1.010 - Brief Notes # 8
Selected Distribution Models

- **The Normal (Gaussian) Distribution:**

Let $X_1, \ldots, X_n$ be independent random variables with common distribution $F_X(x)$. The so called central limit theorem establishes that, under mild conditions on F_X, the sum $Y = X_1 + \ldots + X_n$ approaches as $n \rightarrow \infty$, a limiting distributional form that does not depend on F_X. Such a limiting distribution is called the Normal or Gaussian distribution. It has the probability density function:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(y-m)^2/\sigma^2}$$

where m = mean value of Y
σ = standard deviation of Y

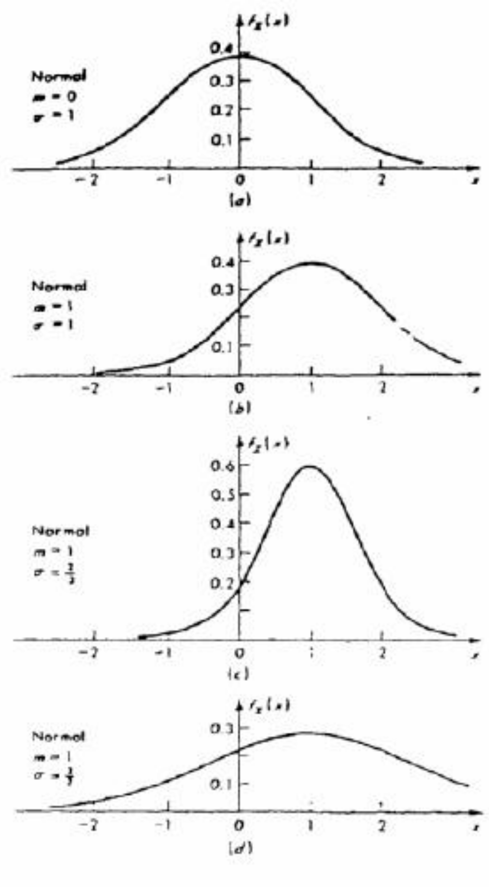
Notice: $m = nm_X$, $\sigma^2 = n{\sigma_X}^2$, where m_X and ${\sigma_X}^2$ are the mean value and variance of X.

- **Properties of the Normal (Gaussian) Distribution:**

1. For most distributions F_X, convergence to the normal distribution is obtained already for n as small as 10.

2. Under mild conditions, the distribution of $\sum_i X_i$ approaches the normal distribution also for dependent and differently distributed X_i.

3. If $X_1, \ldots, X_n$ are independent normal variables, then any linear function $Y = a_0 + \sum_i a_i X_i$ is also normally distributed.



Normal density functions.

- **<u>The Lognormal Distribution:</u>**

  Let $Y = W_1 W_2 \cdots W_n$, where the W_i are iid , positive random variables. Consider:

  $X = \ln Y = \sum_{\text{all } i} \ln W_i$

  For n large, $X \sim N(m_{\ln Y}, \sigma^2_{\ln Y})$

  $Y = e^X$ has a <u>lognormal distribution</u> with PDF:

  $$f_Y(y) = \frac{dx}{dy} f_X[x(y)] = \frac{1}{y} \frac{1}{\sqrt{2\pi}\sigma_{\ln Y}} e^{-\frac{1}{2}(\ln Y - m_{\ln Y})^2/\sigma^2_{\ln Y}}, \quad y \geq 0$$

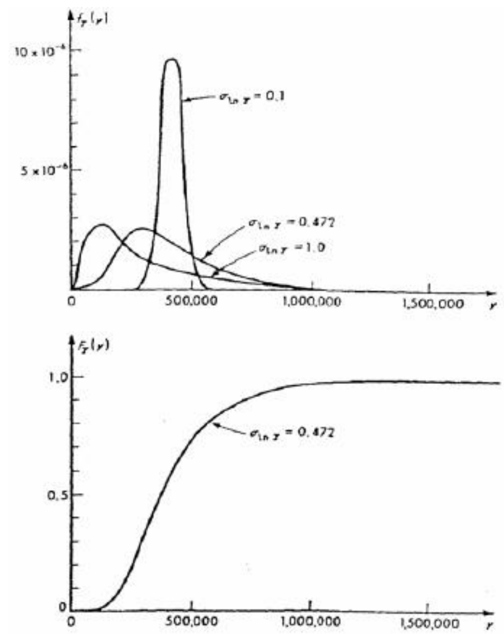
If $X \sim N(m_X, {\sigma_X}^2)$, then $Y = e^X \sim LN(m_Y, {\sigma_Y}^2)$ with mean value and variance given by:

$$\begin{cases} m_Y & = e^{m_X + \frac{1}{2}{\sigma_X}^2} \\ {\sigma_Y}^2 & = e^{2m_X + {\sigma_X}^2}\left(e^{{\sigma_X}^2} - 1\right) \end{cases}$$

Conversely, m_X and ${\sigma_X}^2$ are found from m_Y and ${\sigma_Y}^2$ as follows:

$$\begin{cases} m_X & = 2\ln(m_Y) - \dfrac{1}{2}\ln({\sigma_Y}^2 + m_Y^2) \\ {\sigma_X}^2 & = -2\ln(m_Y) + \ln({\sigma_Y}^2 + m_Y^2) \end{cases}$$

Property: products and ratios of independent lognormal variables are also lognormally distributed.



