X_1|^2 \stackrel{M\to\infty}{\longrightarrow} 0.$$

It remains to apply Lemma 19 from previous section.

□

The covariance matrix $C = \mathrm{Cov}(X)$ is symmetric and non-negative definite, $(Ct,t) = \mathbb{E}(t,X)^2 \geq 0$.

The unique distribution with c.f. in (10.0.1) is called a multivariate normal distribution with covariance C and is denoted by $\mathcal{N}(0,C)$. It can also be defined more constructively as follows. Consider an i.i.d. sequence $g_1,\ldots,g_n$ of $\mathcal{N}(0,1)$ r.v. and let $g = (g_1,\ldots,g_n)^T$. Given a $k\times n$ matrix the covariance matrix of $Ag \in \mathbb{R}^k$ is

$$C = \mathrm{Cov}(Ag) = \mathbb{E}Ag(Ag)^T = A\mathbb{E}gg^TA^T = AA^T.$$

The c.f. of Ag is

$$\mathbb{E}e^{i(t,Ag)} = \mathbb{E}e^{i(A^Tt,g)} = e^{-\frac{1}{2}|A^Tt|^2} = e^{-\frac{1}{2}t^TAA^Tt} = e^{-\frac{1}{2}(Ct,t)}.$$

This means that $Ag \sim \mathcal{N}(0,C)$. Interestingly, the distribution of Ag depends only on AA^T and does not depend on the choice of n and A.

Exercise. Show constructively, using linear algebra, that the distribution of Ag and Bg' is the same if $AA^T = BB^T$.

□

On the other hand, given a covariance matrix C one can always find A such that $C = AA^T$. For example, let $C = QDQ^T$ be its eigenvalue decomposition for orthogonal matrix Q and diagonal D. Since C is nonnegative definite, the elements of D are nonnegative. Then, one can take $n = k$ and $A = C^{1/2} := QD^{1/2}Q^T$ or $A = QD^{1/2}$.

Density in the invertible case. Suppose $\det(C) \neq 0$. Take A such that $C = AA^T$ so that $Ag \sim \mathcal{N}(0,C)$. Since the density of g is

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

for any set $\Omega \subseteq \mathbb{R}^k$ we can write

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

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

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

But since

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

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

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

Therefore, we get

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

This means that the distribution $\mathcal{N}(0, C)$ has the density

$$\Big(\frac{1}{\sqrt{2\pi}}\Big)^k \frac{1}{\sqrt{\det(C)}} \exp\Big(-\frac{1}{2}y^TC^{-1}y\Big).$$

□

General case. Let us take, for example, a vector $X = QD^{1/2}g$ for i.i.d. standard normal vector g so that $X \sim \mathcal{N}(0, C)$. If $q_1, \ldots, q_k$ are the column vectors of Q then

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

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

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

□

Let us look at a couple of properties of normal distributions.

Lemma 21 *If $X \sim \mathcal{N}(0, C)$ on $\mathbb{R}^k$ and $A : \mathbb{R}^k \to \mathbb{R}^m$ is linear then $AX \sim \mathcal{N}(0, ACA^T)$ on $\mathbb{R}^m$.*

Proof. The c.f. of AX is

$$\mathbb{E}e^{i(t,AX)} = \mathbb{E}e^{i(A^Tt,X)} = e^{-\frac{1}{2}(CA^Tt,A^Tt)} = e^{-\frac{1}{2}(ACA^Tt,t)}.$$

□

Lemma 22 *X is normal on $\mathbb{R}^k$ iff (t, X) is normal on $\mathbb{R}$ for all $t \in \mathbb{R}^k$.*

Proof. "$\Longrightarrow$". The c.f. of real-valued random variable (t, X) is

$$f(\lambda) = \mathbb{E}e^{i\lambda(t,X)} = \mathbb{E}e^{i(\lambda t,X)} = e^{-\frac{1}{2}(C\lambda t,\lambda t)} = e^{-\frac{1}{2}\lambda^2(Ct,t)}$$

which means that $(t, X) \sim \mathcal{N}(0, (Ct, t))$.

