8 - Continuous random vectors

- Mean-square deviation
- Mean-variance decomposition
- Gaussian random vectors
- The Gamma function
- The χ^2 distribution
- Confidence ellipsoids
- Marginal density functions
- Example: marginal pdfs for Gaussians
- Degenerate Gaussian random vectors
- Changes of variables for random vectors
- Linear transformations of Gaussians
- Example: Gaussian force on mass
- Distributions and densities in Matlab

Continuous random vectors

The random vector $x : \Omega \to \mathbb{R}^n$ has *induced pdf* $p^x : \mathbb{R}^n \to \mathbb{R}$.

For any subset $A \subset \mathbb{R}^n$, we have

$$\mathbf{Prob}(x \in A) = \int_A p^x(x)\, dx$$

the *mean* or *expected value* of x is

$$\mathbf{E}(x) = \int_{\mathbb{R}^n} x p^x(x)\, dx$$

the *covariance* of x is

$$\mathbf{cov}(x) = \mathbf{E}\big((x-\mu)(x-\mu)^T\big) = \int_{\mathbb{R}^n} (x-\mu)(x-\mu)^T p^x(x)\, dx$$

Mean-square deviation

Suppose $x : \Omega \to \mathbb{R}^n$ is a random variable, with mean μ.

The *mean square deviation from the mean* is given by

$$\mathbf{E}\big(\|x-\mu\|^2\big) = \mathbf{trace}\,\mathbf{cov}(x)$$

Because

$$\begin{aligned}
\mathbf{E}\big(\|x-\mu\|^2\big) &= \mathbf{E}\big((x-\mu)^T(x-\mu)\big) \\
&= \mathbf{E}\,\mathbf{trace}\big((x-\mu)^T(x-\mu)\big) \\
&= \mathbf{E}\,\mathbf{trace}\big((x-\mu)(x-\mu)^T\big) && \text{since } \mathbf{trace}(AB) = \mathbf{trace}(BA) \\
&= \mathbf{trace}\,\mathbf{E}\big((x-\mu)(x-\mu)^T\big) && \text{since } \mathbf{E}\,Ax = A\,\mathbf{E}\,x
\end{aligned}$$

The mean-variance decomposition

The *mean square* of a random variable $x : \Omega \to \mathbb{R}^n$ is

$$\mathbf{E}\big(\|x\|^2\big) = \mathbf{trace}\big(\mathbf{cov}(x)\big) + \|\mathbf{E}\,x\|^2$$

This holds because

$$\begin{aligned}
\mathbf{E}\big(\|x\|^2\big) &= \mathbf{E}\big(\|x - \mu + \mu\|^2\big) \\
&= \mathbf{E}\big(\|x - \mu\|^2 + 2\mu^T(x - \mu) + \|\mu\|^2\big) \\
&= \mathbf{E}\big(\|x - \mu\|^2\big) + 2\mu^T\,\mathbf{E}(x - \mu) + \|\mu\|^2
\end{aligned}$$

Correlation and covariance

The *correlation matrix* of random vector x is

$$\boxed{\mathbf{corr}(x) = \mathbf{E}(xx^T)}$$

- If $\mathbf{E}\, x = 0$ then $\mathbf{corr}(x) = \mathbf{cov}(x)$
- The *mean square* of x is $\mathbf{E}\big(\|x\|^2\big) = \mathbf{trace}\,\mathbf{corr}(x)$

The *correlation-covariance decomposition* is

$$\boxed{\mathbf{corr}(x) = \mathbf{cov}(x) + (\mathbf{E}\, x)(\mathbf{E}\, x^T)}$$

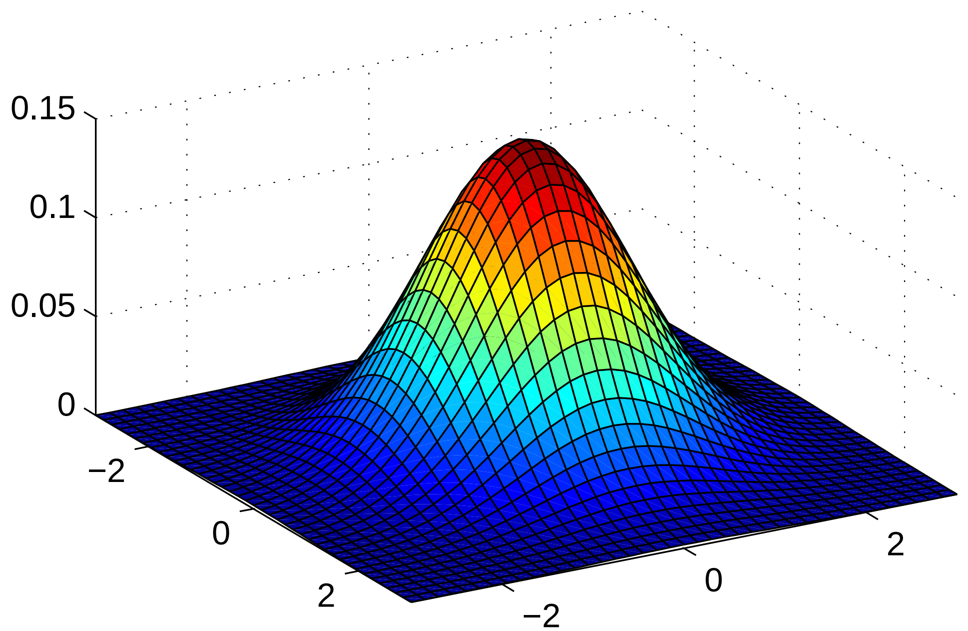
proof takes the same approach as the mean-variance formula

