We now discuss an example of a discrete random vector.

Remark 11.1. While considering a Bernoulli or a Binomial RV, we looked at random experiments with exactly two outcomes. We now consider random experiments with two or more than two outcomes. Suppose a random experiment terminates in one of k possible outcomes $A_1, A_2, \cdots, A_k$ for $k \geq 2$. More generally, we may also consider random experiments which terminate in one of k mutually exclusive and exhaustive events $A_1, A_2, \cdots, A_k$ with $k \geq 2$. Write $p_j = \mathbb{P}(A_j), j = 1, 2, \cdots, k$, which does not change from trial to trial. Then, $p_1 + p_2 + \cdots + p_k = 1$. Suppose n trials are conducted independently and let $X_j, j = 1, 2, \cdots, k$ denote the number of times event A_j has occured, respectively. Then the RVs $X_1, X_2, \cdots, X_k$ satisfy the relation $X_1 + X_2 + \cdots + X_k = n$ and we have

$$\mathbb{P}(X_1 = x_1, \cdots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

for $x_1, \cdots, x_k \in \{0, 1, \cdots, n\}$ with $x_1 + \cdots + x_k = n$. The probability is zero otherwise. Removing the redundancy we have the joint p.m.f. of $(X_1, X_2, \cdots, X_{k-1})$ given by

$$\begin{aligned} & f_{X_1, \cdots, X_{k-1}}(x_1, \cdots, x_{k-1}) \\ &= \frac{n!}{x_1! \cdots x_{k-1}!(n - x_1 - \cdots - x_{k-1})!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} (1 - p_1 - \cdots - p_{k-1})^{n - x_1 - \cdots - x_{k-1}} \end{aligned}$$

for $x_1, \cdots, x_k \in \{0, 1, \cdots, n\}$ with $x_1 + \cdots + x_{k-1} \leq n$ and zero otherwise.

Example 11.2 (Multinomial Distribution)**.** A random vector $X = (X_1, \cdots, X_{k-1})$ is said to follow the Multinomial distribution with parameters n and $p_1, p_2, \cdots, p_k$ if the joint p.m.f. is as in Remark 11.1 above. We now list some properties of the multinomial distribution.

(a) We first compute the joint MGF. For $t_1, t_2, \cdots, t_{k-1} \in \mathbb{R}$,

$$\begin{aligned} & M_X(t_1, t_2, \cdots, t_{k-1}) \\ &= \mathbb{E} \exp(t_1 X_1 + t_2 X_2 + \cdots + t_{k-1} X_{k-1}) \end{aligned}$$

$$= \sum_{\substack{x_1,\cdots,x_k\in\{0,1,\cdots,n\}\\ x_1+\cdots+x_{k-1}\le n}} \frac{n!\exp(t_1x_1+t_2x_2+\cdots+t_{k-1}x_{k-1})}{x_1!\cdots x_{k-1}!(n-x_1-\cdots-x_{k-1})!} p_1^{x_1}\cdots p_{k-1}^{x_{k-1}} p_k^{n-x_1-\cdots-x_{k-1}}$$

$$= \sum_{\substack{x_1,\cdots,x_k\in\{0,1,\cdots,n\}\\ x_1+\cdots+x_{k-1}\le n}} \frac{n!}{x_1!\cdots x_{k-1}!(n-x_1-\cdots-x_{k-1})!} \left(p_1e^{t_1}\right)^{x_1}\cdots\left(p_{k-1}e^{t_{k-1}}\right)^{x_{k-1}} p_k^{n-x_1-\cdots-x_{k-1}}$$

$$= \left(p_1e^{t_1}+p_2e^{t_2}+\cdots+p_{k-1}e^{t_{k-1}}+p_k\right)^n$$

(b) If $t=(t_1,0,\cdots,0)\in\mathbb{R}^{k-1}$, then $M_X(t)=\mathbb{E}\exp(t_1X_1)=M_{X_1}(t_1)$. But, using the above expression for the joint MGF, we have $M_{X_1}(t_1)=M_X(t)=(p_1e^{t_1}+p_2+\cdots+p_{k-1}+p_k)^n=(p_1e^{t_1}+1-p_1)^n$. Therefore, $X_1\sim Binomial(n,p_1)$. Similarly, $X_i\sim Binomial(n,p_i),\forall i=2,\cdots,k-1$. In particular, $\mathbb{E}X_i=np_i, Var(X_i)=np_i(1-p_i)$.

(c) For distinct indices $i,j\in\{1,2,\cdots,k-1\}$,

$$M_{X_i,X_j}(t_i,t_j)=M_X(0,\cdots,0,t_i,0,\cdots,0,t_j,0,\cdots,0)=\left(p_ie^{t_i}+p_je^{t_j}+1-p_i-p_j\right)^n,\forall(t_i,t_j)\in\mathbb{R}^2.$$

Therefore (X_i,X_j) follows the trinomial distribution with the parameters $p_i,p_j,1-p_i-p_j$, i.e. multinomial distribution with the parameters $n=3$ and $p_i,p_j,1-p_i-p_j$.

