Random vectors

Recall that a random vector $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}$ is made up of, say, k random variables.

A random vector has a joint distribution, e.g. a density $f(\mathbf{x})$, that gives probabilities

$$P(\mathbf{X} \in A) = \int_A f(\mathbf{x}) \mathrm{d}\mathbf{x}.$$

Just as a random variable X has a mean $E(X)$ and variance $\mathrm{var}(X)$, a random vector also has a mean *vector* $E(\mathbf{X})$ and a covariance *matrix* $\mathrm{cov}(\mathbf{X})$.

Let $\mathbf{X} = (X_1, \ldots, X_k)$ be a random vector with density $f(x_1, \ldots, x_k)$ or pmf $p(X_1, \ldots, x_k)$. The mean of $\mathbf{X}$ is the vector of marginal means

$$E(\mathbf{X}) = E\left(\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}\right) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_k) \end{bmatrix}.$$

The covariance matrix of $\mathbf{X}$ is given by

$$cov(\mathbf{X}) = \begin{bmatrix} cov(X_1, X_1) & cov(X_1, X_2) & \cdots & cov(X_1, X_k) \\ cov(X_2, X_1) & cov(X_2, X_2) & \cdots & cov(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_k, X_1) & cov(X_k, X_2) & \cdots & cov(X_k, X_k) \end{bmatrix}.$$

Multivariate normal distribution

The normal distribution generalizes to multiple dimensions. We'll first look at two jointly distributed normal random variables, then discuss three or more.

The *bivariate normal density* for (X_1, X_2) is given by $f(x_1, x_2) =$

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}.$$

There are 5 parameters: $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

Read: pp. 81–84, 145–146, 148, 567–568.

- This density jointly defines X_1 and X_2, which live in $\mathbb{R}^2 = (-\infty, \infty) \times (-\infty, \infty)$.
- Marginally, $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$.
- The correlation between X_1 and X_2 is given by $\mathrm{corr}(X_1, X_2) = \rho$.
- For jointly normal random variables, if the correlation is zero then *they are independent*. This is not true in general for jointly defined random variables (e.g. homework 5 problem).
- $E(\mathbf{X}) = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \ \mathrm{cov}(\mathbf{X}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}.$
- Next slide: $\mu_1 = 0, 1$; $\mu_2 = 0, 2$; $\sigma_1^2 = \sigma_2^2 = 1$; $\rho = 0, 0.9, -0.6$.

