CS395T
Computational Statistics with
Application to Bioinformatics

Prof. William H. Press
Spring Term, 2011
The University of Texas at Austin

Lecture 9

The University of Texas at Austin, CS 395T, Spring 2011, Prof. William H. Press

1

Multivariate Normal Distributions

Generalizes Normal (Gaussian) to M-dimensions
Like 1-d Gaussian, completely defined by its mean and (co-)variance
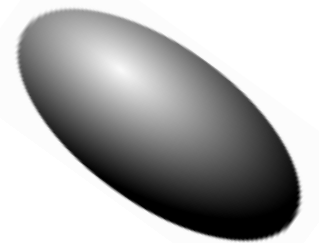Mean is a M-vector, covariance is a M x M matrix

$$N(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{M/2}\det(\boldsymbol{\Sigma})^{1/2}}\exp[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]$$

The mean and covariance of r.v.'s from this distribution are*

$$\boldsymbol{\mu} = \langle\mathbf{x}\rangle \qquad \boldsymbol{\Sigma} = \langle(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\rangle$$

In the one-dimensional case σ is the standard deviation, which can be visualized as "error bars" around the mean.

In more than one dimension Σ can be visualized as an error ellipsoid around the mean in a similar way.

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

*really?

The University of Texas at Austin, CS 395T, Spring 2011, Prof. William H. Press

2

Because mean and covariance are easy to estimate from a data set, it is easy – perhaps too easy – to fit a multivariate normal distribution to data.

$$\boldsymbol{\mu} = \langle \mathbf{x} \rangle \approx \frac{1}{N} \sum_i \mathbf{x}_i \qquad \boldsymbol{\Sigma} = \left\langle (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \right\rangle \approx \frac{1}{N} \sum_i (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T$$

I.e., estimate by sample averages.

But back to "really?" The mean follows from the symmetry argument

$$0 = \int \cdots \int (\mathbf{x} - \boldsymbol{\mu}) \frac{1}{(2\pi)^{M/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp[-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})] \, d^M \mathbf{x}$$

It's *not* obvious that the covariance in fact obtains from the definition of the multivariate Normal. One has to do the multidimensional (and tensor) integral:

$$\mathbf{M}_2 = \int \cdots \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \frac{1}{(2\pi)^{M/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp[-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})] \, d^M \mathbf{x}$$

The University of Texas at Austin, CS 395T, Spring 2011, Prof. William H. Press

3

The only way I know how to do this integral is by trickery involving the Cholesky decomposition ("square root of a positive definite matrix"):

$$\Sigma = \mathbf{L}\mathbf{L}^T \ (\text{Cholesky}), \quad \Sigma^{-1} = (\mathbf{L}^T)^{-1}\mathbf{L}^{-1}, \quad \mathbf{L}\mathbf{y} \equiv \mathbf{x}$$

we're setting μ to 0 for convenience

$$p(\mathbf{y}) = p(\mathbf{x}) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|$$

Jacobian determinant. The transformation law for multivariate probability distributions.

$$= \frac{\det(\mathbf{L})}{(2\pi)^{N/2}\det(\Sigma)^{1/2}} \exp[-\tfrac{1}{2}(\mathbf{y}^T\mathbf{L}^T)(\mathbf{L}^{T-1}\mathbf{L}^{-1})(\mathbf{L}\mathbf{y})]$$

$$= \prod_i (2\pi)^{-1/2} \exp(-\tfrac{1}{2}y_i^2)$$

This is the distribution of N independent univariate Normals N(0,1)!

$$\langle \mathbf{x}\mathbf{x}^T \rangle = \langle \mathbf{L}\mathbf{y}\mathbf{y}^T\mathbf{L}^T \rangle = \mathbf{L}\langle \mathbf{y}\mathbf{y}^T \rangle \mathbf{L}^T = \mathbf{L}\mathbf{L}^T = \Sigma \quad \text{Ha!}$$

