Consider the special case $m = n, \underline{\mu}_X = \{\underline{0}\}, \underline{V}_X = [I]$, that is, $\underline{X}$ is an IID standard normal *n*-vector. Then $\underline{Y}$ is an *n*-dimensional normal with $\underline{\mu}_Y = \underline{A}_0$, $\underline{V}_Y = \underline{A}\underline{A}^T$. This property is used for simulating correlated normals (Section 8.6.2).

7.17 Joint Normal distribution

7.17.1 Bivariate normal

Recall that *X* is said to have a normal distribution with mean μ and variance $\sigma^2 > 0$ if its density function is of the form:

$$N(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma^{1/2}} \exp[-\frac{1}{2} Q_1(x;\mu,\sigma^2)] \tag{7.71}$$

where, $Q_1(x;\mu,\sigma^2) = \frac{1}{\sigma^2}(x-\mu)^2 = (x-\mu)\sigma^{-2}(x-\mu)$. It is related to the standard normal form ϕ through: $N(x;\mu,\sigma^2) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$.

In a parallel manner, a two dimensional RV $\underline{X} = (X_1, X_2)^T$ is said to have a non-singular bivariate normal distribution if its density function is of the form:

$$N_2(\underline{x};\underline{\mu},V) = \frac{1}{2\pi|V|^{1/2}} \exp[-\frac{1}{2} Q_2(\underline{x};\underline{\mu},V)] \tag{7.72}$$

where,

$$Q_2(\underline{x};\underline{\mu},V) = (\underline{x}-\underline{\mu})^T V^{-1}(\underline{x}-\underline{\mu})$$

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

The correlation coefficient is $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}, |\rho| < 1$ since $|\sigma_{12}| < \sigma_1\sigma_2$. The bivariate N.D. is also denoted as $N_2\left(\underline{\mu},V\right)$. Eq (7.72) can be expanded to:

$$\begin{aligned} N_2(x_1,x_2;\mu_1,\mu_2;\sigma_1,\sigma_2;\rho) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\ \exp&\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\} \end{aligned} \tag{7.73}$$

Its standard form (zero means, unit variances) is:

$$\phi_2(x_1, x_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)\right\} \tag{7.74}$$

7.17.1.1 Singular bivariate normal

$\underline{X}$ has a *singular* bivariate normal distribution if

$\begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}$ and $\begin{Bmatrix} \sigma_1 Z + \mu_1 \\ \sigma_2 Z + \mu_2 \end{Bmatrix}$ are identically distributed

Where Z is N (0,1) and $\sigma_1, \sigma_2, \mu_1, \mu_2$ are real

i.e. if $|V| = 0$.

7.17.1.2 Conditional and Marginal distributions

If $f_{X_1 X_2}(x_1, x_2) = N_2(\underline{X}; \underline{\mu}, V)$, the marginal distribution of X_1 regardless of the correlation coefficient is univariate normal: $f_{X_1}(x_1) = N(\mu_1, \sigma_1)$. Likewise, for X_2. This property can be shown by integrating Eq (7.74) with respect to x_2. Since $x_1^2 - 2\rho x_1 x_2 + x_2^2 = x_1^2(1-\rho^2) + (x_2 - \rho x_1)^2$, Eq (7.74) gives us:

$$\begin{aligned} f_{X_1}(x_1) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x_2 - \rho x_1)^2}{2(1-\rho^2)}\right\} \exp\left\{-\frac{x_1^2}{2}\right\} dx_2 \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x_1^2}{2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x_2 - \rho x_1)^2}{2(1-\rho^2)}\right\} dx_2 \end{aligned} \tag{7.75}$$

A subsitution yeilds:

$$\begin{aligned} f_{X_1}(x_1) &= \frac{1}{2\pi} \exp\left\{-\frac{x_1^2}{2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x_2')^2}{2}\right\} dx_2', \text{ where } x_2' = \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}} \\ &= \frac{1}{2\pi} \exp\left\{-\frac{x_1^2}{2}\right\} \sqrt{2\pi} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x_1^2}{2}\right\} \end{aligned} \tag{7.76}$$

which is nothing but the standard normal density function. It is easy to show that the property holds for any arbitrary bivariate normal with non-zero means and non-unit variances.

