CHAPTER 4

4. CONTINUOUS PROBABILITY DISTRIBUTIONS

4.1 CONTINUOUS UNIFORM DISTRIBUTION

NOTATION $X \sim Uniform(a, b)$ **RANGE** $\mathbb{X} = [a, b]$ or (a, b), for $a \leq b$

PDF

$$f_X(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

CDF

$$F_X(x) = \frac{x-a}{b-a} \qquad a \leq x \leq b$$

MGF

$$M_X(t) = \int_a^b \mathrm{e}^{tx} \frac{1}{b-a}\, dx = \frac{(\mathrm{e}^{tb} - \mathrm{e}^{ta})}{t(b-a)} \qquad \text{if } t \neq 0$$

and $M_X(0) = 1$.

rth MOMENT

$$\mathrm{E}_{f_X}[X^r] = \int_a^b x^r \frac{1}{b-a}\, dx = \frac{1}{b-a}\left[\frac{b^{r+1}}{r+1} - \frac{a^{r+1}}{r+1}\right]$$

so therefore

$$\mathrm{E}_{f_X}[X] = \frac{1}{b-a}\left[\frac{b^2-a^2}{2}\right] = \frac{(a+b)}{2}$$

$$\mathrm{E}_{f_X}[X^2] = \frac{1}{b-a}\left[\frac{b^3-a^3}{3}\right] = \frac{(a^2+ab+b^2)}{3}$$

$$\Longrightarrow \mathrm{Var}_{f_X}[X] = \frac{(a^2+ab+b^2)}{3} - \left\{\frac{(a+b)}{2}\right\}^2$$

$$= \frac{(a-b)^2}{12}$$

4.2 EXPONENTIAL DISTRIBUTION

NOTATION $X \sim Exp(\lambda)$ **RANGE** $\mathbb{X} = \mathbb{R}^+$

PDF

$$f_X(x) = \lambda e^{-\lambda x} \qquad x > 0 \qquad \text{where } \lambda > 0.$$

CDF

$$F_X(x) = 1 - e^{-\lambda x} \qquad x > 0$$

MGF

$$M_X(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx = \frac{\lambda}{\lambda - t} \qquad \text{for } t < \lambda$$

rth MOMENT

$$M_X^{(r)}(t) = \frac{r!\lambda}{(\lambda - t)^{r+1}} \Longrightarrow M_X^{(r)}(0) = \frac{r!}{\lambda^r},\ \mathrm{E}_{f_X}[X] = \frac{1}{\lambda},\ \mathrm{E}_{f_X}[X^2] = \frac{2}{\lambda^2} \Longrightarrow \mathrm{Var}_{f_X}[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Notes

(1) Alternative representation uses $\theta = 1/\lambda$ as the parameter of the distribution.

(2) If $X \sim Exp(\lambda)$, then, for all $x, t > 0$,

$$\mathrm{P}[\ X > x + t \mid X > t\ ] = \frac{\mathrm{P}[\ X > x + t, X > t\ ]}{\mathrm{P}[\ X > t\ ]} = \frac{\mathrm{P}[\ X > x + t\ ]}{\mathrm{P}[\ X > t\ ]} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x} = \mathrm{P}[\ X > x\ ]$$

Thus, for all $x, t > 0$, $\mathrm{P}[\ X > x + t \mid X > t\ ] = \mathrm{P}[\ X > x\ ]$ - this is known as the **Lack of Memory Property**, and is unique to the exponential distribution amongst continuous distributions.

(3) Suppose that $X(t)$ is a Poisson process with rate parameter $\lambda > 0$, so that

$$\mathrm{P}[X(t) = n] = \frac{e^{-\lambda t}(\lambda t)^n}{n!}.$$

Let $X_1, ..., X_n$ be random variables defined by $X_1 =$ "time that first event occurs", and, for $i = 2, ..., n$, $X_i =$ "time interval between occurrence of $(i-1)$st and ith events". Then $X_1, ..., X_n$ are i.i.d. $Exp(\lambda)$.

Proof : $X_1, ..., X_n$ are i.i.d. because of the assumption 2. underlying the Poisson process. So consider the distribution of X_1; in particular, consider the probability $\mathrm{P}[X_1 > x]$ for $x > 0$. The event $[X_1 > x]$ is equivalent to the event "No events occur in the interval $(0,x]$", which has probability $e^{-\lambda x}$. But

$$F_{X_1}(x) = \mathrm{P}[\ X_1 \leq x\ ] = 1 - \mathrm{P}[\ X_1 > x\ ] = 1 - e^{-\lambda x} \Longrightarrow X_1 \sim Exp(\lambda).$$

(4) The exponential distribution is used to model failure times in continuous time. It is a continuous **waiting time** distribution, the continuous analogue of the geometric distribution.