(d) For distinct indices $i,j\in\{1,2,\cdots,k-1\}$, consider $t_i=t_j=t\in\mathbb{R}$. Then,

$$M_{X_i+X_j}(t)=M_{X_i,X_j}(t,t)=\left[(p_i+p_j)\,e^t+1-(p_i+p_j)\right]^n,$$

which shows $X_i+X_j\sim Binomial(n,p_i+p_j)$. Then $Var(X_i+X_j)=n(p_i+p_j)(1-p_i-p_j)$. Using the relation

$$Var(X_i+X_j)=Var(X_i)+Var(X_j)+2Cov(X_i,X_j),$$

we have $Cov(X_i,X_j)=-np_ip_j$. Consequently, the correlation between X_i and X_j is

$$\rho(X_i,X_j)=\frac{Cov(X_i,X_j)}{\sqrt{Var(X_i)\,Var(X_j)}}=-\left(\frac{p_ip_j}{(1-p_i)(1-p_j)}\right)^{\frac{1}{2}}.$$

Note 11.3. We now look at distributions that arise in practice from random samples. Such distributions are usually referred to as sampling distributions. More specifically, if $X_1,X_2,\cdots,X_n$

is a random sample from $N(\mu, \sigma^2)$ distribution, we shall look at various statistics involving the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Note 11.4 (Distribution of square of a standard Normal RV)**.** Let $X \sim N(0,1)$. Recall that the p.d.f. of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \forall x \in \mathbb{R}$$

We consider the distribution of $Y = X^2$ by first computing the MGF. We have,

$$M_Y(t) = \mathbb{E}e^{tX^2} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx^2} e^{\left(-\frac{x^2}{2}\right)}\, dx = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\infty} e^{\left(t-\frac{1}{2}\right)x^2}\, dx = (1-2t)^{-\frac{1}{2}}, \forall t < \frac{1}{2}$$

Comparing with the MGF of the $Gamma(\alpha, \beta)$ distribution, we conclude that $X^2 \sim Gamma(\frac{1}{2}, 2)$.

Note 11.5. If $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$ and hence $\left(\frac{X-\mu}{\sigma}\right)^2 \sim Gamma(\frac{1}{2}, 2)$.

Remark 11.6 (Distribution of the sample mean for a random sample from the Normal distribution)*.* If $X_1, X_2, \cdots, X_n$ is a random sample from $N(\mu, \sigma^2)$ distribution, then for $Y = X_1 + X_2 + \cdots + X_n$, using independence of X_i's we have

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \exp(n\mu t + \frac{1}{2} n\sigma^2 t^2)$$

and hence $X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)$. Now, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and consequently, $\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1)$ and $n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 \sim Gamma(\frac{1}{2}, 2)$.

Note 11.7. Let $X_1, X_2, \cdots, X_n$ be independent RVs with $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \cdots, n$. Then $\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2, i = 1, 2, \cdots, n$ are i.i.d. with the common distribution $Gamma(\frac{1}{2}, 2)$. The independence follows from Remark 9.32(h). Using problem set 9, we have

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim Gamma\left(\frac{n}{2}, 2\right).$$

Definition 11.8 (Chi-Squared distribution with n degrees of freedom)**.** Let n be a positive integer. We refer to the $Gamma\left(\frac{n}{2}, 2\right)$ distribution as the Chi-Squared distribution with n degrees of freedom. If an RV X follows the Chi-Squared distribution with n degrees of freedom, we write $X \sim \chi_n^2$.

Note 11.9. Using Note 11.7, we conclude that $\sum_{i=1}^n \left(\frac{X_i-\mu_i}{\sigma_i}\right)^2 \sim \chi^2_n$, where $X_1, X_2, \cdots, X_n$ are independent RVs with $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \cdots, n$.

Note 11.10. As argued in Note 11.7, using Remark 9.32(h) we conclude that $X + Y \sim \chi^2_{m+n}$, where X, Y are independent RVs with $X \sim \chi^2_m$ and $Y \sim \chi^2_n$.

Note 11.11. If $X \sim \chi^2_n$, then using properties of the *Gamma* $\left(\frac{n}{2}, 2\right)$ distribution, we have $\mathbb{E}X = n, Var(X) = 2n$ and $M_X(t) = (1-2t)^{-\frac{n}{2}}, \forall t < \frac{1}{2}$.

Remark 11.12 (Sample Mean and Sample Variance are independent corresponding to a random sample drawn from a Normal distribution). Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ distribution. Consider the sample mean $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ and sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Show that $\bar{X}$ and S_n^2 are independent.

Look at the joint MGF of $(X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X}, \bar{X})$ given by

$$\begin{aligned} M(t_1, t_2, \cdots, t_n, t_{n+1}) &= \mathbb{E}\exp\left(\sum_{j=1}^n t_j(X_j - \bar{X}) + t_{n+1}\bar{X}\right), \forall (t_1, t_2, \cdots, t_n, t_{n+1}) \in \mathbb{R}^{n+1} \\ &= \mathbb{E}\exp\left(\sum_{j=1}^n s_j X_j\right), \end{aligned}$$

where $s_j = t_j + \frac{t_{n+1} - \sum_{i=1}^n t_i}{n}$. Using the independence of X_j's, we have

$$\begin{aligned} M(t_1, t_2, \cdots, t_n, t_{n+1}) &= \prod_{j=1}^n \mathbb{E}\exp(s_j X_j) \\ &= \prod_{j=1}^n \exp\left(\mu s_j + \frac{1}{2}\sigma^2 s_j^2\right) \\ &= \exp\left(\mu \sum_{j=1}^n s_j + \frac{1}{2}\sigma^2 \sum_{j=1}^n s_j^2\right) \\ &= \exp\left(\mu t_{n+1} + \frac{1}{2}\sigma^2 \frac{t_{n+1}^2}{n}\right) \exp\left(\frac{1}{2}\sigma^2 \sum_{j=1}^n \left(t_j - \frac{\sum_{i=1}^n t_i}{n}\right)^2\right) \\ &= M_{\bar{X}}(t_{n+1}) M_{X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X}}(t_1, t_2, \cdots, t_n). \end{aligned}$$

Here, we use the observation that

$$
\begin{aligned}
M_{X_1-\bar{X},X_2-\bar{X},\cdots,X_n-\bar{X}}(t_1,t_2,\cdots,t_n) &= M_{X_1-\bar{X},X_2-\bar{X},\cdots,X_n-\bar{X},\bar{X}}(t_1,t_2,\cdots,t_n,0) \\
&= \exp\left(\frac{1}{2}\sigma^2\sum_{j=1}^{n}\left(t_j-\frac{\sum_{i=1}^{n}t_i}{n}\right)^2\right)
\end{aligned}
$$

and

$$
M_{\bar{X}}(t_{n+1}) = M_{X_1-\bar{X},X_2-\bar{X},\cdots,X_n-\bar{X},\bar{X}}(0,0,\cdots,0,t_{n+1}) = \exp\left(\mu t_{n+1}+\frac{1}{2}\sigma^2\frac{t_{n+1}^2}{n}\right).
$$

Therefore, $(X_1-\bar{X},X_2-\bar{X},\cdots,X_n-\bar{X})$ and $\bar{X}$ are independent. Consequently, the sample variance $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ and $\bar{X}$ are independent.

