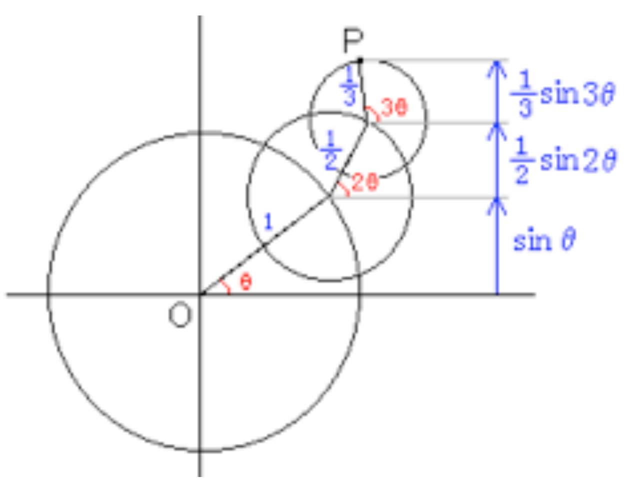
Statistical Rethinking
Winter 2019

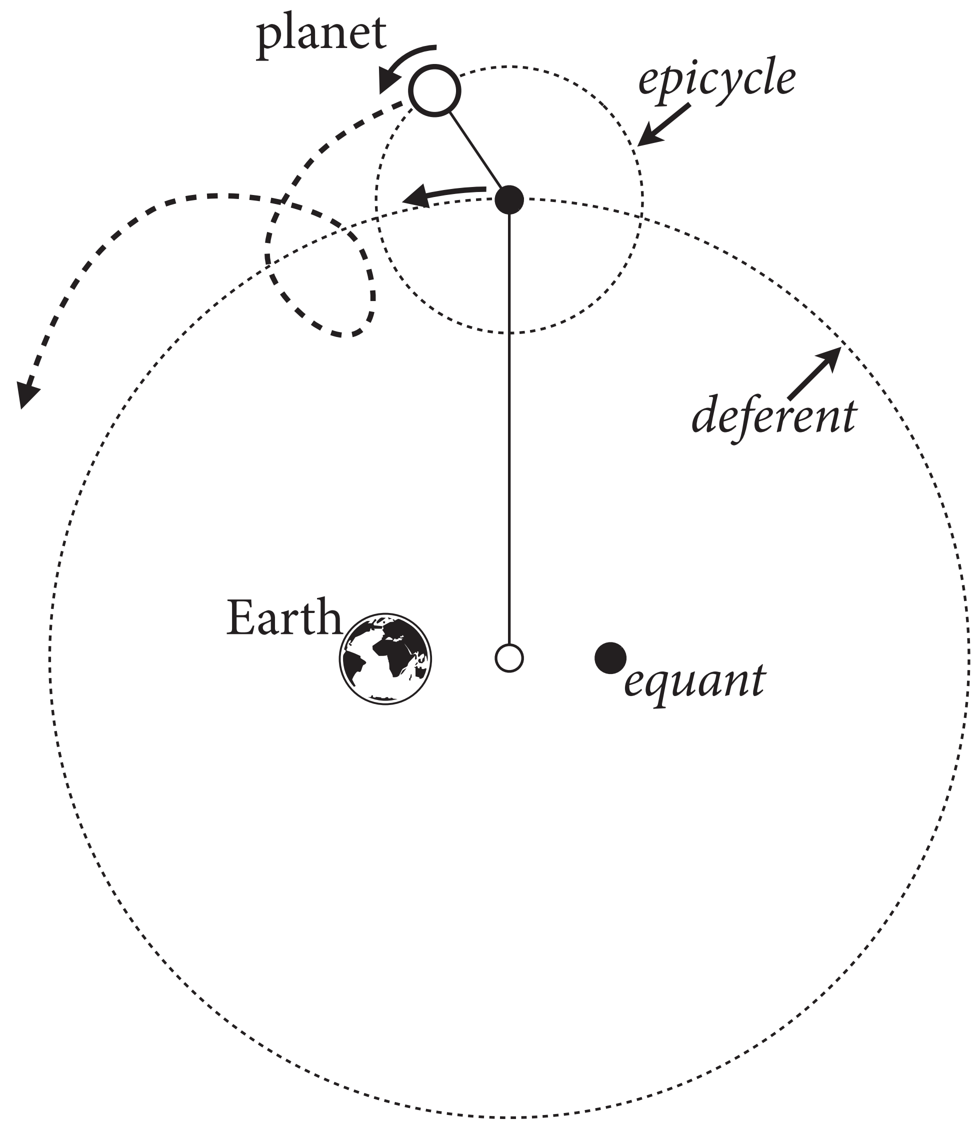
Lecture 03 / Week 2

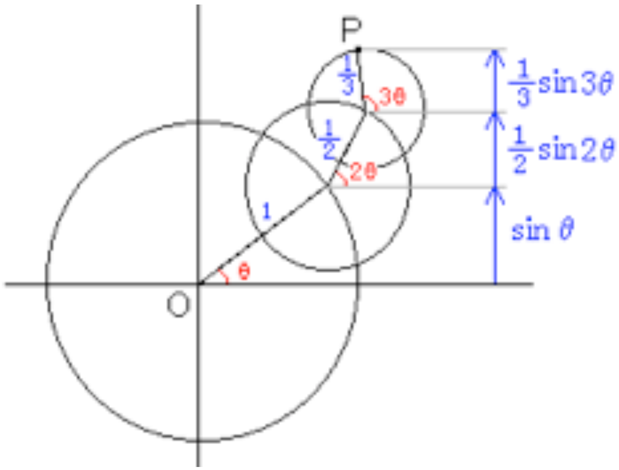
Geocentric Models

Triumph of Geocentrism

- Claudius Ptolemy (90–168)
  - Egyptian mathematician
  - Accurate model of planetary motion
  - Epicycles: orbits on orbits
  - Fourier series

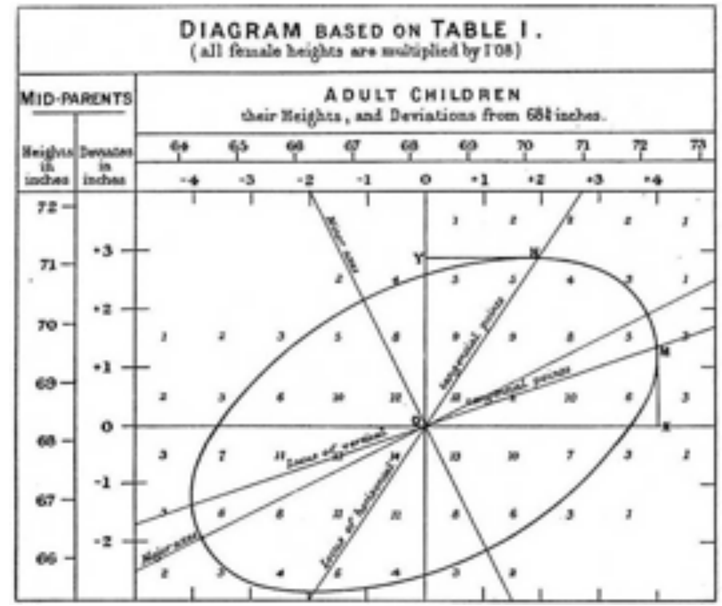






Geocentrism

- Descriptively accurate
- Mechanistically wrong
- General method of approximation
- Known to be wrong

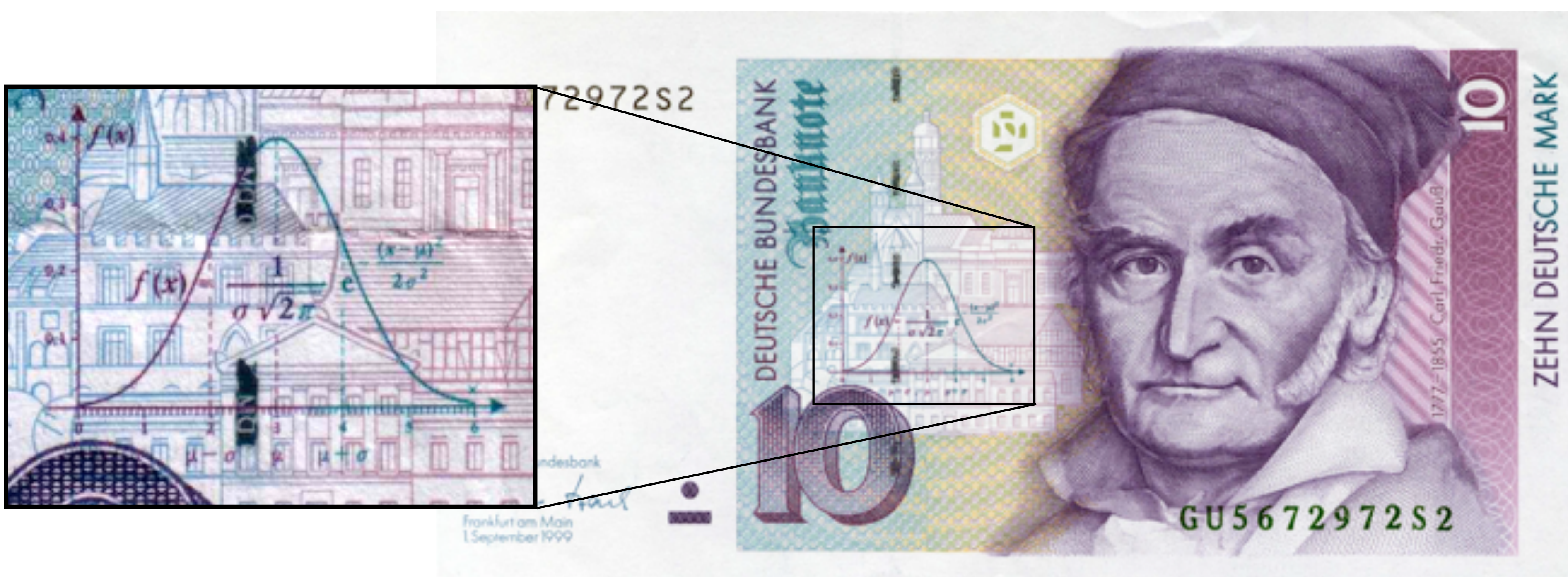


Regression

- Descriptively accurate
- Mechanistically wrong
- General method of approximation
- Taken too seriously

Linear regression

- Simple statistical golems
  - Model of mean and variance of normally (Gaussian) distributed measure
  - Mean as *additive* combination of *weighted* variables
  - Constant variance