(5) If $X \sim Uniform(0, 1)$, and $Y = -\log(1 - X)/\lambda$, then $Y \sim Exp(\lambda)$.

(6) If $X \sim Exp(\lambda)$, then $Y = X^{1/\alpha}$ for $\alpha > 0$ has a (two-parameter) **Weibull** distribution, and

$$f_Y(y) = \alpha\lambda y^{\alpha - 1} e^{-\lambda y^\alpha} \quad y > 0$$

(7) The density function of X can be re-written

$$f_X(x) = \exp\{t(x)q(\lambda) + \log h(x) + \log c(\lambda)\}$$

where $t(x) = -x$, $q(\lambda) = \lambda$, $h(x) = 1$, $c(\lambda) = \lambda$.

4.3 GAMMA DISTRIBUTION

NOTATION $X \sim Ga(\alpha, \beta)$ **RANGE** $\mathbb{X} = \mathbb{R}^+$

PDF

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x > 0 \qquad \text{where } \alpha, \beta > 0.$$

and where, for any real number $\alpha > 0$,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$$

is the **gamma function**, $\Gamma(.)$.

MGF

$$M_X(t) = \int_0^\infty e^{tx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1}\, e^{-(\beta-t)x}\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta-t)^\alpha} = \left(\frac{\beta}{\beta-t}\right)^\alpha$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = \frac{\alpha\beta^\alpha}{(\beta-t)^{\alpha+1}} \qquad M_X^{(2)}(t) = \frac{\alpha(\alpha+1)\beta^\alpha}{(\beta-t)^{\alpha+2}}$$

so that $M_X^{(1)}(0) = \dfrac{\alpha}{\beta}$ and $M_X^{(2)}(0) = \dfrac{\alpha(\alpha+1)}{\beta^2}$, and thus

$$\mathrm{E}_{f_X}[X] = \frac{\alpha}{\beta} \qquad \mathrm{Var}_{f_X}[X] = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}$$

Notes

(1) If $X_1 \sim Ga(\alpha_1, \beta), X_2 \sim Ga(\alpha_2, \beta)$ are independent random variables, and $Y = X_1 + X_2$, then $Y \sim Ga(\alpha_1 + \alpha_2, \beta)$ (directly from properties of mgfs).

(2) $Ga(1, \beta) \equiv Exp(\beta)$.

(3) If $X_1, ..., X_n \sim Exp(\lambda)$ are independent random variables, and $Y = X_1 + ... + X_n$, then $Y \sim Ga(n, \lambda)$ (directly from (1) and (2)).

(4) For $\alpha > 0$,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt = \left[-t^{\alpha-1} e^{-t}\right]_0^\infty + \int_0^\infty (\alpha-1) t^{\alpha-2} e^{-t}\, dt = (\alpha-1) \int_0^\infty t^{\alpha-2} e^{-t}\, dt = (\alpha-1)\Gamma(\alpha-1)$$

so $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$. Thus if $\alpha = 1, 2, ...$, then $\Gamma(\alpha) = (\alpha-1)!$.

(5) **Special Case :** If $\alpha = 1, 2, ...$ the $Ga(\alpha/2, 1/2)$ distribution is also known as the **Chi-squared distribution with α degrees of freedom**

(6) If $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ are independent Chi-squared random variables with n_1 and n_2 degrees of freedom respectively, then random variable defined as the ratio $(X_1/n_1)/(X_2/n_2)$ has an **F-distribution with (n_1, n_2) degrees of freedom**.

(7) The density function of X can be re-written

$$f_X(x) = \exp\left\{\mathbf{t}(\mathbf{x})^T \mathbf{q}(\psi) + \log h(x) + \log c(\psi)\right\}$$

where $\mathbf{t}(\mathbf{x}) = \begin{pmatrix} \log x \\ x \end{pmatrix}$, $\mathbf{q}(\psi) = \begin{pmatrix} \alpha - 1 \\ -\beta \end{pmatrix}$, $h(x) = 1$, $c(\psi) = \dfrac{\beta^\alpha}{\Gamma(\alpha)}$.