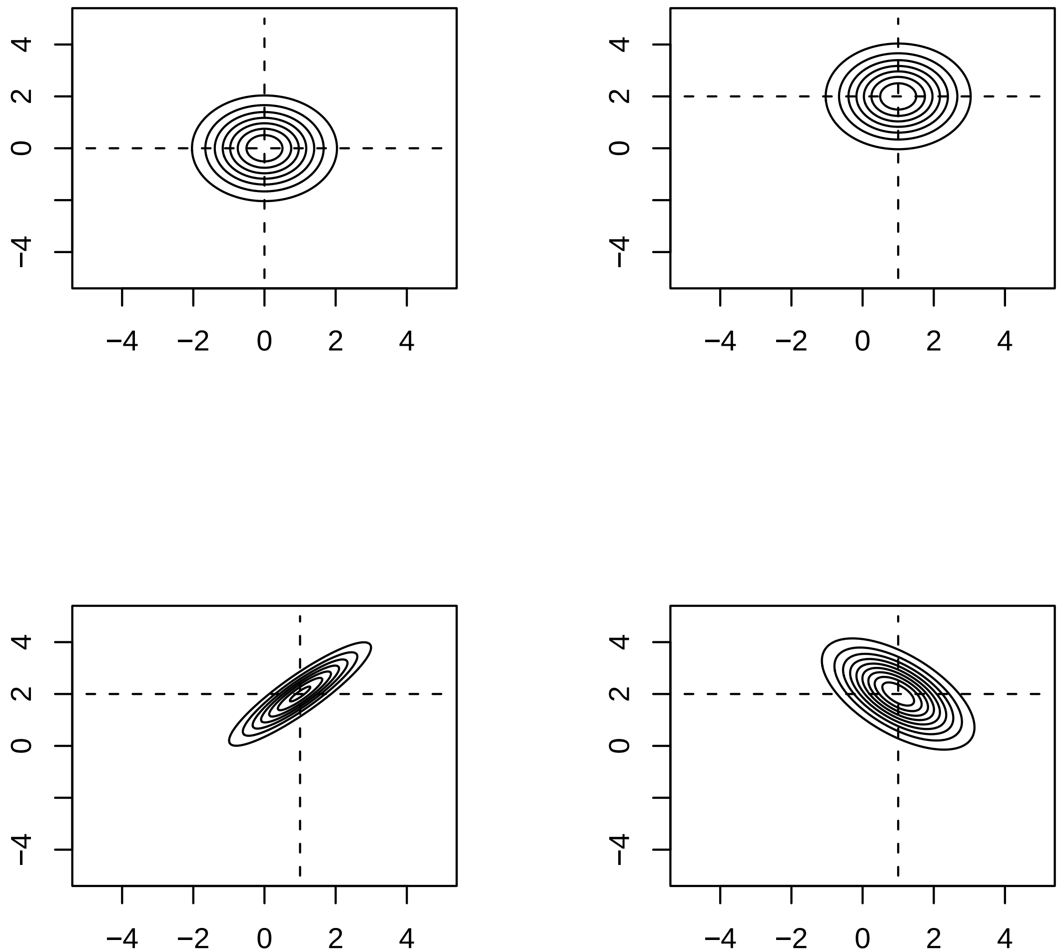

Figure 1: Bivariate normal PDF level curves.

Proof that X_1 indpendent X_2 when $\rho = 0$

When $\rho = 0$ the joint density for (X_1, X_2) simplifies to

$$
\begin{aligned}
f(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{1}{2} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\} \\
&= \left[\frac{1}{\sqrt{2\pi}\sigma_1} e^{-0.5\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2} \right] \left[\frac{1}{\sqrt{2\pi}\sigma_2} e^{-0.5\left(\frac{x_2 - \mu_2 1}{\sigma_2}\right)^2} \right].
\end{aligned}
$$

Since these are each respectively functions of x_1 and x_2 only, and the range of (X_1, X_2) factors into the produce of two sets, X_1 and X_2 are independent and in fact $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$.

Conditional distributions $[X_1|X_2 = x_2]$ **and** $[X_2|X_1 = x_1]$

The conditional distribution of X_1 given $X_2 = x_2$ is

$$[X_1|X_2 = x_2] \sim N\left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right).$$

Similarly,

$$[X_2|X_1 = x_1] \sim N\left(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\right).$$

This ties directly to linear regression:

To predict $X_2|X_1 = x_1$, we have

$$E(X_2|X_1 = x_1) = \left[\mu_2 - \frac{\sigma_2}{\sigma_1}\rho\mu_1\right] + \left[\frac{\sigma_2}{\sigma_1}\rho\right] x_1 = \beta_0 + \beta_1 x_1.$$

Bivariate normal distribution as data model

Here we assume

$$\begin{bmatrix} X_{i1} \\ X_{i2} \end{bmatrix} \overset{iid}{\sim} N_2\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}\right),$$

or succinctly,

$$\mathbf{X}_i \overset{iid}{\sim} N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

If the bivariate normal model is appropriate for paired outcomes, it provides a convenient probability model with some nice properties.

The sample mean $\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$ is the MLE of $\boldsymbol{\mu}$ and the sample covariance matrix $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$ is the MLE for $\boldsymbol{\Sigma}$.

It can be shown that

$$\bar{\mathbf{X}} \sim N_2\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right).$$

The matrix $n\widehat{\boldsymbol{\Sigma}}$ has an "inverted Wishart" distribution.

Say n outcome pairs are to be recorded:
$\{(X_{11}, X_{12}), (X_{21}, X_{22}), \ldots, (X_{n1}, X_{n2})\}$. The i^{th} pair is (X_{i1}, X_{i2}).
The *sample mean vector* is given elementwise by

$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n X_{i1} \\ \frac{1}{n}\sum_{i=1}^n X_{i2} \end{bmatrix},$$

and the *sample covariance matrix* is given elementwise by

$$\mathbf{S} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n (X_{i1} - \bar{X}_1)^2 & \frac{1}{n}\sum_{i=1}^n (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2) \\ \frac{1}{n}\sum_{i=1}^n (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2) & \frac{1}{n}\sum_{i=1}^n (X_{i2} - \bar{X}_2)^2 \end{bmatrix}.$$