1809 Bayesian argument for normal error and least-squares estimation

THEORIA

MOTVS CORPORVM

COELESTIVM

IN

SECTIONIBVS CONICIS SOLEM AMBIENTIVM

AVCTORE

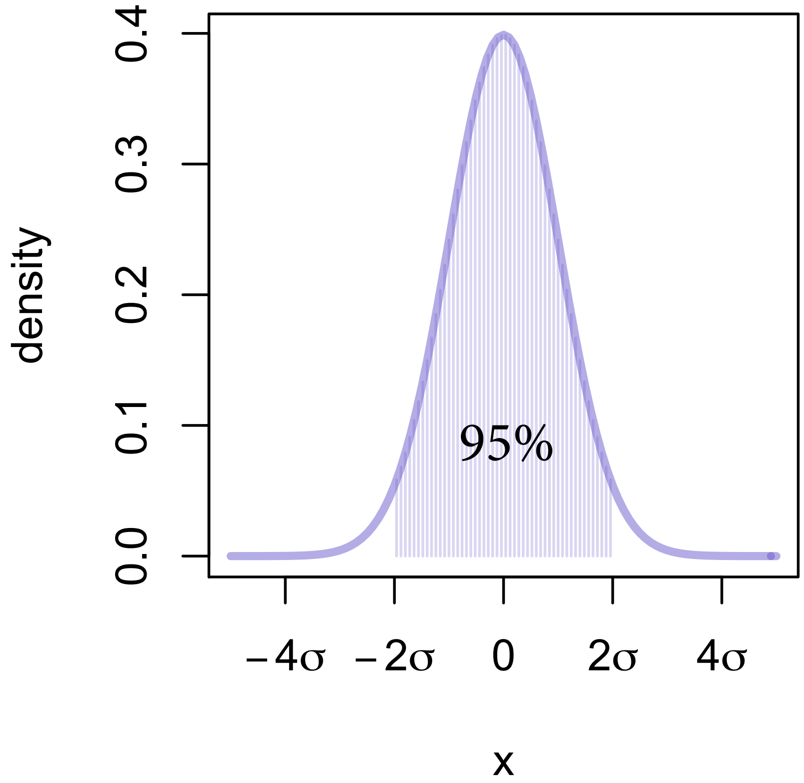
CAROLO FRIDERICO GAVSS

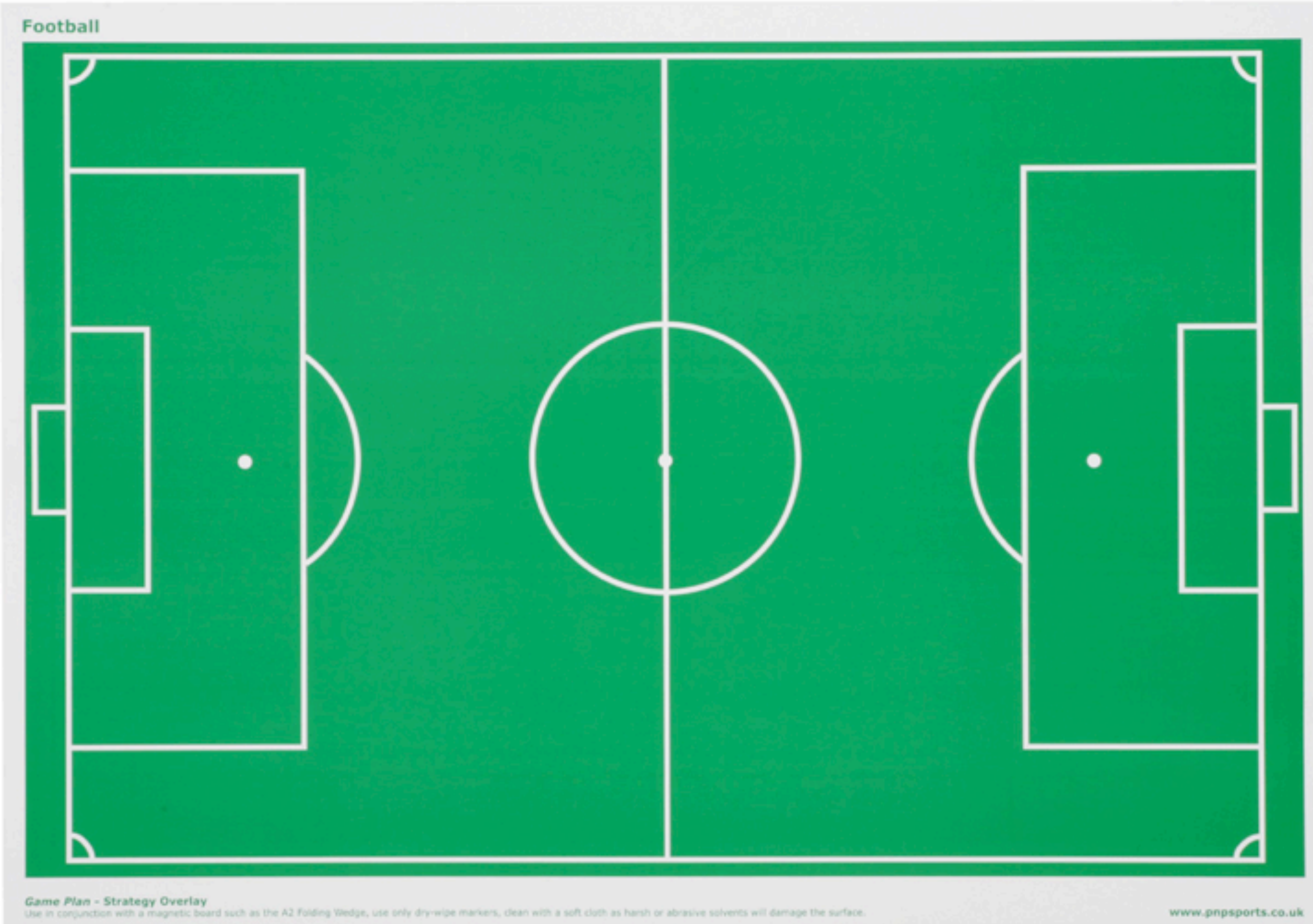
HAMBVRGI SVMTIBVS FRID. PERTHES ET I. H. BESSER

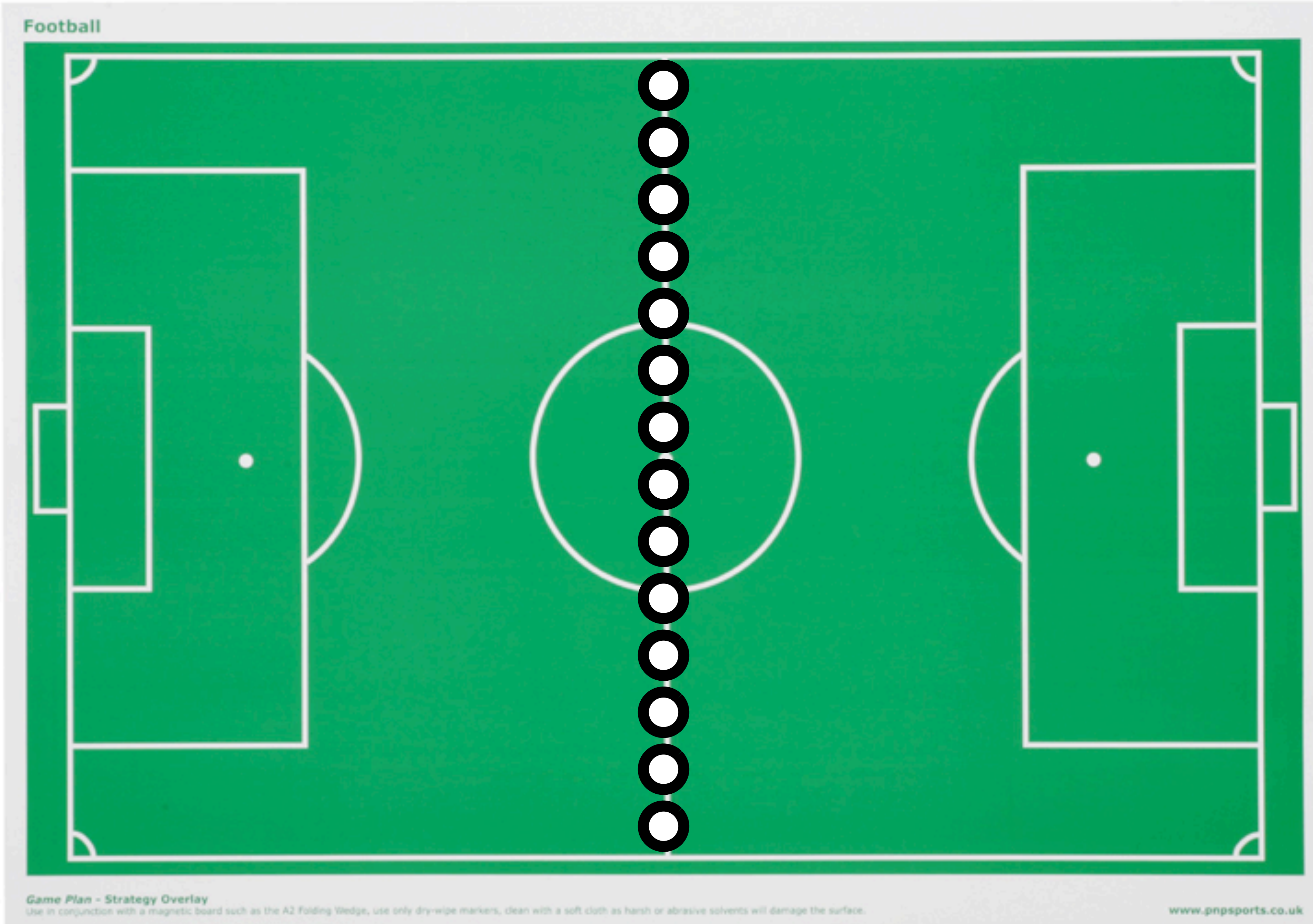
1809.

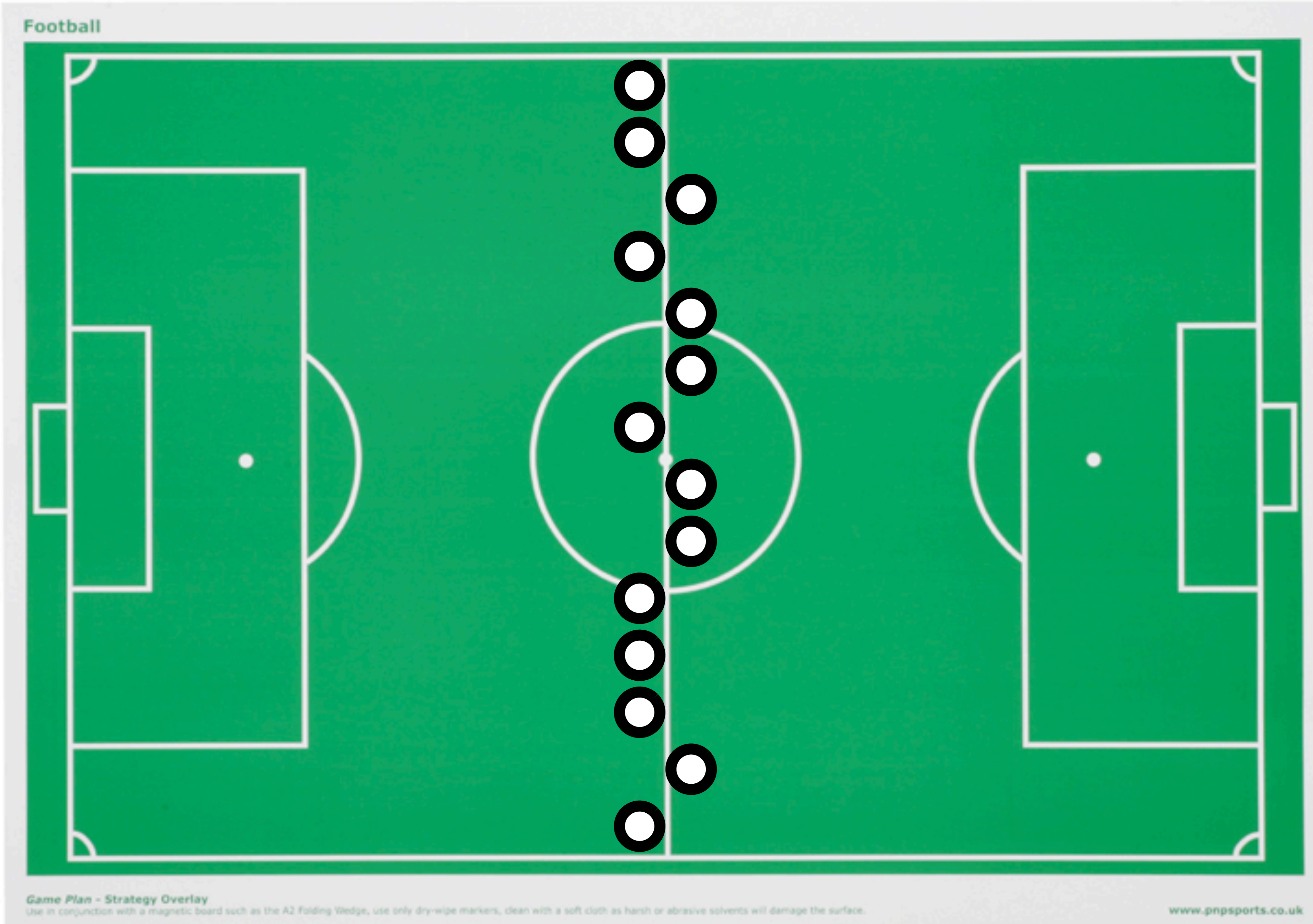
Why normal?

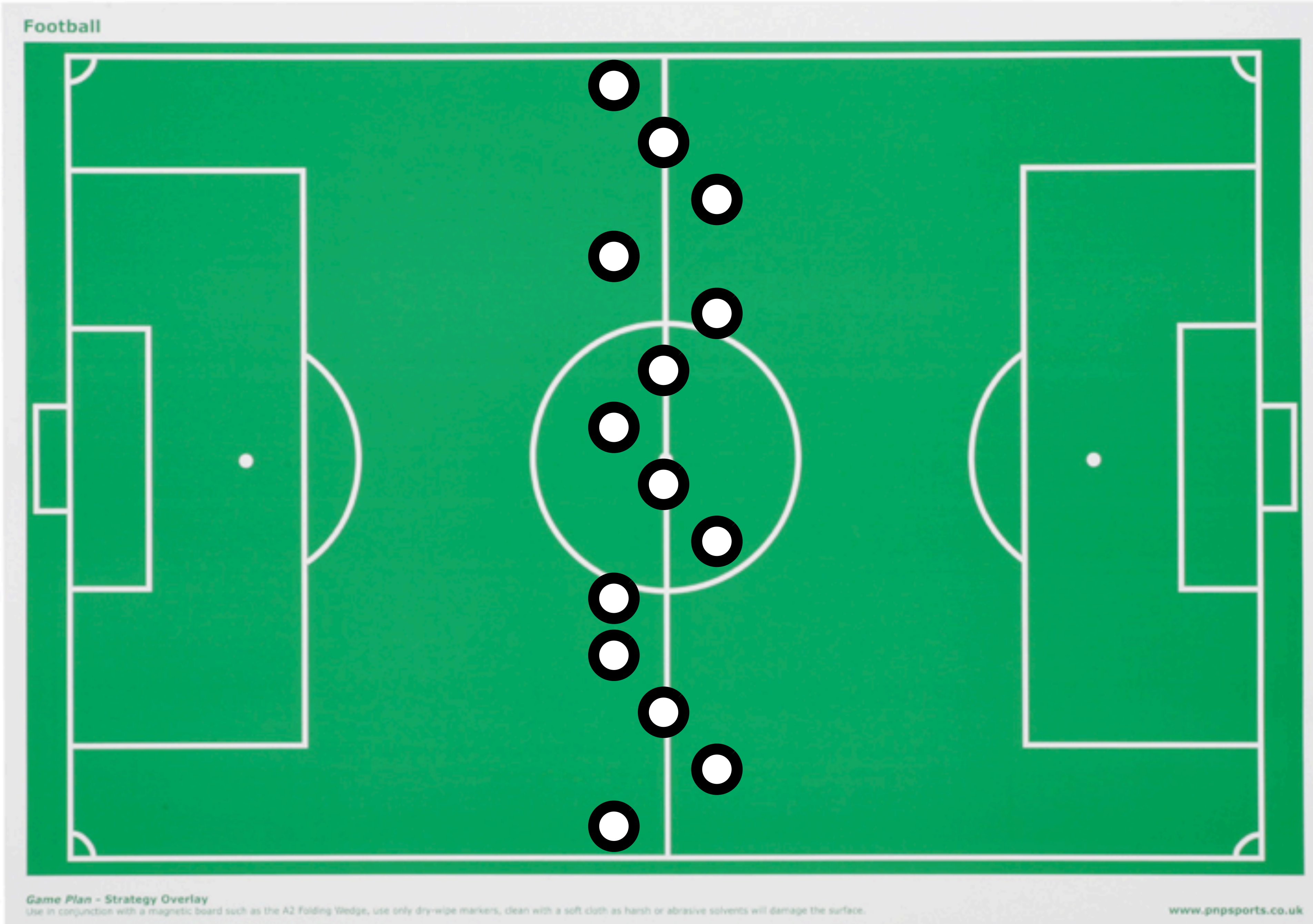
- Why are normal (Gaussian) distributions so common in statistics?
  1. Easy to calculate with
  2. Common in nature
  3. Very conservative assumption