Gaussian random vectors

The random variable $x : \Omega \rightarrow \mathbb{R}^n$ is called *Gaussian* if it has induced pdf

$$p^x(x) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right)$$

write this as $x \sim \mathcal{N}(\mu, \Sigma)$, here $\Sigma = \Sigma^T$ and $\Sigma > 0$



Gaussian random vectors

Suppose $x \sim \mathcal{N}(\mu, \Sigma)$. Then

- The mean of x is

$$\mathbf{E}\, x = \mu$$

- The covariance of x is

$$\mathbf{cov}(x) = \Sigma$$

Ellipsoids

the Gaussian pdf is constant on the surface of the ellipsoids

$$S_\alpha = \left\{ x \in \mathbb{R}^n \mid (x-\mu)^T \Sigma^{-1} (x-\mu) \leq \alpha \right\}$$

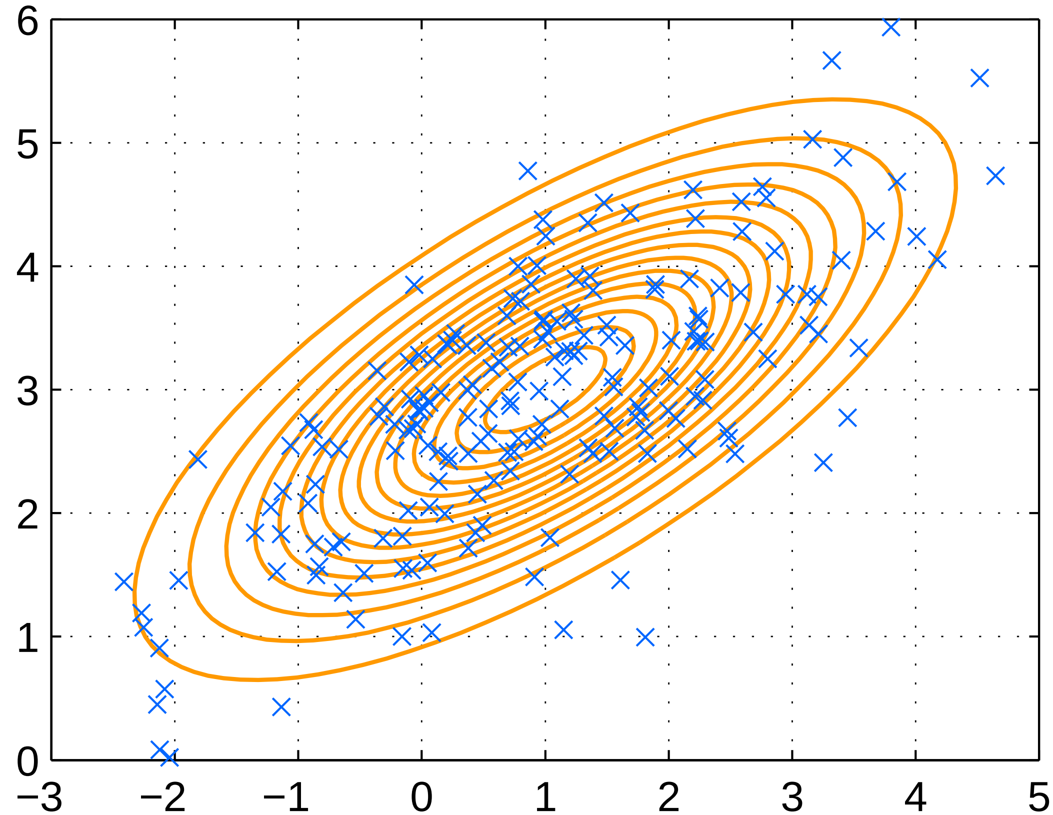
center is at μ, semiaxis lengths are $\sqrt{\alpha \lambda_i(\Sigma)}$.

