Multivariate Normal Distribution

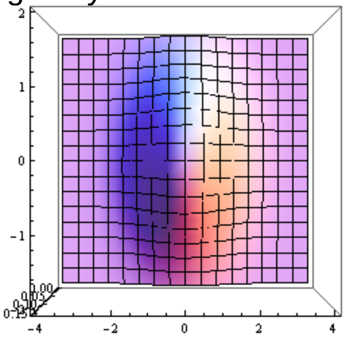
Rebecca Jennings, Mary Wakeman-Linn, Xin Zhao
November 11, 2010

1. Basics

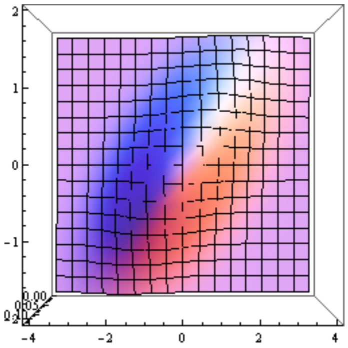
1.1 Parameters

We say $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with parameters $\boldsymbol{\mu} = [E[X_1], \ldots . E[X_n]]'$ and $\boldsymbol{\Sigma} = \mathrm{Cov}[X_i X_j]$ i=1..n, j=1…n. The expected value of $\mathbf{X}$ is $\boldsymbol{\mu}$. There is no restriction on $\boldsymbol{\mu}$, but $\boldsymbol{\Sigma}$ must be at least negative semidefinite, and it must be negative definite in order for a closed form of the pdf to exist.

Notice that, unlike the univariate normal, it is not sufficient to include only the variance as a parameter for the multivariate normal. This can be seen with the following example, where both variables have the same mean and variance, but the covariance varies greatly.



Covariance = 0



Covariance = 0.9

1.2 PDF, CDF, MGF

If $\mathbf{X}$ is an n-dimensional random vector that has a multivariate normal distribution, then it has the following properties [1]:

PDF: $\frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right) \; for \; \boldsymbol{X} \in \mathcal{R}^n$

CDF: Just like the Univariate Normal, there is no closed form for the cdf, although it can be computed numerically.

MGF: $exp(\boldsymbol{t}'\boldsymbol{\mu})exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\right) \; for \; all \; \boldsymbol{t} \in \mathcal{R}^n$

The pdf is maximized at the expected value, which is analogous to the mode of a univariate random variable.

---

[1] Vectors are denoted in bold and transposed vectors have a prime (‘).

2. Why do we care?

2.1 History

The bivariate normal distribution (the two variable version of the multivariate normal) was brought to prominence by Francis Galton (a cousin to Charles Darwin) in order to explain issues of variation between generations brought up by Darwin in *On the Origin of Species*. In a presentation to the British Association for the Advancement of Science in 1885, Galton used multivariate analysis, and the bivariate normal distribution specifically, to examine the relationship between the height of fathers and their eldest adult son. Earlier appearances of the multivariate normal distribution include Robert Adrain in 1808 and Laplace in 1812.

2.2 Applications

Early uses of the multivariate normal applied to biological studies, particularly evolution. However, the distribution can be used to examine ~~at~~ the relationship between any number of normally distributed variables. For economists, multivariate distributions are useful because many elements in nature are distributed with a normal distribution and are related to other normally distributed variables. Some economists have used multivariate distributions to model rainfall in adjacent geographic zones in order to examine the effect on agricultural productivity. [illegible]e multivariate normal distribution can frequently model data gathered from the same individual. For example, we often think that height and weight, when observed on the same individual, are distributed bivariate normal. We could extend this to foot size, or any other number of related physical characteristics, and these measurements together could be modeled as multivariate normal. It can also be used to model measurements that are normally distributed between individuals who are related, which was Galton's use of the distribution. Height, weight, and intelligence are just three examples where we would expect a multivariate normal to satisfactorily model data observed between related individuals.