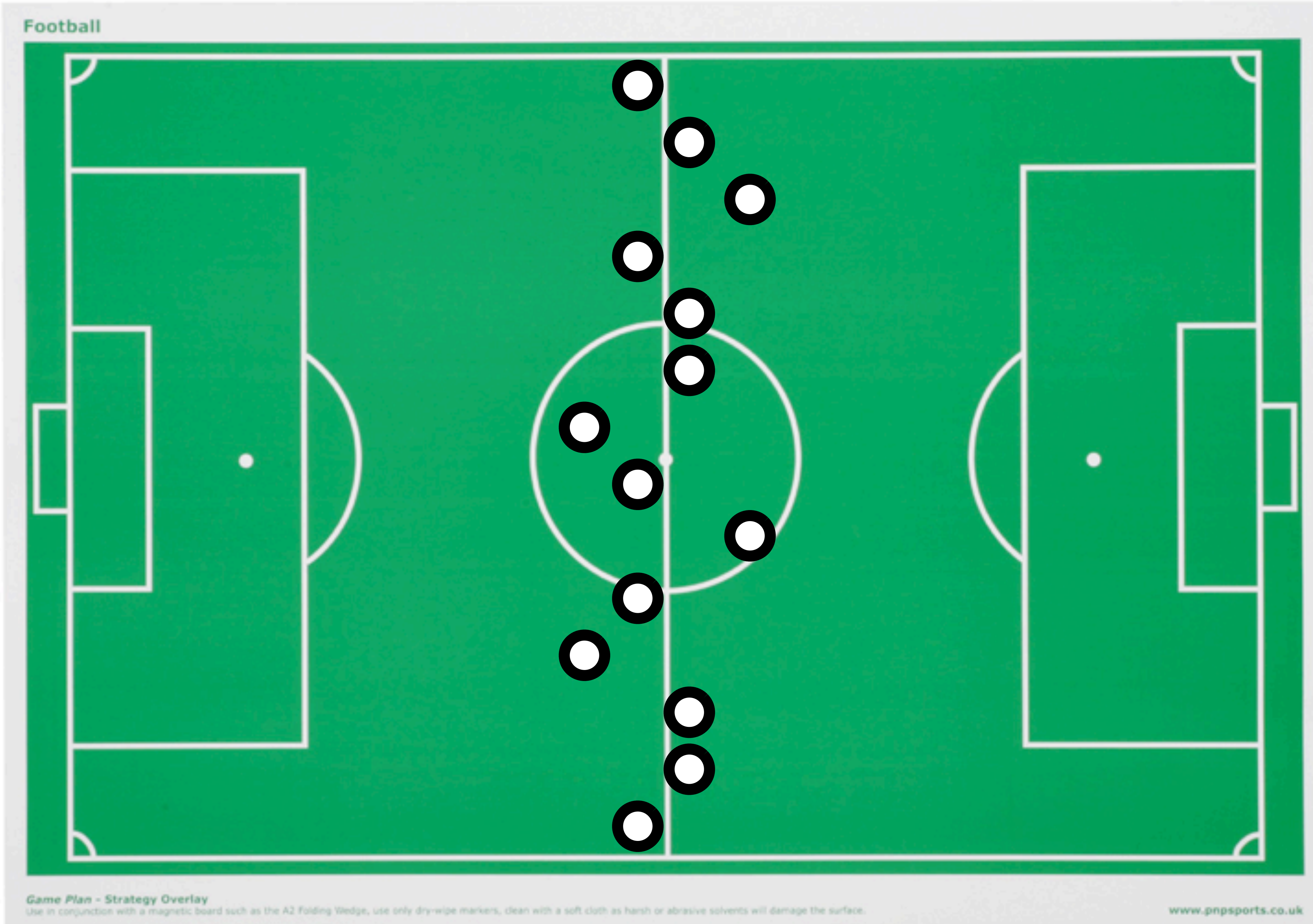


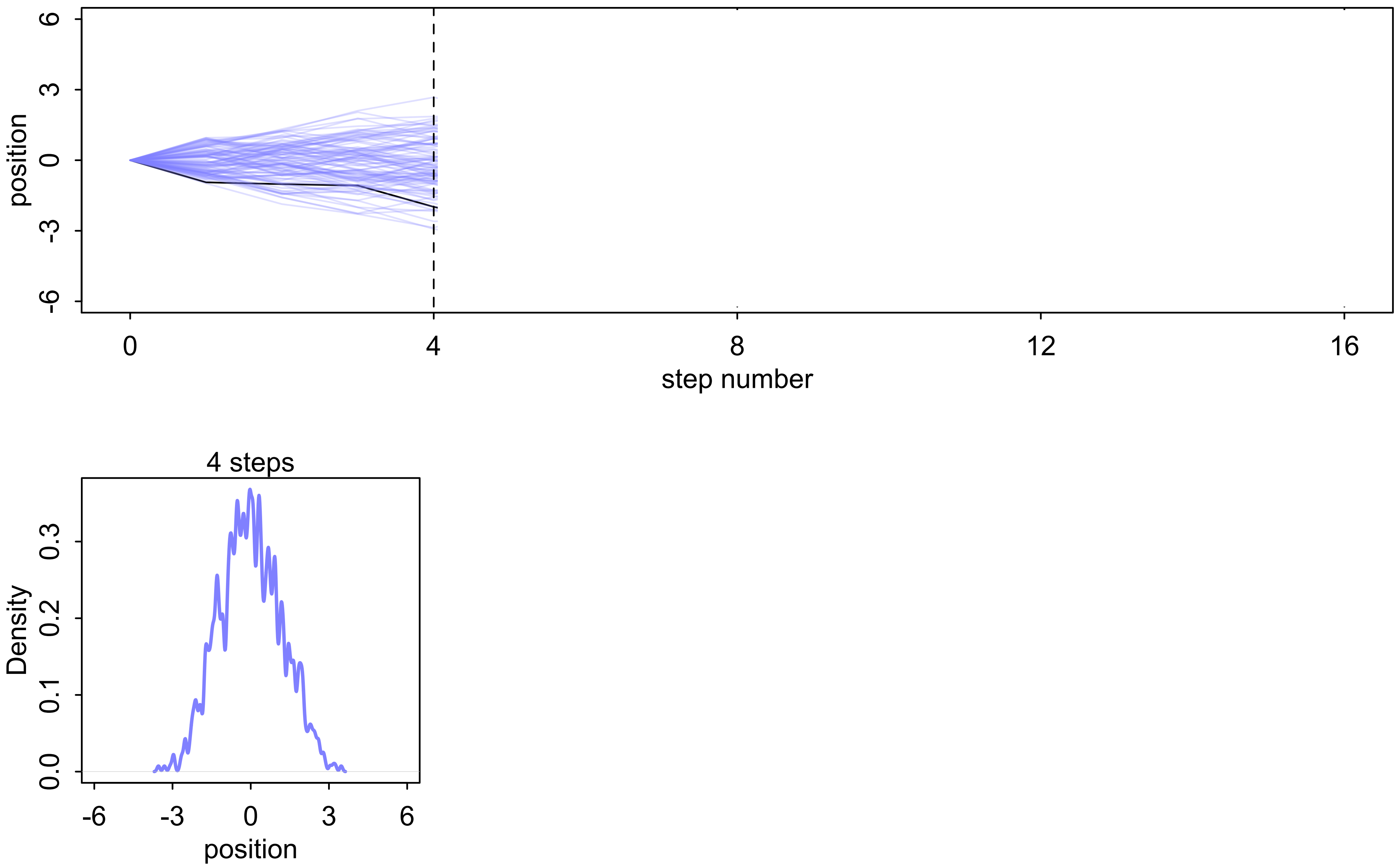

Football
Game Plan - Strategy Overlay
Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.
www.pnpsports.co.uk


Football
Game Plan - Strategy Overlay
Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.
www.pnpsports.co.uk


Football
Game Plan - Strategy Overlay
Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.
www.pnpsports.co.uk


Football
Game Plan - Strategy Overlay
Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.
www.pnpsports.co.uk


Football
Game Plan - Strategy Overlay
Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.
www.pnpsports.co.uk



position
6
3
0
-3
-6
0
4
8
12
16
step number
4 steps
Density
0.0
0.1
0.2
0.3
-6
-3
0
3
6
position

Figure 4.2

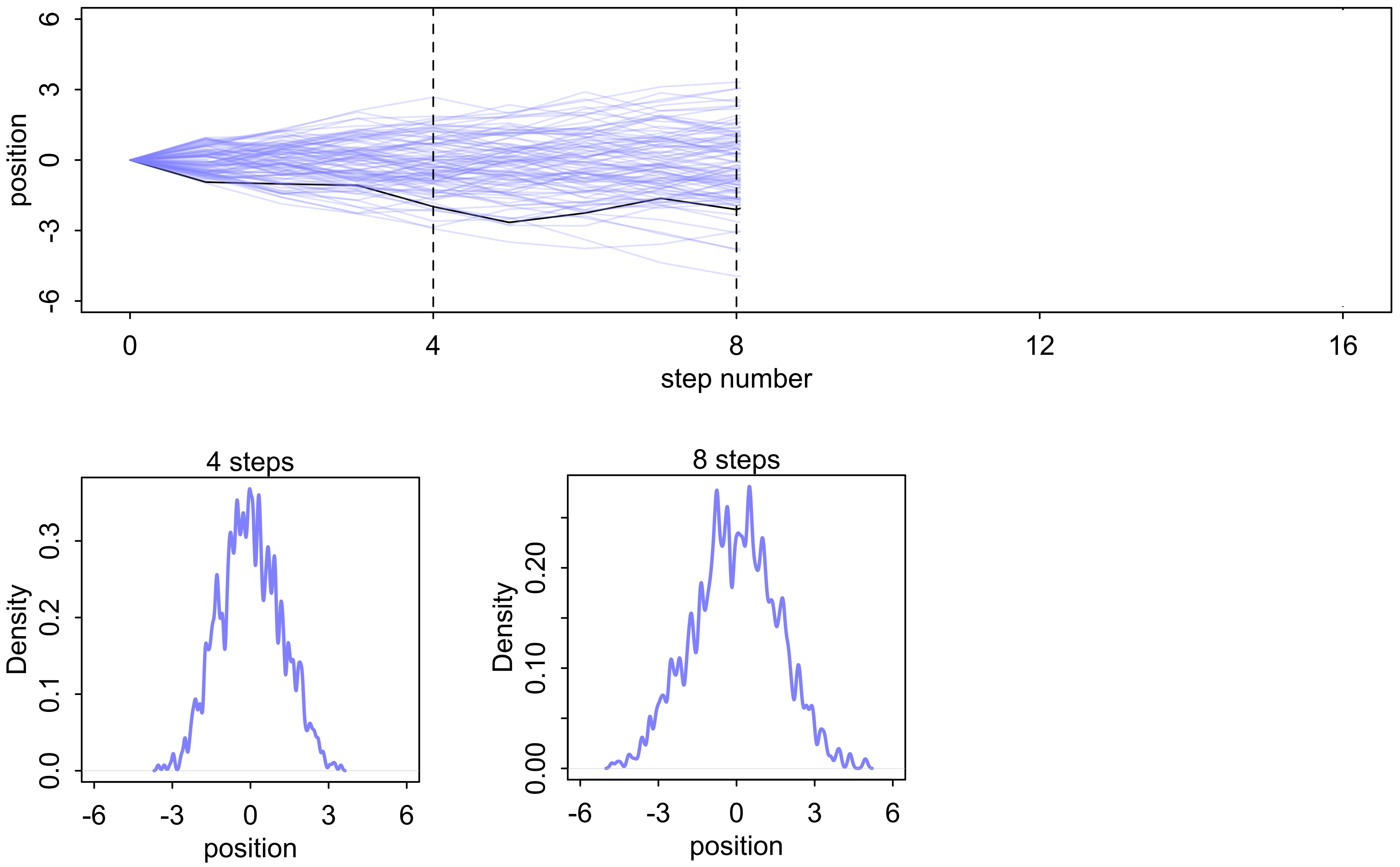
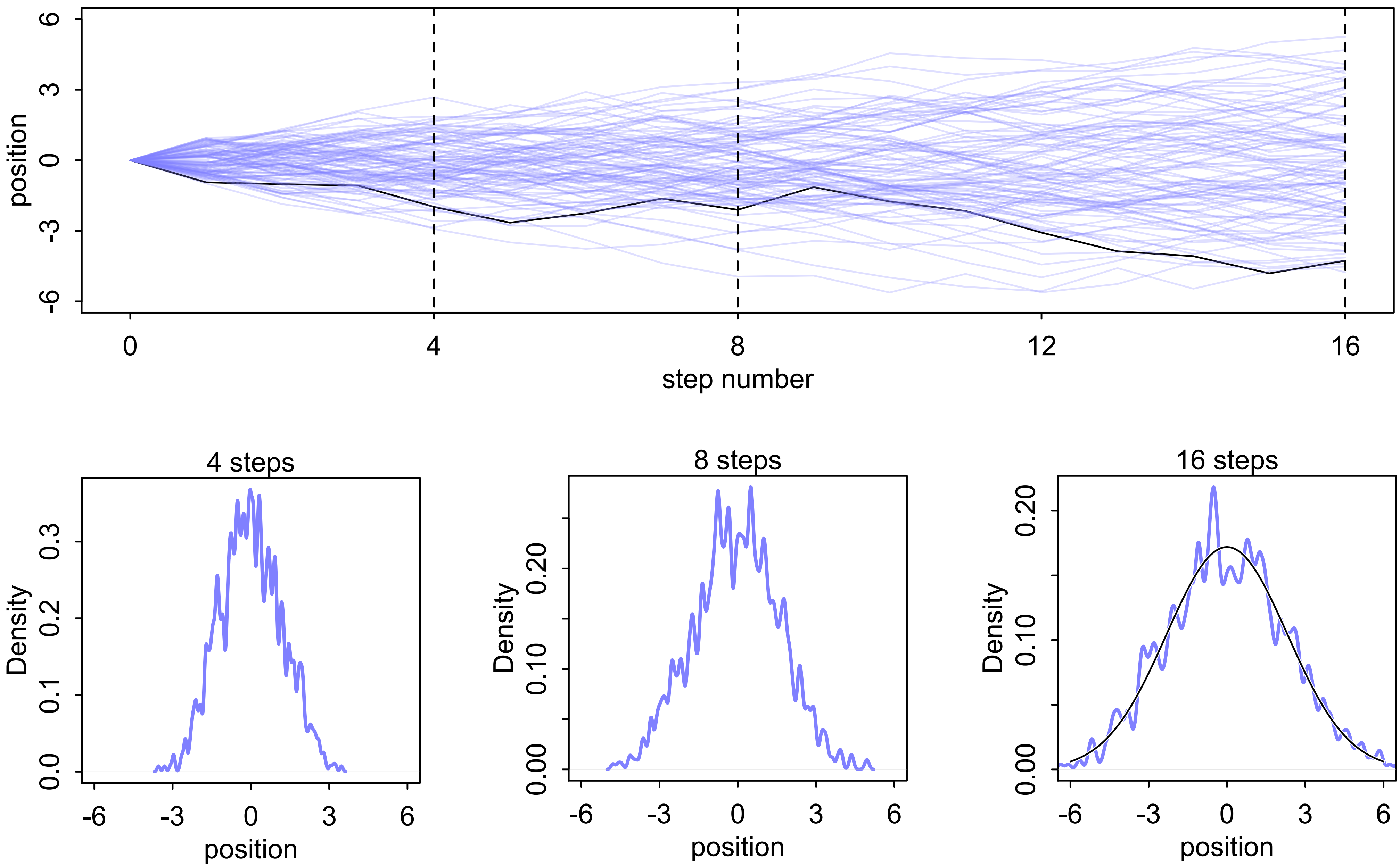