"$\Longleftarrow$". If (t, X) is normal then

$$\mathbb{E}e^{i(t,X)} = e^{-\frac{1}{2}(Ct,t)}$$

because the variance of (t, X) is (Ct, t).

□

Lemma 23 *Let $Z = (X, Y)$ where $X = (X_1, \ldots, X_i)$ and $Y = (Y_1, \ldots, Y_j)$ and suppose that Z is normal on $\mathbb{R}^{i+j}$. Then X and Y are independent iff $\mathrm{Cov}(X_m, Y_n) = 0$ for all m, n.*

Proof. One way is obvious. The other way around, suppose that

$$C = \mathrm{Cov}(Z) = \begin{bmatrix} D & 0 \\ 0 & F \end{bmatrix}.$$

Then the c.f. of Z is

$$\mathbb{E}e^{i(t,Z)} = e^{-\frac{1}{2}(Ct,t)} = e^{-\frac{1}{2}(Dt_1,t_1)-\frac{1}{2}(Ft_2,t_2)} = \mathbb{E}e^{i(t_1,X)}\mathbb{E}e^{i(t_2,Y)},$$

where $t = (t_1, t_2)$. By uniqueness, X and Y are independent.

□

Lemma 24 *(Continuous Mapping.) Suppose that $\mathbb{P}_n \to \mathbb{P}$ on X and $G : X \to Y$ is a continuous map. Then $\mathbb{P}_n \circ G^{-1} \to \mathbb{P} \circ G^{-1}$ on Y. In other words, if r.v. $Z_n \to Z$ weakly then $G(Z_n) \to G(Z)$ weakly.*

Proof. This is obvious, because for any $f \in C_b(Y)$, we have $f \circ G \in C_b(X)$ and therefore,

$$\mathbb{E}f(G(Z_n)) \to \mathbb{E}f(G(Z)).$$

□

Lemma 25 *If $\mathbb{P}_n \to \mathbb{P}$ on $\mathbb{R}^k$ and $\mathbb{Q}_n \to \mathbb{Q}$ on $\mathbb{R}^m$ then $\mathbb{P}_n \times \mathbb{Q}_n \to \mathbb{P} \times \mathbb{Q}$ on $\mathbb{R}^{k+m}$.*

Proof. By Fubini theorem, The c.f.

$$\int e^{i(t,x)} d\mathbb{P}_n \times \mathbb{Q}_n(x) = \int e^{i(t_1,x_1)} d\mathbb{P}_n \int e^{i(t_2,x_2)} d\mathbb{Q}_n \to \int e^{i(t_1,x_1)} d\mathbb{P} \int e^{i(t_2,x_2)} d\mathbb{Q} = \int e^{i(t,x)} d\mathbb{P} \times \mathbb{Q}.$$

By Lemma 19 it remains to show that $(\mathbb{P}_n \times \mathbb{Q}_n)$ is uniformly tight. By Theorem 21, since $\mathbb{P}_n \to \mathbb{P}$, $(\mathbb{P}_n)$ is uniformly tight. Therefore, there exists a compact K on $\mathbb{R}^k$ such that $\mathbb{P}_n(K) > 1 - \varepsilon$. Similarly, for some compact K' on $\mathbb{R}^m$, $\mathbb{Q}_n(K') > 1 - \varepsilon$. We have,

$$\mathbb{P}_n \times \mathbb{Q}_n(K \times K') > 1 - 2\varepsilon$$

and $K \times K'$ is a compact on $\mathbb{R}^{k+m}$.

□

Corollary 1 *If $\mathbb{P}_n \to \mathbb{P}$ and $\mathbb{Q}_n \to \mathbb{Q}$ both on $\mathbb{R}^k$ then $\mathbb{P}_n * \mathbb{Q}_n \to \mathbb{P} * \mathbb{Q}$.*

Proof. Since a function $G : \mathbb{R}^{k+k} \to \mathbb{R}^k$ given by $G(x, y) = x + y$ is continuous, by continuous mapping lemma,

$$\mathbb{P}_n * \mathbb{Q}_n = (\mathbb{P}_n \times \mathbb{Q}_n) \circ G^{-1} \to (\mathbb{P} \times \mathbb{Q}) \circ G^{-1} = \mathbb{P} * \mathbb{Q}.$$

□