2.3 Galton's Work

Let us consider Galton's study of height of fathers and sons. Galton found that fathers and sons height, F and S respectively, are normally distributed with equal means of 68 inches and varian[illegible] of 3 inches. He also found that if he only considered fathers of a certain given height, f, the height of sons was normally distributed and that the average height was a linear function of f. In more general terms, this means that if we have two variables that are bivariate normal, say x and y, the conditional distribution of one given the other (X|Y) can be represented as a normal distribution where the parameters of X will be functions of the given Y. This result will be explored more thoroughly when we discuss the conditional distribution of the multivariate normal.

Further, Galton found that if the father is taller (shorter) than average the son will also be taller (shorter) than average, but that in general, sons will tend more to the average than fathers.

2.4 Multivariate Central Limit Theorem

Of particular importance to us is the extension of the Central Limit Theorem for multivariate analysis. We can say that if $\{X_n\}$ is a sequence of iid random vectors with a common mean vector μ and a variance-covariance matrix that is positive definite, we can say that

$$Y_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu})$$

converges in distribution to a N(**0, ∑)** distribution.

3 Derivation of Multivariate Normal from Univariate Normal

We pause briefly to derive results that are necessary in order to completely discuss the relationship between variables that are multivariate normal.

3.1 Moment Generating Function

Consider the random vector **Z** = $(Z_1, Z_2, \ldots Z_n)'$ where $Z_1,\ldots Z_n$ are iid N(0,1). Then the density of **Z** is the product of the densities of each Z_i, and is

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_i^2\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^{\frac{n}{2}} exp\left(-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{\frac{n}{2}} exp\left(-\frac{1}{2}\boldsymbol{z'z}\right)\ for\ \boldsymbol{z} \in \mathcal{R}^n$$

Because the Z_i's have mean 0, variance 1, and are uncorrelated, the mean and covariance matrix of **Z** are $E[\boldsymbol{Z}] = \boldsymbol{0}$ and $Cov[\boldsymbol{Z}] = \boldsymbol{I}_n$ where $\boldsymbol{I}_n$ is the identity matrix of order n. Because the Z_i's are independent, the mgf of **Z** is

$$M_{\boldsymbol{Z}}(\boldsymbol{t}) = E[exp(\boldsymbol{t'Z})] = E\left[\prod_{i=1}^{n} exp(t_i Z_i)\right] = \prod_{i=1}^{n} E[exp(t_i Z_i)]$$

$$= exp\left(\frac{1}{2}\sum_{i=1}^{n} t_i^2\right) = exp\left(\frac{1}{2}\boldsymbol{t't}\right)$$

We say that **Z** ~ $N_n(0, I_n)$ if Z has the previous mgf and pdf. While this is only for the case when Z is made up of iid N(0,1) variables, this result can be used to derive the general case.

Let **∑** be a positive semidefinite and symmetric nxn matrix. Let **μ** be a nx1 vector of constants. Define $\boldsymbol{X} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} + \boldsymbol{\mu}$. Then $E[\boldsymbol{X}] = \boldsymbol{\mu}$ and $Cov[\boldsymbol{X}] = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$. [2] Further, we can find the mgf of **X**

$$M_{\boldsymbol{X}}(t) = E\big(exp(\boldsymbol{t'X})\big) = E\left[exp\left(\boldsymbol{t}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} + \boldsymbol{t'\mu}\right)\right]$$

---

[2] This follows from $E[\boldsymbol{Z}] = \boldsymbol{0}$ and a theorem stating that if **X** = $(X_1, \ldots X_n)$ such that $Var(X_i) < \infty$, then Cov(**AX) = A**Cov(**X**)**A**'.

$$= \exp(\boldsymbol{t}'\boldsymbol{\mu})\, E\left[exp\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{t}\right)'\boldsymbol{Z}\right]$$

$$= \exp(\boldsymbol{t}'\boldsymbol{\mu})\exp\left(\frac{1}{2}\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{t}\right)'\left(\boldsymbol{\Sigma}^{1/2}\boldsymbol{t}\right)\right)^{3}$$

$$= exp(\boldsymbol{t}'\boldsymbol{\mu})exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\right)$$

3.2 PDF

Previously we had required that ∑ be positive semidefinite, but if it is instead positive definite, then the inverse of both $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{1/2}$ exist, and we can derive the pdf of X[4].