If $\underline{X}$ is bivariate normal, then the conditional density of X_1 given a fixed value of $X_2=b$ is, once again, normal. Applying the definition of the conditional density function, and using the above result that the marginal density of X_2 is normal:

$$f_{X_1|X_2=b}(x_1) = \frac{f_{X_1X_2}(x_1,b)}{f_{X_2}(b)}$$
$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \frac{\exp\left\{\frac{x_1^2 - 2\rho x_1 b + b^2}{2(1-\rho^2)}\right\}}{\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{b^2}{2}\right\}} \quad (7.77)$$
$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{\frac{(x_1 - \rho b)^2}{2(1-\rho^2)}\right\}$$
$$= N(\rho b, \sqrt{1-\rho^2})$$

For any arbitrary bivariate normal, the above result generalizes to the result that the conditional density of X_1 given $X_2 = x_2$ is Normal with conditional mean:

$$\mu_{1|2} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

and conditional variance:

$$\sigma_{1|2}^2 = \sigma_1^2(1-\rho^2)$$

Clearly, if $\rho = 0$, then $\mu_{1|2} = \mu_1$, $\sigma_{1|2}^2 = \sigma_1^2$ and $f_{1|2}(x_1;\mu_{1|2},\sigma_{1|2}) = f_1(x_1;\mu_1,\sigma_1)$, i.e., X_1 and X_2 are independent.

7.17.1.3 Linear combinations

The linear combination of any bivariate Normal is again bivariate normal:

If $\underline{Y} = C\underline{X} + \underline{b}$ where $\underline{X} \sim N_2(\underline{\mu}, V)$, then $\underline{Y} \sim N_2(C\mu + b, CVC^T)$

7.17.1.4 Standard uncorrelated bivariate normal

Like the standard normal random variable (zero mean and unit variance), we can define the standard uncorrelated normal. Its mean vector is $(0,0)^T$ and its covariance matrix is the identity matrix.

$$\underline{Z} \text{ is } N_2(\underline{0}, I_2) \quad \text{where} \quad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Any bivariate Normal $\underline{X} \sim N_2(\underline{\mu}, V)$ can be created from the standard uncorrleated bivariate normal by defining:

$$\underline{X} = C\underline{Z} + \underline{b}$$

such that its mean and covariance matrix are: $\mu_X = b$ and $V_X = CI_2C^T = CC^T$.

Standard bivariate Normal:

The standard bivariate normal has its mean vector as $(0,0)^T$, the diagonal elements of its covariance matrix as unity, but the off diagonal terms are not necessarily zero.

$\underline{Y}$ is $N_2(\underline{0}, V_Y)$ where $V_Y = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

Any bivariate Normal $\underline{X} \sim N_2(\underline{\mu}, V_X)$ with the same correlation structure $V_X(i,j) = \sigma_i \sigma_j V_Y(i,j)$ can be created from $\underline{Y}$ by defining:

$X_i = \sigma_i Y_i + \mu_i$

7.17.2 Multivariate normal distributions

An *n* dimensional random variable $\underline{X}$ with mean $\underline{\mu}$ and cov. matrix V is said to have a nonsingular multivariate normal distribution if V is positive definite and the joint PDF of $\underline{X}$ is :

$$f_{\underline{X}}(\underline{x};\underline{\mu},V) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^T V^{-1} (\underline{x}-\underline{\mu}) \right), |V| > 0 \qquad (9.78)$$

which is symbolically written as $\underline{X} \sim N_n(\underline{\mu}, V)$.

7.17.2.1 Properties

(1) If $X \sim N_n(\underline{\mu}, V)$

and $\underline{Y} = C\underline{X} + \underline{b}$ where $C = \left[c_{ij}\right]_{n \times n}$, $b = \{b_i\}_n$

then $\underline{Y}$ is n dimensional normal: $Y \sim N_n(C\underline{\mu} + \underline{b}, CVC^T)$

(2) If $\underline{X} \sim N_n(\underline{\mu}, V)$

and $\underline{Y} = C\underline{X} + \underline{b}$ where $C = \left[c_{ij}\right]_{m \times n}$, $b = \{b_i\}_m$

then $\underline{Y} = \{Y_i\}_m$ is m dimensional normal: $\underline{Y} \sim N_m(C\underline{\mu} + \underline{b}, CVC^T)$

(3) $\underline{Z}$ is the standard independent multivariate normal variable, $\underline{Z} \sim N_n(\underline{0}, I_n)$

If we have a nonsingular $n \times n$ matrix C, then $\underline{X}$ and $(C\underline{Z} + \underline{\mu})$ are identically distributed, with $\underline{\mu}_X = \underline{\mu}$ and $V_X = CC^T$.