Example:

$$\mu = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

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

contours at $p(x) = 0.01, 0.02, \ldots$



Gamma function

the *gamma function* is

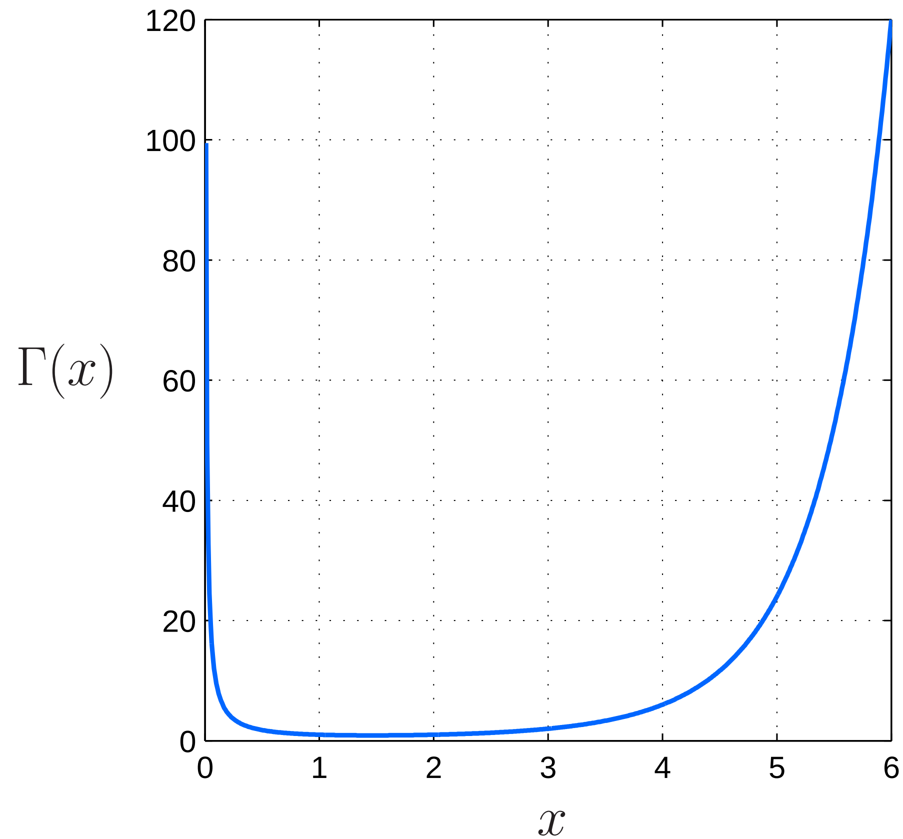
$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt \qquad \text{for } x > 0$$