The sample mean vector $\bar{\mathbf{X}}$ estimates $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and the sample covariance matrix $\mathbf{S}$ estimates

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}.$$

We will place hats on parameter *estimators* based on the data. So

$$\hat{\mu}_1 = \bar{X}_1,\ \hat{\mu}_2 = \bar{X}_2,\ \hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)^2,\ \hat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_{i2} - \bar{X}_2)^2.$$

Also,

$$\widehat{cov}(X_1, X_2) = \frac{1}{n}\sum_{i=1}^{n}(X_{i1} - \bar{X})(X_{i2} - \bar{X}_2).$$

So a natural estimate of ρ is then

$$\hat{\rho} = \frac{\widehat{cov}(X_1, X_2)}{\hat{\sigma}_1 \hat{\sigma}_2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)^2}}.$$

This is in fact the MLE estimate based on the bivariate normal model. It is also a "plug-in" estimator based on the method-of-moments too. It is commonly referred to as the Pearson correlation coefficient. You can get it as, e.g., `cor(age,Gesell)` in `R`.

This estimate of correlation can be unduly influenced by outliers in the sample. An alternative measure of linear association is the Spearman correlation based on ranks. This correlation estimates something a bit different than the Pearson correlation.

```
> cor(age,Gesell)
[1] -0.64029
> cor(age,Gesell,method="spearman")
[1] -0.3166224
```

Gesell data:

Recall: X is age in months a child speaks his/her first word and let Y is Gesell adaptive score, a measure of a child's aptitude. *Question*: how does the child's aptitude *change* with how long it takes them to speak? Here, $n = 21$.

In `R` we find $\widehat{E}(\mathbf{X}) = \begin{bmatrix} 14.38 \\ 93.67 \end{bmatrix}$. Also, $\widehat{\text{cov}}(\mathbf{X}) = \begin{bmatrix} 60.14 & -67.78 \\ -67.78 & 186.32 \end{bmatrix}$.

Assuming a bivariate model, we plug in the MLEs and obtain the estimated PDF for (X, Y):

$$f(x, y) = \exp(-60.22 + 1.3006x - 0.0134x^2 + 0.9520y - 0.0098xy - 0.0043y^2).$$