4.4 BETA DISTRIBUTION

NOTATION $X \sim Be(\alpha, \beta)$ **RANGE** $\mathbb{X} = (0,1)$

PDF

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} \qquad 0 < x < 1 \text{ where } \alpha, \beta > 0.$$

rth MOMENT

For $r = 1, 2, ...,$

$$\begin{aligned} \mathrm{E}_{f_X}[X^r] &= \int_0^1 x^r \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{r+\alpha-1}\, (1-x)^{\beta-1}\, dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(r+\alpha)\Gamma(\beta)}{\Gamma(r+\alpha+\beta)} \end{aligned}$$

$$\Longrightarrow \mathrm{E}_{f_X}[X] = = \frac{\alpha}{\alpha+\beta} \qquad \mathrm{E}_{f_X}[X^2] = = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

$$\Longrightarrow \mathrm{Var}_{f_X}[X] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Notes

(1) The beta distribution arises naturally in the context of order statistics; if $X_1, ..., X_k$ are i.i.d. random variables with cdf F_X, say, consider first the random variables $U_1, ..., U_k$ defined by $U_i = F_X(X_i)$ for $i = 1, ..., k$. It can be shown that $U_1, ..., U_k$ are i.i.d. $Uniform(0,1)$ random variables. Now, consider the order statistics $Y_1, ..., Y_k$ derived from $U_1, ..., U_k$; using previous results, it can be shown that the marginal distribution of the jth order statistic is $Be(j, k-j+1)$, for $j = 1, ..., k$.

(2) If $X_1 \sim Ga(\alpha_1, \beta)$, $X_2 \sim Ga(\alpha_2, \beta)$ are independent random variables, and $Y = X_1/(X_1+X_2)$, then $Y \sim Be(\alpha_1, \alpha_2)$ (using standard multivariate transformation techniques).

(3) Suppose that random variables X and Y have a joint probability distribution such that the **conditional** distribution of X, given $Y = y$ for $0 < y < 1$, is binomial, $Bin(n, y)$, and the **marginal** distribution of Y is beta, $Be(\alpha, \beta)$, so that

$$f_{X|Y}(x|y) = \binom{n}{x} y^x (1-y)^{n-x} \quad x = 0, 1, ..., n \qquad f_Y(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \quad 0 < y < 1.$$

Then the **marginal** distribution of X is given by

$$f_X(x) = \int_0^1 f_{X|Y}(x|y) f_Y(y)\, dy = \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)} \qquad x = 0, 1, 2...n$$

(3) If $\alpha = \beta = 1$, $Be(\alpha, \beta) \equiv Uniform(0,1)$.

(4) The density function of X can be re-written

$$f_X(x) = \exp\left\{\mathbf{t}(x)^T \mathbf{q}(\psi) + \log h(x) + \log c(\psi)\right\}$$

where $\mathbf{t}(x) = \begin{pmatrix} \log x \\ \log(1-x) \end{pmatrix}$, $\mathbf{q}(\psi) = \begin{pmatrix} \alpha - 1 \\ \beta - 1 \end{pmatrix}$, $h(x) = 1$, $c(\psi) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$.

4.5 NORMAL DISTRIBUTION

NOTATION $X \sim N(\mu, \sigma^2)$ **RANGE** $\mathbb{X} = \mathbb{R}$

PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad x \in \mathbb{R} \quad \text{where } \mu \in \mathbb{R}, \sigma > 0.$$

MGF

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2 + tx\right\} dx \\ &= \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(x - (\mu + t\sigma^2)\right)^2\right\} dx \\ &= \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \end{aligned}$$

because the integrand is a pdf, and thus the integral is equal to one.

***r*th MOMENT**

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = (\mu + t\sigma^2)\exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \qquad M_X^{(2)}(t) = (\mu^2 + 2t\sigma^2\mu + t^2\sigma^4 + \sigma^2)\exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\}$$

so that $M_X^{(1)}(0) = \mu$ and $M_X^{(2)}(0) = \mu^2 + \sigma^2$, and thus

$$\mathrm{E}_{f_X}[X] = \mu \qquad \mathrm{Var}_{f_X}[X] = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

Notes

(1) Special Case : If $\mu = 0$, $\sigma^2 = 1$, then X has a **Standard** or **Unit** normal distribution. Usually, the pdf of the unit normal is written $\phi(x)$, and the cdf is written $\Phi(x)$.