By the previous transformation of $\boldsymbol{X} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} + \boldsymbol{\mu}$, we can define the inverse transformation as $\boldsymbol{Z} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})$ with Jacobian[5] $\left|\boldsymbol{\Sigma}^{-1/2}\right| = |\boldsymbol{\Sigma}|^{-1/2}$. By the transformation method, $f_{\boldsymbol{X}}(x) = f_Z(z)|\boldsymbol{\Sigma}|^{-1/2}$. Which gives

$$f_{\boldsymbol{X}}(x) = |\boldsymbol{\Sigma}|^{-1/2}\left(\frac{1}{2\pi}\right)^{n/2} exp\left(-\frac{1}{2}\left(\boldsymbol{\Sigma}^{1/2}(\boldsymbol{X}-\boldsymbol{\mu})\right)'\left(\boldsymbol{\Sigma}^{1/2}(\boldsymbol{X}-\boldsymbol{\mu})\right)\right)$$

$$= \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right) for\ \boldsymbol{X} \epsilon \mathcal{R}^n$$

So **X** ~ N_n(**μ, ∑)** with **μ** = [E[X_1],….E[X_n]]’ and **∑ =** Cov[X_iX_j] i=1..n, j=1…n

4 Some Useful results

4.1 Linear Transform of Multivariate Normal Distribution

We can generalize the previous steps to show that a linear transformation of a multivariate normal random vector has a multivariate normal distribution:

> Suppose **X** ~ N_n(**μ**, ∑). Let **Y** = **AX** + **b** where **b** ϵ $\mathcal{R}^m$ and **A** is an mxn matrix. Then **Y** ~ N_m(**Aμ** + **b**, **A∑A’**).[6]

4.2 Marginal Distribution

The linear transform result can be used to find the marginal distribution of a multivariate normal distribution. This result is not terribly useful in the bivariate case, but can be useful as the number of variables increases and one wants to examine the distribution of s selection of them.

---

[3] This follows from the fact that E[exp{**t’Z}**] **=** exp{(1/2)**t’t**}.

[4] The positive definiteness of ∑ is required for the pdf to have a functional form, not for the distribution to exist.

[5] The Jacobian is the determinant of the first partial derivatives of the inverse transformation. $|J| = \left|{}^{dz}/_{dx}\right|$

[6] This follows from the mgf of **Y**: $E[exp(\boldsymbol{tY})] = exp\left(\boldsymbol{t}'(\boldsymbol{A\mu}+\boldsymbol{b}) + (1/2)\boldsymbol{t}'\boldsymbol{A\Sigma A}'\boldsymbol{t}\right)$

For example, consider a clothing company that wants to release coordinated sets of different types of clothing. If they have data gathered from individuals, they can assume that an individual's observations of height, weight, shoe size, waist size, and other physical measurements will be multivariate normal. However, as they will not need to coordinate all of these observations in order to release their clothing, they would use the marginal distribution of only the variables they care about.

> Let $\mathbf{X}_1$ be a subvector of $\mathbf{X}$, of dimension m < n. The remainder of $\mathbf{X}$ will be the subvector $\mathbf{X}_2$, with dimension p = n-m. Further, partition the mean and covariance in the same way.
>
> $$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$
>
> $\sum_{11}$ is the covariance of $\mathbf{X}_1$ while $\sum_{12}$ is the covariance between components of $\mathbf{X}_1$ and $\mathbf{X}_2$. Define $\boldsymbol{A} = \left[\boldsymbol{I}_m \ \vdots \ \boldsymbol{O}_{mp}\right]$ where $\boldsymbol{O}_{mp}$ is an mxp matrix of zeroes. Then $\mathbf{X}_1 = \mathbf{AX}$ and we can apply the earlier result to say that $\mathbf{X}_1 \sim N_m(\boldsymbol{\mu}_1, \sum_{11})$.