Remark 11.13 (Distribution of the sample variance for a random sample from the Normal distribution). Let $X_1,X_2,\cdots,X_n$ be a random sample from $N(\mu,\sigma^2)$ distribution. By looking at the joint MGF of $X_1-\bar{X},\cdots,X_n-\bar{X}$ and $\bar{X}$, Remark 11.12 gives $\sum_{i=1}^{n}(X_i-\bar{X})^2$ and $\bar{X}$ are independent. Now,

$$
\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X}+\bar{X}-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2+\frac{n(\bar{X}-\mu)^2}{\sigma^2}
$$

where $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\sim\chi_n^2$ and $\frac{n(\bar{X}-\mu)^2}{\sigma^2}\sim\chi_1^2$. Since, $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ and $\frac{n(\bar{X}-\mu)^2}{\sigma^2}$ are independent, we conclude $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2\sim\chi_{n-1}^2$. Taking the sample variance as $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$, we conclude that $\frac{(n-1)S_n^2}{\sigma^2}\sim\chi_{n-1}^2$.

Note 11.14. Given a random sample $X_1,X_2,\cdots,X_n$ from $N(\mu,\sigma^2)$ distribution, the sample mean $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\sim N\left(\mu,\frac{\sigma^2}{n}\right)$ and sample variance $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ has the property that $\frac{(n-1)S_n^2}{\sigma^2}\sim\chi_{n-1}^2$. The distribution of $\frac{\bar{X}-\mu}{S_n}$ is of interest.

Definition 11.15 (Student's t-distribution with n degrees of freedom)**.** Let n be a positive integer. Let $X\sim N(0,1)$ and $Y\sim\chi_n^2$ be independent RVs. Then,

$$
T=\frac{X}{\sqrt{\frac{Y}{n}}}
$$

is said to follow the t-distribution with n degrees of freedom. In this case, we write $T \sim t_n$. The p.d.f. is given by

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)\sqrt{n}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \forall t \in \mathbb{R}.$$

Here, $\mathbb{E}T^k$ exists if $k < n$. Since, the distribution is symmetric about 0 and hence $\mathbb{E}T^k = 0$ for all k odd with $k < n$. If k is even and $k < n$, then

$$\mathbb{E}T^k = n^{\frac{k}{2}}\frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{n-k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}.$$

In particular, if $n > 2$, then $\mathbb{E}T = 0$ and $Var(T) = \frac{n}{n-2}$. The t-distribution appears in the tests for statistical significance.

Note 11.16. If $X_1, X_2, \cdots, X_n$ is a random sample from $N(\mu, \sigma^2)$ distribution, then $\sqrt{n}\frac{\bar{X}-\mu}{S_n} \sim t_{n-1}$.

Note 11.17. Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distribution, respectively. Consider the sample variances $S_1^2 := \frac{1}{m-1}\sum_{i=1}^m (X_i - \bar{X})^2$ and $S_2^2 := \frac{1}{n-1}\sum_{j=1}^n (Y_j - \bar{Y})^2$. The distribution of $\frac{S_1^2}{S_2^2}$ is of interest. Note that $\frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2_{m-1}$ and $\frac{(n-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n-1}$.

Definition 11.18 (F-distribution with degrees of freedom m and n)**.** Let m and n be positive integers. Let $X \sim \chi^2_m$ and $Y \sim \chi^2_n$ be independent RVs. Then,

$$F = \frac{\frac{X}{m}}{\frac{Y}{n}}$$

is said to follow the F-distribution with degrees of freedom m and n. In this case, we write $F \sim F_{m,n}$. The p.d.f. is given by

$$f_F(x) = \begin{cases} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\frac{m}{n}\left(\frac{m}{n}x\right)^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{B\left(\frac{m}{2},\frac{n}{2}\right)} \frac{m}{n} \left(\frac{m}{n}x\right)^{\frac{m}{2}-1} \left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}, \text{ if } x>0,\\ 0, \text{ otherwise.}\end{cases}$$

Note 11.19. If $F \sim F_{m,n}$, then $\frac{1}{F} \sim F_{n,m}$.

Note 11.20. Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distribution, respectively. Consider the sample variances $S_1^2 = \frac{1}{m-1}\sum_{i=1}^m (X_i - \bar{X})^2$ and $S_2^2 = \frac{1}{n-1}\sum_{j=1}^n (Y_j - \bar{Y})^2$. The distribution of $\frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F_{m-1,n-1}$.

We now look at the Normal distribution and construct some generalization in higher dimensions. This will give us an example of a continuous random vector, which under additional hypothesis becomes absolutely continuous.

We give multiple ways of defining this generalization, all of which, at the end, turn out to be equivalent.