for $x > 0$

$$\Gamma(x+1) = x\Gamma(x)$$

$\Gamma(1) = 1$, so for integer $x > 1$

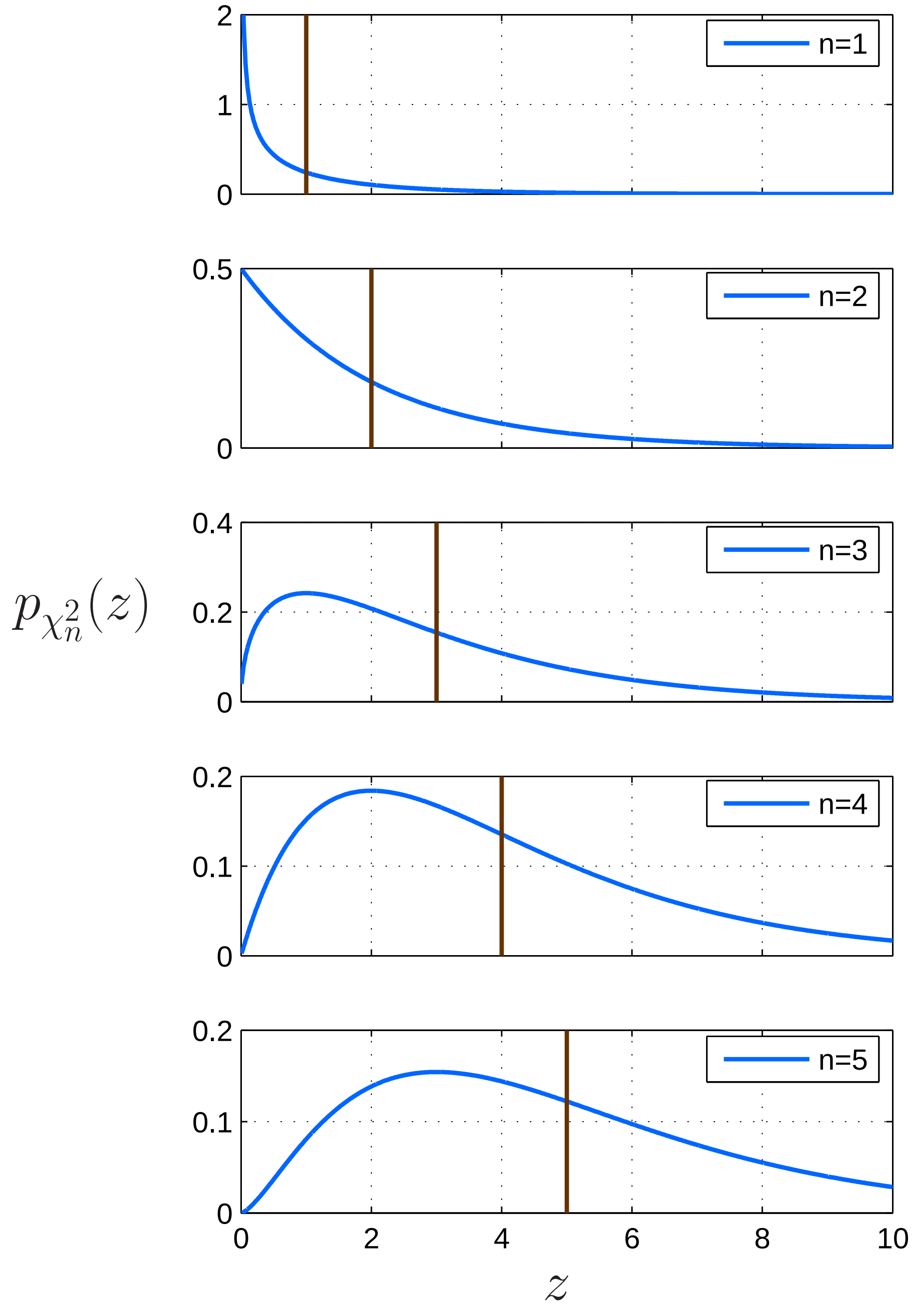
$$\Gamma(x) = (x-1)!$$



The χ^2 distribution

The χ^2_n probability density function is

$$p_{\chi^2_n}(z) = \frac{1}{2^{\frac{n}{2}}\Gamma(n/2)} z^{\frac{n}{2}-1} e^{-\frac{z}{2}}$$



- A family of pdfs, one for each $n > 0$
- If $z \sim \chi^2_n$, then $\mathbf{E}\, z = n$

Gaussian random vectors and confidence ellipsoids

Suppose $x : \Omega \to \mathbb{R}^n$ is Gaussian, i.e., $x \sim \mathcal{N}(\mu, \Sigma)$. Define the random variable

$$z = (x-\mu)^T \Sigma^{-1} (x-\mu)$$

which is a measure of the distance of x from μ

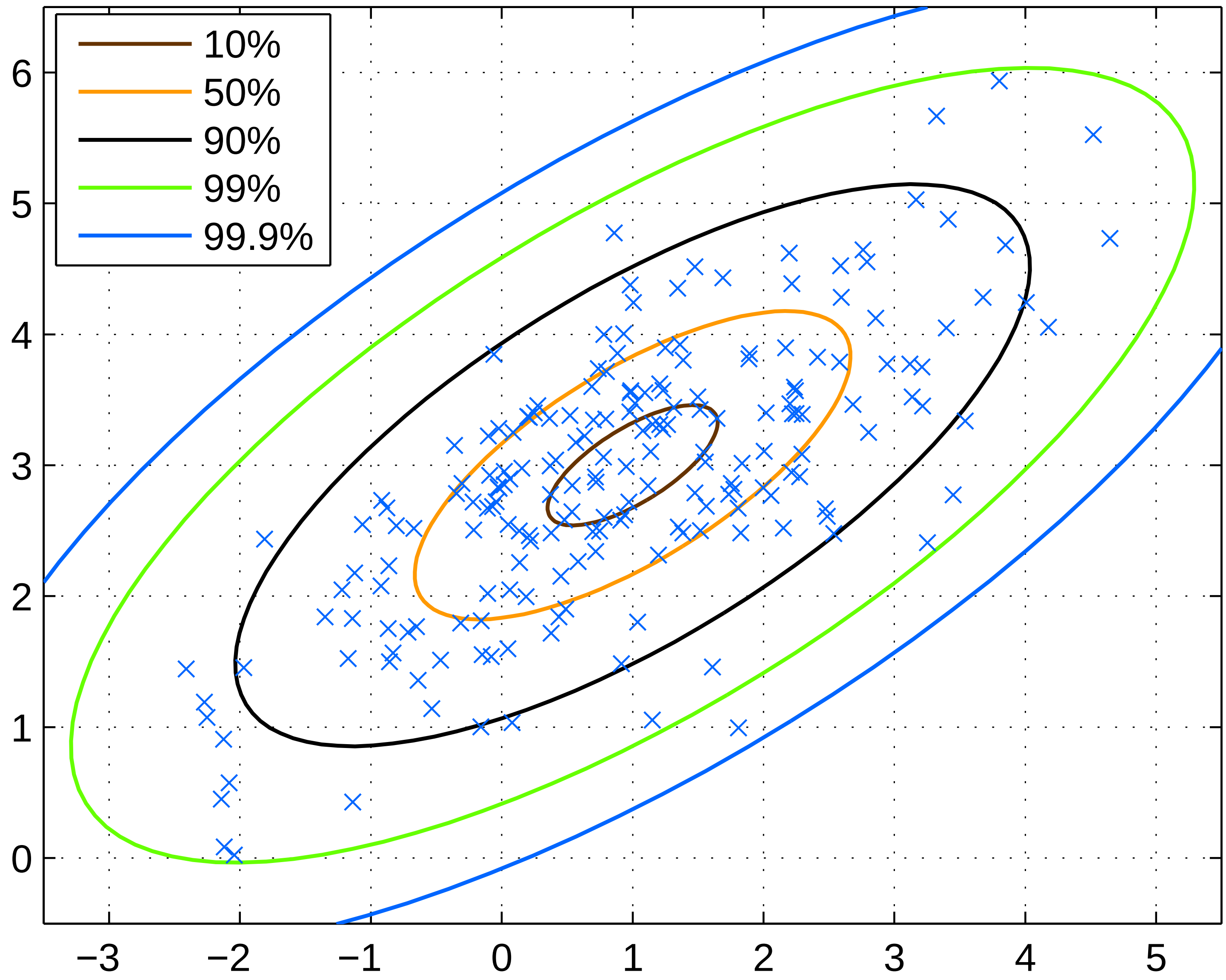
- z has a χ_n^2 distribution
- Hence prob. that x lies in the ellipsoid $S_\alpha = \left\{ x \in \mathbb{R}^n \;\middle|\; (x-\mu)^T \Sigma^{-1} (x-\mu) \leq \alpha \right\}$