4.3 Independence

Recall that, in general, independent random variables will have a covariance of 0, but a covariance of 0 does not imply that two random variables are independent. In the case of the multivariate normal distribution, however, we can show that a covariance of 0 does imply independence.

> Suppose $\mathbf{X}$ is divided into subvectors as before. $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\sum_{12} = \mathbf{O}$.

This result follows from an examination of the joint and marginal mgfs of $\mathbf{X}_1$ and $\mathbf{X}_2$. The joint mgf is

$$M_{X_{\mathbf{1}},X_{\mathbf{2}}}(\boldsymbol{t}_1, \boldsymbol{t}_2) = exp\left(\boldsymbol{t}_1'\boldsymbol{\mu}_1 + \boldsymbol{t}_2'\boldsymbol{\mu}_2 + \frac{1}{2}(\boldsymbol{t}_1'\boldsymbol{\Sigma}_{11}\boldsymbol{t}_1 + \boldsymbol{t}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{t}_2 + \boldsymbol{t}_2'\boldsymbol{\Sigma}'_{12}\boldsymbol{t}_1 + \boldsymbol{t}_1'\boldsymbol{\Sigma}_{12}\boldsymbol{t}_2)\right)$$

And the product of their marginal distributions is

$$M_{\boldsymbol{X}_1}(\boldsymbol{t}_1)M_{\boldsymbol{X}_2}(\boldsymbol{t}_2) = exp\left(\boldsymbol{t}_1'\boldsymbol{\mu}_1 + \boldsymbol{t}_2'\boldsymbol{\mu}_2 + \frac{1}{2}(\boldsymbol{t}_1'\boldsymbol{\Sigma}_{11}\boldsymbol{t}_1 + \boldsymbol{t}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{t}_2)\right)$$

These can clearly only be equal if the covariance between their components are all 0. In other words, only if $\boldsymbol{\Sigma}_{12} = \boldsymbol{O}$.

4.4 Conditional Distribution of Multivariate Normal

Recall the earlier mention of conditional distribution of a sons heights given fathers. This is a continuation of that discussion.

Assume that $\mathbf{X}$ is partitioned as above. Assume that $\sum$ is positive definite (so that the pdf has a functional form). Then

$$\boldsymbol{X}_1|\boldsymbol{X}_2 \sim N_m(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}')$$

5 An example using these results

5.1 PDF

Let's assume that in some population of married couples, the height the husband (H) and wife (W) have a bivariate normal distribution with parameters μ_H = 5.8 feet; μ_W = 5.3 feet; $\sigma_H = \sigma_W$ = 0.2 foot; ρ = 0.6 (recall that $\sigma_{12} = \rho\sigma_1\sigma_2$). Accordingly, we can say that (H,W) has a $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where

$$\mu = \begin{bmatrix} 5.8 \\ 5.3 \end{bmatrix} \; and \; \Sigma = \begin{bmatrix} 0.04 & 0.024 \\ 0.024 & 0.04 \end{bmatrix}$$

$$f(h,w) = \frac{1}{2\pi\sigma_H\sigma_W(1-\rho^2)} exp\left(\frac{-1}{2(1-\rho^2)}\left(\left(\frac{h-\mu_H}{\sigma_H}\right)^2 - 2\rho\left(\frac{h-\mu_H}{\sigma_H}\right)\left(\frac{w-\mu_w}{\sigma_W}\right) + \left(\frac{w-\mu_W}{\sigma_W}\right)^2\right)\right)$$

$$= \frac{1}{0.0512\pi} exp\left(\frac{-1}{1.28}\left(\left(\frac{h-5.8}{0.2}\right)^2 - 1.2\left(\frac{h-5.8}{0.2}\right)\left(\frac{w-5.3}{0.2}\right) + \left(\frac{w-5.3}{0.2}\right)^2\right)\right)$$

5.2 Conditional Distribution

Using the previous result, we can see that the conditional pdf of any Y given X = x will be distributed

$$N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2(1-\rho^2)\right)^{7}$$

So if we want to find the conditional pdf of W given H = 6.3, we know it is normal with

$$\mu_W = 5.3 + (0.6)(6.3 - 5.8) = 5.6 \text{ and } \sigma_W = (0.2)\sqrt{1-0.36} = 0.16$$

If we wanted to be more general, we could say that $W \sim N(5.3 + 0.6(H - 5.8), 0.16)$, where we now have W distributed univariate normal with its mean a function of H. This many not always be useful, as it imposes a sort of one-way causal relationship on a variable that may not be valid.