position
step number
4 steps
Density
position
8 steps
Density
position

Figure 4.2

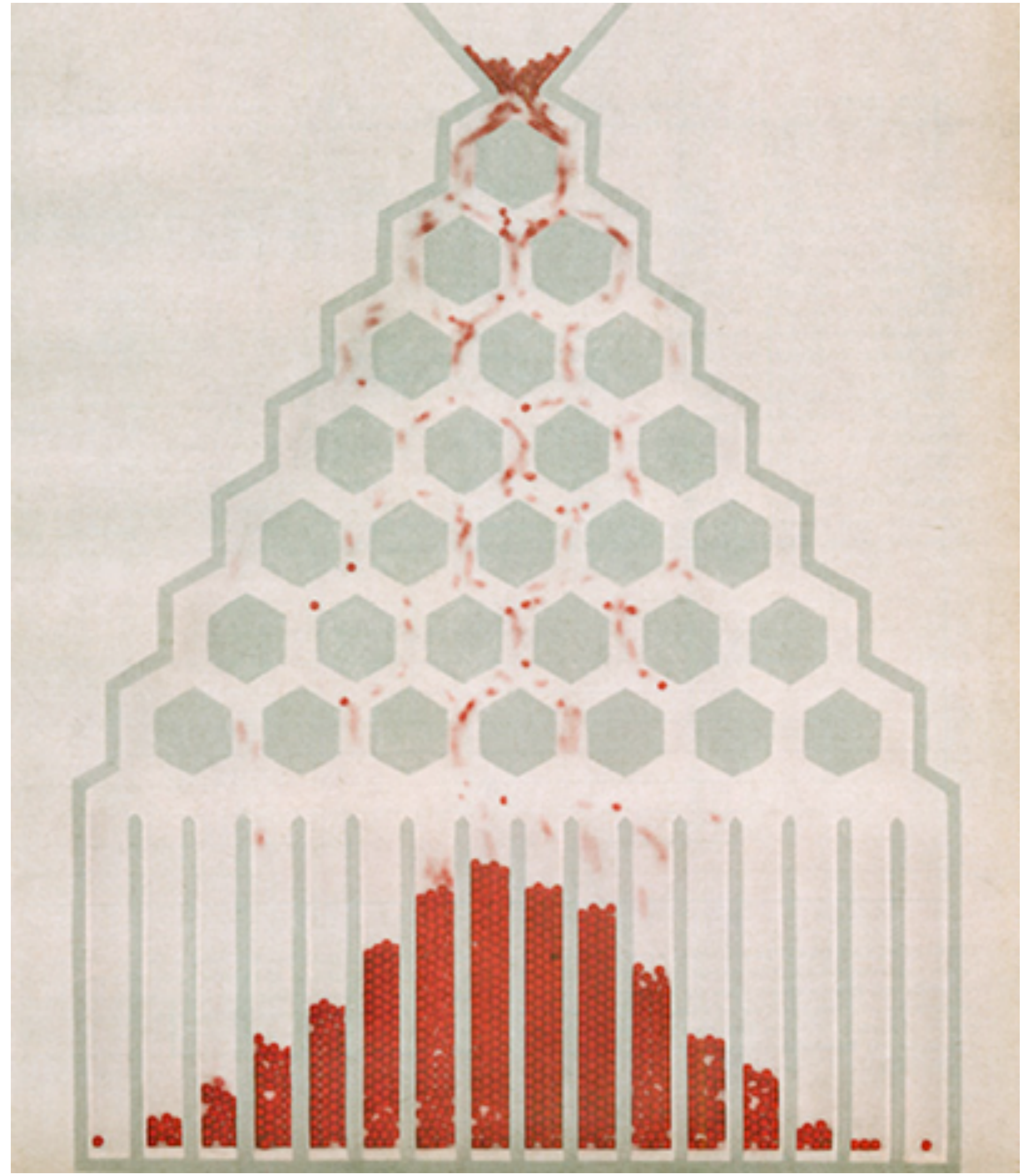


position
step number
0
4
8
12
16
6
3
0
-3
-6
4 steps
8 steps
16 steps
Density
position
-6
-3
0
3
6
0.0
0.1
0.2
0.3
0.00
0.10
0.20

Figure 4.2

Why normal?

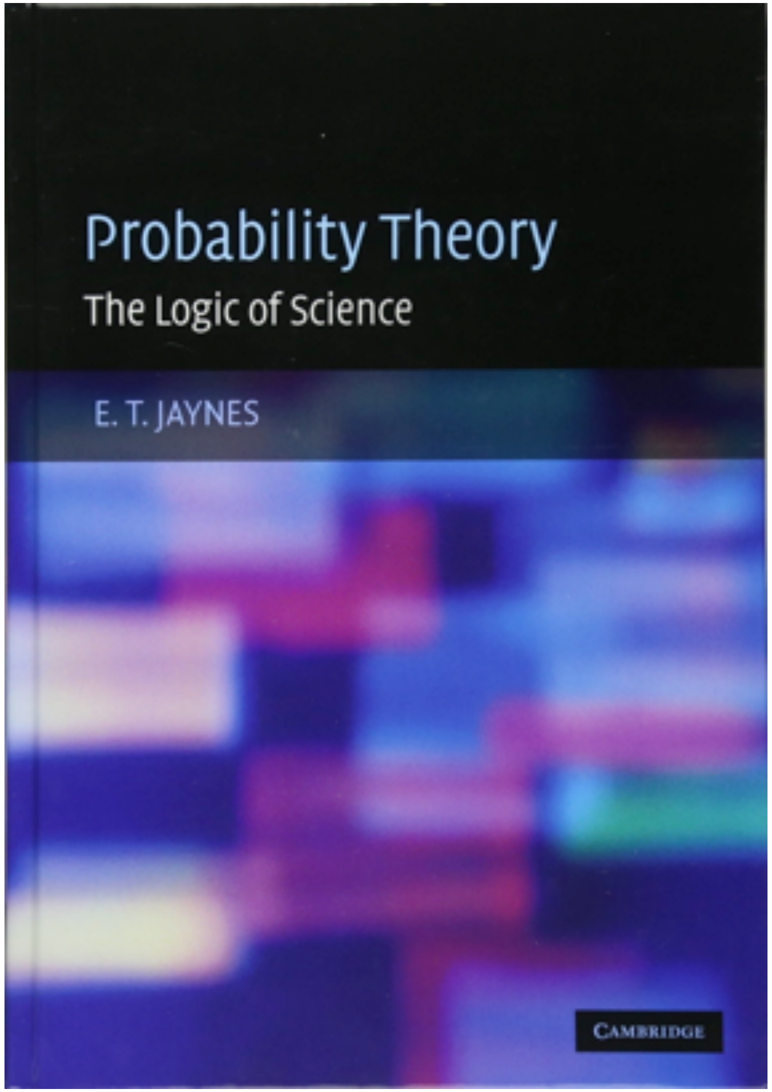
- Processes that produce normal distributions
  - Addition
  - Products of small deviations
  - Logarithms of products



Francis Galton's 1894 "bean machine" for simulating normal distributions

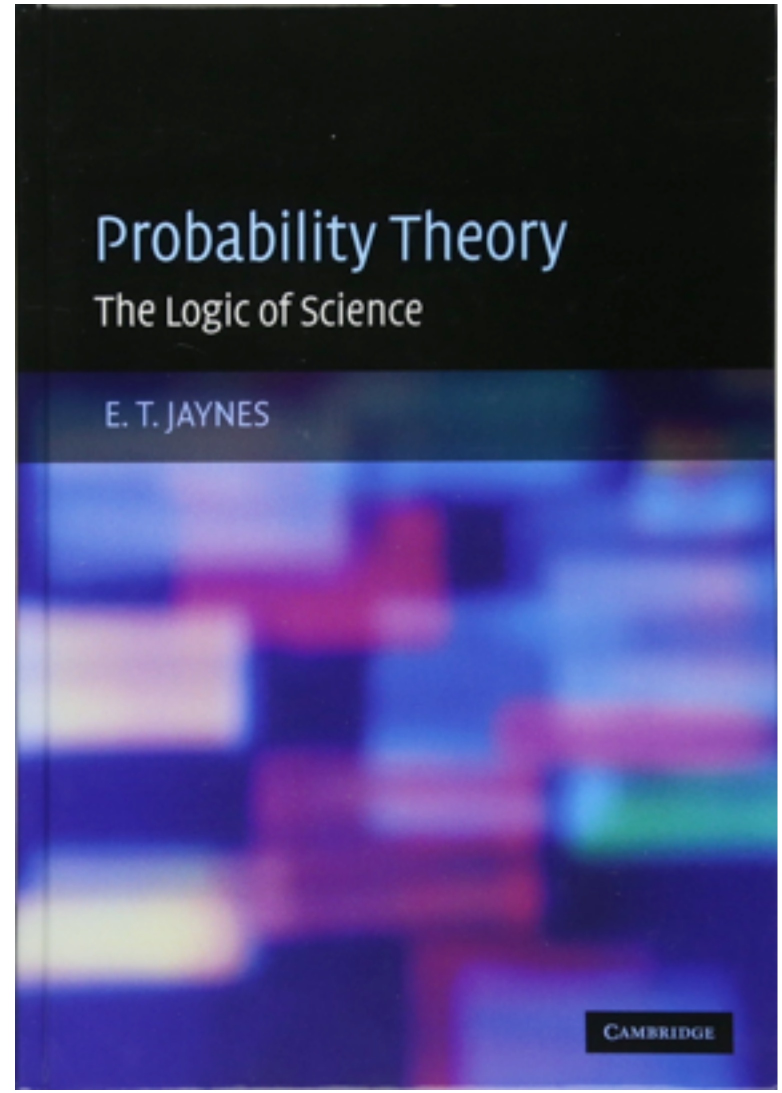
Why normal?

- Ontological perspective
    - Processes which add fluctuations result in dampening
    - Damped fluctuations end up Gaussian
    - No information left, except mean and variance
    - Can't infer process from distribution!



Why normal?

- Ontological perspective
  - Processes which add fluctuations result in dampening
  - Damped fluctuations end up Gaussian
  - No information left, except mean and variance
  - Can't infer process from distribution!
- Epistemological perspective
  - Know only *mean* and *variance*
  - Then least surprising and most conservative (*maximum entropy*) distribution is Gaussian
  - Nature likes maximum entropy distributions



Linear models

- Models of normally distributed data common
  - "General Linear Model": *t*-test, single regression, multiple regression, ANOVA, ANCOVA, MANOVA, MANCOVA, yadda yadda yadda
  - All the same thing
- Learn strategy, not procedure

Willard Boepple

Language for modeling

- Revisit globe tossing model:

"is distributed"

outcome

data distribution (likelihood)

$$W \sim \text{Binomial}(N, p)$$

$$p \sim \text{Uniform}(0, 1)$$

parameter to estimate

prior distribution

Language for modeling

Define the variables:

$$y_i$$

$$\mu_i$$

$$\beta$$

$$\sigma$$

$$x_i$$

Language for modeling

Give them definitions:

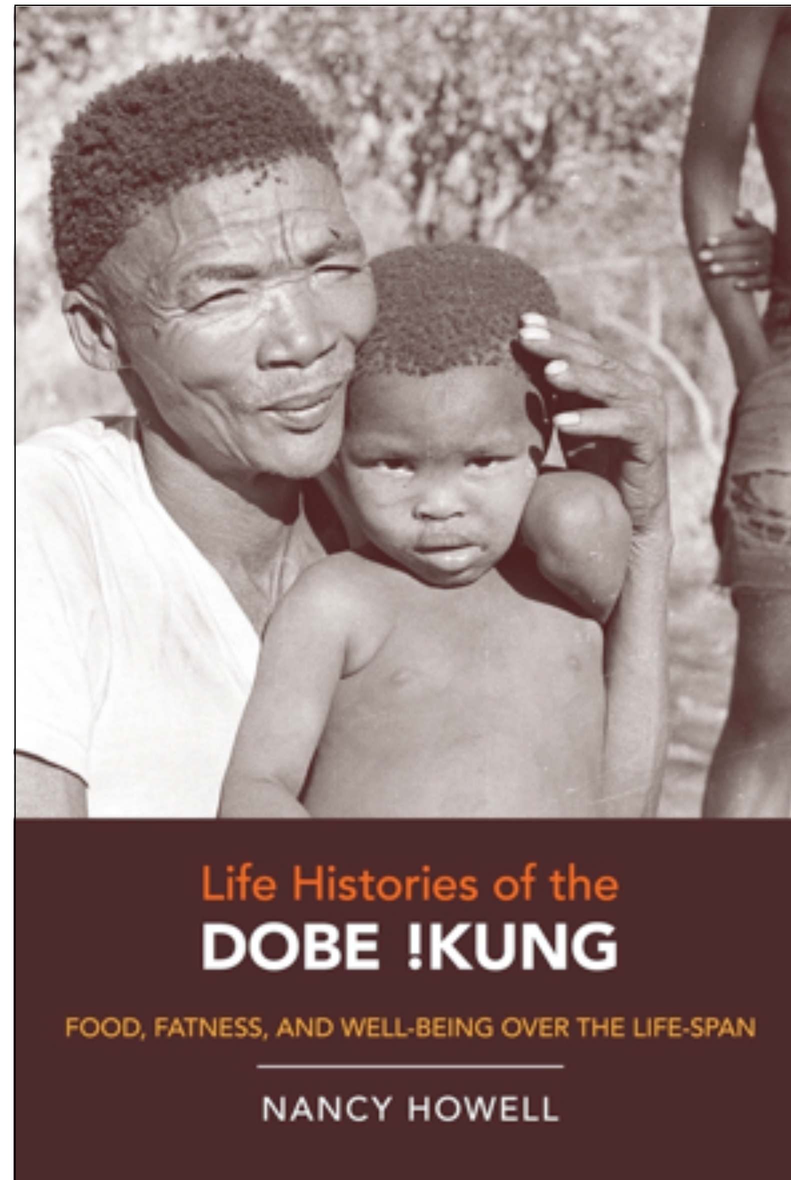
$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \beta x_i \\
\beta &\sim \text{Normal}(0, 10) \\
\sigma &\sim \text{Exponential}(1) \\
x_i &\sim \text{Normal}(0, 1)
\end{aligned}
$$

Some data: Kalahari foragers

```
library(rethinking)
data(Howell1)
d <- Howell1
```

```
precis( d )
```

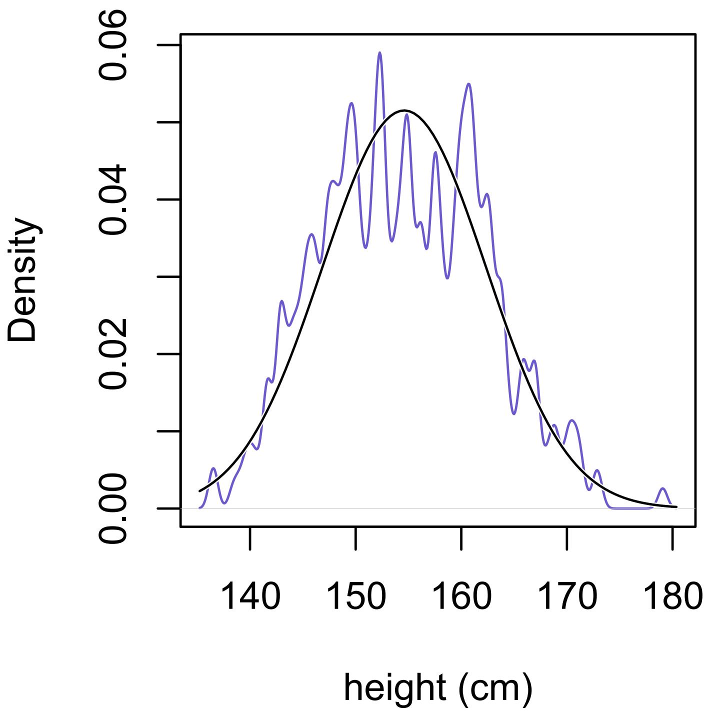
```
'data.frame': 544 obs. of 4 variables:
         mean    sd  5.5%  94.5%     histogram
height 138.26 27.60 81.11 165.74
weight  35.61 14.72  9.36  54.50
age     29.34 20.75  1.00  66.13
male     0.47  0.50  0.00   1.00
```



Gaussian model

- A first model:

$$h_i \sim \text{Normal}(\mu, \sigma)$$



Gaussian model

mean

"is distributed"

outcome

standard deviation

$$h_i \sim \text{Normal}(\mu, \sigma)$$

Height h_i of an individual i is distributed normally, with mean *mu* and standard deviation *sigma*.