$$\boxed{\mathbf{Prob}(x \in S_\alpha) = F_{\chi_n^2}(\alpha)}$$

- for example $F_{\chi_n^2}(\alpha) \approx \begin{cases} \frac{1}{2} & \text{if } \alpha = n \\ 0.9 & \text{if } \alpha = n + 2\sqrt{n} \end{cases}$ 90% confidence ellipsoid

Confidence ellipsoids

The plot shows the confidence ellipsoids and 200 sample points.



Marginal probability density functions

Suppose $x : \Omega \to \mathbb{R}^n$ is an RV with pdf $p^x : \mathbb{R}^n \to \mathbb{R}$, and $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, where $x_1 \in \mathbb{R}^r$.

Define the *marginal pdf* of x_1 to be the function p^{x_1} such that

$$\mathbf{Prob}(x_1 \in W) = \int_W p^{x_1}(z)\,dz \qquad \text{for all } W \subset R^r$$

We also know that

$$\mathbf{Prob}(x_1 \in W) = \int_W \int_{x_2 \in \mathbb{R}^{n-r}} p^x(x_1, x_2)\,dx_2\,dx_1$$

Since these are equal, we have

$$\boxed{p^{x_1}(x_1) = \int_{x_2 \in \mathbb{R}^{n-r}} p^x(x_1, x_2)\,dx_2}$$

The marginal pdf of a Gaussian

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

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

Let's look at the component x_1

- Since $x_1 = \begin{bmatrix} I & 0 \end{bmatrix} x$, we have the mean

$$\mathbf{E}\, x_1 = \begin{bmatrix} I & 0 \end{bmatrix} \mu = \mu_1$$

  and also the covariance

$$\mathbf{cov}(x_1) = \begin{bmatrix} I & 0 \end{bmatrix} \Sigma \begin{bmatrix} I \\ 0 \end{bmatrix} = \Sigma_{11}$$

- In fact, the random variable x_1 is *Gaussian*; this is not obvious

Proof: the marginal pdf of a Gaussian

Assume for convenience that $\mathbf{E}\, x = 0$. The marginal pdf of x_1 is

$$p^{x_1}(x_1) = \int_{x_2} c_1 \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) dx_2$$

We have, by the completion of squares formula

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

and so, setting $S = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^T \Sigma_{11}^{-1} x_1 \ + \ (x_2 - \Sigma_{21}\Sigma_{11}^{-1} x_1)^T S^{-1} (x_2 - \Sigma_{21}\Sigma_{11}^{-1} x_1)$$

Proof: the marginal pdf of a Gaussian

Hence we have

$$p^{x_1}(x_1) = c_1 \exp\left(-\frac{1}{2}x_1^T \Sigma_{11}^{-1} x_1\right) \int_{x_2} \exp\left(-\frac{1}{2}(x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1)^T S^{-1}(x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1)\right) dx_2$$