Definition 11.21 (A constructive approach to Multivariate Normal distribution)**.** The constructive approach is split into two steps. First, consider $X_1, X_2, \cdots, X_d$ to be independent $N(0,1)$ RVs defined on the same probability space. Then, the common p.d.f. is given by

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \forall x \in \mathbb{R}, i = 1, 2, \cdots, d.$$

Now, consider the random vector $X = (X_1, X_2, \cdots, X_d)^t$. Since the RVs X_i's are independent, the random vector X is absolutely continuous (see Theorem 9.9) with the joint p.d.f. given by

$$f_{X_1,\cdots,X_d}(x_1, \cdots, x_d) = (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^d x_i^2\right), \forall x_1, x_2, \cdots, x_d \in \mathbb{R}.$$

We define X to be a standard Normal d-dimensional random vector or a standard multivariate d-dimensional Normal random vector.

Next, we say that Y is a multivariate d-dimensional Normal random vector, if there exist $b = (b_1, \cdots, b_d)^t \in \mathbb{R}^d$ and a $d \times d$ real matrix $A = (a_{ij})_{d\times d}$ such that $Y = AX + b$, with X as in step one.

Remark 11.22. We now discuss basic properties of the RV Y constructed in Definition 11.21.

(i) First, we identify the marginal distributions. For any $i = 1, 2, \cdots, d$, we have $Y_i = \sum_{j=1}^{d} a_{ij} X_j + b_i$ and hence, we have $Y_j \sim N(b_i, \sum_{j=1}^{d} a_{ij}^2)$.

(ii) Fix $i, j = 1, 2, \cdots, d$ with $i \neq j$. Then

$$Cov(Y_i, Y_j) = Cov(\sum_{k=1}^{d} a_{ik} X_k + b_i, \sum_{k=1}^{d} a_{jk} X_k + b_j) = \sum_{k=1}^{d} a_{ik} a_{jk}.$$

In the above simplification, we have used the fact that $Cov(X_i, X_j) = 0$, due to the independence of X_i and X_j.

(iii) Consider a $d \times d$ real matrix $K = (k_{ij})_{d \times d}$ with $k_{ij} = Cov(Y_i, Y_j), \forall i, j$. Then, by the computations above, we conclude

$$k_{ii} = Cov(Y_i, Y_i) = Var(Y_i) = \sum_{l=1}^{d} a_{il}^2, i = 1, 2, \cdots, d$$

and for $i \neq j$,

$$k_{ij} = Cov(Y_i, Y_j) = \sum_{l=1}^{d} a_{il} a_{jl}.$$

The matrix K shall be referred to as the variance-covariance matrix of the random vector Y or simply, the covariance matrix. It is also called the dispersion matrix.

(iv) Continue with the matrix K defined above. For $i, j = 1, 2, \cdots, d$, observe that

$$k_{ij} = \sum_{l=1}^{d} a_{il} a_{jl} = (AA^t)_{ij},$$

where A^t denotes the transpose of the matrix A. Then, $K = AA^t$. Consequently, K is a real symmetric matrix. Hence, there exists an orthogonal matrix U and a diagonal matrix D such that $U^t K U = D$ or $K = UDU^t$. In particular, K is also non-negative definite.

(v) We now compute the joint MGF of Y, if it exists. For $u_1, u_2, \cdots, u_d \in \mathbb{R}$, write $u = (u_1, \cdots, u_d)^t$ and observe that

$$\mathbb{E} \exp(u_1 Y_1 + u_2 Y_2 + \cdots + u_d Y_d) = \exp\left(\sum_{j=1}^{d} u_j b_j\right) \mathbb{E} \exp\left(\sum_{i=1}^{d} u_i \sum_{j=1}^{d} a_{ij} X_j\right)$$

$$= \exp(u^t b)\ \mathbb{E}\exp\left(\sum_{j=1}^{d}(\sum_{i=1}^{d} a_{ij}u_i)X_j\right).$$

Here, $u^t b = \sum_{j=1}^{d} u_j b_j$. Since X_i's are independent, and the marginal MGFs $M_{X_i} = \exp\left(\frac{1}{2}x^2\right), \forall x \in \mathbb{R}, i = 1, 2, \cdots, d$, we have

$$\begin{aligned}
\mathbb{E}\exp(u_1Y_1 + u_2Y_2 + \cdots + u_dY_d) &= \exp(u^t b)\ \ \exp\left(\frac{1}{2}\sum_{j=1}^{d}(\sum_{i=1}^{d} a_{ij}u_i)^2\right) \\
&= \exp(u^t b)\ \ \exp\left(\frac{1}{2}\sum_{j=1}^{d}\sum_{i,l=1}^{d} a_{ij}u_i a_{lj}u_l\right) \\
&= \exp(u^t b)\ \ \exp\left(\frac{1}{2}\sum_{i,l=1}^{d} k_{il}u_i u_l\right) \\
&= \exp(u^t b + \frac{1}{2}u^t K u).
\end{aligned}$$

Hence, we have computed the joint MGF of Y. Note that this MGF exists for all points $u = (u_1, \cdots, u_d)^t \in \mathbb{R}^d$.

(vi) We now compute the Characteristic function of Y. We have

$$\Phi_Y(u) = \exp\left(iu^t b - \frac{1}{2}u^t K u\right), \forall u \in \mathbb{R}^d$$

Notation 11.23. We usually refer to Y, as defined in Definition 11.21, as a multivariate Normal random vector with mean vector b and variance-covariance matrix K and write $Y \sim N_d(b, K)$.

Example 11.24. Consider the two-dimensional case. Then with $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $b = \binom{b_1}{b_2}$, we have

$$Y_1 = b_1 + a_{11}X_1 + a_{12}X_2, \quad Y_2 = b_2 + a_{21}X_1 + a_{22}X_2.$$

Moreover, $\mathbb{E}Y_1 = b_1, Var(Y_1) = a_{11}^2 + a_{12}^2, \mathbb{E}Y_2 = b_2, Var(Y_2) = a_{21}^2 + a_{22}^2$ and $Cov(Y_1, Y_2) = a_{11}a_{21} + a_{12}a_{22}$. Here,

$$K = \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{11}a_{21} + a_{12}a_{22} & a_{11}^2 + a_{12}^2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} = AA^t.$$

Note 11.25. By Remark 11.22(iv), $K = AA^t$. Then, $det(K) = det(AA^t) = (det(A))^2$. Therefore, $det(K) \neq 0$ if and only if $det(A) \neq 0$. In other words, K is invertible if and only if A is invertible.