We can further find from $Y \overset{\bullet}{\sim} N(93.67, 186.32)$,

$$f_Y(y) = \exp(-3.557 - 0.00256(y - 93.67)^2).$$

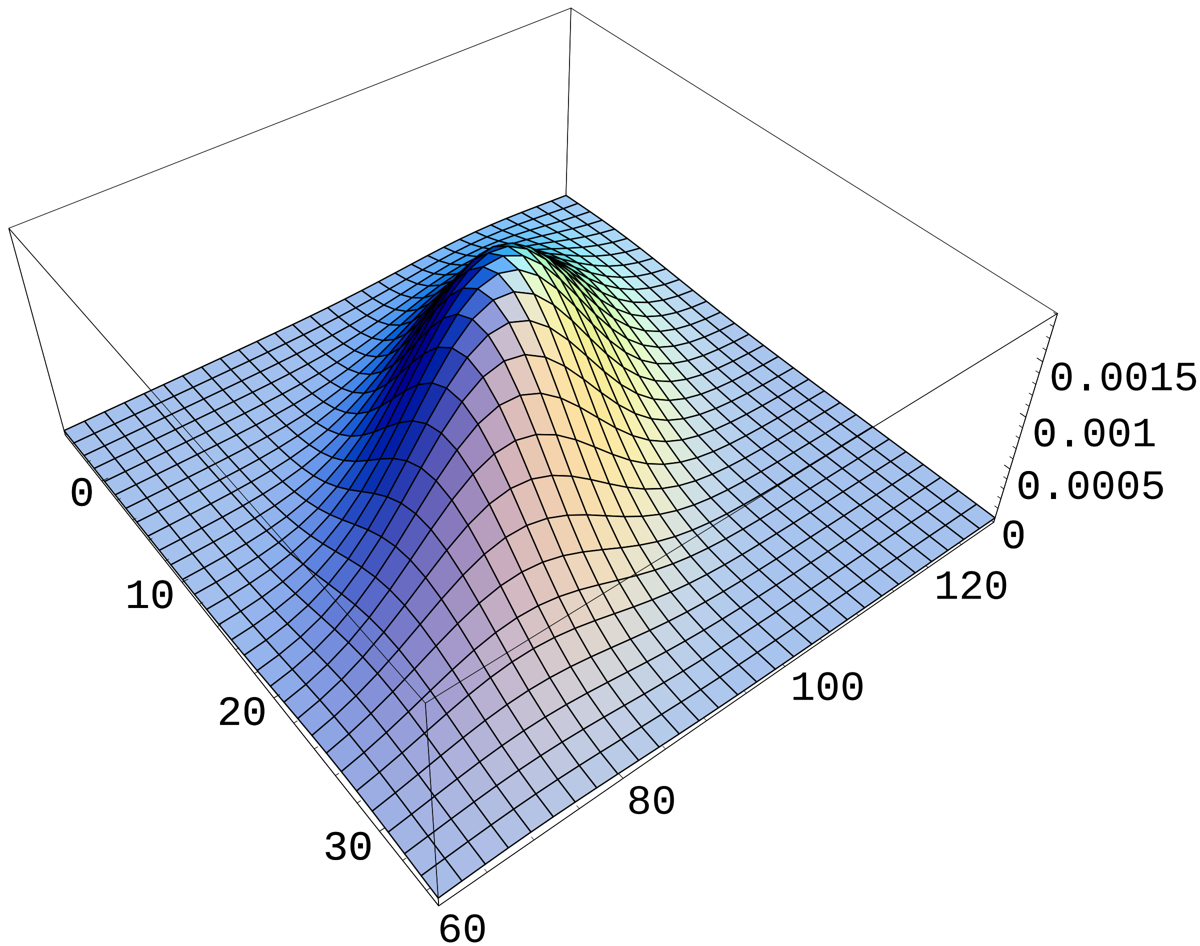

Figure 2: 3D plot of $f(x,y)$ for (X,Y) based on MLEs.

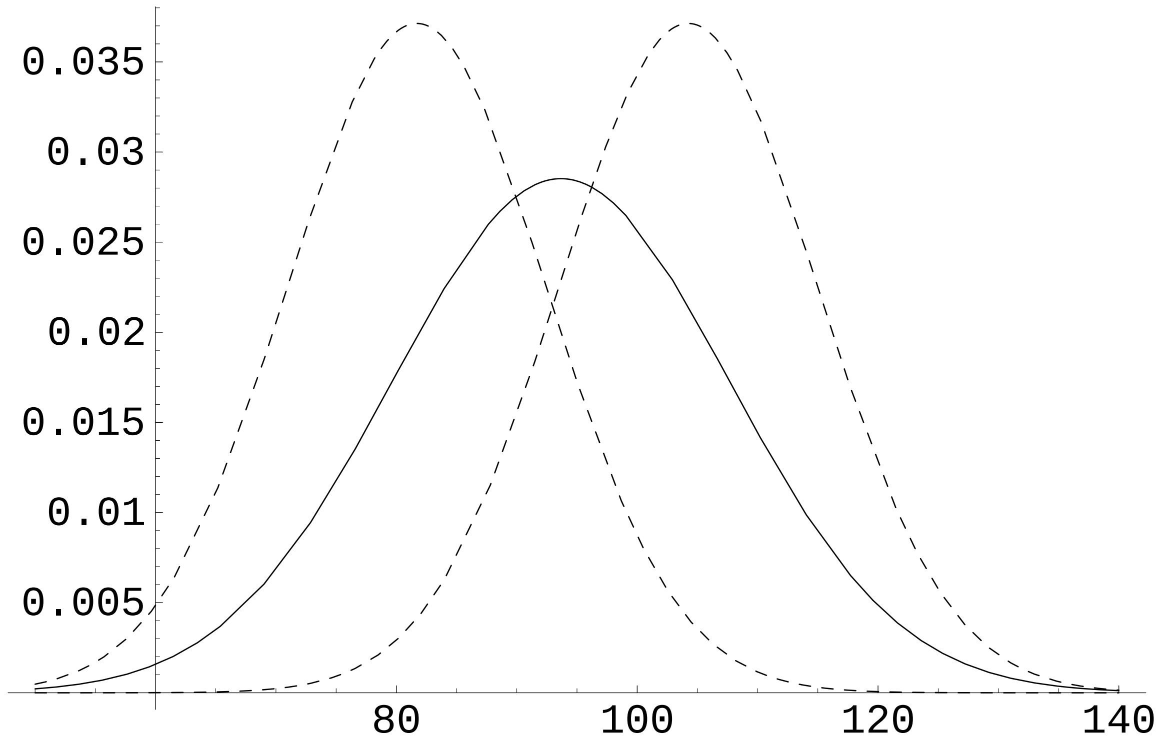

Figure 3: Solid is $f_Y(y)$; left dashed is $f_{Y|X}(y|25)$ the right dashed is $f_{Y|X}(y|10)$. As the age in months of first words $X = x$ increases, the distribution of Gesell Adaptive Scores Y decreases.

