the density of a bivariate random vector

$(X, Y)$ has a bivariate normal distribution
if it has a density given by

\[ f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1 - \rho^2}} \exp \left( -\frac{1}{2(1 - \rho^2)} \left( \frac{(x - \mu_X)^2}{\sigma_X^2} + \frac{(y - \mu_Y)^2}{\sigma_Y^2} - 2\rho \frac{(x - \mu_X)(y - \mu_Y)}{\sigma_X\sigma_Y} \right) \right). \]
transformation rule

If $X$ is a random variable and $Y = g(X)$, with $g : \mathbb{R} \rightarrow \mathbb{R}$ monotone and differentiable with inverse $h$, then

$$f_Y(y) = \frac{f_X(h(y))}{|g'(h(y))|}.$$ 

If $X$ is an $n$-dimensional random vector and $Y = g(X)$, where $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is invertible (with inverse $h$) and differentiable, then

$$f_Y(y) = \frac{f_X(h(y))}{|J(h(y))|},$$

where

$$J(x) = \det \begin{pmatrix} \frac{\partial}{\partial x_1} g_1(x) & \cdots & \frac{\partial}{\partial x_n} g_1(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} g_n(x) & \cdots & \frac{\partial}{\partial x_n} g_n(x). \end{pmatrix}$$
**Important implication:** if $X$ and $Y$ are independent, then \( \text{Cov}(X, Y) = 0 \), so they are uncorrelated.

**BUT**, if $X$ and $Y$ are uncorrelated, they are not necessarily independent.

Example:

<table>
<thead>
<tr>
<th>$Y$ \ $X$</th>
<th>$-1$</th>
<th>$0$</th>
<th>$+1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0$</td>
<td>$1/4$</td>
<td>$0$</td>
<td>$1/4$</td>
</tr>
<tr>
<td>$1$</td>
<td>$0$</td>
<td>$1/2$</td>
<td>$0$</td>
</tr>
<tr>
<td></td>
<td>$1/4$</td>
<td>$1/2$</td>
<td>$1/4$</td>
</tr>
</tbody>
</table>

We see that $\mathbb{E}X = 0$, $\mathbb{E}(XY) = 0$, so $\text{Cov}(X, Y) = 0$,
but $\mathbb{P}(X = 0, Y = 0) \neq \mathbb{P}(X = 0)\mathbb{P}(Y = 0)$.

However........
Remember that in general
\[ \rho = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}. \]

Let \((X, Y)\) have a bivariate normal distribution with parameters \(\mu_X\) (the expected value of \(X\)), \(\mu_Y\), \(\sigma_X\) (the standard deviation of \(X\)), \(\sigma_Y\) and (correlation coefficient) \(\rho\).

We have seen that IN THIS CASE \(X\) and \(Y\) are independent iff \(\rho = 0\).

**Hence for bivariate normal \((X, Y)\) independence is equivalent to being uncorrelated!**

**Warning:** If \(X\) is normal and \(Y\) is normal, then it does NOT necessarily follow that \((X, Y)\) is bivariate normal. However, if one also knows that \(X\) and \(Y\) are independent, then \((X, Y)\) is bivariate normal.
If $X = (X_1, \ldots, X_m)^\top$ and $Y = (Y_1, \ldots, Y_n)^\top$ are random vectors, then $\text{Cov}(X, Y)$ is the $m \times n$ matrix with elements

$$\text{Cov}(X, Y)_{ij} = \text{Cov}(X_i, Y_j).$$

For $X = Y$ we write $\text{Cov}(X)$ instead of $\text{Cov}(X, Y)$.

**Proposition**

- $\text{Cov}(X)$ is a symmetric nonnegative definite matrix.
- If a sub-vector of $X$ is independent of a sub-vector of $Y$, then their corresponding covariance matrix is the zero matrix.
- If $X$ has expectation vector $\mu$ and covariance matrix $\Sigma$, then $Y = AX + b$ has expectation vector $A\mu + b$ and covariance matrix $A\Sigma A^\top$. 

Peter Spreij  
Statistics M.Phil. course Tinbergen Institute
Let a random $n$-vector $X$ have expectation vector $\mu$ and covariance matrix $\Sigma$. Assume that $\Sigma$ is invertible. Then $X$ is said to have multivariate normal distribution if the density of $X$ is

$$
\frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp \left( -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right).
$$

**Proposition**

- Two non-overlapping sub-vectors of $X$ are independent iff their covariance matrix is zero.
- If $X$ has a multivariate normal distribution with expectation vector $\mu$ and covariance matrix $\Sigma$, then $Y = AX + b$ ($A$ a square invertible matrix, $b$ a vector) also has a multivariate normal distribution, with expectation vector $A\mu + b$ and covariance matrix $A\Sigma A^\top$. A subvector of $Y$ also has a normal distribution.
Proposition

Let $X, X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be random variables, $c$ a real constant.

1. If $X_n \xrightarrow{P} X$, then also $X_n \xrightarrow{d} X$.
2. If $X_n \xrightarrow{d} c$, then also $X_n \xrightarrow{P} c$.
3. If $X_n \xrightarrow{P} c$, then also $g(X_n) \xrightarrow{P} g(c)$, if $g$ is a continuous function at $c$.
   
