The moment generating function (mgf) of X at α is defined as $\phi_X(\alpha) = E(\exp(\alpha' X))$. This uniquely determines the probability distribution of X. Note that $\phi_X((t_1, 0)')E(\exp(t_1 X_1)) = \phi_{X_1}(t_1)$. If X and Y are independent, $\phi_{X+Y}(t) = E(\exp(t'(X+Y))) = E(\exp(t'X)\exp(t'Y))$
$= E(\exp(t'X))E(\exp(t'Y)) = \phi_X(t)\phi_Y(t)$.

Theorem (Cramer-Wold device). If X is a random vector, its probability distribution is completely determined by the distribution of all linear functions, $\alpha' X$, $\alpha \in \mathcal{R}^p$.

Proof. The mgf of $\alpha' X$, for any $\alpha \in \mathcal{R}^p$ is $\phi_{\alpha' X}(t) = E(t\alpha' X)$. Suppose this is known for all $\alpha \in \mathcal{R}^p$. Now, for any α, note $\phi_X(\alpha) = E(\exp(\alpha' X) = \phi_{\alpha' X}(1)$, which is then known.

Remark. To define the joint multivariate distribution of a random vector, it is enough to specify the distribution of all its linear functions.

Multivariate Normal Distribution

Definition. $X_{p\times 1}$ is p-variate normal if for every $\alpha \in \mathcal{R}^p$, the distribution of $\alpha' X$ is univariate normal.

Result. If X has the p-variate normal distribution, then both $\mu = E(X)$ and $\Sigma = Cov(X)$ exist and the distribution of X is determined by μ and Σ.

Proof. Let $X = (X_1, \ldots, X_p)'$. Then for each i, $X_i = \alpha_i' X$ where $\alpha_i = (0, \ldots, 0, 1, 0, \ldots, 0)'$. Therefore, $X_i = \alpha_i' X \sim N(.,.)$. Hence, $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_{ii}$ exist. Also, since $|\sigma_{ij}| = |Cov(X_i, X_j)| \leq \sqrt{\sigma_{ii}\sigma_{jj}}$, σ_{ij} exists. Set $\mu = (\mu_1, \ldots, \mu_p)'$ and $\Sigma = ((\sigma_{ij}))$. Further, $E(\alpha' X) = \alpha' \mu$ and $Var(\alpha' X) = \alpha' \Sigma \alpha$, so

$$\alpha' X \sim N(\alpha' \mu, \alpha' \Sigma \alpha), \text{ for all } \alpha \in \mathcal{R}^p.$$

Since $\{\alpha' X, \alpha \in \mathcal{R}^p\}$ determine the distribution of X, μ and Σ suffice.
Notation: $X \sim N_p(\mu, \Sigma)$.

Result. If $X \sim N_p(\mu, \Sigma)$, then for any $A_{k\times p}$, $b_{k\times 1}$,
$Y = AX + b \sim N_k(A\mu + b, A\Sigma A')$.

Proof. Consider linear functions, $\alpha' Y = \alpha' AX + \alpha' b = \beta' X + c$, which are univariate normal. Therefore Y is k-variate normal. $E(Y) = A\mu + b$, $Cov(Y) = Cov(AX) = A\Sigma A'$.

Theorem. $X_{p\times 1} \sim N_p(\mu, \Sigma)$ iff $X_{p\times 1} = C_{p\times r} Z_{r\times 1} + \mu$ where $Z = (Z_1, \ldots, Z_r)'$, Z_i i.i.d $N(0,1)$, $\Sigma = CC'$, $r = \text{rank}(\Sigma) = \text{rank}(C)$.

Proof. if part: If $X = CZ + \mu$ and $Z \sim N_r(0, I_r)$, then $X \sim N_p(\mu, CC' = \Sigma)$.

Z is multivariate normal since linear functions of Z are linear combinations of Z_i's, which are univarite normal (as can be shown using the change of variable (jacobian) formula for joint densities, or using the mgf of normal).

Only if: If $X \sim\sim N_p(\mu, \Sigma)$, and $\text{rank}(\Sigma) = r \le p$, then consider the spectral decomposition, $\Sigma = H\Delta H'$, H orthogonal, $\Delta = \begin{pmatrix} \Delta_1 & 0 \\ 0 & 0 \end{pmatrix}$,
$\Delta_1 = \text{diagonal}(\delta_1, \ldots, \delta_r)$, $\delta_i > 0$. Now, $X - \mu \sim N(0, \Sigma)$, and
$H'(X-\mu) \sim N(0, \Delta)$. Let $H'(X-\mu) = \begin{pmatrix} Y_{r\times 1} \\ T_{(p-r)\times 1} \end{pmatrix}$. Then,

$$\begin{pmatrix} Y_{r\times 1} \\ T_{(p-r)\times 1} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Delta_1 & 0 \\ 0 & 0 \end{pmatrix}\right).$$

Therefore, $T = 0$ w.p. 1. Let $Z = \Delta_1^{-1/2} Y$. Then $Z \sim N_r(0, I_r)$. Therefore, w.p. 1, $H'(X-\mu) = \begin{pmatrix} \Delta_1^{1/2} Z \\ 0 \end{pmatrix}$. Further, w.p. 1,

$$X - \mu = H \begin{pmatrix} \Delta_1^{1/2} Z \\ 0 \end{pmatrix} = (H_1 | H_2) \begin{pmatrix} \Delta_1^{1/2} Z \\ 0 \end{pmatrix} = H_1 \Delta_1^{1/2} Z = CZ.$$

Also, $CC' = H_1 \Delta_1^{1/2} \Delta_1^{1/2} H_1' = H_1 \Delta_1 H_1'$ and

$$\Sigma = H\Delta H' = (H_1 | H_2) \begin{pmatrix} \Delta_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} H_1' \\ H_2' \end{pmatrix} = H_1 \Delta_1 H_1'.$$

Recall that if $Z_1 \sim N(0,1)$, its mgf is $\phi_{Z_1}(t) = E(\exp(tZ_1)) = \exp(t^2/2)$. Therefore, if $Z \sim N_r(0, I_r)$ then

$$\phi_Z(u) = E(\exp(u'Z)) = E(\exp(\sum_{j=1}^{r} u_j Z_j)) = \exp(\sum_{j=1}^{r} u_j^2/2) = \exp(\frac{1}{2}u'u).$$

Then, if $X \sim\sim N_p(\mu, \Sigma)$, its mgf is:

$$\phi_X(t) = \exp(t'\mu + \frac{1}{2} t'\Sigma t),$$

since $E(\exp(t'X)) = E(\exp(t'(CZ + \mu))) = \exp(t'\mu) E(\exp(t'CZ)) =$ $\exp(t'\mu)\exp(t'CC't/2) = \exp(t'\mu + t'\Sigma t/2)$.