Thus, if given $\underline{X} \sim N_n(\underline{\mu}, V)$ we can find C such that $CC^T = V$,

then we can uncouple this distribution into $Z \sim N_n(0, I_n)$

where $\underline{Z} = C^{-1}(\underline{X} - \underline{\mu})$

7.17.2.2 Marginal distributions

Consider the following partitions

$$\{X\}_n = \begin{Bmatrix} X_{1_k} \\ X_{2_{n-k}} \end{Bmatrix}, \quad \{\mu\} = \begin{Bmatrix} \mu_{1_k} \\ \mu_{2_{(n-k)}} \end{Bmatrix},$$

$$V = \left[\frac{V_{11} \,|\, V_{12}}{V_{21} \,|\, V_{22_{(n-k)(n-k)}}} \right]$$

If $\underline{X}$ is $N_n(\underline{\mu}, V)$

Then

$\underline{X}_1 \sim N_k(\underline{\mu}_1, V_{11})$

$\underline{X}_2 \sim N_{n-k}(\underline{\mu}_2, V_{22})$

7.17.2.3 Independence

$\underline{X}, \underline{X}_1, \underline{X}_2$ are all as defined above.

Then $\underline{X}_1$ & $\underline{X}_2$ are independent if and only if

$V_{12} = \underline{\underline{0}}$

In this case (that $\underline{X}_1, \underline{X}_2$ are independent)

$$f(\underline{X}; \underline{\mu}, V) = f_1(\underline{X}_1; \underline{\mu}_1, V_1)\, f_2(\underline{X}_2; \underline{\mu}_2, V_2)$$

7.17.2.4 Conditional distribution

consider the partition as above :

$$\underline{X} = \begin{Bmatrix} \underline{X}_{1k} \\ \underline{X}_2 \end{Bmatrix}, \ \underline{\mu} = \begin{Bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{Bmatrix}, \ V = \left[\frac{V_{11}}{V_{21}} \Big| \frac{V_{12}}{V_{22}} \right]$$

If $\underline{X} \sim N_n(\underline{\mu}, V)$

Then conditional distribution of $\underline{X}_1$, given $\underline{X}_2 = \underline{x}_2$

is $N_k\left(\underline{\mu}_{1|2}, V_{11|2}\right)$

where $\underline{\mu}_{1|2} = \mu_1 + V_{12}\, V_{22}^{-1}(\underline{x}_2 - \underline{\mu}_2)$

and $V_{11|2} = V_{11} - V_{12}\, V_{22}^{-1}\, V_{21}$ is independent of $\underline{x}_2$.

As an example, take $n = 2$:

$$\underline{X} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}, \underline{\mu} = \begin{Bmatrix} \mu_1 \\ \mu_2 \end{Bmatrix}, V = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2{}^2 \end{bmatrix}$$

$$\phi_{X_1|X_2=x_2} = \phi(\mu_{1|2}, \sigma_{1|2}^2)$$

$$\mu_{1|2} = \mu_1 + \mathrm{cov}(X_1 X_2)\frac{1}{\sigma_2^2}(x_2 - \mu_2)$$

$$= \mu_1 + \rho_{12}\frac{\sigma_1\sigma_2}{\sigma_2}(x_2 - \mu_2) = \mu_1 + \frac{\rho_{12}\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

Recall, $\sigma_{12} = \rho_{12}\sigma_1\sigma_2$

then, $\sigma_{1|2}^2 = \sigma_1^2 - \sigma_{12}\frac{1}{\sigma_2^2}\sigma_{21}$

$$= \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}$$

$$= \sigma_1^2 - \rho_{12}^2\frac{\sigma_1^2 \cancel{\sigma_2^2}}{\cancel{\sigma_2^2}}$$

$$= \sigma_1^2(1 - \rho_{12}^2)$$

7.17.2.5 MGF for multivariate Normal

$G\underline{X}(s) = E\left(e^{\underline{s}^T\underline{X}}\right)$

$\left(x-\mu-As\right)^T A^{-1}\left(x-\mu-As\right)$

$=\left(x_0-As\right)^T A^{-1}\left(x_0-As\right) \qquad where\ x_0 = x-\mu$

$=\left(x_0-As\right)^T \left(A^{-1}x_0-s\right)$

$=x_0{}^T\left(A^{-1}x_0-s\right)-s^TA^T\left(A^{-1}x_0-s\right)$

$=x_0{}^TA^{-1}x_0-x_0^Ts-s^TAA^{-1}x_0+s^TAs$

$=x_0{}^TA^{-1}x_0-x_0^Ts-s^Tx_0+s^TAs$

$=x_0{}^TA^{-1}x_0-2s^Tx_0-s^TAs$

$=x_0{}^TA^{-1}x_0-2s^Tx+2s^T\mu+s^TAs$

$\therefore \frac{1}{2}x_0{}^TA^{-1}x_0-s^Tx=\frac{1}{2}\left(x-\mu-As\right)^T A^{-1}\left(x-\mu-As\right)-s^T\mu-\frac{1}{2}s^TAs$