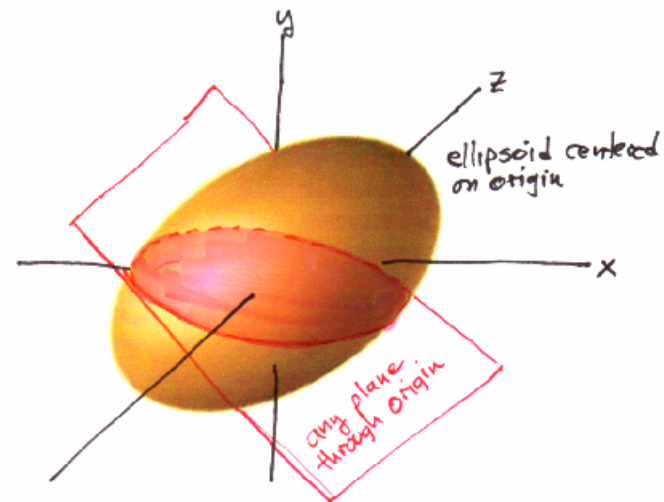(I don't know an elementary proof, i.e., without some matrix decomposition. Can you find one?)

The University of Texas at Austin, CS 395T, Spring 2011, Prof. William H. Press

4

Reduced dimension properties of multivariate normal

$$N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{M/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp[-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})]$$

1. Any slice through a m.v.n. is a m.v.n ("constraint" or "conditioning")

2. Any projection of a m.v.n. is a m.v.n ("marginalization")

You can prove both assertions by "completing the square" in the exponential, producing an exponential in (only) the reduced dimension times an exponential in (only) the lost dimensions. Then the second exponential is either constant (slice case) or can be integrated over (projection case).

ellipsoid centered on origin

any plane through origin

The University of Texas at Austin, CS 395T, Spring 2011, Prof. William H. Press

5

How to generate multivariate normal deviates $N(\mu,\Sigma)$:

Cholesky: $\quad \Sigma = \mathbf{L}\mathbf{L}^T$

Fill $\mathbf{y}$ with independent Normals: $\quad \mathbf{y} = \{y_i\} \sim \mathrm{N}(0,1)$

Transform: $\quad \mathbf{x} = \mathbf{L}\mathbf{y} + \boldsymbol{\mu}$ $\quad$ That's it! x is the desired m.v.n.

Proof: $\quad \langle \mathbf{y}\mathbf{y}^T \rangle = \mathbf{1}$

$$\langle (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \rangle = \langle (\mathbf{L}\boldsymbol{y})(\mathbf{L}\boldsymbol{y})^T \rangle$$
$$= \langle \mathbf{L}(\mathbf{y}\mathbf{y}^T)\mathbf{L}^T \rangle = \mathbf{L} \langle \mathbf{y}\mathbf{y}^T \rangle \mathbf{L}^T$$
$$= \mathbf{L}\mathbf{L}^T = \Sigma$$

Even easier: MATLAB has a built-in function mvnrnd(MU,SIGMA). But be sure you get a bunch of m.v.n.'s all in one call, because it (probably) re-does the Cholesky decomposition on each call!

Notice that the proof never used Normality. You can fill $\mathbf{y}$ with anything with zero mean and variance one, and you'll reproduce Σ. But the result won't be Normal!

The University of Texas at Austin, CS 395T, Spring 2011, Prof. William H. Press

6

So, easy operations are:

1. Fitting a multivariate normal to a set of points (just compute the sample mean and covariance!)

2. Sampling from the fitted m.v.n.

```
mu  = mean([len1 len2])
sig = cov(len1,len2)
mu  =
    3.2844
    3.2483
sig =
    0.6125    0.2476
    0.2476    0.5458
rsamp = mvnrnd(mu,sig,1000);
```

In MATLAB, for example, these are one-line operations.

