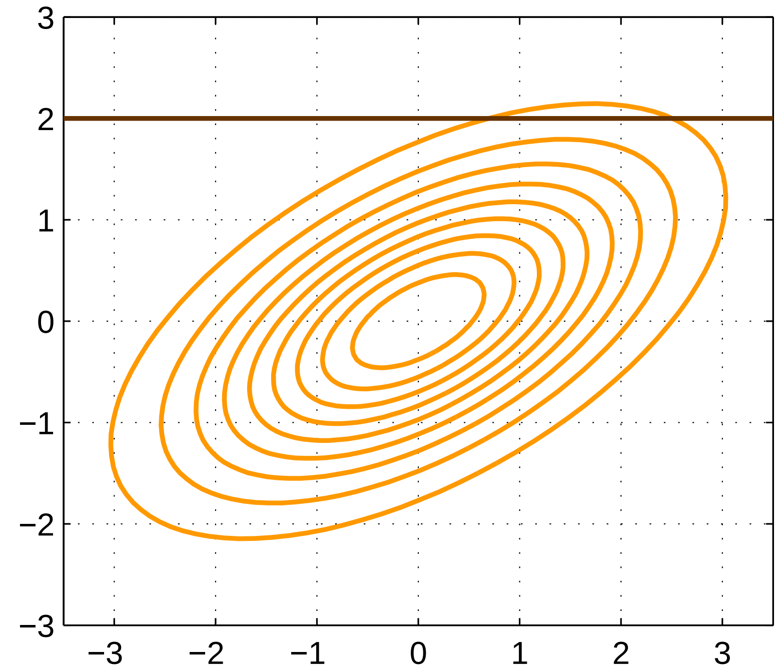
9 - Conditional density

- Conditional density
- Conditional pdf as a limit
- Conditional mean and covariance
- Conditional notation
- Conditional pdf for a Gaussian
- Conditional confidence intervals
- Independence
- Adding independent random variables
- Uncorrelated Gaussians
- Adding Gaussian random vectors

Conditioning

- Given a joint pdf $p(x, y)$ on $\mathbb{R}^2$
- Measure y; we would like to find the conditional pdf of x



Frequency Interpretation

- Collect data (x, y)
- Discard those pairs with

$$y \notin [y_{\text{meas}}, y_{\text{meas}} + \varepsilon]$$

  for small $\varepsilon > 0$

- Then the density of x will approximate the conditional pdf

Conditioning

We would like to define the conditional probability

$$\mathbf{Prob}(A \mid B) = \frac{\mathbf{Prob}(A \cap B)}{\mathbf{Prob}(B)}$$

where

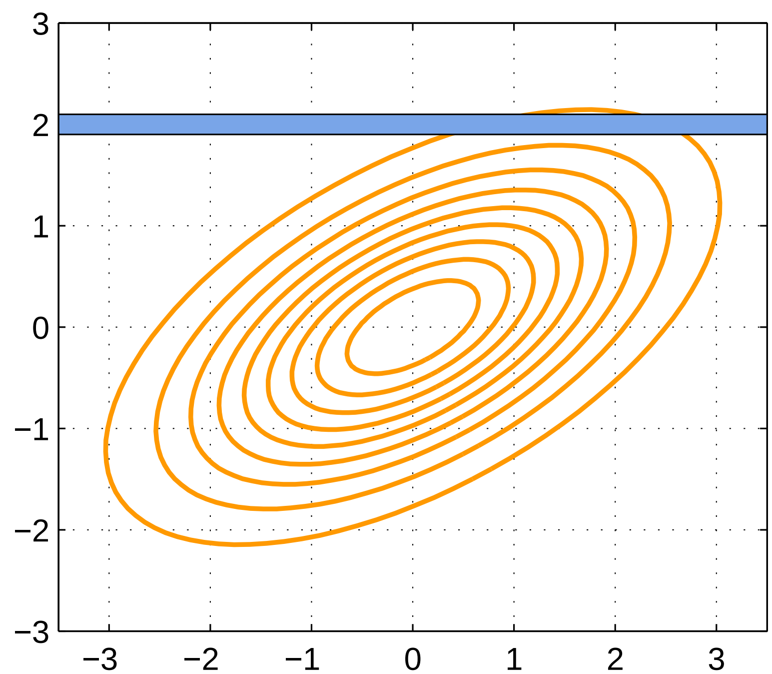
$$A = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \;\middle|\; x \in [a_1, a_2] \right\} \qquad B = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \;\middle|\; y = y_{\text{meas}} \right\}$$

We cannot do this, since $\mathbf{Prob}(B) = 0$.