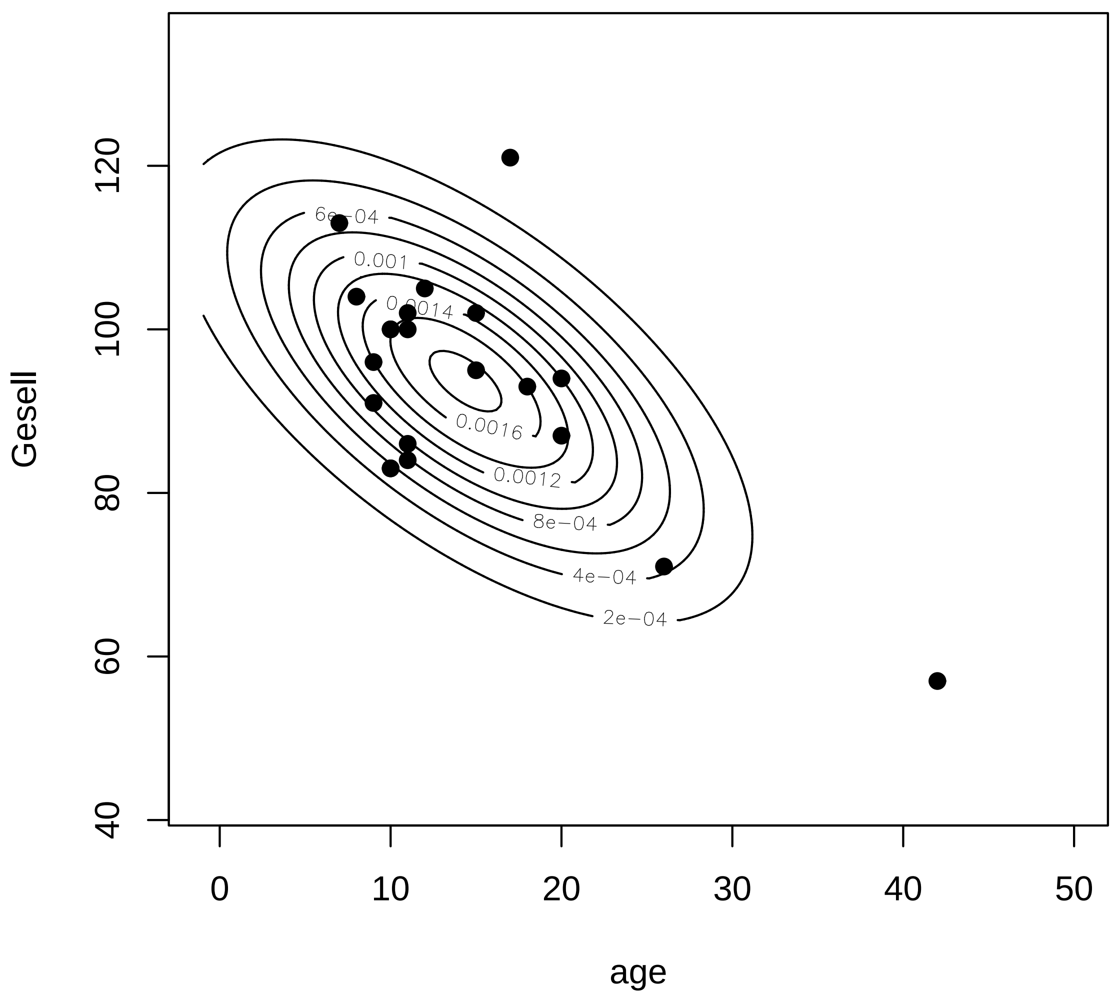

Figure 4: MLE estimate of density with actual data.