Gaussian model

- Add priors:

$$h_i \sim \text{Normal}(\mu, \sigma)$$
$$\mu \sim \text{Normal}(178, 20)$$
$$\sigma \sim \text{Uniform}(0, 50)$$

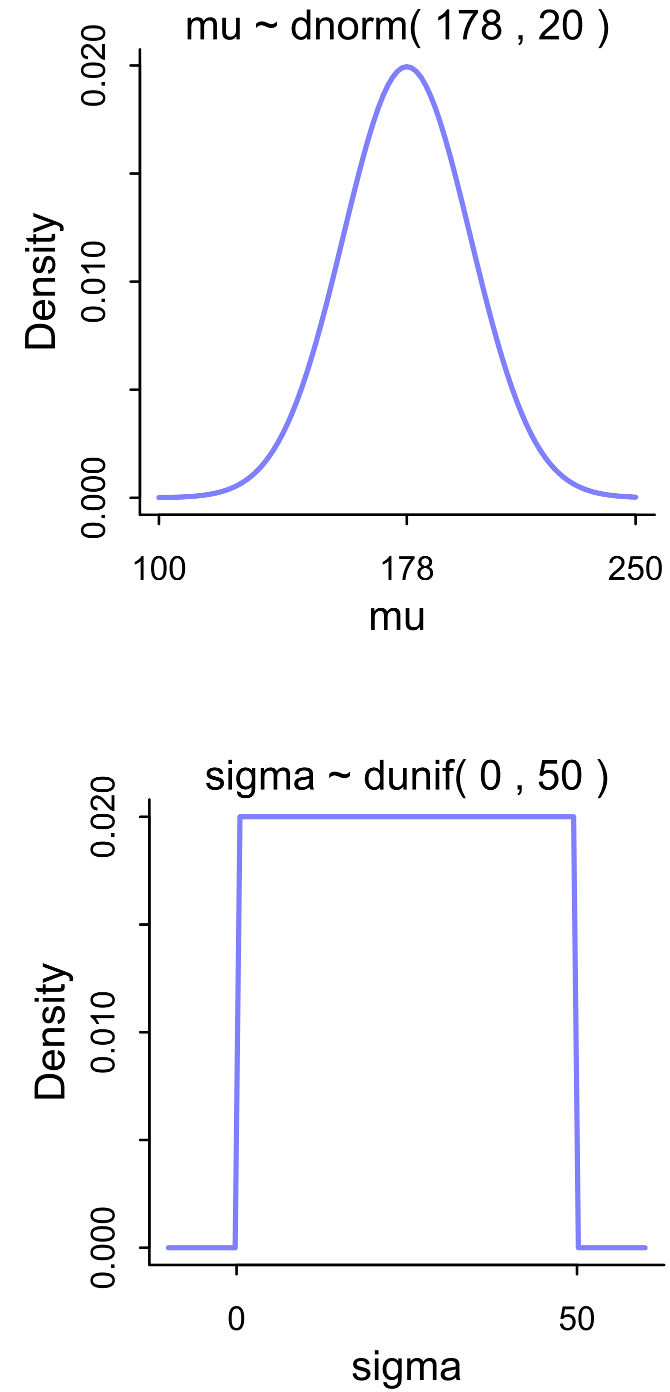


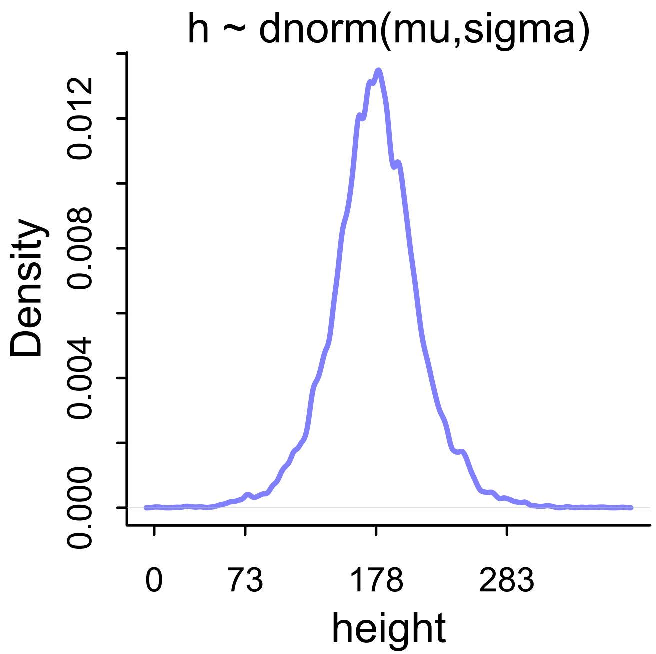
Figure 4.3

Prior predictive distribution

- What do these priors imply about height, before we see data? Simulate! => *prior predictive distribution*

```
sample_mu <- rnorm( 1e4 , 178 , 20 )
sample_sigma <- runif( 1e4 , 0 , 50 )
prior_h <- rnorm( 1e4 , sample_mu , sample_sigma )
dens( prior_h )
```

R code 4.14



100 cm = 3.3 feet
200 cm = 6.5 feet

Figure 4.3

Prior predictive distribution

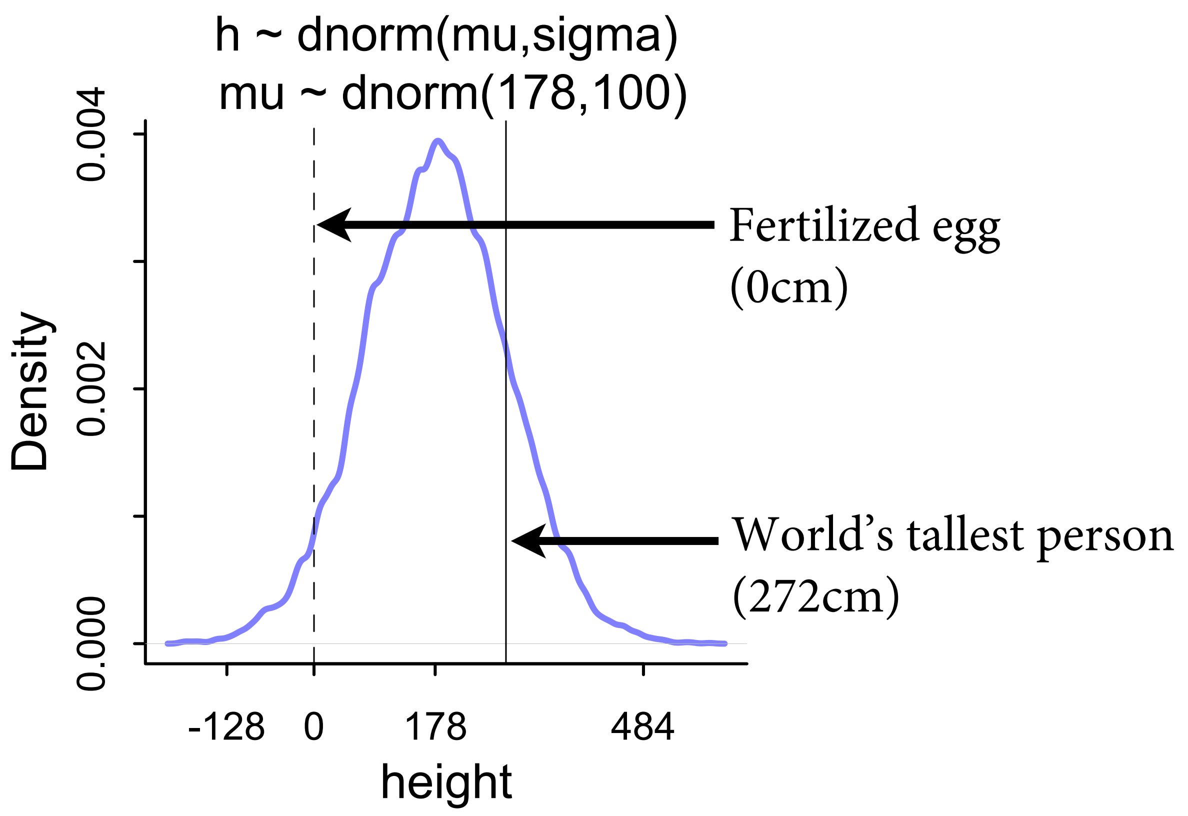
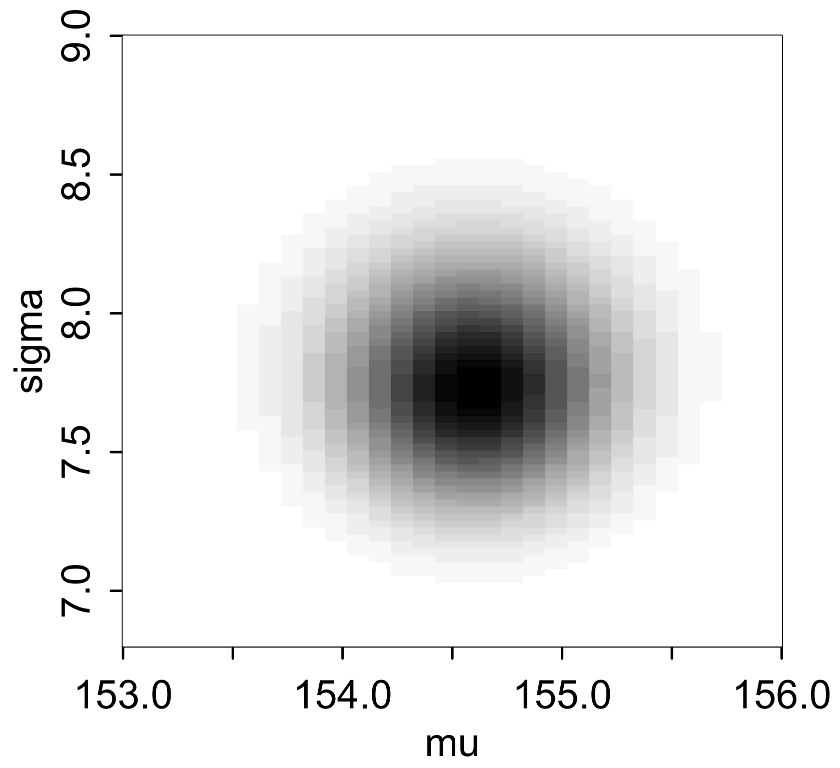


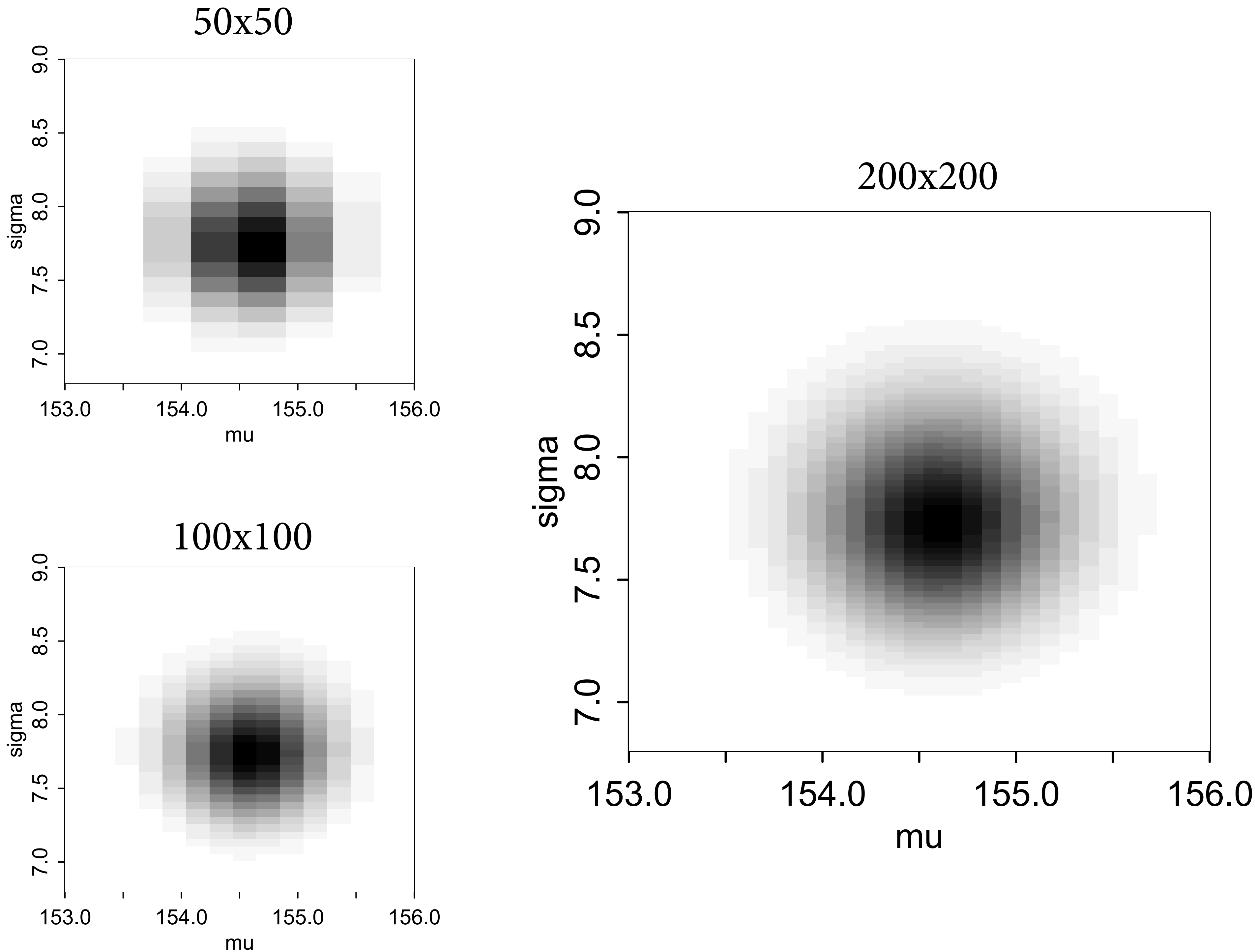
Figure 4.3

Computing the posterior

- Aim for the posterior distribution, which is now 2-dimensional

- Grid approximation: Compute posterior for many combinations of *mu* and *sigma*