$\therefore G_{\underline{X}}(s)=E\left[e^{S^TX}\right]=\int_{-\infty}^{\infty}....\int ce^{\underline{s}^T\underline{x}}e^{-\frac{1}{2}\left(\underline{x}_0{}^T\underline{A}^{-1}x_0\right)}d\underline{x}$

$=\int_{-\infty}^{\infty}...\int ce^{s^T\mu+\frac{1}{2}s^TAs}e^{-\frac{1}{2}(x-\mu-As)^TA^{-1}(x-\mu-As)}d\underline{x}$

$e^{s^T\mu+\frac{1}{2}s^TAs}\times 1.$

7.17.3 Examples involving joint normal random variables

7.17.3.1 Product of lognormals

Consider the product,

$$Y = a_0Y_1^{\alpha_1}Y_2^{\alpha_2}......Y_n^{\alpha_n} \qquad (7.79)$$

If Y_i is LN for all *i*, the product *Y* is lognormal as well.

The mean and variance of *Y* can be found the following way. Let mi be the median of Yi, and let Vi be the c.o.v. of Yi.

$\ln Y = \ln a_0 + \alpha_1 \ln Y_1 + \alpha_2 \ln Y_2 + ...$

$\mu \ln Y = \ln a_0 + \alpha_1\mu_{\ln Y_1} + ...$

$\ln mY = \ln a_0 + \alpha_1 \ln m_1 +$

$m_Y = a_0 m_1^{\alpha_1} m_2^{\alpha_2}...m_n^{\alpha_n}$

If Y_i's are independent.

$1+V_Y{}^2 = (1+V_1{}^2)^{\alpha_1{}^2}(1+V_2{}^2)^{\alpha_2{}^2}...(1+V_n{}^2)^{\alpha_n{}^2}$

7.17.3.2 Example: power demand

The peak daily power demand, D, in Los Angeles is a Normal variable with mean 5 GW and coefficient of variation (c.o.v.) 40%. The power supply network for Los Angeles has a capacity, C, which is also a Normal random variable with mean 8 GW and c.o.v. 20%. C and D are independent of each other.

A "brownout" is said to occur if D exceeds C.

a) Find the probability of a brownout in Los Angeles on a given day.

b) The probability of daily brownout needs to be reduced to 0.023. This is possible by bringing up the capacity to C_{new} . What should be the mean of C_{new} (assume that C_{new} remains a Normal variable and its c.o.v. is still 20%) ?

7.17.3.3 Independent vs uncorrelated Gaussian variables

Examples

7.18 Convergence of a sequence of RVs

A sequence[8] of real numbers is called convergent if it has a limit. A real number *l* is the limit of a sequence if for each positive ε there is an N such that for all $n \geq N$ we have $|x_n - l| < \varepsilon$. A sequence can have at most one limit, and conventionally, +/- ∞ is not considered a valid limit. Also, a sequence is convergent if and only if it is a Cauchy sequence. A sequence is a Cauchy sequence if given ε > 0 there is an N such that for all n≥N and all m≥N, we have $|x_n - x_m| < \varepsilon$.

Now consider a sequence of random variables $\{X_1, X_2, ..., X_n\}$. Not all sequences or random variables converge to anything. But in some cases we know they do, as in the mean of *n* iid random variables. Can we generalized this? The question whether a sequence of RVs converge arises naturally in cases of differentiation and integration of a stochastic process, *X*(*t*).

Define the "derivative" of process *X* the usual way as:

$$Y(t,h) = \frac{X(t+h) - X(t)}{h}, \; h \to 0 \tag{7.80}$$

What does it mean? Is *Y* a legitimate stochastic process? If so, in what sense? What and how does it converge to? We will deal with these questions in Sections 10.7 and 10.8. First, we describe the four different ways that convergence of a sequence of random variable can be talked of.

[8] A sequence is a function of positive integers, i.e., <x_n> is a function that maps each natural number n into the real number x_n.