   Similar statement for $\xrightarrow{d}$.
4. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, then $g(X_n, Y_n) \xrightarrow{d} g(X, c)$, if $g$ is a continuous function (on $\mathbb{R}^2$).
Proposition

Let $X, X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be random variables, $c$ a real constant.

1. If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then also $X_n \pm Y_n \xrightarrow{P} X \pm Y$.

2. If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then also $X_nY_n \xrightarrow{P} XY$, and also $X_n/Y_n \xrightarrow{P} X/Y$ provided $P(Y \neq 0) = 1$.

3. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$, then also $X_n \pm Y_n \xrightarrow{d} X \pm c$.

4. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$, then also $X_nY_n \xrightarrow{d} Xc$, and $X_n/Y_n \xrightarrow{d} X/c$ provided $c \neq 0$. 
A random variable $X$ is said to have a $\chi^2$ distribution with $n$ degrees of freedom ($\chi^2_n$ distribution) if it has the same distribution as $\sum_{i=1}^{n} Z_i^2$, where the $Z_i$ are iid standard normal random variables:

$$X \overset{d}{=} \sum_{i=1}^{n} Z_i^2.$$
A random variable $X$ is said to have a $t$ distribution with $n$ degrees of freedom ($t_n$ distribution) if

$$X \overset{d}{=} \frac{Z}{\sqrt{W/n}},$$

where $Z$ and $W$ are independent random variables, $Z$ having a standard normal distribution and $W$ having a $\chi^2_n$ distribution.

For large $n$, the $t_n$ distribution is approximately normal (see the tables in Rice for an illustration).
Theorem

Let $X_1, \ldots, X_n$ be a sample from a $N(\mu, \sigma^2)$ distribution. Then

- $\bar{X}$ and $\sum_{i=1}^{n} (X_i - \bar{X})^2$ are independent.
- $\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2$ has a $\chi^2_{n-1}$ distribution.
- The statistic
  \[
  \frac{\sqrt{n}(\bar{X} - \mu)}{S_n}
  \]
  has a $t_{n-1}$ distribution, where

\[
S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.
\]
The Gauss test statistic for $\mu$ when we deal with a sample from the $N(\mu, \sigma^2)$ distribution is

$$\frac{\sqrt{n}(X - \mu)}{\sigma},$$

which we can only use when $\sigma$ is known. If this is not the case, we replace it in the above statistic with $S = \left(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2\right)^{1/2}$. The resulting statistic

$$\frac{\sqrt{n}(X - \mu)}{S},$$

has a $t_{n-1}$ distribution.
Recall that (under some assumptions, including $\hat{\theta}_n \xrightarrow{P} \theta_0$)

$$\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, 1).$$

Hence $(1 - \alpha)$-confidence interval for $\theta_0$ would have limits

$$\hat{\theta} \pm \frac{z(\alpha/2)}{\sqrt{nI(\theta)}}.$$

But, since $\theta_0$ is unknown this does not work. Instead we take the calculable confidence interval

$$\hat{\theta} \pm \frac{z(\alpha/2)}{\sqrt{nI(\hat{\theta})}}.$$

Justification: if $I$ is continuous, then $I(\hat{\theta}_n) \xrightarrow{P} I(\theta_0)$. Hence also

$$\sqrt{nI(\hat{\theta})(\hat{\theta} - \theta_0)} \xrightarrow{d} N(0, 1).$$
(generalized) likelihood ratio test

Neyman-Pearson test to testing $H_0 : \theta = \theta_0$ against $H_A : \theta = \theta_A$ rejects $H_0$ for small values of

$$\frac{f_{\theta_0}(X)}{f_{\theta_A}(X)},$$

when $X$ is observed and where the $f_{\theta}$ are ‘densities’.

For composite hypotheses testing this approach is generalized as follows. We consider $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_A$, where $\Theta_0 \cap \Theta_A = \emptyset$. Let $\Theta = \Theta_0 \cup \Theta_A$. The GLR test rejects the null-hypothesis for small values of

$$\Lambda = \Lambda(X) = \frac{\sup_{\theta \in \Theta_0} f_{\theta}(X)}{\sup_{\theta \in \Theta} f_{\theta}(X)}.$$

Remark: notice that the denominator is maximized by the Maximum likelihood estimator (if it exists).
distribution of the GLR test statistic $\Lambda$

To find the rejection region, one needs the distribution of $\Lambda$ (under the null-hypothesis). Usually $\Lambda$ and its distribution are difficult to handle. Therefore one uses an asymptotic result for the case when we observe a large sample $X = (X_1, \ldots, X_n)$.

Under certain conditions one has the following result:

The distribution of $L = -2 \log \Lambda(X)$ (under $H_0$!) is approximately $\chi^2_{d-d_0}$, where $d = \dim \Theta$ and $d_0 = \dim \Theta_0$.

Hence the rejection set $R$ is approximated by the set \[ \{ x : -2 \log \Lambda(x) \geq \chi^2_{d-d_0}(\alpha) \} . \]
Alternatively, you can compute an approximation of the \( p \)-value, when you observe \( X = x \). The \( p \)-value is

\[
\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(L \geq -2 \log \Lambda(x)),
\]

which you approximate by

\[
P(\chi^2_{d-d_0} > -2 \log \Lambda(x)).
\]

Reject \( H_0 \) if this is smaller than \( \chi^2_{d-d_0}(\alpha) \).