`R` code to get estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, plot level curves of PDF, and data.

```
age<-c(15,26,10,9,15,20,18,11,8,20,7,9,10,11,11,10,12,42,17,11,10)
Gesell<-c(95,71,83,91,102,87,93,100,104,94,113,96,83,84,102,100,105,57,121,86,100)
data=matrix(c(age,Gesell),ncol=2) # make data matrix
m=c(mean(age),mean(Gesell)) # mean vector s=cov(data) # covariance
matrix

x1=seq(min(age)-sd(age),max(age)+sd(age),length=200) # grid of representative age values
x2=seq(min(Gesell)-sd(Gesell),max(Gesell)+sd(Gesell),length=200) # grid of Gesell values

f=function(x,y){
     r=s[1,2]/sqrt(s[1,1]*s[2,2])
     term1=(x-m[1])/sqrt(s[1,1]); term2=(y-m[2])/sqrt(s[2,2]); term3=-2*r*term1*term2
     exp(-0.5*(term1^2+term2^2+term3)/(1-r^2))/(2*3.141*sqrt(s[1,1]*s[2,2]*(1-r^2)))
} # function that gives the best fitting bivariate normal density to the data

z=outer(x1,x2,f) # compute the joint pdf over the grid
contour(x1,x2,z,nlevels=7,xlab="age",ylab="Gesell") # make contour plot
points(age,Gesell,pch=19) # superimpose filled points
```

Checking bivariate normality

There are no "multivariate" boxplots. There are multivariate histograms (and smoothed histograms), but they're typically not that useful for checking normality.

Easiest thing to do is to check that marginal boxplots/histograms of $X_{11}, \ldots, X_{1n}$ and $X_{21}, \ldots, X_{2n}$ are approximately normal. This does not guarantee joint normality though.

One can additionally look at a simple scatterplot of the data, checking to see that it shows an approximately 'elliptical' cloud of points with no glaringly obvious outlying values.

Multivariate normal distribution

In general, a k-variate normal is defined through the mean and covariance matrix:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix} \sim N_k\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix} \right).$$

Succinctly,

$$\mathbf{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Recall that if $Z \sim N(0,1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$. The definition of the multivariate normal distribution just extends this idea.

Instead of one standard normal, we have a list of k independent standard normals $\mathbf{Z} = (Z_1, \ldots, Z_k)$, and consider the same sort of transformation in the multivariate case using matrices and vectors.

Let $Z_1, \ldots, Z_k \overset{iid}{\sim} N(0, 1)$. The joint pdf of $(Z_1, \ldots, Z_k)$ is given by

$$f(z_1, \ldots, z_k) = \prod_{i=1}^{k} \exp(-0.5 z_i^2)/\sqrt{2\pi}.$$

Let

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix},$$

where $\boldsymbol{\Sigma}$ is symmetric (i.e. $\boldsymbol{\Sigma}^T = \boldsymbol{\Sigma}$, which implies $\sigma_{ij} = \sigma_{ji}$ for all $1 \leq i, j \leq k$).

Let $\mathbf{\Sigma}^{1/2}$ be any matrix such that $\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma}$. Then $\mathbf{X} = \boldsymbol{\mu} + \mathbf{\Sigma}^{1/2}\mathbf{Z}$ is said to have a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, written

$$\mathbf{X} \sim N_k(\boldsymbol{\mu}, \mathbf{\Sigma}).$$

Written in terms of matrices

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} + \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix}^{1/2} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{bmatrix}.$$

Using some math, it can be shown that the pdf of the new vector $\mathbf{X} = (X_1, \ldots, X_k)$ is given by

$$f(x_1, \ldots, x_k | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\{-0.5(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}.$$

In the one-dimensional case, this simplifies to our old friend

$$f(x_1 | \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\{-0.5(x - \mu)(\sigma^2)^{-1}(x - \mu)\},$$

the pdf of a $N(\mu, \sigma^2)$ random variable X.

There are several important properties of multivariate normal vectors...

Let

$$\mathbf{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Then

1. For each X_i in $\mathbf{X} = (X_1, \ldots, X_k)$, $E(X_i) = \mu_i$ and $var(X_i) = \sigma_{ii}$. That is, marginally, $X_i \sim N(\mu_i, \sigma_{ii})$.