Instead we look at

$$B_\varepsilon = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \;\middle|\; y \in [y_{\text{meas}}, y_{\text{meas}} + \varepsilon] \right\}$$

and take the limit as $\varepsilon \to 0$



Conditional pdf as a limit

$$\mathbf{Prob}(A \mid B) = \lim_{\varepsilon \to 0} \mathbf{Prob}(A \mid B_\varepsilon)$$

$$= \lim_{\varepsilon \to 0} \frac{\displaystyle\int_{y_{\text{meas}}}^{y_{\text{meas}}+\varepsilon} \int_{a_1}^{a_2} p(x,y)\,dx\,dy}{\displaystyle\int_{y_{\text{meas}}}^{y_{\text{meas}}+\varepsilon} \int_{-\infty}^{\infty} p(x,y)\,dx\,dy}$$

$$= \lim_{\varepsilon \to 0} \frac{\varepsilon \displaystyle\int_{a_1}^{a_2} p(x, y_{\text{meas}})\,dx \;+\; \text{terms of order } \varepsilon^2 \text{ or higher}}{\varepsilon \displaystyle\int_{-\infty}^{\infty} p(x, y_{\text{meas}})\,dx \;+\; \text{terms of order } \varepsilon^2 \text{ or higher}}$$

$$= \int_{a_1}^{a_2} \frac{p(x, y_{\text{meas}})}{p^y(y_{\text{meas}})}\,dx$$

where p^y is the marginal pdf of y.

Conditional pdf

We write this as

$$\mathbf{Prob}\big(x \in [a_1, a_2] \,|\, y = y_{\text{meas}}\big) = \int_{a_1}^{a_2} \frac{p(x, y_{\text{meas}})}{p^y(y_{\text{meas}})}\, dx$$

We define the conditional pdf $p^{|y}$ of x given $y = y_{\text{meas}}$ by

$$\mathbf{Prob}\big(x \in [a_1, a_2] \,|\, y = y_{\text{meas}}\big) = \int_{a_1}^{a_2} p^{|y}(x, y_{\text{meas}})\, dx$$

Since this holds for all a_1, a_2, we must have

$$\boxed{p^{|y}(x, y_{\text{meas}}) = \frac{p(x, y_{\text{meas}})}{p^y(y_{\text{meas}})}}$$

Again, we can think of the denominator as simply normalizing the pdf.

Conditional mean and covariance

The *conditional mean* of x given y is

$$\mathbf{E}(x \mid y = y_{\text{meas}}) = \int x\, p^{|y}(x, y_{\text{meas}})\, dx$$

This is a *function of y_{meas}*

The *conditional covariance* of x given y is

$$\mathbf{cov}(x \mid y = w) = \mathbf{E}\Big(\big(x - f(w)\big)\big(x - f(w)\big)^T \,\Big|\, y = w \Big)$$

Here $f(w) = \mathbf{E}(x \mid y = w)$. The conditional covariance is also a function of y_{meas}

Conditional notation

- conditional expectation defines a function; e.g.,

$$f(w) = \mathbf{E}(x \,|\, y = w)$$

  defines a function $f : \mathbb{R}^m \to \mathbb{R}^n$

- We often take expectations of the form

$$\mathbf{E}\big(f(y)\big)$$

  For the above f we have $\mathbf{E}\big(f(y)\big) = \mathbf{E}\, x$

- This is often written

$$\mathbf{E}\big(\mathbf{E}(x \,|\, y)\big)$$

Conditional notation

- Another common notation

$$h(w) = \mathbf{cov}(x \mid y = w)$$

- Again, you see

$$\mathbf{E}\,\mathbf{trace}\,\mathbf{cov}(x \mid y)$$

which means $\mathbf{E}\,\mathbf{trace}\big(h(y)\big)$

Conditional pdf for a Gaussian

Suppose $x \sim \mathcal{N}(0, \Sigma)$, and

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Suppose we measure $x_2 = y$. We would like to find the conditional pdf of x_1 given $x_2 = y$

- Is it Gaussian?
- What is the *conditional mean* $\mathbf{E}\big(x_1 \,|\, x_2 = y\big)$ of x_1 given x_2
- What is the *conditional covariance* $\mathbf{cov}(x_1 \,|\, x_2 = y)$ of x_1 given x_2?

Conditional pdf for a Gaussian

The pdf of x is

$$p^x(x) = c_1 \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right)$$

By the completion of squares formula

$$\Sigma^{-1} = \begin{bmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{bmatrix} \begin{bmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix}$$

Hence

$$x^T \Sigma^{-1} x = (x_1 - Lx_2)^T T^{-1} (x_1 - Lx_2) + x_2^T \Sigma_{22}^{-1} x_2$$

where

$$T = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \qquad L = \Sigma_{12}\Sigma_{22}^{-1}$$

Conditional pdf for a Gaussian

The conditional pdf of x_1 given x_2 is therefore

$$p^{|x_2}(x_1, y) = \frac{c_1}{p^{x_2}(y)} \exp\left(-\frac{1}{2}(x_1 - Ly)^T T^{-1}(x_1 - Ly) - \frac{1}{2} y^T \Sigma_{22}^{-1} y \right)$$

$$= \frac{c_1}{p^{x_2}(y)} \exp\left(-\frac{1}{2} y^T \Sigma_{22}^{-1} y \right) \exp\left(-\frac{1}{2}(x_1 - Ly)^T T^{-1}(x_1 - Ly) \right)$$

Hence $p^{|x_2}(x_1, y)$ is *Gaussian*

$$p^{|x_2}(x_1, y) = c_2(y) \exp\left(-\frac{1}{2}(x_1 - Ly)^T T^{-1}(x_1 - Ly) \right)$$

where $c_2(y)$ is such that $\int p^{|x_2}(x_1, y)\, dx_1 = 1$

Conditional pdf for a Gaussian

- If $x \sim \mathcal{N}(0, \Sigma)$, then the conditional pdf of x_1 given $x_2 = y$ is *Gaussian*