$$= c_2 \exp\left(-\frac{1}{2}x_1^T \Sigma_{11}^{-1} x_1\right)$$

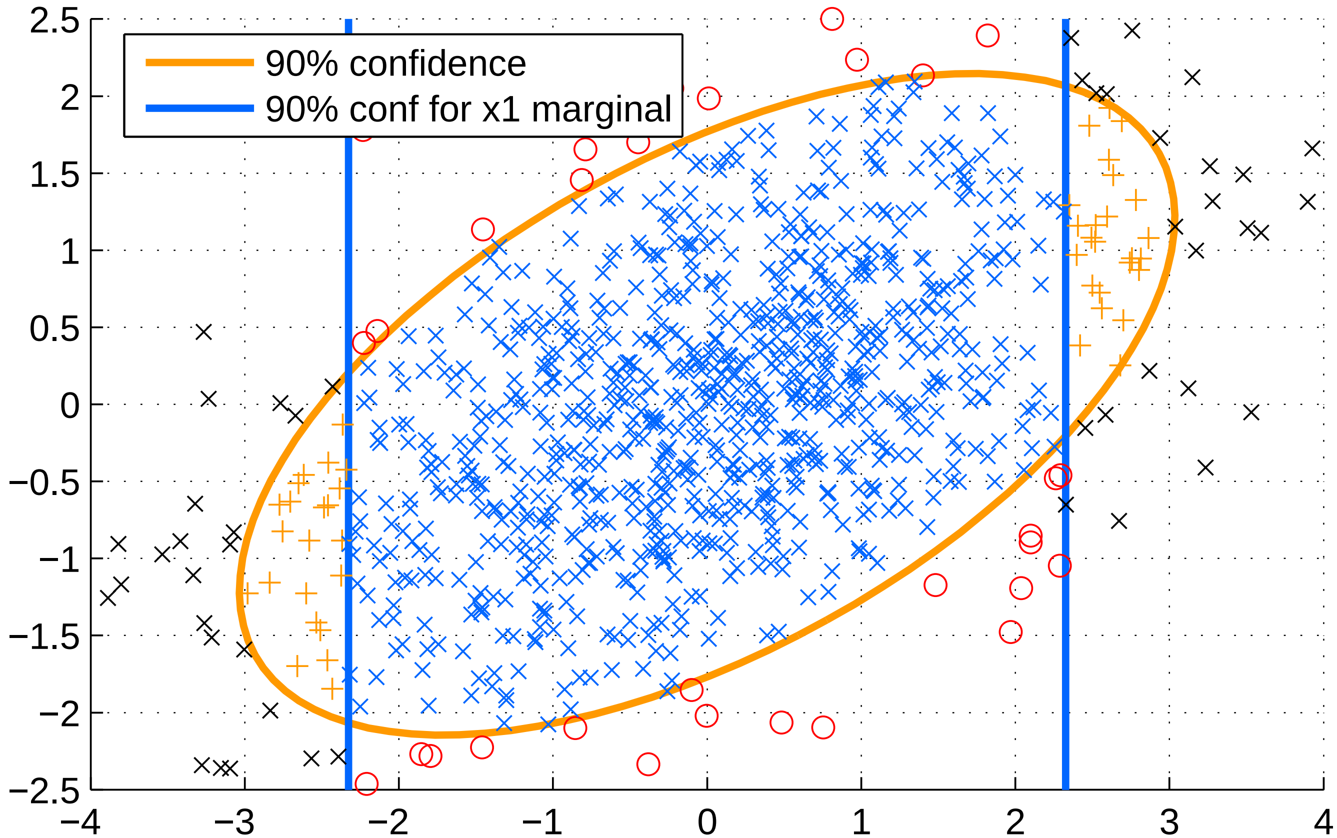
Now c_2 is determined, because $\displaystyle\int p^{x_1}(z)\, dz = 1$, so we don't need to calculate it explicitly.

Therefore, if $x \sim \mathcal{N}(0, \Sigma)$ the marginal pdf of x_1 is *Gaussian*, and

$$x_1 \sim \mathcal{N}(0, \Sigma_{11})$$

Example: marginal pdf for Gaussians

Suppose $\Sigma = \begin{bmatrix} 2 & 0.8 \\ 0.8 & 1 \end{bmatrix}$ and $x \sim \mathcal{N}(0, \Sigma)$. A simulation of 1000 points is below



- all blue and orange points (908) are within 90% confidence ellipsoid for x
- all blue and red points (899) are within 90% confidence interval for x_1

Degenerate Gaussian random vectors

- it's convenient to allow Σ singular, but still $\Sigma = \Sigma^T$ and $\Sigma \geq 0$
  this means that in some directions, *x is not random at all*

- obviously density formula does not hold; instead write

$$\Sigma = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^T$$

  where $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ is orthogonal, and $\Sigma_1 > 0$

  columns of Q_1 are orthonormal basis for $\mathbf{range}(\Sigma)$
  columns of Q_2 are orthonormal basis for $\mathbf{null}(\Sigma)$

- let $\begin{bmatrix} z \\ w \end{bmatrix} = Q^T x$; then

$$z \sim \mathcal{N}(Q_1^T \mu, \Sigma_1) \text{ is non-degenerate Gaussian}$$

$$w = Q_2^T \mu \text{ is not random}$$

Changes of variables for random vectors

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is continuous, and $h : \mathbb{R}^n \to \mathbb{R}^n$ satisfies

- h is one-to-one and onto; i.e., h is invertible
- Both h and h^{-1} are differentiable, with continuous derivative

The derivative of h at x is $Dh(x)$, the *Jacobian* matrix

$$\bigl(Dh(x)\bigr)_{ij} = \frac{\partial h_i}{\partial x_j}(x)$$

Then for any $A \subset \mathbb{R}^n$

$$\boxed{\int_{h(A)} f(x)\, dx = \int_A f\bigl(h(y)\bigr) \left|\det Dh(y)\right| dy}$$