2. For any $r \times k$ matrix $\mathbf{M}$,
$$\mathbf{MX} \sim N_r(\mathbf{M}\boldsymbol{\mu}, \mathbf{M}\boldsymbol{\Sigma}\mathbf{M}').$$

3. For any two (X_i, X_j) where $1 \leq i < j \leq k$, $cov(X_i, X_j) = \sigma_{ij}$. The off-diagonal elements of $\boldsymbol{\Sigma}$ give the covariance between two elements of $(X_1, \ldots, X_k)$. Note then $\rho(X_i, X_j) = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$.

4. For any $k \times 1$ vector $\mathbf{m} = (m_1, \ldots, m_k)$ and $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{m} + \mathbf{Y} \sim N_k(\mathbf{m} + \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

As a simple example, let

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_3\left(\begin{bmatrix} -2 \\ 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 4 \end{bmatrix} \right).$$

Define

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & -1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{MX} = \begin{bmatrix} 1 & 0 & -1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}.$$

Then $X_2 \sim N(5, 3)$, $cov(X_2, X_3) = -1$ and

$$\begin{bmatrix} 1 & 0 & -1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim$$

$$N_2\left(\begin{bmatrix} 1 & 0 & -1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} -2 \\ 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & -1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 4 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & \frac{1}{3} \\ -1 & \frac{1}{3} \end{bmatrix} \right),$$

or simplifying,

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_2\left(\begin{bmatrix} -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & \frac{11}{9} \end{bmatrix} \right).$$

Note that for the transformed vector $\mathbf{Y} = (Y_1, Y_2)$, $cov(Y_1, Y_2) = 0$ and therefore Y_1 and Y_2 are uncorrelated, i.e. $\rho(Y_1, Y_2) = 0$.

We will use these properties later on, when we discuss the large sample distribution of MLE vectors

$$\widehat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_p \end{bmatrix} \overset{\bullet}{\sim} N_p\left(\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix}, \ \widehat{\text{cov}}(\widehat{\boldsymbol{\theta}}) \right).$$

For the linear model (e.g. simple linear regression or the two-sample model) $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, the MLE that describes the mean parameters has exactly a multivariate normal distribution (p. 574)

$$\widehat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2).$$

$p = 2$ is the number of mean parameters. The MLE $\hat{\sigma}^2$ has a gamma distribution.

Let's try an experiment where we know $\boldsymbol{\beta} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$

We will generate 200 independent experiments, each with a sample size of $n = 20$. The model is

$$Y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \ldots, 20,$$

where $\beta_0 = 5$, $\beta_1 = 3$, and $\sigma = 0.2$. Take x_i to be equally spaced from $x_1 = 0$ to $x_{20} = 1$. The errors satisfy $e_1, \ldots, e_{20} \overset{iid}{\sim} N(0, 0.2^2)$.

Here's `R` code to sample the $m = 200$ $\widehat{\boldsymbol{\beta}}$'s, each obtained from a sample of size $n = 20$:

```
sigma=0.2
m=200 # number of beta.hat's sampled
n=20  # sample size going into each experiment

y=rep(0,n)

beta0.hat=rep(0,m)
beta1.hat=rep(0,m)

x=seq(0,1,length=n) # predictor values evenly spaced over [0,1]
beta=c(5,3) # true (but usually unknown) beta vector
mean=beta[1]+beta[2]*x # true mean function on grid of x values

for(i in 1:m){
 y=mean+rnorm(n,0,sigma)
 fit=lm(y~x)
 beta0.hat[i]=fit$coefficients[1]
 beta1.hat[i]=fit$coefficients[2]
}

X.design=matrix(c(rep(1,n),x),ncol=2)
m=beta # exact mean vector
s=solve(t(X.design)%*%X.design)*sigma^2 # exact covariance

x1=seq(4.7,5.3,length=200) # grid of representative beta.hat0 values
x2=seq(2.5,3.4,length=200) # grid of representative beta.hat1 values
```

```
f=function(x,y){
     r=s[1,2]/sqrt(s[1,1]*s[2,2])
     term1=(x-m[1])/sqrt(s[1,1]); term2=(y-m[2])/sqrt(s[2,2]); term3=-2*r*term1*term2
     exp(-0.5*(term1^2+term2^2+term3)/(1-r^2))/(2*3.141*sqrt(s[1,1]*s[2,2]*(1-r^2)))
} # exact bivariate normal density according to theory

g0=rep(beta[1],200); g1=rep(beta[2],200)
z=outer(x1,x2,f) # compute the joint pdf over the grid
contour(x1,x2,z,nlevels=15,xlab="beta0.hat",ylab="beta1.hat") # make contour plot
points(beta0.hat,beta1.hat) # superimpose beta.hat's from m experiments of sample size n
lines(x1,g1,lty=2); lines(g0,x2,lty=2) # dotted lines crossing at true beta
```

Exactly, the theory tells us

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \sim N_2\left(\begin{bmatrix} 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 0.00743 & -0.01086 \\ -0.01086 & 0.02171 \end{bmatrix} \right).$$

Let's see how a plot of the 200 $\widehat{\boldsymbol{\beta}}$'s looks superimposed on the exact sampling distribution...