- The *conditional mean* is

$$\boxed{\mathbf{E}(x_1 \,|\, x_2 = y) = \Sigma_{12}\Sigma_{22}^{-1} y}$$

  It is a *linear function* of y

- The *conditional covariance* is

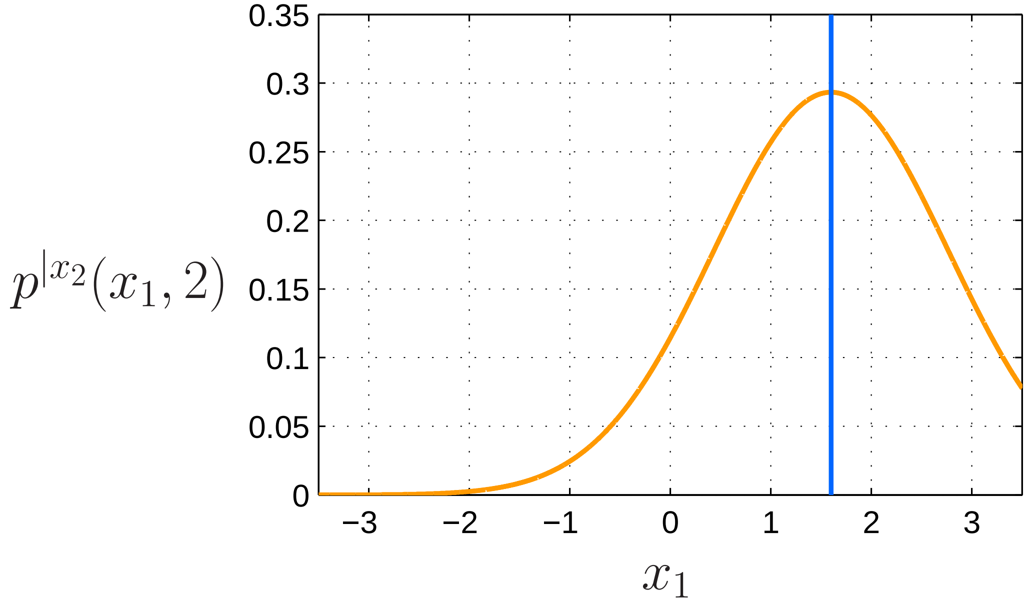
$$\boxed{\mathbf{cov}(x_1 \,|\, x_2 = y) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}$$

  It is *not* a function of y

Conditional pdf for a Gaussian

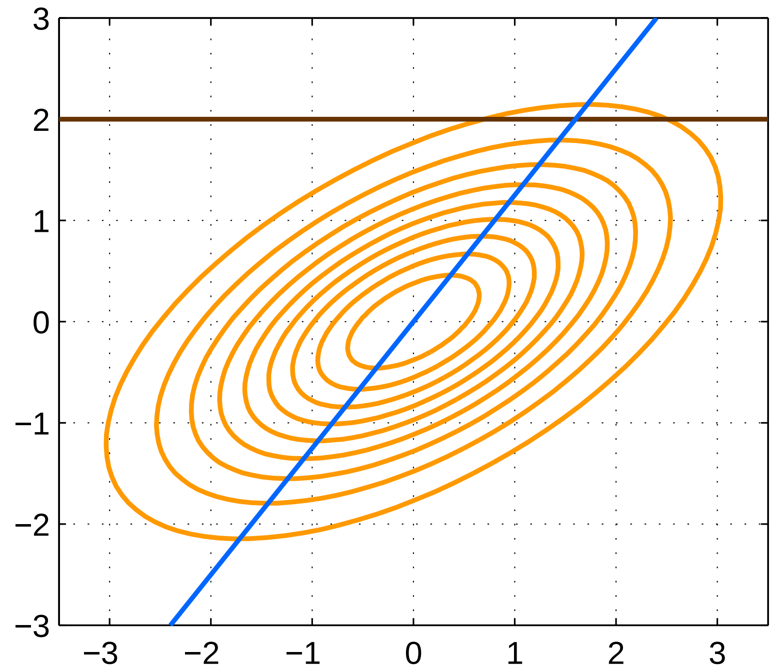
- Here

$$\mathbf{cov}(x) = \begin{bmatrix} 2 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$