Changes of variables for random vectors

Suppose $x : \Omega \to \mathbb{R}^n$ is a *random vector*, and $y = g(x)$, where g is invertible, and g and g^{-1} are continuously differentiable. Then

$$p^y(y) = \frac{p^x\bigl(g^{-1}(y)\bigr)}{\bigl|\det(Dg)\bigl(g^{-1}(y)\bigr)\bigr|}$$

As in the scalar case, this holds because

$$\begin{aligned}
\mathbf{Prob}(y \in A) &= \int_A p^y(y)\,dy \\
&= \int_{g^{-1}(A)} p^x(x)\,dx \\
&= \int_A \frac{p^x\bigl(g^{-1}(y)\bigr)}{\bigl|\det(Dg)\bigl(g^{-1}(y)\bigr)\bigr|}\,dy
\end{aligned}$$

where $D(g^{-1})(y) = \Bigl((Dg)\bigl(g^{-1}(y)\bigr)\Bigr)^{-1}$

Example: linear transformations

Consider $y = Ax + b$, where $A \in \mathbb{R}^{n \times n}$ is *invertible*. Then

$$p^y(y) = \frac{p^x\big(A^{-1}(y-b)\big)}{|\det A|}$$

Linear transformations of Gaussians

a linear function of a Gaussian random vector is a Gaussian random vector

Suppose $x \sim \mathcal{N}(\mu_x, \Sigma_x)$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Consider the linear function of x

$$y = Ax + b$$

- we already know how means and covariances transform; we have

$$\mathbf{E}(y) = A\,\mathbf{E}\,x + b \qquad\qquad \mathbf{cov}(y) = A\,\mathbf{cov}(x)A^T$$

- The amazing fact is that y is *Gaussian*

Linear transformations of Gaussians

To show this, first suppose $A \in \mathbb{R}^{n\times n}$ is *invertible*. Let $\mu_y = A\mu_x + b$ and $\Sigma_y = A\Sigma_x A^T$.

We know

$$p^x(x) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det \Sigma_x)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma_x^{-1}(x-\mu)\right)$$

So

$$p^y(y) = \frac{p^x\big(A^{-1}(y-b)\big)}{|\det A|}$$

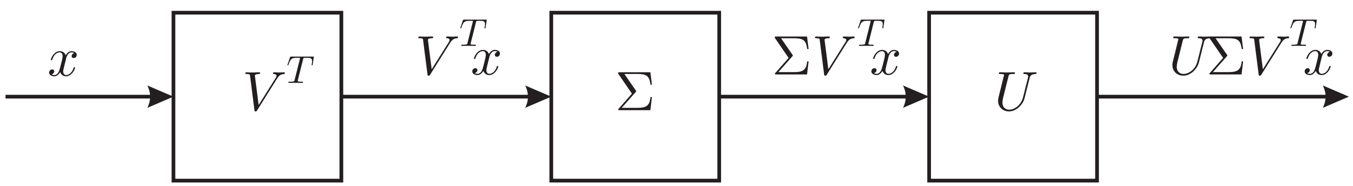
$$= \frac{1}{|\det A|(2\pi)^{\frac{n}{2}}(\det \Sigma_x)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y-b-A\mu_x)^T (A^{-1})^T \Sigma_x^{-1} A^{-1}(y-b-A\mu_x)\right)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}(\det \Sigma_y)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y-\mu_y)^T \Sigma_y^{-1}(y-\mu_y)\right)$$

Non-invertible linear transformations of Gaussians

Suppose $A \in \mathbb{R}^{m \times n}$, and $y = Ax$ where $x \sim \mathcal{N}(0, \Sigma_x)$. The SVD of A is

$$A = U \Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$



This decomposes the map into

$$y = Uw \qquad \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \Sigma z \qquad \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = V^T x$$

Non-invertible linear transformations of Gaussians

Since V is invertible, we know $z \sim \mathcal{N}(0, \Sigma_z)$, where

$$\Sigma_z = \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \Sigma_x \begin{bmatrix} V_1 & V_2 \end{bmatrix}$$

We know z is Gaussian, hence the marginal z_1 is Gaussian

$$z_1 \sim \mathcal{N}(0, V_1^T \Sigma_x V_1)$$

Also $w_2 = 0$, and since Σ_1 is invertible, w_1 is Gaussian

$$w_1 \sim \mathcal{N}(0, \Sigma_1 V_1^T \Sigma_x V_1 \Sigma_1)$$

Since $w = U^T y$, we have y is a *degenerate Gaussian random vector* where

- $w_1 = U_1^T y$ are the components of y that are Gaussian
- $w_2 = 0$ are the components of y that are not random

Full-rank case

When $\mathbf{range}(A) = \mathbb{R}^m$, i.e., A is full row rank, we have

$$y \sim \mathcal{N}(0, A\Sigma_x A^T)$$

Because the SVD of A is

$$A = U \begin{bmatrix} \Sigma_1 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