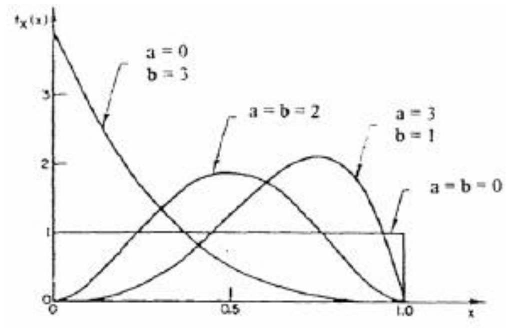
Lognormal distributions showing influence of $\sigma_{\ln Y}$.

- **The Beta Distribution:**

The Beta distribution is commonly used to describe random variables with values in a finite interval. The interval may be normalized to be [0, 1]. The Beta density can take on a wide variety of shapes. It has the form:

$f_Y(y) \propto y^a(1-y)^b$

where a and b are parameters. For $a = b = 0$, the Beta distribution becomes the uniform distribution.



Standard Beta PDF

- **Multivariate Normal Distribution:**

Consider $\underline{Y} = \sum_{i=1}^{n} \underline{X}_i$, where the $\underline{X}_i$ are iid random vectors.
As n becomes large, the joint probability density of $\underline{Y}$ approaches a form of the type:

$$f_{\underline{Y}}(\underline{y}) = \frac{(\det \underline{\Sigma})^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\underline{y}-\underline{m})^T \underline{\Sigma}^{-1}(\underline{y}-\underline{m})}$$

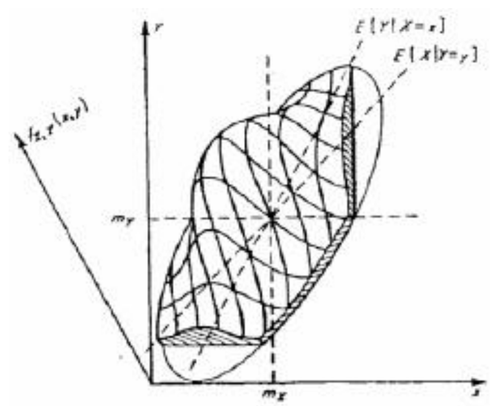
where $\underline{m}$ and $\underline{\Sigma}$ are the mean vector and covariance matrix of $\underline{Y}$.

Properties

1. Contours of $f_{\underline{Y}}$ are ellipsoids centered at $\underline{m}$.
2. If the components of $\underline{Y}$ are uncorrelated, then they are independent.
3. The vector $\underline{Z} = \underline{a} + \underline{B}\,\underline{Y}$, where $\underline{a}$ is a given vector and $\underline{B}$ is a given matrix, has jointly normal distribution $N(\underline{a} + \underline{B}\,\underline{m}, \underline{B}\,\underline{\Sigma}\,\underline{B}^T)$.

Let $\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \end{bmatrix}$ be a partition of $\underline{Y}$, with associated partitioned mean vector $\begin{bmatrix} \underline{m}_1 \\ \underline{m}_2 \end{bmatrix}$ and covariance matrix $\begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix}$. Then:

4. $\underline{Y}_i$ has jointly normal distribution: $N(\underline{m}_i, \underline{\Sigma}_{ii})$.
5. $(\underline{Y}_1 | \underline{Y}_2 = \underline{y}_2)$ has normal distribution $N(\underline{m}_1 + \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}(\underline{y}_2 - \underline{m}_2), \underline{\Sigma}_{11} - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{\Sigma}_{12}^T)$.



The bivariate normal distribution.

Relationships between Mean and Variance of Normal and Lognormal Distributions

If $X \sim N(m_X, {\sigma_X}^2)$, then $Y = e^X \sim LN(m_Y, {\sigma_Y}^2)$ with mean value and variance given by:

$$\begin{cases} m_Y &= e^{m_X + \frac{1}{2}{\sigma_X}^2} \\ {\sigma_Y}^2 &= e^{2m_X + {\sigma_X}^2}\left(e^{{\sigma_X}^2} - 1\right) \end{cases}$$

Conversely, m_X and ${\sigma_X}^2$ are found from m_Y and ${\sigma_Y}^2$ as follows:

$$\begin{cases} m_X &= 2\ln(m_Y) - \dfrac{1}{2}\ln({\sigma_Y}^2 + m_Y^2) \\ {\sigma_X}^2 &= -2\ln(m_Y) + \ln({\sigma_Y}^2 + m_Y^2) \end{cases}$$