We can now find the probability that the wife has a height between 5.28 and 5.92 feet:

$$P(5.28 < W < 5.92 || H = 6.3) = \Phi(2) - \Phi(-2) = 0.954^{8}$$

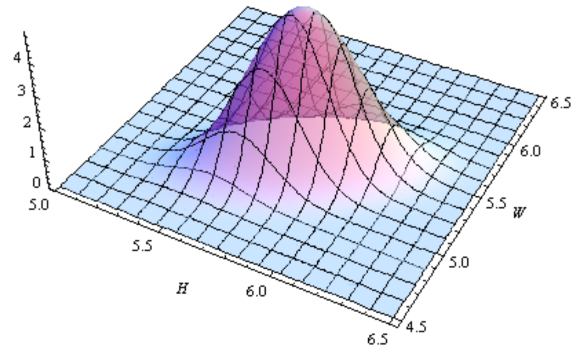
5.3 Graphs

---

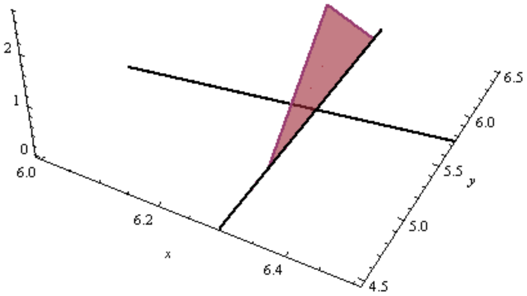
[7] Notice that while the conditional mean of Y depends on the value of x, the variance is the same for all real values of x. Using values from a table of CDF values of the normal distribution, we can say that give X=x, the conditional probability that Y is within$(2.576)\sigma_Y\sqrt{1-\rho^2}$ units of the conditional mean is 0.99 no matter the values of x. As ρ^2 approaches 1 we can see that this area shrinks.

[8] This can be found using a table of values for the CDF of the standard normal distribution.

Because this is a bivariate example, we can look at some graphs of the pdf and the conditional pdf using the parameters given in the problem.



Bivariate Normal of H and W

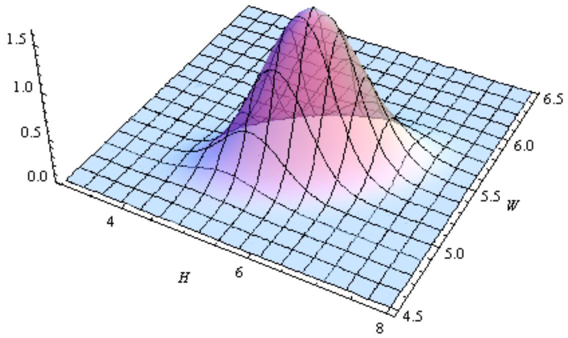


Conditional Distribution of W given H = 6.3

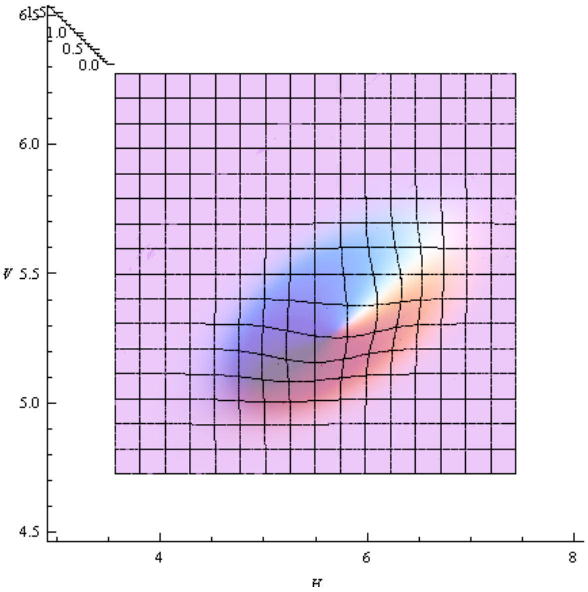
What happens to the density function if we change the correlation coefficient to -0.6?