Example 11.26. Continue with Example 11.24. Assume that $\sigma_1^2 := Var(Y_1) > 0, \sigma_2^2 := Var(Y_1) > 0$ and set $\rho := \frac{Cov(Y_1,Y_2)}{\sigma_1\sigma_2}$. If $|\rho| \neq 1$, then we have K is invertible. Furthermore, in this case, we can show that Y is absolutely continuous by computing the joint p.d.f..

So far, we gave a constructive definition of a multivariate Normal random vector. We have also computed the corresponding MGF. Recall that the distribution of any random vector is determined by its MGF, provided the MGF exists. This leads us to an alternative definition of a multivariate Normal random vector.

Definition 11.27 (Multivariate Normal distribution through its MGF)**.** Let $b = (b_1, \cdots, b_d)^t \in \mathbb{R}^d$ and let K be a $d \times d$ real symmetric non-negative definite matrix. A d-dimensional random vector Y is said to be multivariate Normal if its MGF is given by

$$M_Y(u) = \exp(u^t b + \frac{1}{2} u^t K u), \forall u \in \mathbb{R}^d.$$

Note 11.28. If K is as in the definition above, then there exists a $d \times d$ real matrix A such that $K = AA^t$. Then taking $X = (X_1, X_2, \cdots, X_d)^t$ with X_i's as independent $N(0,1)$ RVs and setting $Y = AX + b$ gives us the MGF above. This is the connection between the constructive definition 11.21 and the definition 11.27 through its MGF.

Note 11.29. In terms of notation 11.23, the random vector Y in definition 11.27 above is a multivariate Normal random vector with mean vector b and variance-covariance matrix K, i.e. $Y \sim N_d(b, K)$.

Example 11.30. Let $Y \sim N_d(b, K)$. Then for any $c \in \mathbb{R}^n$ and any $n \times d$ real matrix B, consider the n-dimensional random vector $Z = c + BY$. A straight-forward verification yields $Z \sim N_n(c + Bb, BKB^t)$.

Remark 11.31. Considering the case $n = 1$ in the above example 11.30, any linear combination $\sum_{j=1}^d a_j Y_j$, with $a_1, a_2, \cdots, a_d \in \mathbb{R}$ and $Y = (Y_1, \cdots, Y_d)^t \sim N_d(b, K)$, is univariate normal.

With the above observation at hand, we are ready to prove the following important result on multivariate Normal distributions. We shall see that this result yields yet another definition of multivariate Normal distribution.

Theorem 11.32. *A random vector $Y = (Y_1, \cdots, Y_d)^t$ is multivariate Normal if and only if any linear combination $\sum_{j=1}^d a_j Y_j$, with $a_1, a_2, \cdots, a_d \in \mathbb{R}$, is univariate Normal.*

Proof. The 'only if' part follows from the Remark 11.31 above.

Now, consider the 'if' part of the statement. Assume that $\sum_{j=1}^d a_j Y_j$ is univariate Normal for all choices of $a_1, a_2, \cdots, a_d \in \mathbb{R}$. In particular, Y_j's are univariate Normal RVs.

Suppose that $Y_j \sim N(\mu_j, \sigma_j^2)$. Then, with $a = (a_1, a_2, \cdots, a_d)^t$ and $\mu = (\mu_1, \mu_2, \cdots, \mu_d)^t$,

$$\mathbb{E}\left(\sum_{j=1}^d a_j Y_j\right) = \sum_{j=1}^d a_j \mu_j = a^t \mu$$

and

$$Var\left(\sum_{j=1}^d a_j Y_j\right) = a^t K a,$$

where $K = (k_{ij})_{d\times d}$ is defined by $k_{ij} = Cov(Y_i, Y_j), \forall 1 \le i, j \le d$.

Since $a^t Y = \sum_{j=1}^d a_j Y_j$ is univariate Normal, we have

$$\mathbb{E}\exp\left(a^t Y u\right) = \exp\left(a^t \mu u + \frac{1}{2} a^t K a u^2\right), \forall u \in \mathbb{R}.$$

In particular, for $u = 1$, we have $\mathbb{E}\exp(a^t Y) = \exp\left(a^t \mu + \frac{1}{2} a^t K a\right)$. Since, the MGF, if it exists, uniquely determines the distribution, we have $Y \sim N_d(\mu, K)$. This completes the proof. $\square$

Motivated by Theorem 11.32, we make the following alternative definition for multivariate Normal distribution.

Definition 11.33. We say that a d-dimensional random vector Y follows multivariate Normal distribution if $a^t Y$ is univariate Normal for all $a \in \mathbb{R}^d$.

Remark 11.34. Recall the notion of Characteristic function of a random vector. Like the MGFs, the Characteristic functions also determine the distributions uniquely. One may also define multivariate Normal distributions by specifying the Characteristic function.

We have discussed multiple equivalent ways of defining the multivariate Normal distribution. Now, we shall focus on the question of existence of a joint p.d.f. of a multivariate Normal random vector.

Theorem 11.35. *Let $Y \sim N_d(b, K)$. Suppose K is invertible. Then the following statements hold.*

(i) $\sum_{j=1}^d a_j(Y_j - b_j) = 0$ with probability 1 if and only if $a_j = 0, \forall j = 1, 2, \cdots, d$, i.e., the components $Y_j - b_j, j = 1, 2, \cdots, d$ are linearly independent.