- We have

$$L = 0.8 \qquad T = 1.36$$



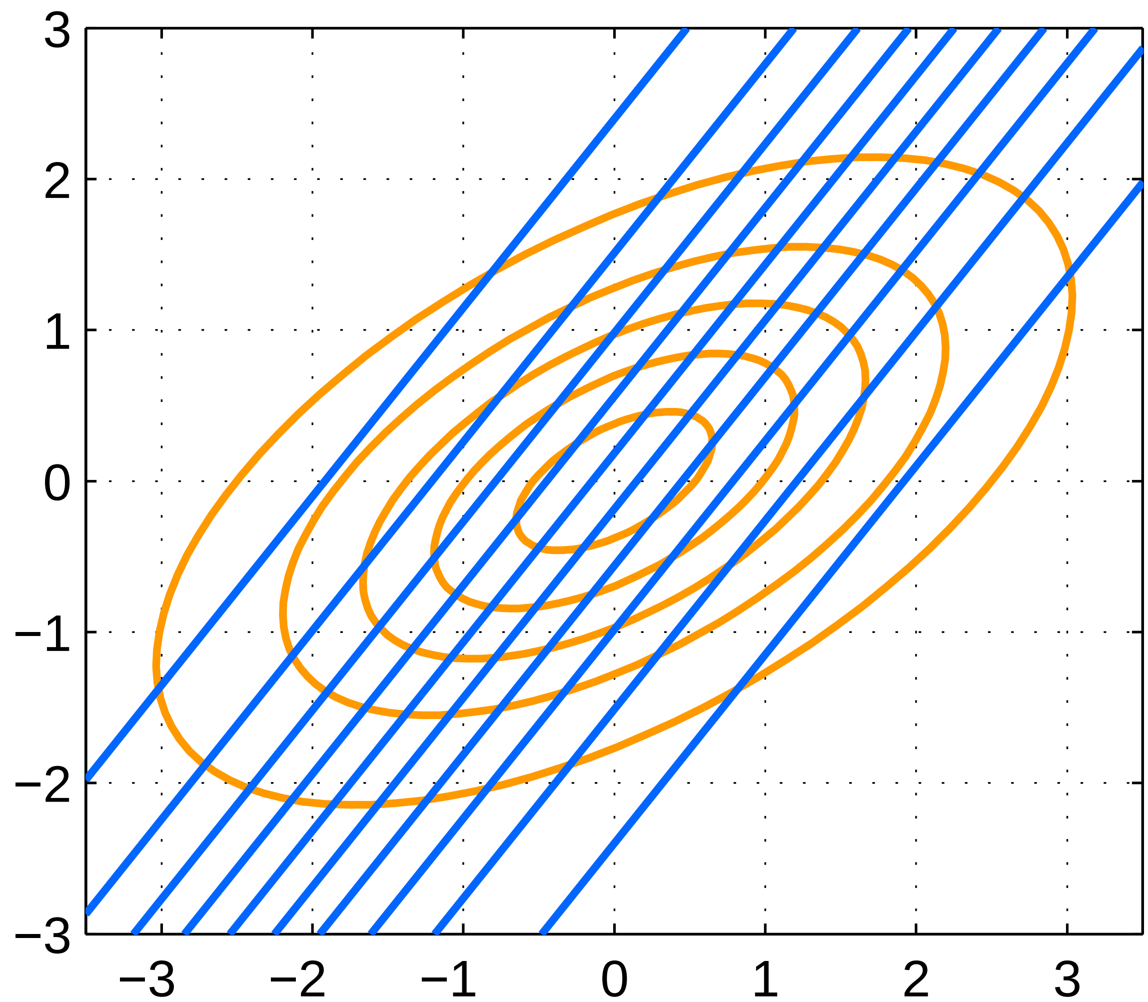
- Hence

$$\mathbf{E}(x_1 \,|\, x_2 = 2) = 1.6$$

$$\mathbf{cov}(x_1 \,|\, x_2 = 2) = 0.8$$

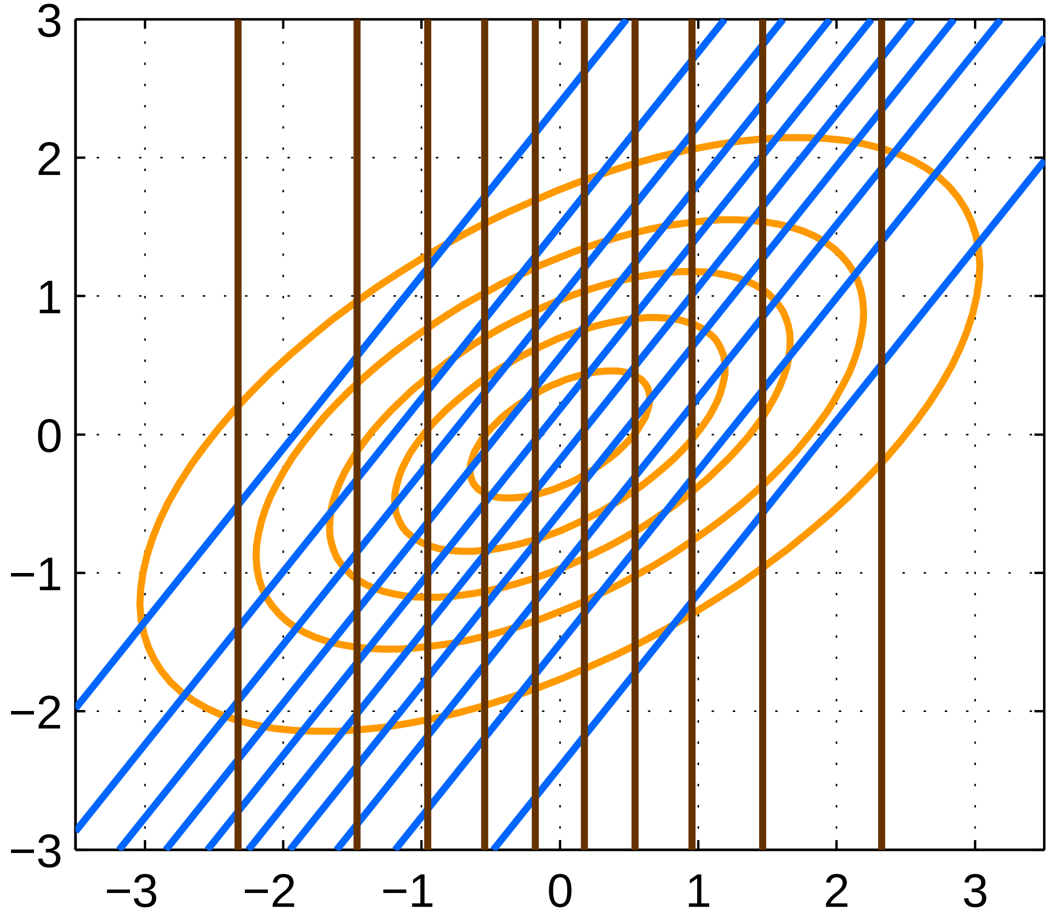
Conditional pdf for a Gaussian

- Here $\Sigma = \begin{bmatrix} 2 & 0.8 \\ 0.8 & 1 \end{bmatrix}$
- The contours of the *conditional pdf* $p^{|x_2}(x_1, x_2)$ as a function of x_1 and x_2 are shown in blue.
- Both confidence ellipsoids and conditional contours correspond to confidence levels of $0.1, 0.3, 0.5, 0.7, 0.9$
- The conditional pdf is constant on lines $x_1 = \Sigma_{12}\Sigma_{22}^{-1}x_2 + a$



Conditional confidence intervals

Compare with the marginal pdf confidence intervals



Notice that conditional confidence intervals are narrower, since $T = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ hence $T \leq \Sigma_{11}$, i.e., measuring x_2 gives *information* about x_1

Independence

suppose $x : \Omega \to s\mathbb{R}^n$ is a random variable with induced pdf $p^x : \mathbb{R}^n \to \mathbb{R}$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

each component x_i is a random variable, with *marginal pdf* $p^{x_i} : \mathbb{R} \to \mathbb{R}$ given by integrating over all other components, e.g. in $\mathbb{R}^3$,