Then $y = Uw_1$, and since U is invertible, we have

$$y \sim \mathcal{N}(0, U\Sigma_1 V_1^T \Sigma_x V_1 \Sigma_1 U^T)$$

Example: simulating Gaussian random vectors

In Matlab, its easy to generate $x \sim \mathcal{N}(0, I)$

```
x=randn(n,1)
```

to generate $y \sim \mathcal{N}(\mu, \Sigma)$, we can use

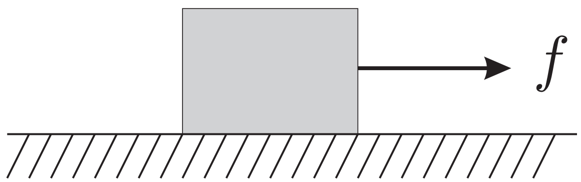
$$y = \Sigma^{\frac{1}{2}} x + \mu$$

extremely useful for simulation

Example: Gaussian random force on mass

- x is the sequence of applied forces, so $f(t) = x_j$ for t in the interval $[j-1, j]$.
- y_1, y_2 are final position and velocity
- $y = Ax$ where $A = \begin{bmatrix} 9.5 & 8.5 & 7.5 & 6.5 & 5.5 & 4.5 & 3.5 & 2.5 & 1.5 & 0.5 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$
- suppose the forces are Gaussian, and the vector $x \sim \mathcal{N}(0, \Sigma)$, where

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



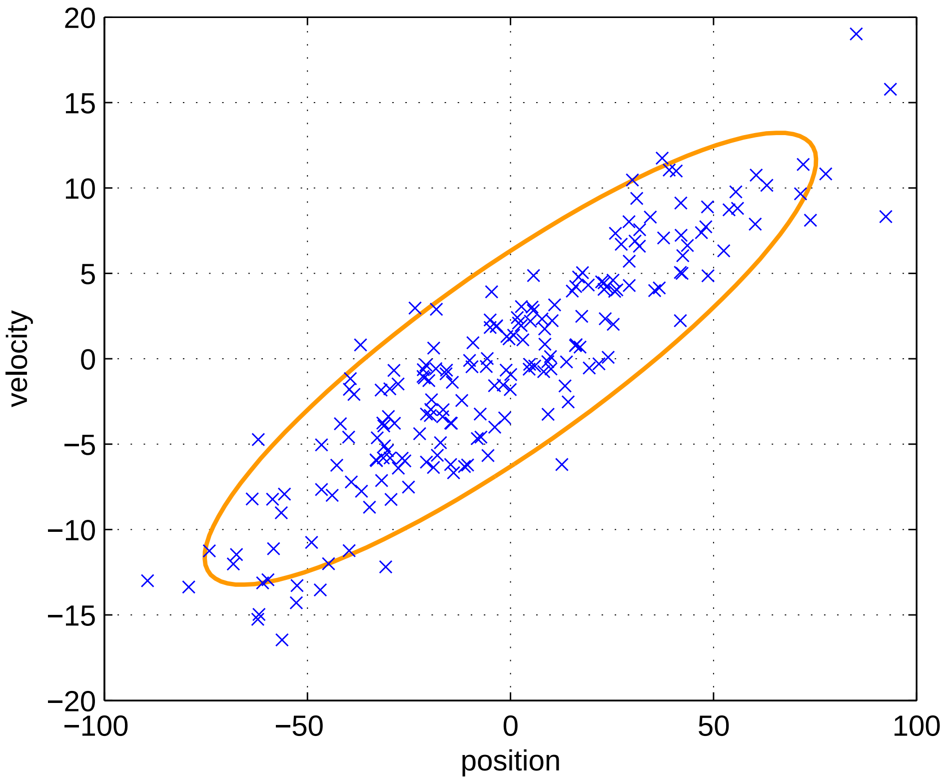
Example: Gaussian random force on mass

the covariance of y is

$$\Sigma_y = A \Sigma A^T$$

the 90% confidence ellipsoid is

$$\left\{ y \in \mathbb{R}^2 \mid y^T \Sigma_y^{-1} y \leq F_{\chi_n^2}^{-1}(0.9) \right\}$$



Components of a Gaussian random vector

Suppose $x : \Omega \to \mathbb{R}^n$ and $x \sim \mathcal{N}(0, \Sigma)$, and let $c \in \mathbb{R}^n$ be a unit vector

Let $y = c^T x$

- y is the component of x in the direction c
- y is Gaussian, with $\mathbf{E}\, y = 0$ and $\mathbf{cov}(y) = c^T \Sigma c$
- So $\mathbf{E}(y^2) = c^T \Sigma c$
- The unit vector c that minimizes $c^T \Sigma c$ is the eigenvector of Σ with the smallest eigenvalue. Then

$$\mathbf{E}(y^2) = \lambda_{\text{min}}$$

Distributions and densities in Matlab

Matlab has useful functions in the statistics toolbox:

`chi2pdf`	Chi square density
`normpdf`	Gaussian density
`chi2cdf`	Chi square cdf
`normcdf`	Gaussian cdf
`chi2inv`	Chi square inverse cdf
`norminv`	Gaussian inverse cdf

as well as `gamma` and `erf`