50x50
100x100
200x200
sigma
mu
9.0
8.5
8.0
7.5
7.0
153.0
154.0
155.0
156.0

Drawing samples to work with

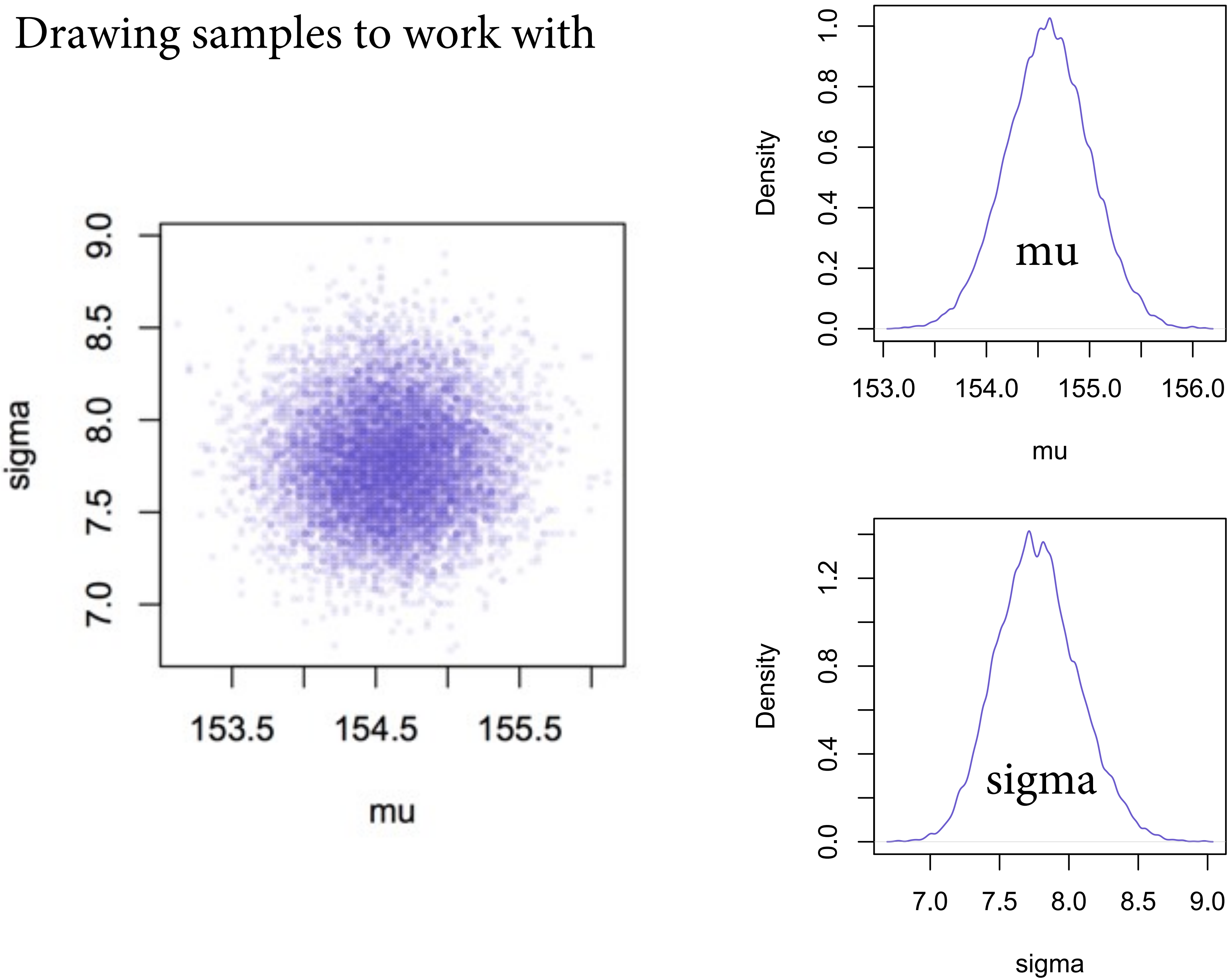


Figure 4.4

Quadratic approximation

- Approximate posterior as Gaussian
- Can estimate with two things:
  - Peak of posterior, *maximum a posteriori* (MAP)
  - Standard deviations & correlations of parameters (covariance matrix)
- With flat priors, same as conventional *maximum likelihood estimation*

Using quap

R code 4.27

```
flist <- alist(
    height ~ dnorm( mu , sigma ) ,
    mu ~ dnorm( 178 , 20 ) ,
    sigma ~ dunif( 0 , 50 )
)
```

$$h_i \sim \text{Normal}(\mu, \sigma)$$
$$\mu \sim \text{Normal}(178, 20)$$
$$\sigma \sim \text{Uniform}(0, 50)$$

Using quap

R code 4.27

```
flist <- alist(
    height ~ dnorm( mu , sigma ) ,
    mu ~ dnorm( 178 , 20 ) ,
    sigma ~ dunif( 0 , 50 )
)
```

$$h_i \sim \text{Normal}(\mu, \sigma)$$
$$\mu \sim \text{Normal}(178, 20)$$
$$\sigma \sim \text{Uniform}(0, 50)$$

R code 4.28

```
m4.1 <- quap( flist , data=d2 )
```

Using quap

R code 4.27

```
flist <- alist(
    height ~ dnorm( mu , sigma ) ,
    mu ~ dnorm( 178 , 20 ) ,
    sigma ~ dunif( 0 , 50 )
)
```

$$h_i \sim \text{Normal}(\mu, \sigma)$$
$$\mu \sim \text{Normal}(178, 20)$$
$$\sigma \sim \text{Uniform}(0, 50)$$

R code 4.28

```
m4.1 <- quap( flist , data=d2 )
```

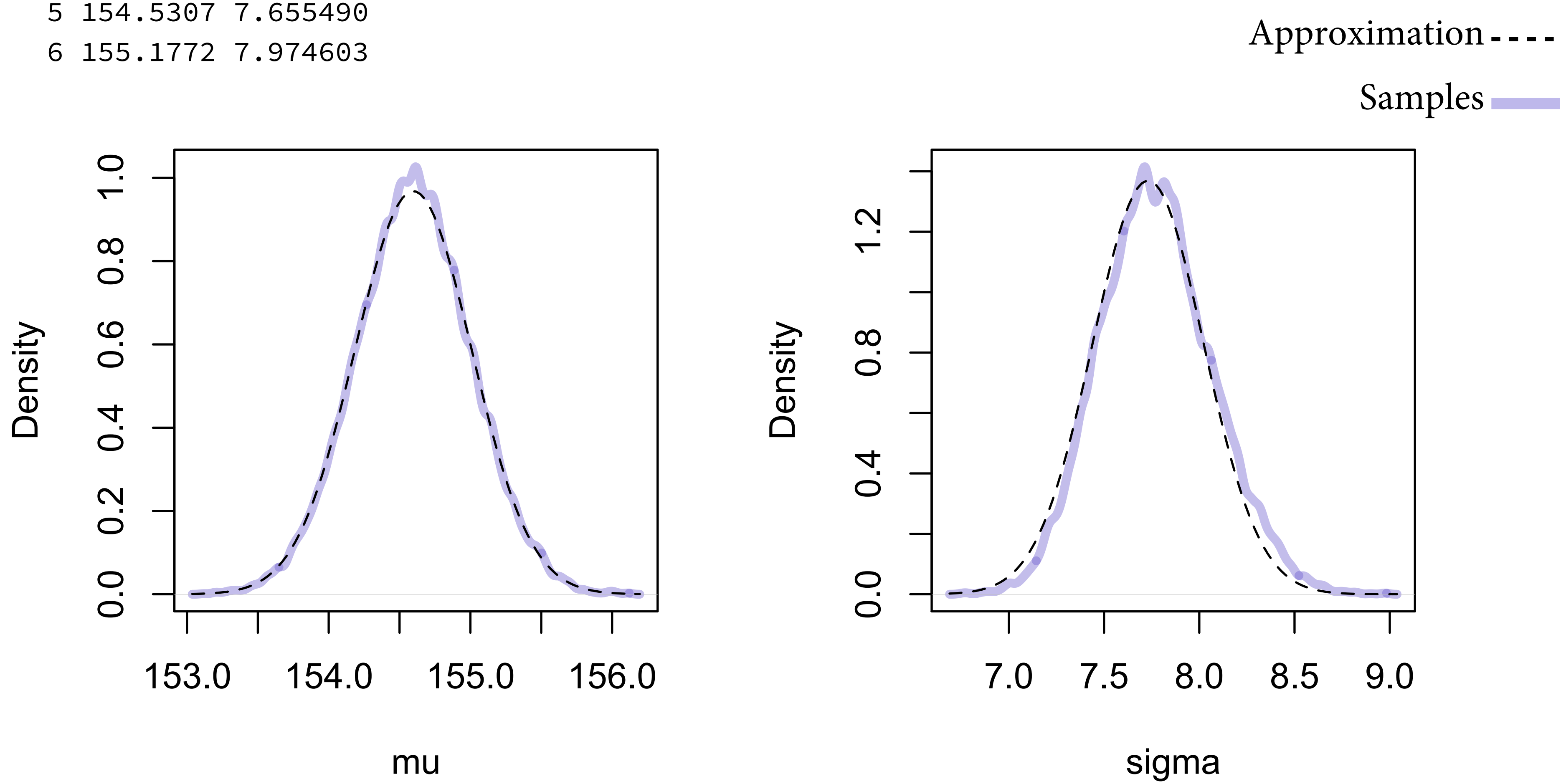
```
precis( m4.1 )
```

```
        Mean StdDev   5.5%  94.5%
mu    154.61   0.41 153.95 155.27
sigma   7.73   0.29   7.27   8.20
```

```
library(rethinking)
post <- extract.samples( m4.1 , n=1e4 )
head(post)
```

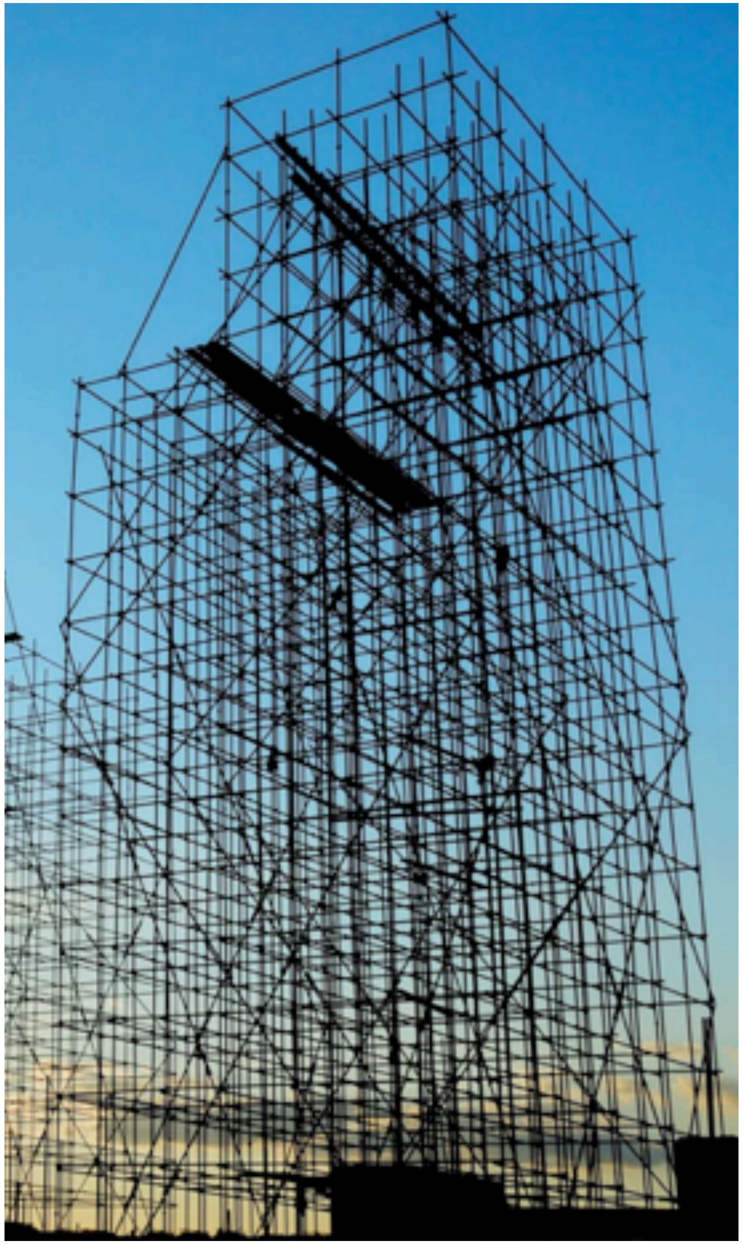
R code 4.34

```
        mu    sigma
1 155.0031 7.443893
2 154.0347 7.771255
3 154.9157 7.822178
4 154.4252 7.530331
5 154.5307 7.655490
6 155.1772 7.974603
```



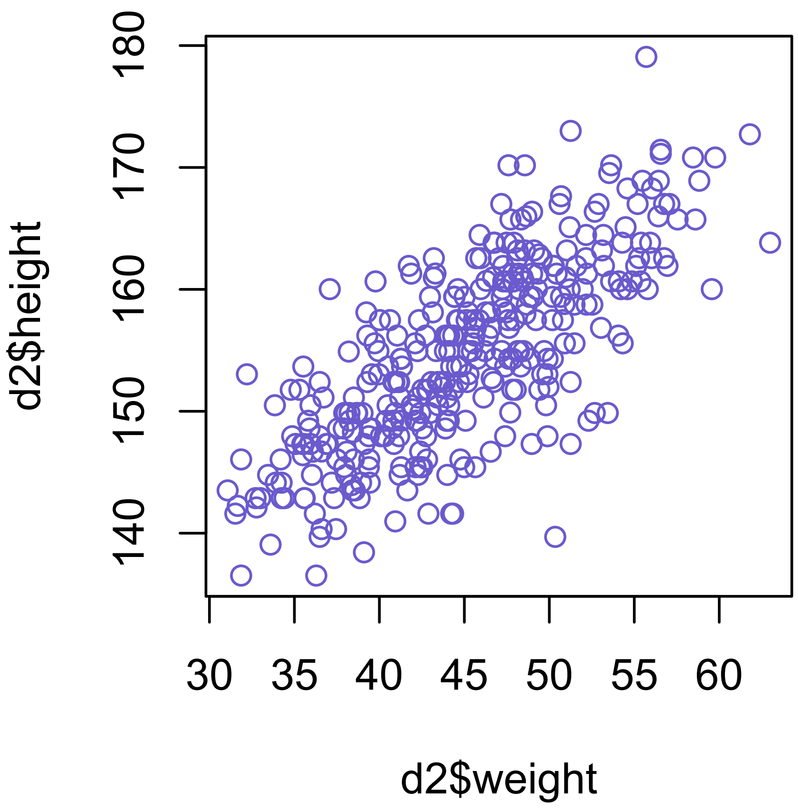
Scaffolds

- quap is a scaffold
  - Forces full specification of model, so you learn it
  - Works with a very wide class of models
  - Same as *penalized maximum likelihood*
  - Not always a good way to approximate posterior



Adding a predictor variable

- How does weight describe height?