(ii) Y has a joint p.d.f. given by

$$f_Y(y) = (2\pi)^{-\frac{d}{2}}(det(K))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y-b)^t K^{-1}(y-b)\right), \forall y \in \mathbb{R}^d.$$

Proof. For the first statement, if $a_j = 0, \forall j$, then we have $\sum_{j=1}^d a_j(Y_j - b_j) = 0$ with probability 1.

To prove the converse. Since K is a $d \times d$ real symmetric positive-definite matrix, it is diagonalizable. There exists a $d \times d$ orthogonal matrix A such that $D = A^t K A$ is a diagonal matrix with eigen-values of K as the diagonal entries. Note that the eigen-values are strictly positive, since K is positive-definite.

Consider $Z = A^t(Y - b)$. By Example 11.30, $Z \sim N_d(0, D)$. If with probability 1, $a^t(Y - b) = 0$, then $Var(a^t(Y - b)) = \mathbb{E}\left(\sum_{j=1}^d a_j^2 (Y_j - b_j)^2\right) = 0$. Again,

$$\mathbb{E}\left(\sum_{j=1}^d a_j^2 (Y_j - b_j)^2\right) = \mathbb{E}\left(a^t(Y-b)(Y-b)^t a\right) = a^t K a > 0,$$

unless $a_j = 0, \forall j$, since K is positive-definite. Hence, we must have $a_j = 0, \forall j$. This proves the first statement.

To obtain the joint p.d.f. as stated in the second statement. Continue with the notations of the proof of the first part. If $\lambda_1, \cdots, \lambda_d$ are the eigen-values of K, using the constructive definition of multivariate Normal distribution, $Z \sim N_d(0, D)$ is given by the joint p.d.f.

$$f_Z(z) = (2\pi)^{-\frac{d}{2}}(det(D))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{d}\frac{z_j^2}{\lambda_j}\right), \forall z = (z_1, \cdots, z_d)^t \in \mathbb{R}^d.$$

But, $Y = AZ + b$ with $det(A) = \pm 1$, since A is orthogonal. Using change of variables, we have the p.d.f. of Y as stated above. This completes the proof. $\square$

Corollary 11.36. *If $Y = (Y_1, \cdots, Y_d)^t \sim N_d(b, K)$ with Y_j's uncorrelated, then $Y_1, \cdots, Y_d$ are independent.*

Proof. If Y_j's are uncorrelated, then $K = (k_{ij})_{d\times d}$ is a diagonal matrix. If a diagonal entry $k_{ii} = 0$, then $Var(Y_i) = k_{ii} = 0$. This implies Y_i is degenerate.

If the non-degenerate Y_j's are independent, then along with the degenerate Y_i's we obtain the required independence. Thus without loss of generality, we assume that the diagonal entries are non-zero.

Since the diagonal entries are the eigen-values of K, we get $det(K) \neq 0$, i.e., K is invertible. By Theorem 11.35, the joint p.d.f. of Y is

$$f_Y(y) = (2\pi)^{-\frac{d}{2}}(det(K))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{d}\frac{(y_j - b_j)^2}{\sqrt{k_{ii}}}\right), \forall y = (y_1, \cdots, y_d)^t \in \mathbb{R}^d.$$

The joint p.d.f. can be split with terms separately depending on different y_j's and hence Y_j's are independent with $Y_j \sim N(b_j, k_{jj})$. This completes the proof. $\square$

Remark 11.37. If K is a singular matrix, then using arguments similar to Theorem 11.35, we can show that $(Y_j - b_j), j = 1, \cdots, d$ are linearly dependent. In fact, any linearly independent subcollection of $(Y_j - b_j), j = 1, \cdots, d$ is a multivariate Normal random vector of the appropriate dimension with a joint p.d.f..

We now discuss a special case of 2-dimensional Normal random vector with a p.d.f.. We now revisit most of the results which we already studied in the general d-dimensional framework.

Definition 11.38. A bivariate random vector $X = (X_1, X_2)$ is said to follow bivariate Normal distribution $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ for $\mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}, \sigma_1 > 0, \sigma_2 > 0, \rho \in (-1, 1)$, if the joint p.d.f. is given by

$$
\begin{aligned}
&f_{X_1,X_2}(x_1,x_2)\\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right],
\end{aligned}
$$

for all $(x_1,x_2) \in \mathbb{R}^2$. To check that f_{X_1,X_2}, as above, is a p.d.f., first note that $f_{X_1,X_2}(x_1,x_2) \geq 0, \forall (x_1,x_2) \in \mathbb{R}^2$. Now, changing variables to $y_1 = \frac{x_1-\mu_1}{\sigma_1}, y_2 = \frac{x_2-\mu_2}{\sigma_2}$, we have

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\,dx_1dx_2\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{y_2^2}{2}\right)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}(y_1-\rho y_2)^2\right]\,dy_1dy_2\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{y_2^2}{2}\right)\,dy_2\\
&= 1.
\end{aligned}
$$

This completes the verification that f_{X_1,X_2} is a joint p.d.f..

Remark 11.39. Let $X = (X_1,X_2) \sim N_2(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ for some $\mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}, \sigma_1 > 0, \sigma_2 > 0, \rho \in (-1,1)$.

(a) The marginal p.d.f. of X_2 is given by

$$
\begin{aligned}
&f_{X_2}(x_2)\\
&= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\,dx_1, \forall x_2 \in \mathbb{R}\\
&= \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\\
&\times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sigma_1^2(1-\rho^2)}\left\{x_1 - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right)\right\}^2\right]dx_1\\
&= \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]
\end{aligned}
$$

and hence $X_2 \sim N(\mu_2, \sigma_2^2)$. Similarly, $X_1 \sim N(\mu_1, \sigma_1^2)$. Thus the parameters $\mu_1 = \mathbb{E}X_1, \sigma_1^2 = Var(X_1), \mu_2 = \mathbb{E}X_2, \sigma_2^2 = Var(X_2)$ have their own interpretation.