$$p^{x_3}(x_3) = \int_{x_2=-\infty}^{\infty} \int_{x_1=-\infty}^{\infty} p^x(x)\, dx_1 dx_2$$

random variables $x_1, \dots, x_n \in \mathbb{R}$ are called *independent* if

$$p^x(x) = p^{x_1}(x_1) p^{x_2}(x_2) \dots p^{x_n}(x_n)$$

Adding independent random variables

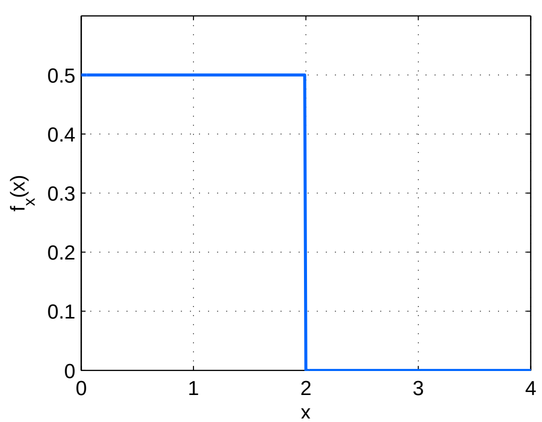
$x, y \in \mathbb{R}$ are *independent* random variables, with induced pdf's p^x and p^y

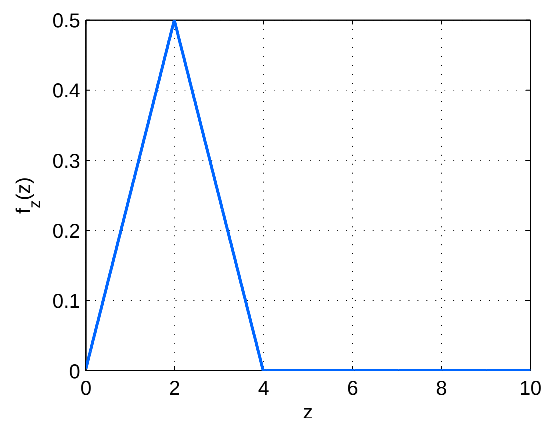
let $z = x + y$; the induced pdf of z is the *convolution* of p^x and p^y

$$p^z(z) = \int_{-\infty}^{\infty} p^x(x) p^y(z - x) \, dx$$

example: convolution of a uniform pdf with itself

if x, y are both uniform on $[0, 2]$, and independent, let $z = x + y$





the distribution of the sum of two uniform rv's is not uniform.

Uncorrelated Gaussians

Uncorrelated Gaussians are independent

Suppose $x : \Omega \to \mathbb{R}^n$. If $x \sim \mathcal{N}(\mu, \Sigma)$ and $\Sigma_{12} = 0$, then x_1 and x_2 are independent.

To see this, notice that

$$
\begin{aligned}
p^x(x_1, x_2) &= c_1 \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right) \\
&= c_1 \exp\left(-\frac{1}{2} x_1^T \Sigma_{11}^{-1} x_1\right) \exp\left(-\frac{1}{2} x_2^T \Sigma_{22}^{-1} x_2\right) \\
&= p^{x_1}(x_1) p^{x_2}(x_2)
\end{aligned}
$$

where p^{x_1} and p^{x_2} are Gaussian pdfs.