As you can see, changing the sign of the correlation coefficient changes the orientation of the density function, but not the general shape.

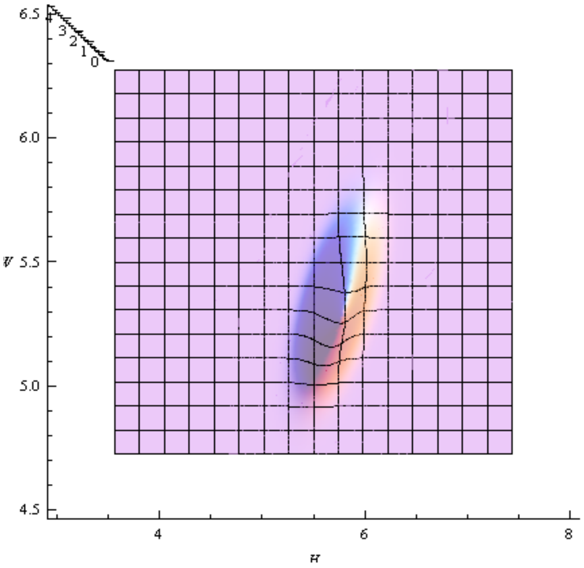
If we triple the standard deviation of H (from 0.2 to 0.6) we get the following density graph



which looks much like the first graph, but consider the H axis. This density is considerably spread. Seen from above and with the same axis, one can more easily see the difference.

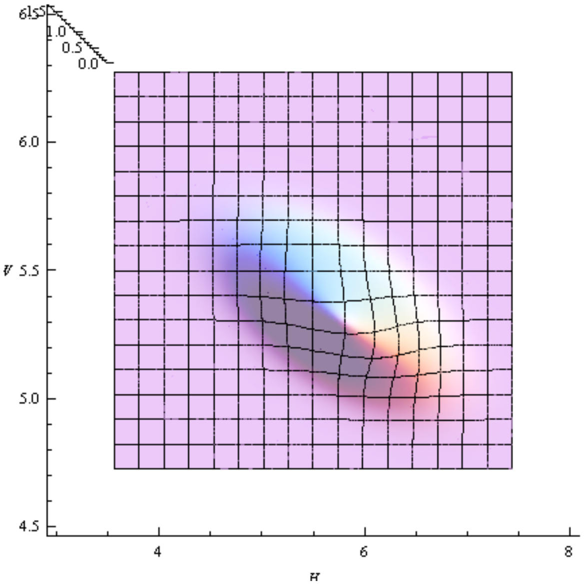


Standard Deviation of H = 0.6

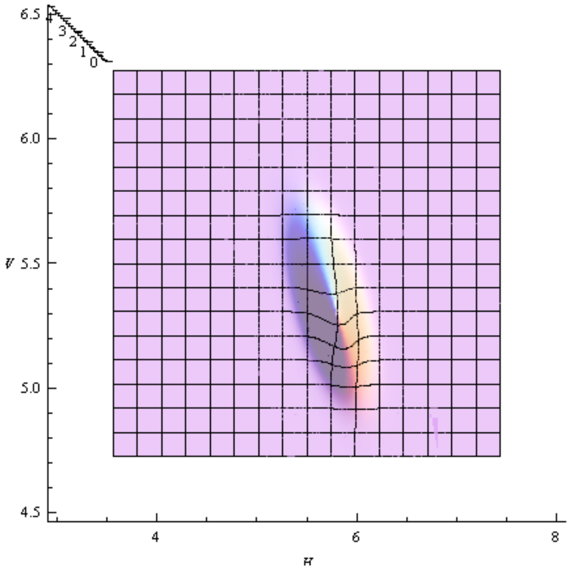


Standard Deviation of H = 0.2

What happens if we change the sign of the correlation coefficient?