(b) The covariance $Cov(X_1, X_2)$ is given by

$$\begin{aligned}Cov(X_1, X_2) &= \mathbb{E}(X_1 - \mu_1)(X_2 - \mu_2)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1 - \mu_1)(x_2 - \mu_2) f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2\end{aligned}$$

By changing variables to $y_1 = \frac{x_1-\mu_1}{\sigma_1}, y_2 = \frac{x_2-\mu_2}{\sigma_2}$, and simplifying the above expression, we have $Cov(X_1, X_2) = \rho\sigma_1\sigma_2$. Consequently, the correlation $\rho(X_1, X_2) = \rho$. We now have the interpretation of the parameter ρ.

(c) The conditional distribution of X_1 given $X_2 = x_2 \in \mathbb{R}$ is described by the conditional p.d.f.

$$\begin{aligned}f_{X_1|X_2}(x_1 \mid x_2) &= \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}\\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sigma_1^2(1-\rho^2)}\left\{x_1 - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)\right)\right\}^2\right], \forall x_1 \in \mathbb{R}\end{aligned}$$

and hence $X_1 \mid X_2 = x_2 \sim N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2))$. Similarly, for $x_1 \in \mathbb{R}$, $X_2 \mid X_1 = x_1 \sim N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2))$.

(d) Using the conditional distributions obtained above, we conclude

$$\begin{aligned}\mathbb{E}[X_1 \mid X_2 = x_2] &= \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),\\ Var[X_1 \mid X_2 = x_2] &= \sigma_1^2(1-\rho^2),\\ \mathbb{E}[X_2 \mid X_1 = x_1] &= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1),\\ Var[X_2 \mid X_1 = x_1] &= \sigma_2^2(1-\rho^2).\end{aligned}$$

(e) If X_1 and X_2 are independent, then they are uncorrelated and in particular $\rho = \rho(X_1, X_2) = 0$. Conversely, if $\rho = 0$, then

$$\begin{aligned}f_{X_1,X_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{1}{2}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right]\\ &= f_{X_1}(x_1) f_{X_2}(x_2), \forall (x_1, x_2) \in \mathbb{R}^2\end{aligned}$$

and hence X_1 and X_2 are independent.

(f) Consider the Variance-Covariance matrix, Dispersion matrix or simply, the Covariance matrix

$$\Sigma = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & Var(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Observe that this is a symmetric matrix with $det(\Sigma) = \sigma_1^2\sigma_2^2(1-\rho^2) > 0$ and hence the matrix is invertible. In fact, this matrix is positive-definite and its eigen-values are positive. The joint p.d.f. can be rewritten as

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left[-\frac{1}{2}(x_1-\mu_1, x_2-\mu_2)\Sigma^{-1}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}\right], \forall (x_1, x_2) \in \mathbb{R}^2.$$

(g) We now compute the joint MGF of X. We have

$$\begin{aligned}
&M_X(t_1, t_2) \\
&= \mathbb{E}\exp(t_1X_1 + t_2X_2) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(t_1x_1 + t_2x_2) f_{X_1,X_2}(x_1, x_2)\, dx_1dx_2 \\
&= \int_{-\infty}^{\infty} \exp(t_2x_2) f_{X_2}(x_2) \int_{-\infty}^{\infty} \exp(t_1x_1) f_{X_1|X_2}(x_1 \mid x_2)\, dx_1dx_2 \\
&= \int_{-\infty}^{\infty} \exp(t_2x_2) f_{X_2}(x_2) \exp(\mu_1t_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)t_1 + \frac{1}{2}\sigma_1^2(1-\rho^2)t_1^2)\, dx_2 \\
&= \exp(\mu_1t_1 + \rho\frac{\sigma_1}{\sigma_2}(-\mu_2)t_1 + \frac{1}{2}\sigma_1^2(1-\rho^2)t_1^2) \int_{-\infty}^{\infty} \exp(t_2x_2 + \rho\frac{\sigma_1}{\sigma_2}t_1x_2) f_{X_2}(x_2)\, dx_2 \\
&= \exp(\mu_1t_1 + \rho\frac{\sigma_1}{\sigma_2}(-\mu_2)t_1 + \frac{1}{2}\sigma_1^2(1-\rho^2)t_1^2) \exp\left(\mu_2\left\{t_2 + \rho\frac{\sigma_1}{\sigma_2}t_1\right\} + \frac{1}{2}\sigma_2^2\left\{t_2 + \rho\frac{\sigma_1}{\sigma_2}t_1\right\}^2\right) \\
&= \exp(\mu_1t_1 + \mu_2t_2 + \frac{1}{2}\sigma_1^2t_1^2 + \frac{1}{2}\sigma_2^2t_2^2 + \rho\sigma_1\sigma_2t_1t_2), \forall (t_1, t_2) \in \mathbb{R}^2.
\end{aligned}$$

(h) Let $c_1, c_2 \in \mathbb{R}$ such that at least one of c_1, c_2 is not zero and take $Y = c_1X_1 + c_2X_2$. Now,

$$M_Y(t) = \mathbb{E}\exp(c_1tX_1 + c_2tX_2) = \exp\left[(\mu_1c_1 + \mu_2c_2)t + \left(\frac{1}{2}\sigma_1^2c_1^2 + \frac{1}{2}\sigma_2^2c_2^2 + \rho\sigma_1\sigma_2c_1c_2\right)t^2\right], \forall t \in \mathbb{R}.$$

Looking at the structure of the MGF, we conclude that $Y = c_1X_1 + c_2X_2 \sim N(c_1\mu_1 + c_2\mu_2, c_1^2\sigma_1^2 + c_2\sigma_2^2 + 2\rho c_1c_2\sigma_1\sigma_2)$.