Adding a predictor variable

- Use a linear model of the mean, *mu*:

$$
\begin{aligned}
h_i &\sim \text{Normal}(\mu_i, \sigma) && \text{[likelihood]} \\
\mu_i &= \alpha + \beta(x_i - \bar{x}) && \text{[linear model]} \\
\alpha &\sim \text{Normal}(178, 20) && [\alpha \text{ prior}] \\
\beta &\sim \text{Normal}(0, 10) && [\beta \text{ prior}] \\
\sigma &\sim \text{Uniform}(0, 50) && [\sigma \text{ prior}]
\end{aligned}
$$

Adding a predictor variable

$$h_i \sim \text{Normal}(\mu_i, \sigma)$$

mean on row i

$$\mu_i = \alpha + \beta(x_i - \bar{x})$$

mean weight

weight on row i

mean when $x_i = 0$
"intercept"

change in mean, per unit change x_i
"slope"

Prior predictive distribution

- What do these priors mean?
- Let's simulate to find out!

$$h_i \sim \text{Normal}(\mu_i, \sigma) \quad \text{[likelihood]}$$

$$\mu_i = \alpha + \beta(x_i - \bar{x}) \quad \text{[linear model]}$$

$$\alpha \sim \text{Normal}(178, 20) \quad [\alpha \text{ prior}]$$

$$\beta \sim \text{Normal}(0, 10) \quad [\beta \text{ prior}]$$

$$\sigma \sim \text{Uniform}(0, 50) \quad [\sigma \text{ prior}]$$

Prior predictive distribution

```
set.seed(2971)
N <- 100                        # 100 lines
a <- rnorm( N , 178 , 20 )
b <- rnorm( N , 0 , 10 )
```

$$\alpha \sim \text{Normal}(178, 20)$$

$$\beta \sim \text{Normal}(0, 10)$$

R code 4.38

Prior predictive distribution

```
set.seed(2971)
N <- 100                          # 100 lines
a <- rnorm( N , 178 , 20 )
b <- rnorm( N , 0 , 10 )
```

$$\alpha \sim \text{Normal}(178, 20)$$
$$\beta \sim \text{Normal}(0, 10)$$

R code 4.38

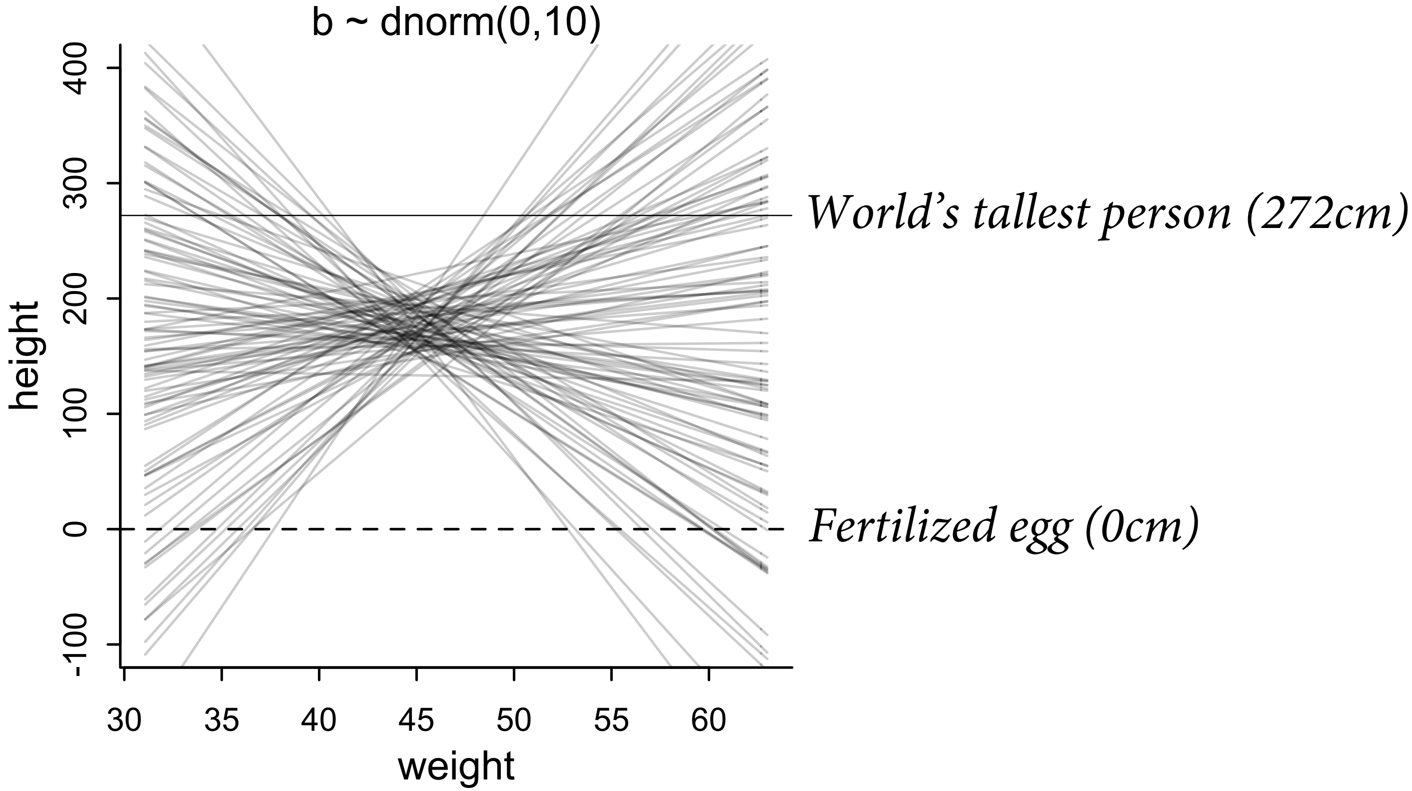


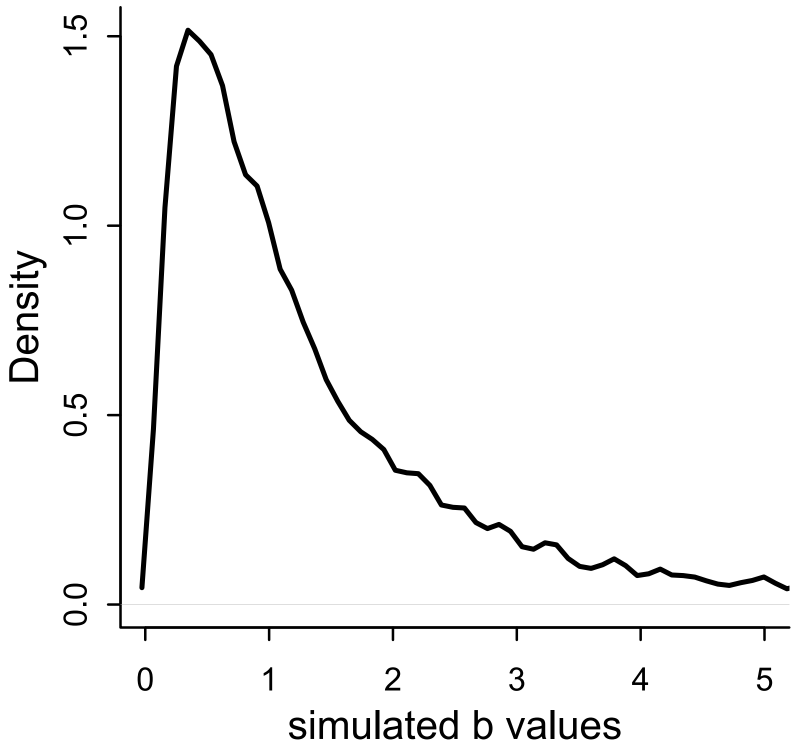
Figure 4.5

Prior predictive distribution

R code 4.40

```
b <- rlnorm( 1e4 , 0 , 1 )
dens( b , xlim=c(0,5) , adj=0.1 )
```

$$\beta \sim \text{Log-Normal}(0, 1)$$



Prior predictive distribution

R code 4.41

```
set.seed(2971)
N <- 100                    # 100 lines
a <- rnorm( N , 178 , 20 )
b <- rlnorm( N , 0 , 1 )
```

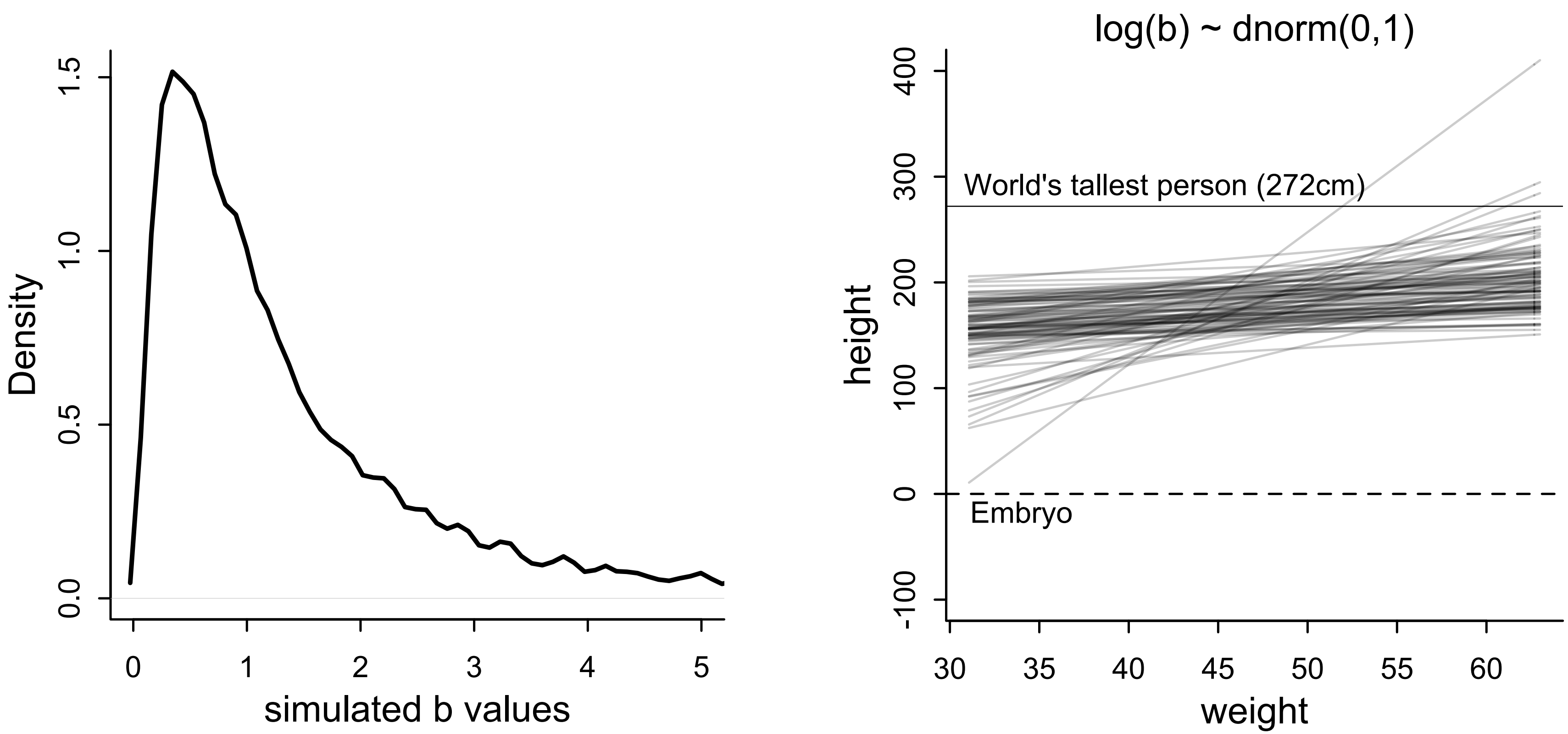


Figure 4.5

Approximate the posterior

R code 4.42

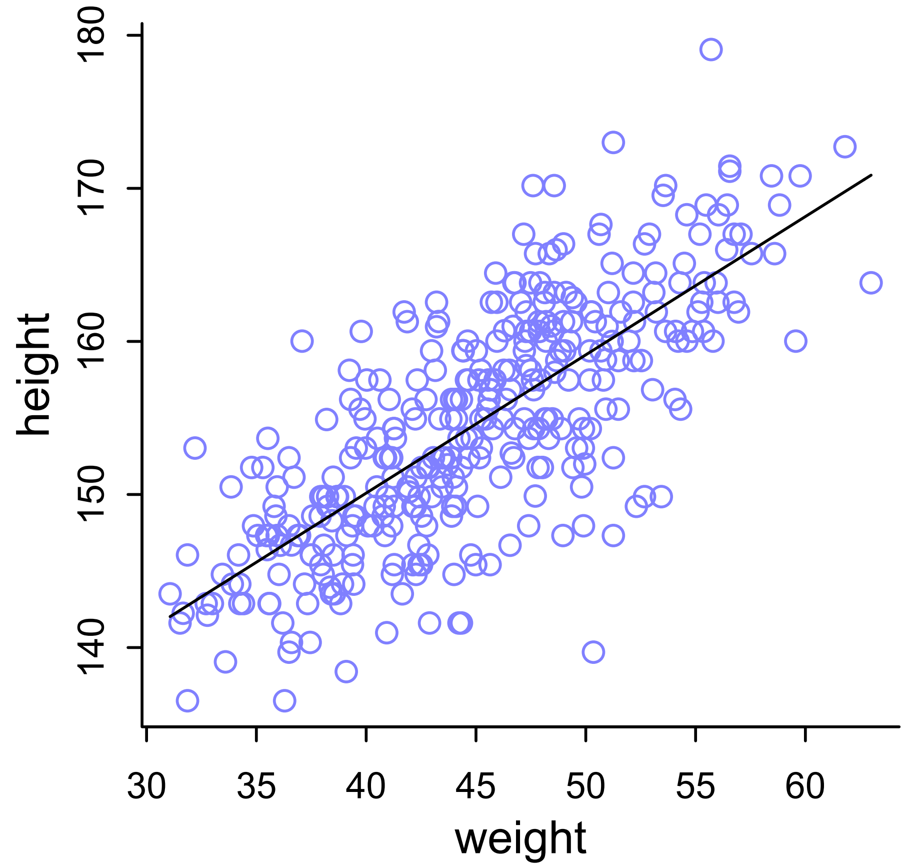
```
# load data again, since it's a long way back
library(rethinking)
data(Howell1)
d <- Howell1
d2 <- d[ d$age >= 18 , ]

# define the average weight, x-bar
xbar <- mean(d2$weight)

# fit model
m4.3 <- quap(
    alist(
        height ~ dnorm( mu , sigma ) ,
        mu <- a + b*( weight - xbar ) ,
        a ~ dnorm( 178 , 20 ) ,
        b ~ dlnorm( 0 , 1 ) ,
        sigma ~ dunif( 0 , 50 )
    ) ,
    data=d2 )
```

$$h_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta(x_i - \bar{x})$$
$$\alpha \sim \text{Normal}(178, 20)$$
$$\beta \sim \text{Log-Normal}(0, 1)$$
$$\sigma \sim \text{Uniform}(0, 50)$$