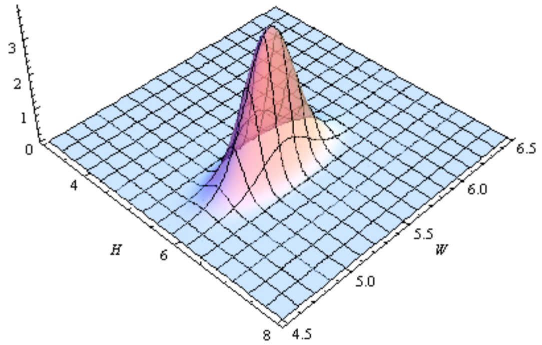
SD of H = 0.6 and correlation = - 0.6



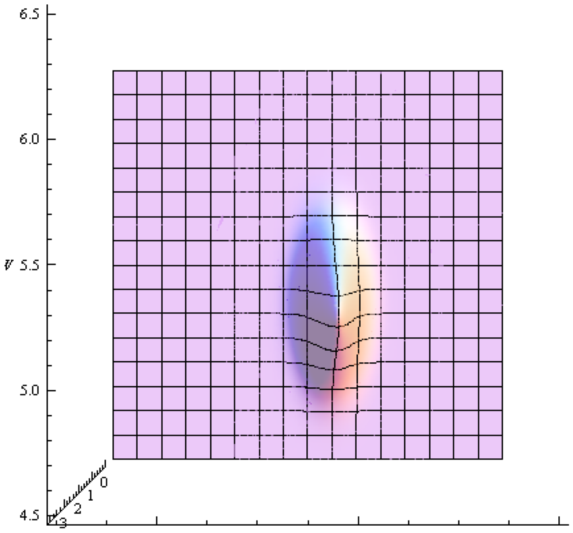
SD of H = 0.2 and correlation = - 0.6

As you can see, changing the sign of the correlation coefficient changes the orientation of the density function. Correlation tells us of the relationship between the two variables. In this example, a positive correlation tells us we will tend to observe taller husband with taller wives. A negative correlation will mean we tend to observe shorter husbands with taller wives. The size of the correlation coefficient tells us the size of the relationship. A number close to one tells that there is little predictive power in knowing the height of a husband when trying to predict the height of the wife, while a coefficient closer to 1 says that knowing the husband’s height gives us a good deal of information about the wife.

What about changing the value of the correlation coefficient? When ρ = 0, we get a density function that looks like the combination of two normal, which it should.