Remark 11.40. The above statement for the linear combination of X_1, X_2 actually characterizes the bivariate Normal distribution. If $X = (X_1, X_2)$ is such that $\mathbb{E}X_1 = \mu_1, \mathbb{E}X_2 = \mu_2, Var(X_1) = \sigma_1^2 > 0, Var(X_2) = \sigma_2^2 > 0, \rho(X_1, X_2) = \rho \in (-1, 1)$ and $c_1X_1 + c_2X_2 \sim N(c_1\mu_1 + c_2\mu_2, c_1^2\sigma_1^2 + c_2\sigma_2^2 + 2\rho c_1c_2\sigma_1\sigma_2)$ for all $(c_1, c_2) \neq (0, 0)$, then

$$\begin{aligned} M_X(t_1, t_2) &= \mathbb{E}\exp(t_1X_1 + t_2X_2) \\ &= M_{t_1X_1+t_2X_2}(1) \\ &= \exp(\mu_1t_1 + \mu_2t_2 + \frac{1}{2}\sigma_1^2t_1^2 + \frac{1}{2}\sigma_2^2t_2^2 + \rho\sigma_1\sigma_2t_1t_2), \forall(t_1, t_2) \in \mathbb{R}^2. \end{aligned}$$

Since an MGF determines the distribution, we conclude $X = (X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

Remark 11.41 (Interpretation of parameters appearing in the p.d.f. of a Continuous RV). In the examples of continuous RVs discussed in this course, we have seen that certain parameters appear in the description of p.d.fs. If we specify the values of these parameters, then we obtain a specific example of distribution from a family of possible distributions. In certain cases, we have already been able to interpret them in terms of properties of the distribution of the RV. For example, if $X \sim N(\mu, \sigma^2)$, then $\mu = \mathbb{E}X$ and $\sigma^2 = Var(X)$. We list some interpretation of these parameters.

(a) (Location parameter) If we have a family of p.d.f.s $f_\theta, \theta \in \Theta$, where θ is a real valued parameter (i.e. $\Theta \subseteq \mathbb{R}$) and if $f_\theta(x) = f_0(x - \theta), \forall x \in \mathbb{R}$, then we say that θ is a location parameter for the family of distributions given by the p.d.f.s f_θ. In this case, the family is called a location family and the p.d.f. f_0 is free of θ, i.e. does not depend on θ. We can restate this fact in terms of the corresponding RVs X_θ as follows: the p.d.f./distribution of $X_\theta - \theta$ does not depend on θ.

(b) (Scale parameter) If we have a family of p.d.f.s f_θ, where θ is a real valued parameter (i.e. $\Theta \subseteq \mathbb{R}$) and if $f_\theta(x) = \frac{1}{\theta}f_1(\frac{x}{\theta}), \forall x \in \mathbb{R}$, then we say that θ is a scale parameter for the family of distributions given by the p.d.f.s f_θ. In this case, the family is called a scale family and the p.d.f. f_1 is free of θ, i.e. does not depend on θ. We can restate this fact in terms of the corresponding RVs X_θ as follows: the p.d.f./distribution of $\frac{1}{\theta}X_\theta$ does not depend on θ.

(c) (Location-scale parameter) If we have a family of p.d.f.s $f_{\mu,\sigma}$ with $\sigma > 0$ and if $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right) = f_{0,1}(x), \forall x \in \mathbb{R}$, then we say that (μ, σ) is a location-scale parameter for the family of distributions given by the p.d.f.s $f_{\mu,\sigma}$. In this case, the family is called a location-scale family and the p.d.f. $f_{0,1}$ is free of (μ, σ), i.e. does not depend on (μ, σ). We can restate this fact in terms of the corresponding RVs $X_{\mu,\sigma}$ as follows: the p.d.f./distribution of $\frac{X_{\mu,\sigma}-\mu}{\sigma}$ does not depend on (μ, σ).

(d) (Shape parameter) Some family of p.d.f.s also has a shape parameter, where changing the value of the parameter affects the shape of the graph of the p.d.f..

Example 11.42. (a) The family of RVs $X_{\mu,\theta} \sim \mathit{Cauchy}(\mu, \theta), \mu \in \mathbb{R}, \theta > 0$ with the p.d.f.

$$f_{\mu,\theta}(x) = \frac{\theta}{\pi} \frac{1}{\theta^2 + (x-\mu)^2}, \forall x \in \mathbb{R}$$

is a location-scale family with location parameter μ and scale parameter θ.

(b) For the family of RVs $X_\alpha \sim \mathit{Gamma}(\alpha, 1), \alpha > 0$ with the p.d.f.

$$f_\alpha(x) = \begin{cases} \frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}, \text{ if } x > 0, \\ 0, \text{ otherwise} \end{cases}$$

α is a shape parameter.

Definition 11.43 (Weibull distribution)**.** We say that an RV X follows the Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, if its p.d.f. is given by

$$f_X(x) = \begin{cases} \frac{\alpha}{\beta^\alpha}x^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right], \text{ if } x > 0, \\ 0, \text{ otherwise.} \end{cases}$$

Note 11.44. Let $X \sim \mathit{Exponential}(\beta^\alpha)$ for some $\alpha, \beta > 0$. Then $Y = X^{\frac{1}{\alpha}}$ follows the Weibull distribution with shape parameter α and scale parameter β.

Definition 11.45 (Pareto distribution)**.** We say that an RV X follows the Pareto distribution with scale parameter $\theta > 0$ and shape parameter $\alpha > 0$, if its p.d.f. is given by

$$f_X(x) = \begin{cases} \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$