Example: correlation

Suppose $x \sim \mathcal{N}(0, \Sigma)$ where

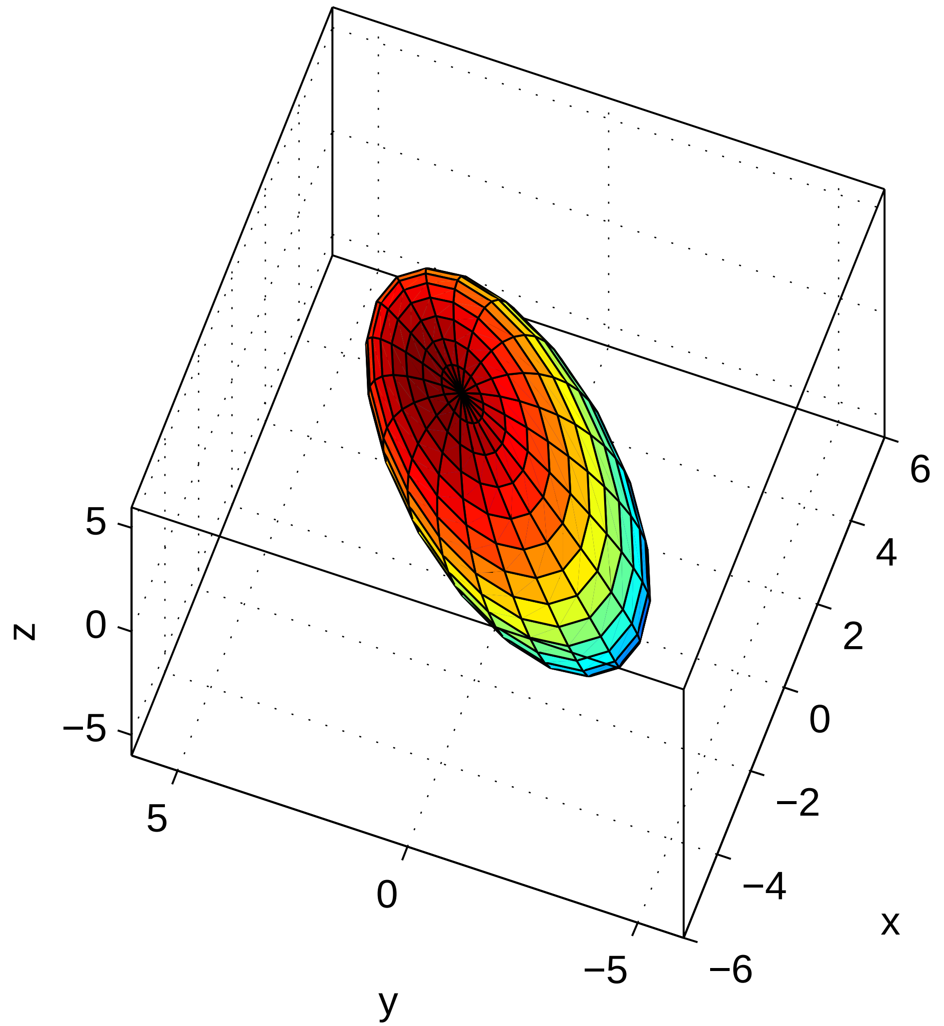
$$\Sigma = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Then

- x and y are correlated
- z and y are correlated
- z and x are uncorrelated
- The eigenvalues of Σ are $0.59, 2, 3.4$