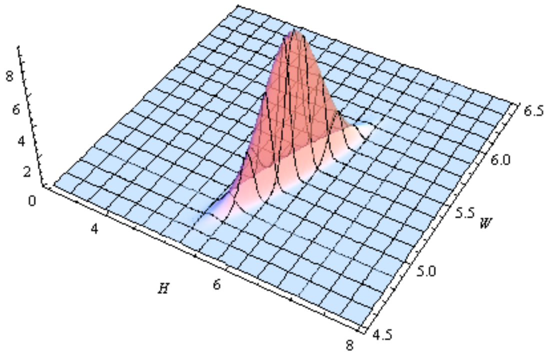
Correlation = 0 between H and W



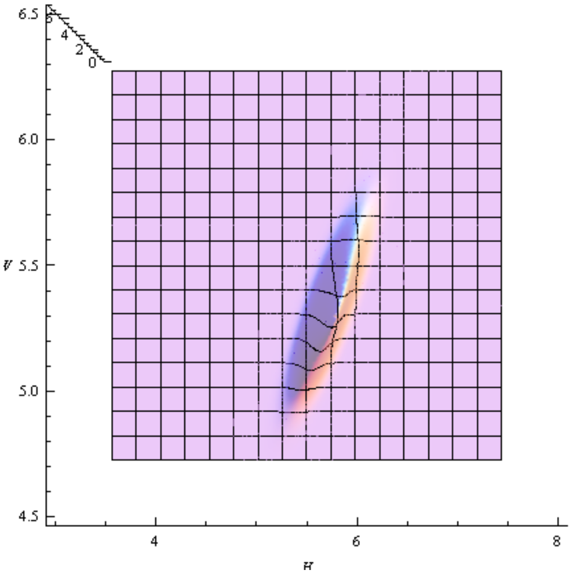
Seen from above

Notice that each individual normal has the mean of the corresponding variable.

What about making ρ closer to 1?



Correlation = 0.9



Seen from above

So we can see that the correlation coefficient changes the severity of the orientation of the density, not just the direction.

6 Relationship with Chi Squared

Hopefully everybody is aware that if X is distributed N(μ, σ^2) with $\sigma^2 > 0$, then the random variable V = $\frac{(X-\mu)^2}{\sigma^2}$ is distributed $\chi^2(1)$. A similar result can be found for the multivariate normal distribution.

> Suppose **X** has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution where **∑** is positive definite.. Then the random variable W = (**X** - **μ**)' $\boldsymbol{\Sigma}^{-1}$ (**X** - **μ**) has a $\chi^2(n)$ distribution.[9]

7 Testing for Multivariate Normalcy

While we often assume that variables are either individually or jointly normal, sometimes it is necessary to prove assumptions, and there are tests to see if a vector is distributed normally. In general, it is much easier to determine if a random vector is not distributed multivariate normal than proving that it is. If any of the individual variables have marginal distributions that are not normal, then the random vector cannot be distributed multivariate normal. However, the reverse is not true unless the variables are also independent. We can show that if we take two variables that are marginally normal, their joint pdf is not necessarily bivariate normal.

For example

$$f(x,y) = \frac{1}{2\pi} exp\left(\frac{-1}{2}(x^2 + y^2)\right)\left\{1 + xyexp\left(-\frac{1}{2}(x^2 + y^2 - 2)\right)\right\} \; for\ x \in \mathcal{R}\ and\ y \in \mathcal{R}$$

is not bivariate normal, but x and y are both distributed normally. Tests for multivariate normalcy involve testing a null hypothesis of normalcy against an alternative of non-normalcy. A sufficiently small p-value will allow the rejection of the null, that the data set is not normal. There are various statistical tests for this, including the Cox-Small test, and an adaptation of the Friedman-Rafsky test.

We can see the probability that a given observation is pulled from a multivariate normal distribution by using either the marginal or conditional. For example, given a bivariate normal distribution with $\mu_x = 2.8, \mu_y = 110, \sigma_x^2 = 0.16, \sigma_y^2 = 100, \rho = 0.6$, the probability that Y is between 106 and 124 is 0.574. Or, given an value for X, say X = 3.2, the probability that Y is between 106 and 124 is 0.735. Thus, using given parameters of a distribution, we can test the probability that any observation is taking from that distribution using various statistical tests at our disposal.

---

[9] This result is derived from the fact that W can be written **Z'Z** where **Z** ~ $N_n(0, \mathbf{I}_n)$. From the univariate result, Z_i^2 is distibuted $\chi^2(1)$ and because the Z_i are independent, their sum (which is W) is distributed $\chi^2(n)$.

8. Citations:

Griffiths, William. “A Gibbs’ Sampler for the Parameters of a Truncated Multivariate Normal Distribution.” Contemporary Issues in Economics:Theory and Application, ed. Stan Hurn and Ralph Becker, Edward Elgar Publishing, 2004

Hogg, Robert V. and Allen T. Craig. Introduction to Mathematical Statistics.

Stigler, Stephen M. “Darwin, Galton, and the Statistical Enlightenment.” Journal of the Royal Statistical Society, 173, Part 3, pp. 469–482, first published online 14 May 2010