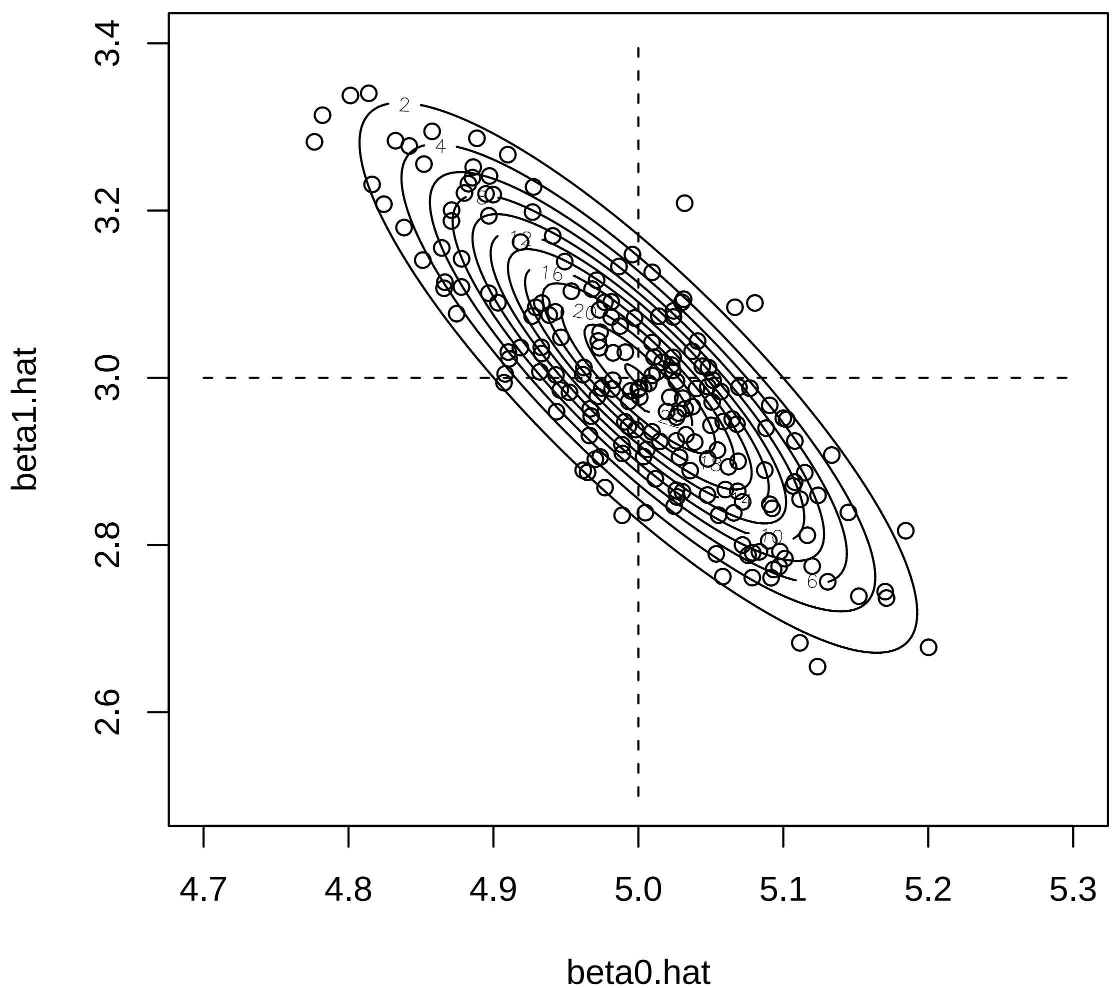

Figure 5: Theoretical distribution (pdf) with 200 sampled MLE's.