(2) If $X \sim N(0,1)$, and $Y = \sigma X + \mu$, then $Y \sim N(\mu, \sigma^2)$. Re-expressing this result, if $X \sim N(\mu, \sigma^2)$, and $Y = (X - \mu)/\sigma$, then $Y \sim N(0,1)$. (using transformation or mgf techniques)

(3) **The Central Limit Theorem.**

Suppose $X_1, ..., X_n$ are i.i.d. random variables with mgf M_X, with $\mathrm{E}_{f_X}[X_i] = \mu$ and $\mathrm{Var}_{f_X}[X_i] = \sigma^2$ that is, the mgf and the expectation and variance of the X_is are specified, but the pdf is not. Let the random variable Z_n be defined by

$$Z_n = \frac{\displaystyle\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}}$$

and let Z_n have mgf M_{Z_n}. Then, as $n \longrightarrow \infty$, $M_{Z_n}(t) \longrightarrow \exp\left\{t^2/2\right\}$ **irrespective of the distribution of the X_is**, that is, the distribution of Z_n tends to a **unit normal** distribution as n tends to infinity. (for full details, see Central Limit Theorem handout).

(4) If $X \sim N(0,1)$, and $Y = X^2$, then $Y \sim \chi^2_1$, so that the square of a unit normal random variable has a chi-square distribution with 1 degree of freedom.

(5) If $X \sim N(0,1)$, and $Y \sim N(0,1)$ are independent random variables, and Z is defined by $Z = X/Y$, the Z has a **Cauchy** distribution

$$f_Z(z) = \frac{1}{\pi}\frac{1}{1+z^2} \qquad z \in \mathbb{R}$$

(6) If $X \sim N(0,1)$, and $Y \sim Ga(n/2, 1/2)$ for $n = 1, 2, ...$ (so that $Y \sim \chi^2_n$), are independent random variables, and T is defined by

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

then T has a **Student-t** distribution with n degrees of freedom, $T \sim St(n)$,

$$f_T(t) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)}\left(\frac{1}{n\pi}\right)^{1/2}\left\{\frac{1}{1+t^2/n}\right\}^{(n+1)/2} \qquad t \in \mathbb{R}$$

Taking limiting cases of the Student-t distribution

$$\begin{array}{ll} n \longrightarrow \infty & St(n) \longrightarrow N(0,1) \\ n \longrightarrow 1 & St(n) \longrightarrow Cauchy \end{array}$$

(7) The density function of X can be re-written

$$f_X(x) = \exp\left\{\mathbf{t}(\mathbf{x})^T\mathbf{q}(\psi) + \log h(x) + \log c(\psi)\right\}$$

where $\mathbf{t}(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$, $\mathbf{q}(\psi) = \begin{pmatrix} \mu/\sigma^2 \\ -1/2\sigma^2 \end{pmatrix}$, $h(x) = 1$, and $c(\psi) = (2\pi\sigma^2)^{1/2}\exp\left\{\dfrac{-\mu^2}{2\sigma^2}\right\}$.

MULTIVARIATE PROBABILITY DISTRIBUTIONS

THE MULTINOMIAL DISTRIBUTION

The **multinomial** distribution is a multivariate generalization of the **binomial** distribution. Recall that the binomial distribution arose from an infinite Urn model with two types of objects being sampled without replacement. Suppose that the proportion of "Type 1" objects in the urn is θ (so $0 \leq \theta \leq 1$) and hence the proportion of "Type 2" objects in the urn is $1-\theta$. Suppose that n objects are sampled, and X is the random variable corresponding to the number of "Type 1" objects in the sample. Then $X \sim \text{Bin}(n, \theta)$, and

$$f_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad \text{for } x \in \{0, 1, 2, ..., n\}$$

Now consider a generalization; suppose that the Urn contains $k+1$ types of objects ($k = 1, 2, ...$), with

$$\text{"Proportion of type } i\text{" objects} \ = \ \theta_i$$

for $i = 1, ..., k+1$. Let X_i be the random variable corresponding to the number of type i objects in a sample of size n, for $i = 1, ..., k$. Then the joint distribution of vector $\mathbf{X} = (X_1, ..., X_k)$ is given by

$$f_{X_1,...,X_k}(x_1, ..., x_k) \ = \ \frac{n!}{x_1!...x_k!x_{k+1}!} \theta_1^{x_1} ....\theta_k^{x_k} \theta_{k+1}{}^{x_{k+1}} \ = \ \frac{n!}{\left\{ \prod\limits_{i=1}^{k+1} x_i! \right\}} \prod_{i=1}^{k+1} \theta_i^{x_i}$$

where $0 \leq \theta_i \leq 1$ for all i, and $\theta_1 + ... + \theta_k + \theta_{k+1} = 1$, and where x_{k+1} is defined by $x_{k+1} = n - (x_1 + ... + x_k)$. This is the mass function for the **multinomial distribution** which reduces to the binomial if $k = 1$. It can also be shown that the marginal distribution of X_i is $\text{Bin}(n, \theta_i)$.