R code 4.44

```
precis( m4.3 )
```

```
       mean   sd   5.5%  94.5%
a     154.60 0.27 154.17 155.03
b       0.90 0.04   0.84   0.97
sigma   5.07 0.19   4.77   5.38
```



R code 4.46

```
plot( height ~ weight , data=d2 , col=rangi2 )
post <- extract.samples( m4.3 )
a_map <- mean(post$a)
b_map <- mean(post$b)
curve( a_map + b_map*(x - xbar) , add=TRUE )
```

Figure 4.6

Sampling from the posterior

- Want to get uncertainty onto that graph
- Again, sample from posterior
  1. Use mean and standard deviation to approximate posterior
  2. Sample from *multivariate normal* distribution of parameters
  3. Use samples to generate predictions that "integrate over" the uncertainty

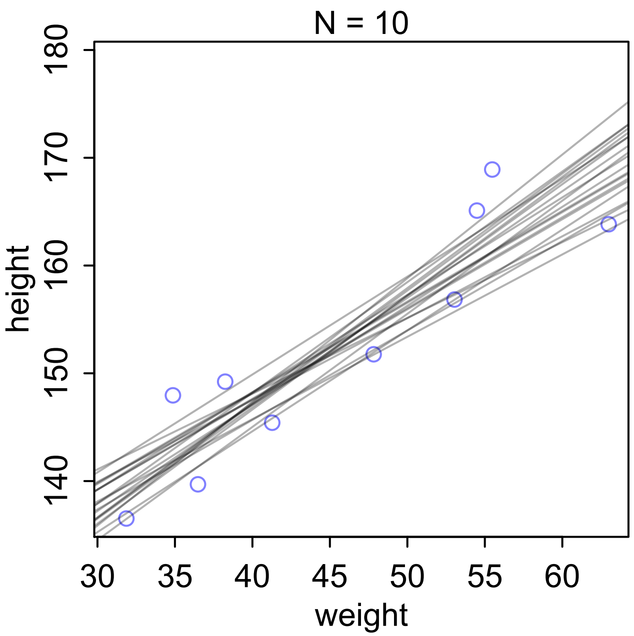
Sampling from the posterior

R code
4.47

```
post <- extract.samples( m4.3 )
post[1:5,]
```

```
         a         b    sigma
1 154.5505 0.9222372 5.188631
2 154.4965 0.9286227 5.278370
3 154.4794 0.9490329 4.937513
4 155.2289 0.9252048 4.869807
5 154.9545 0.8192535 5.063672
```

Posterior is full of lines



R code 4.47

```
post <- extract.samples( m4.3 )
post[1:5,]
```

```
         a         b    sigma
1 154.5505 0.9222372 5.188631
2 154.4965 0.9286227 5.278370
3 154.4794 0.9490329 4.937513
4 155.2289 0.9252048 4.869807
5 154.9545 0.8192535 5.063672
```

Figure 4.7

Posterior is full of lines

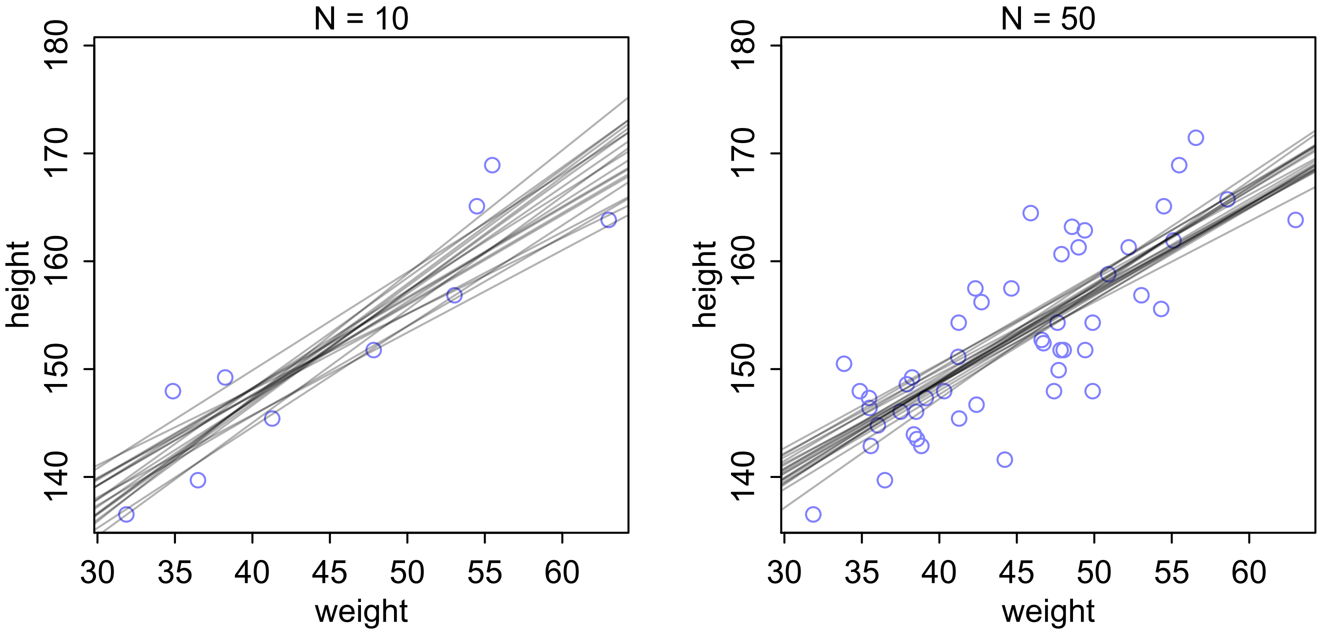


Figure 4.7

Posterior is full of lines

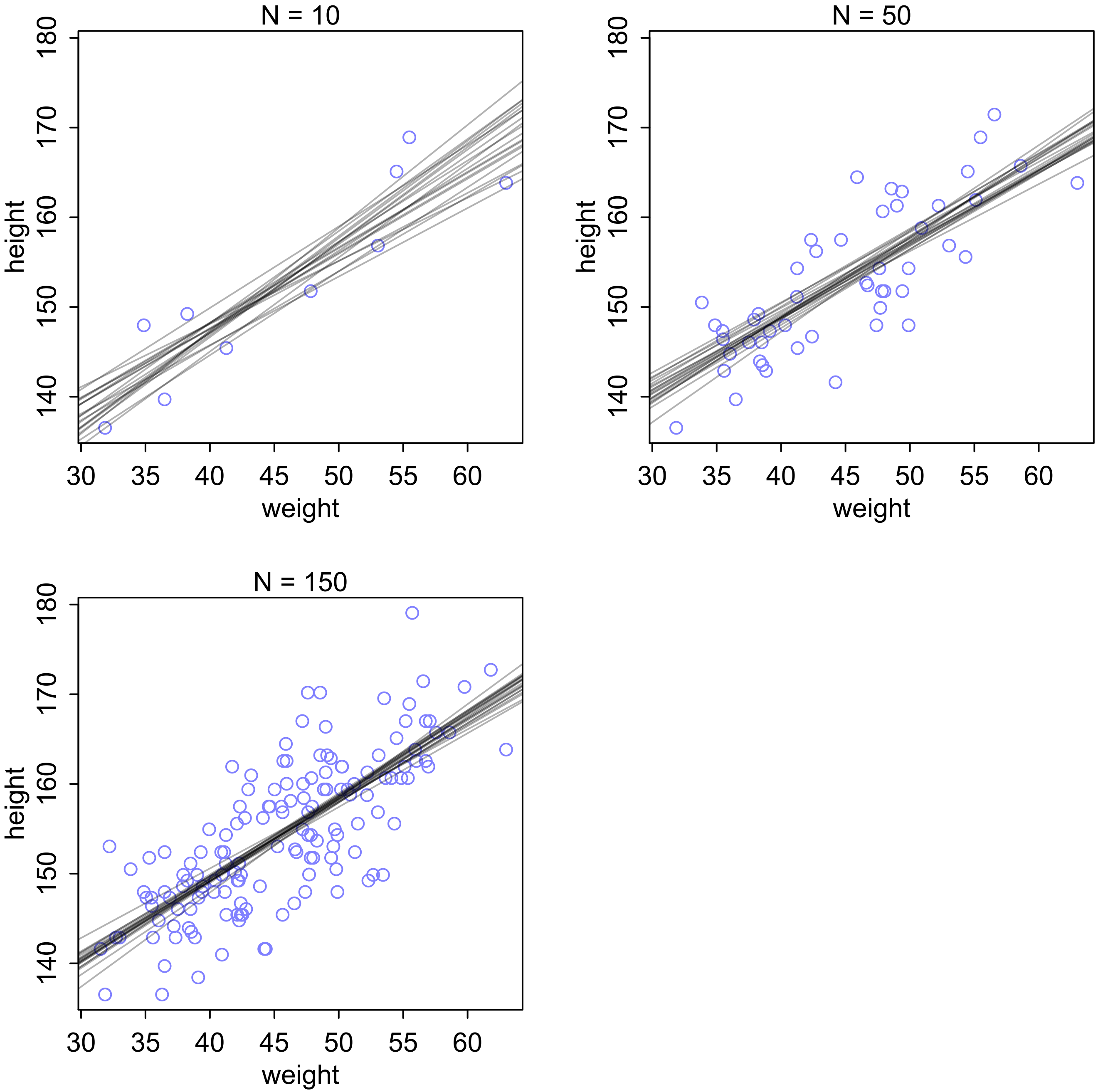


Figure 4.7

Posterior is full of lines

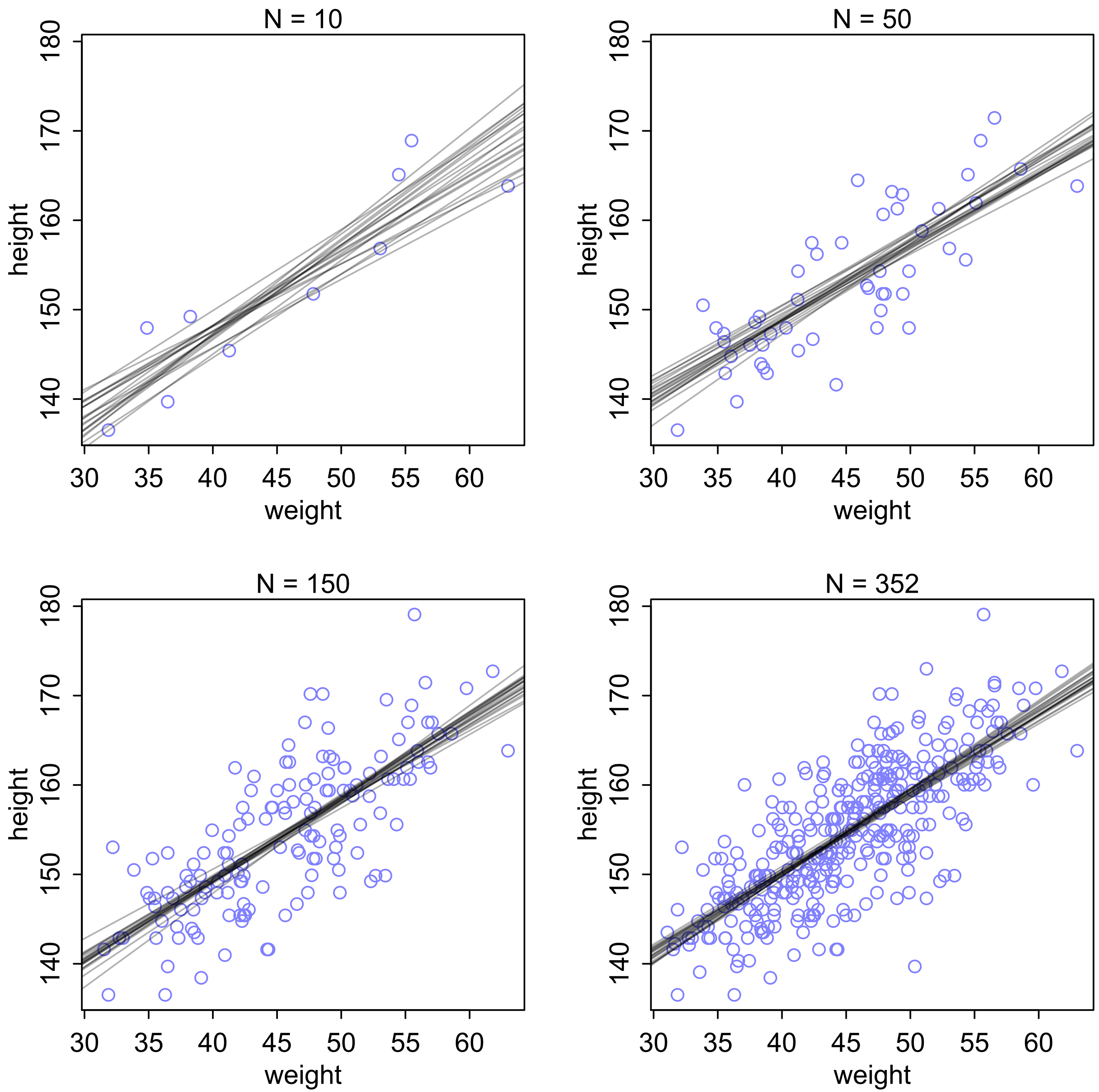


Figure 4.7