Example:
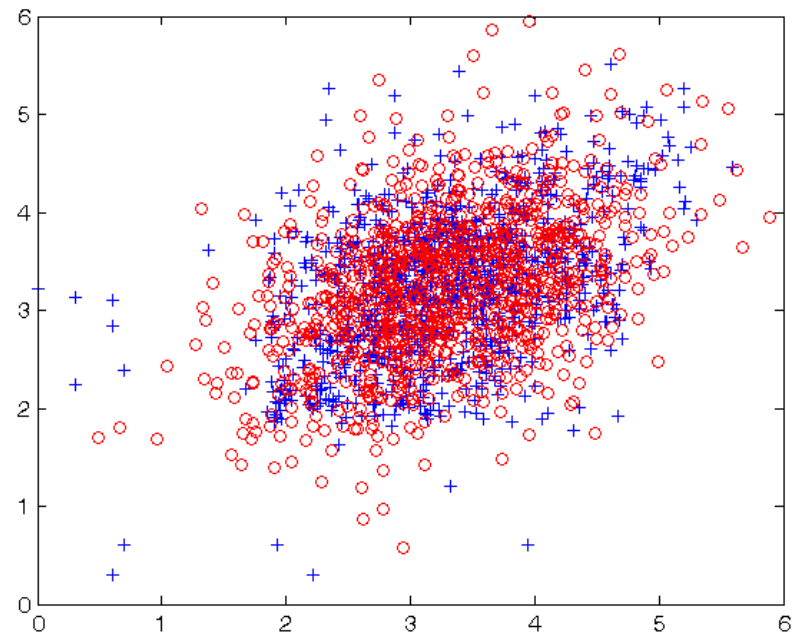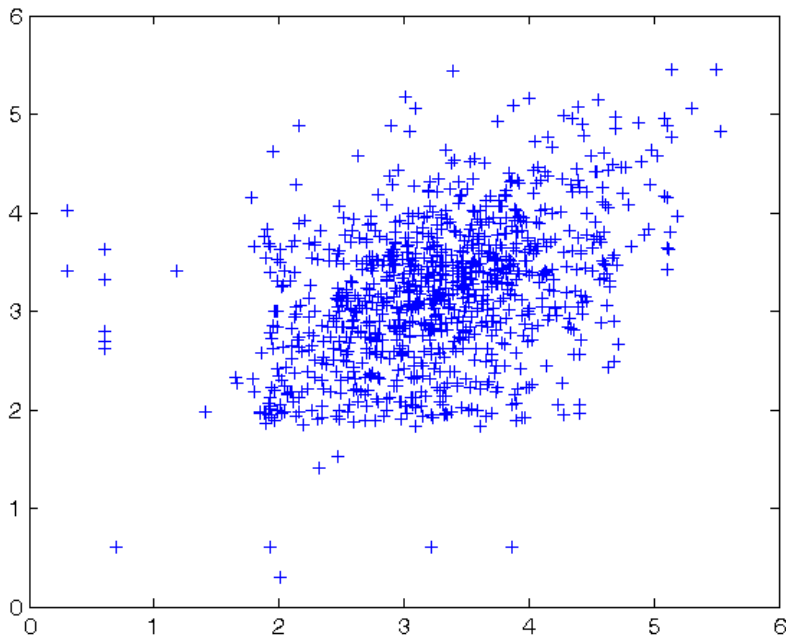
A related, useful, Cholesky trick is to draw error ellipses (ellipsoids, …)