EXAMPLE :
A dice is rolled n times; let $X_i =$"total number of i scores". Then $\mathbf{X} = (X_1, ..., X_5)$ has a multinomial distribution with $\theta_i = 1/6$ for $i = 1, ..., 6$.

THE DIRICHLET DISTRIBUTION

The **Dirichlet** distribution is a multivariate generalization of the **beta** distribution. Recall that the beta distribution arose a s follows; suppose that V_1 and V_2 are independent Gamma random variables with $V_1 \sim Ga(\alpha_1, \beta), V_2 \sim Ga(\alpha_2, \beta)$. Then if X is defined by

$$X = \frac{V_1}{V_1 + V_2}$$

we have that $X \sim Be(\alpha_1, \alpha_2)$, and

$$f_X(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1} \ 0 < x < 1$$

Now consider a generalization; suppose that $V_1, ..., V_{k+1}$ are independent Gamma random variables with $V_i \sim Ga(\alpha_i, \beta)$, for $i = 1, ..., k+1$. Define

$$X_i = \frac{V_i}{V_1 + ... + V_{k+1}}$$

for $i = 1, ..., k$. Then the joint distribution of vector $\mathbf{X} = (X_1, ..., X_k)$ is given by

$$f_{X_1,...,X_k}(x_1, ..., x_k) \ = \ \frac{\Gamma(\alpha)}{\Gamma(\alpha_1)...\Gamma(\alpha_k)\Gamma(\alpha_{k+1})} x_1^{\alpha_1 - 1} ... x_k^{\alpha_k - 1} x_{k+1}{}^{\alpha_{k+1} - 1}$$

for $0 \leq x_i \leq 1$ for all i such that $x_1 + ... + x_k + x_{k+1} = 1$, where $\alpha = \alpha_1 + ... + \alpha_{k+1}$ and where x_{k+1} is defined by $x_{k+1} = 1 - (x_1 + ... + x_k)$. This is the density function which reduces to the beta distribution if $k = 1$. It can also be shown that the marginal distribution of X_i is Beta(α_i, α).

EXAMPLE :
The composition of a mineral sample is determined in terms of percentage composition of five compounds. Let X_i be the percentage content of compound i, for $i = 1, ...4$. Then $\mathbf{X} = (X_1, ..., X_4)$ is a vector random variable whose joint probability structure could be described using a distribution.

THE MULTIVARIATE NORMAL DISTRIBUTION

The **multivariate normal** distribution is a multivariate generalization of the **normal** distribution which can be generated in the following way. Suppose that $X_1, ..., X_k$ are i.i.d. $N(0, \sigma^2)$ random variables. Using vector notation, we can write the joint density function of $X_1, ..., X_k$ as

$$f_{X_1,...,X_k}(x_1, ..., x_k) \;=\; \left(\frac{1}{2\pi\sigma^2}\right)^{k/2} \exp\left\{-\frac{1}{2\sigma^2}\,\mathbf{x}^T\mathbf{x}\right\}$$

where $\mathbf{x} = (x_1, ..., x_k)$. Now consider the multivariate transformation from $(X_1, ..., X_k)$ to $(Y_1, ..., Y_k)$ (that is, from $\mathbf{X}$ to $\mathbf{Y}$) defined by

$$\mathbf{Y} = A^T\mathbf{X} + \mu$$

where A is a $k \times k$ invertible matrix of real numbers, and μ is a $k \times 1$ vector. This is a 1-1 transformation, so using the usual multivariate transformation formula, we can obtain the joint density function of $(Y_1, ..., Y_k)$ as

$$f_{Y_1,...,Y_k}(y_1, ..., y_k) \;=\; \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mu)^T\Sigma^{-1}(\mathbf{y} - \mu)\right\}$$

where $\Sigma = \sigma^2 A^T A$. This is the pdf of the **multivariate normal distribution**. It can be shown that any marginal, joint marginal, or conditional distribution of a subset of $Y_1, ..., Y_k$ is normal or multivariate normal.