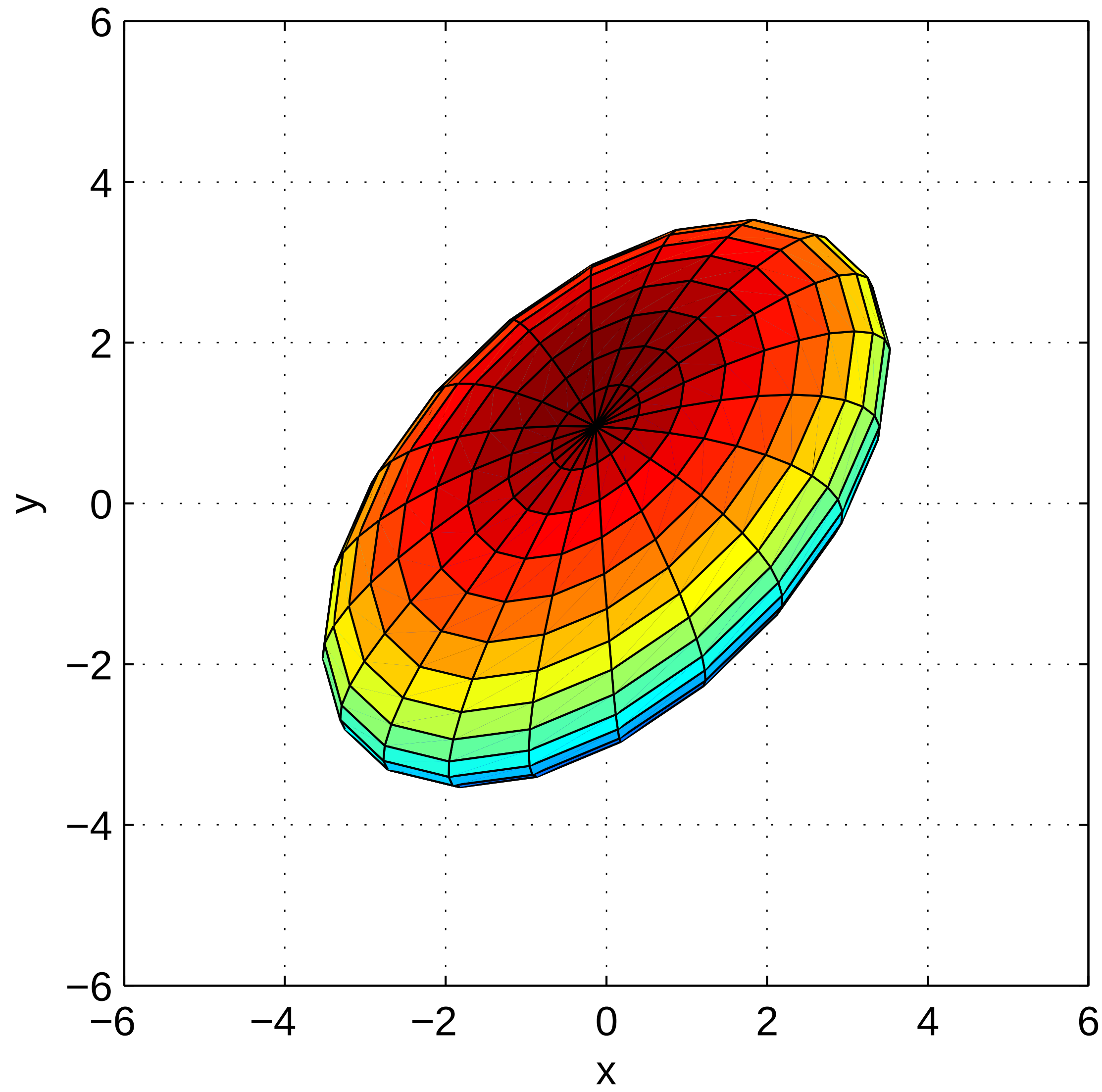
Example: correlation

The 90% confidence ellipsoid is



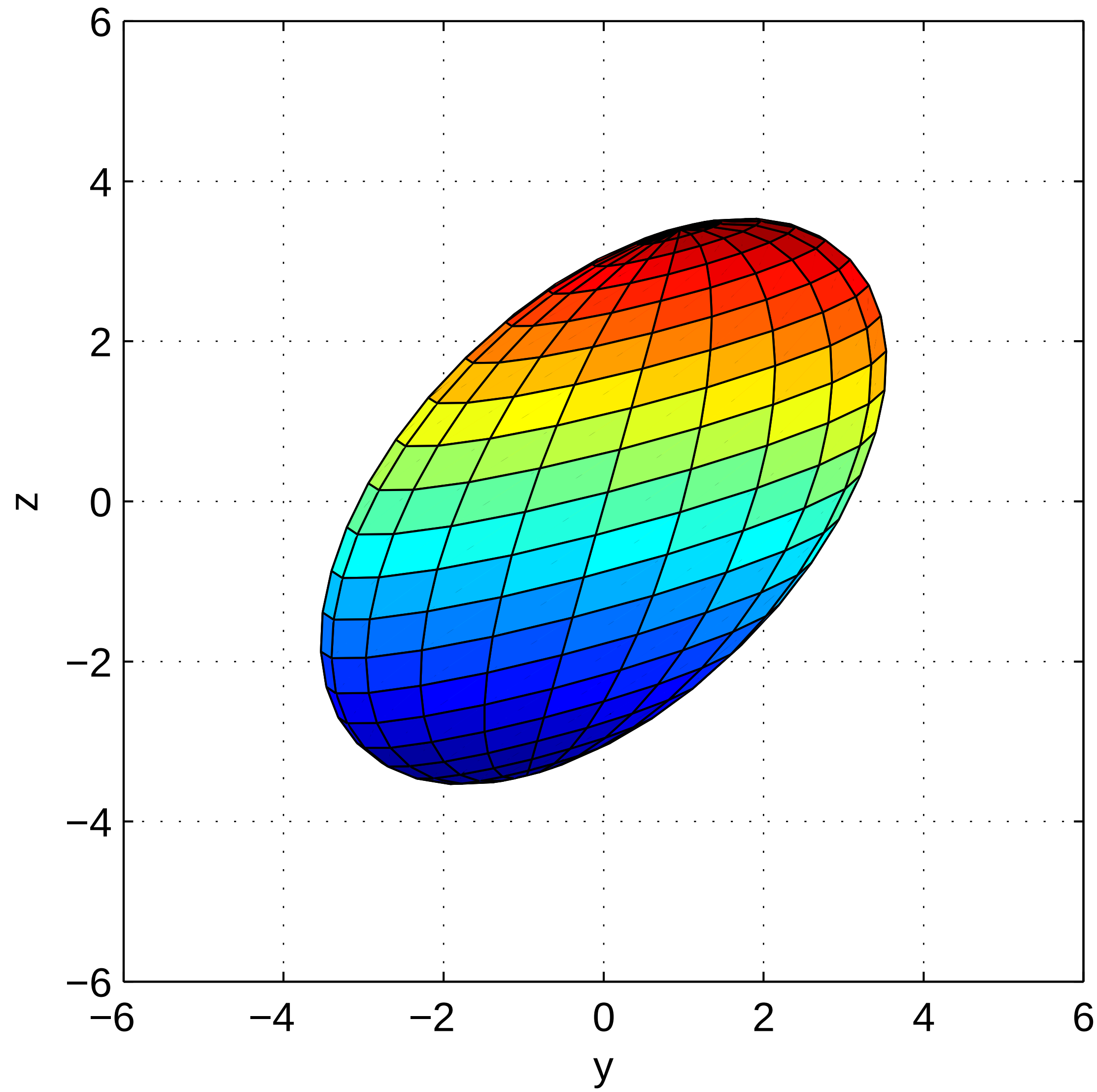
Example: correlation

The 90% confidence ellipsoid is



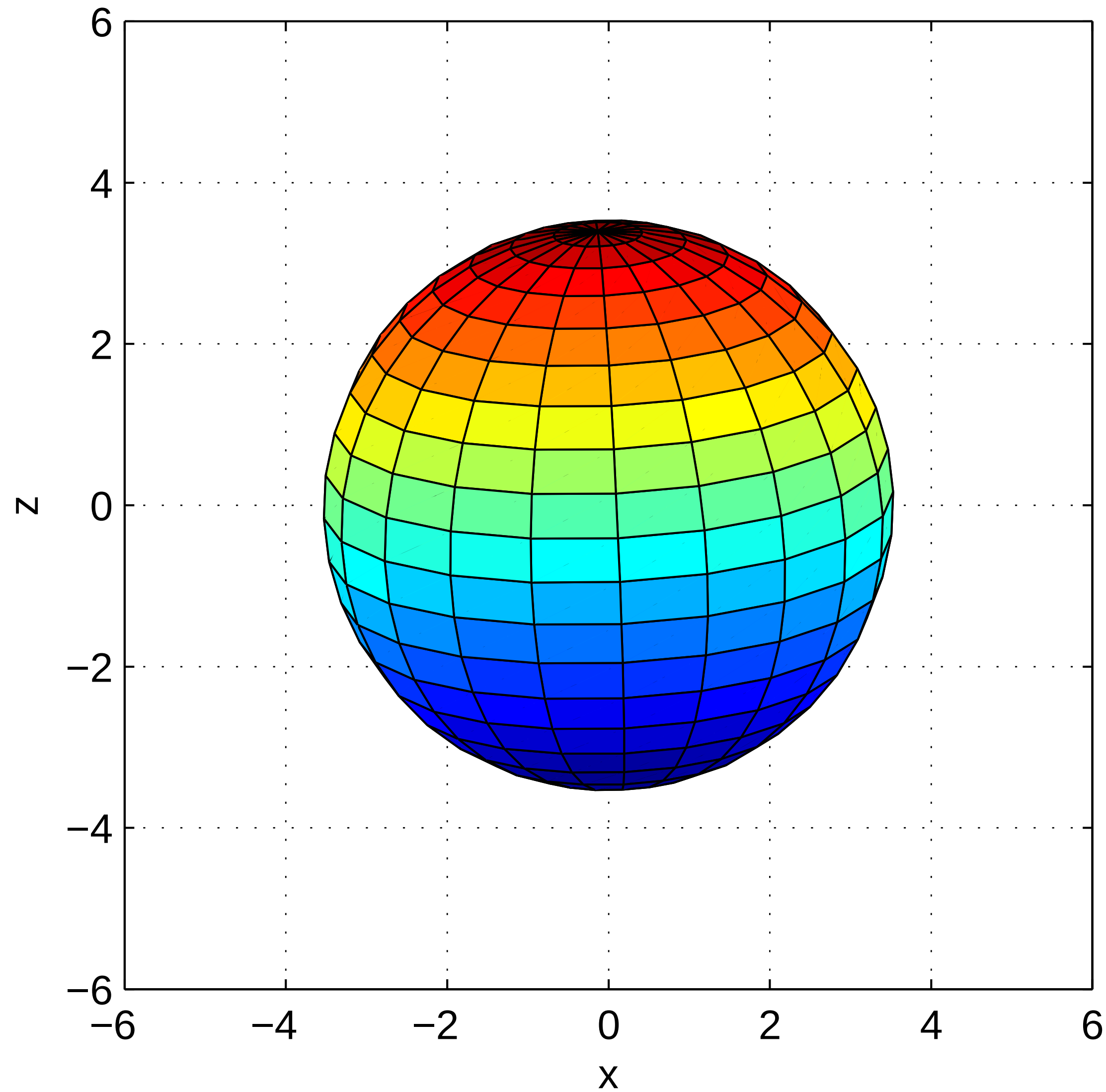
Example: correlation

The 90% confidence ellipsoid is



Example: correlation

The 90% confidence ellipsoid is



example: adding Gaussian random vectors

suppose $x \in \mathbb{R}^2$, and $x \sim \mathcal{N}(\mu, \Sigma)$; let

$$\begin{aligned} z &= \begin{bmatrix} 1 & 1 \end{bmatrix} x \\ &= x_1 + x_2 \end{aligned}$$

then

$$\mathbf{E}\, z = \mu_1 + \mu_2$$

and

$$\mathbf{cov}(z) = \begin{bmatrix} 1 & 1 \end{bmatrix} \Sigma \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

if $\Sigma = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}$ is diagonal, so x_1 and x_2 are independent, then

$$\mathbf{cov}(z) = \Sigma_{11} + \Sigma_{22}$$