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T$$

So, locus of points at 1 standard deviation is

$$1 = (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad \Rightarrow \quad \left| \mathbf{L}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right| = 1$$
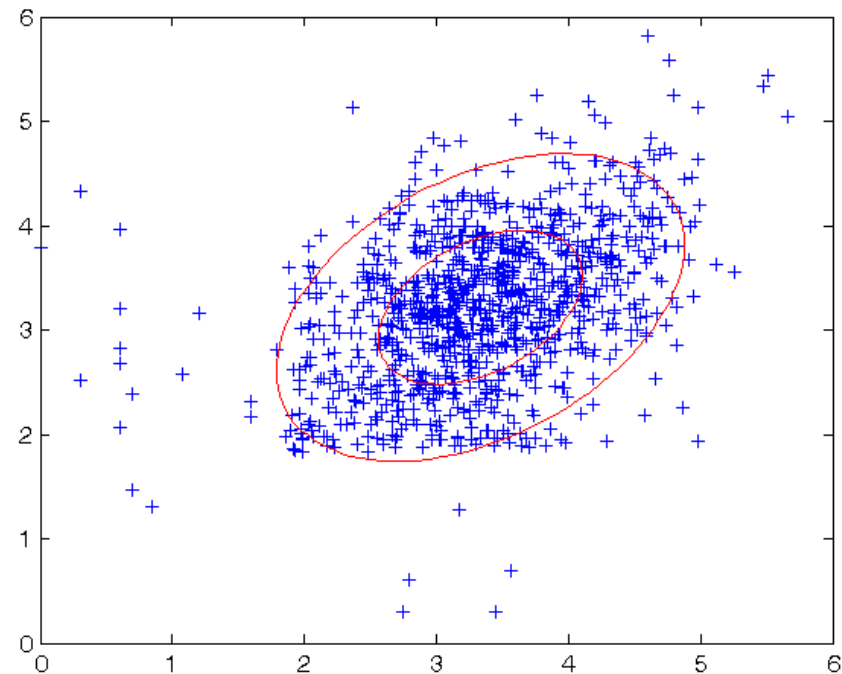
So, if z is on the unit circle (sphere, …) then
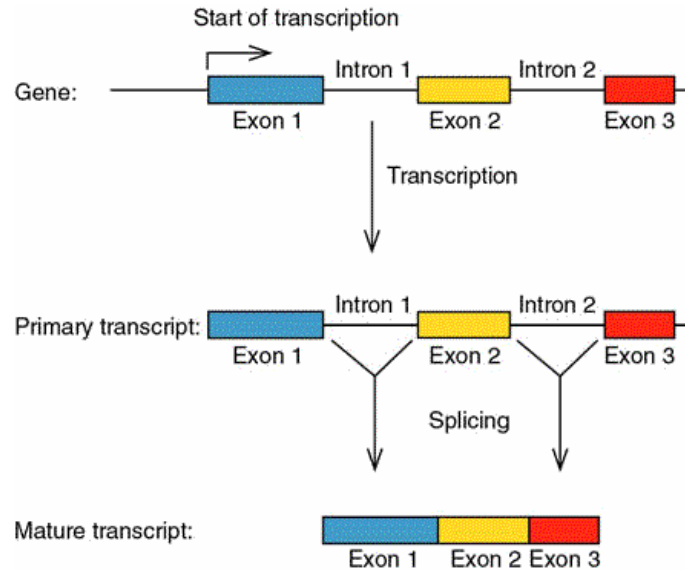
$$\mathbf{x} = \mathbf{L}\mathbf{z} + \boldsymbol{\mu}$$

will be on the error ellipse.

my coding of this idea looks like this

```
function [x y] = errorellipse(mu,sigma,stdev,n)
L = chol(sigma,'lower');
circle =
 [cos(2*pi*(0:n)/n); sin(2*pi*(0:n)/n)].*stdev;
ellipse = L*circle + repmat(mu,[1,n+1]);
x = ellipse(1,:);
y = ellipse(2,:);
```


The University of Texas at Austin, CS 395T, Spring 2011, Prof. William H. Press

8

The distribution we have been looking at has some interesting biology in it!

file genestats.dat (on course web site) contains 20694 lines like this:

```
ENST00000341866     17470      3262 0.00002 4 1290 349 1412 211 169 678 13361 <EOL>
ENST00000314348     22078      1834 0.00001 7 100 166 113 178 165 262 850 5475
385 3273 1149 2070 7892
ENST00000313081     13858      1160 0.00001 6 496 150 107 85 151 171 2068 76 2063
674 7817
ENST00000298622     80000      6487 0.00001 24 135 498 216 120 147 132 36 60 129
129 84 63 99 99 54 66 69 78 204 66 73 1081 397 2452 12133 15737 1513 769 942
103 829 2272 1340 3058 327 2371 1361 471 2922 735 85 9218 1257 2247 897 822
12104
```

number of exons N

N exon lengths

N-1 intron lengths

total length

ignore for now

gene name

total length of exons

The University of Texas at Austin, CS 395T, Spring 2011, Prof. William H. Press

9

Log$_{10}$ of size of 1st and 2nd introns for 1000 genes:

This is kind of fun, because it's not just the usual featureless scatter plot

notice the "hard edges"
this is biology!
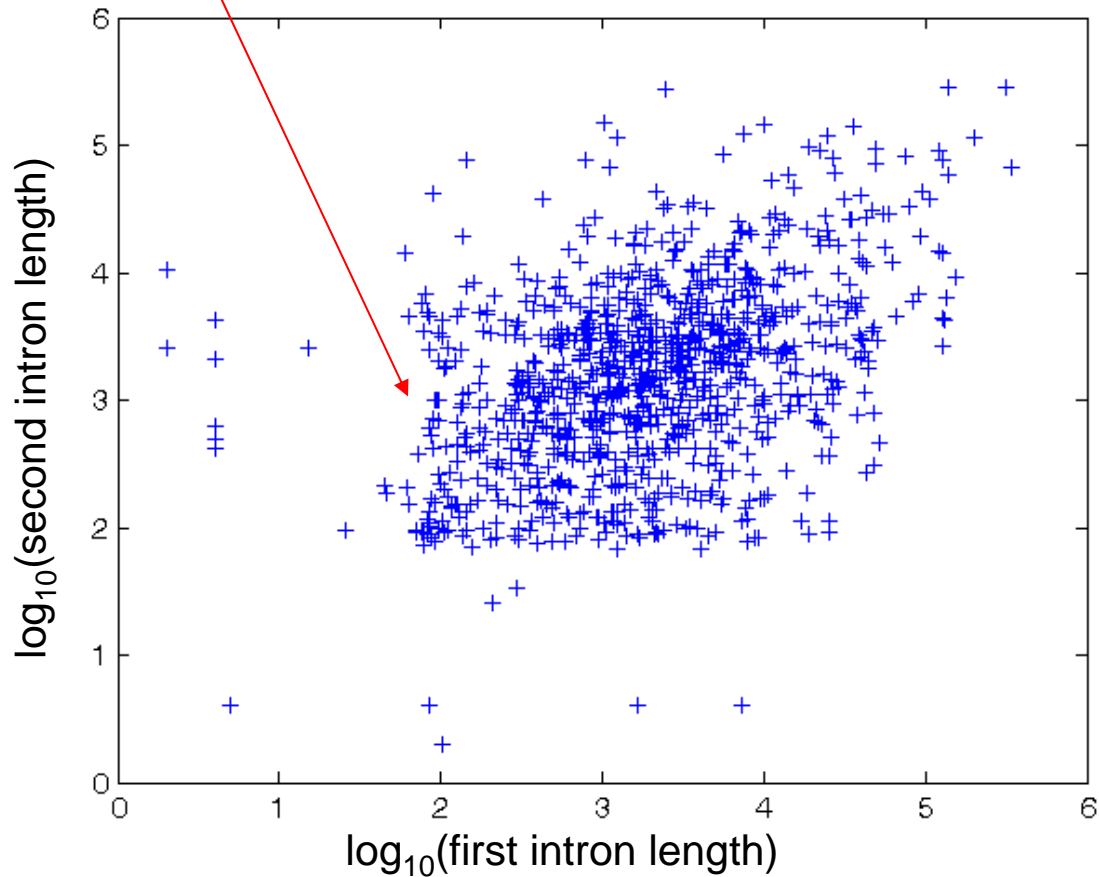
Is there a significant correlation here? If the first intron is long, does the second one also tend to be? Or is our eye being fooled by the non-Gaussian shape?

The University of Texas at Austin, CS 395T, Spring 2011, Prof. William H. Press

10