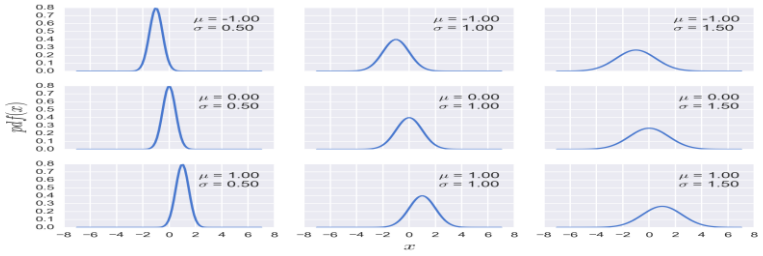
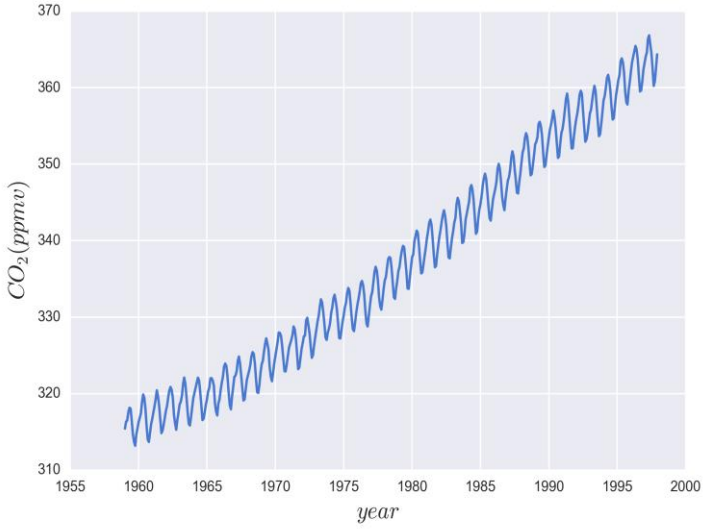
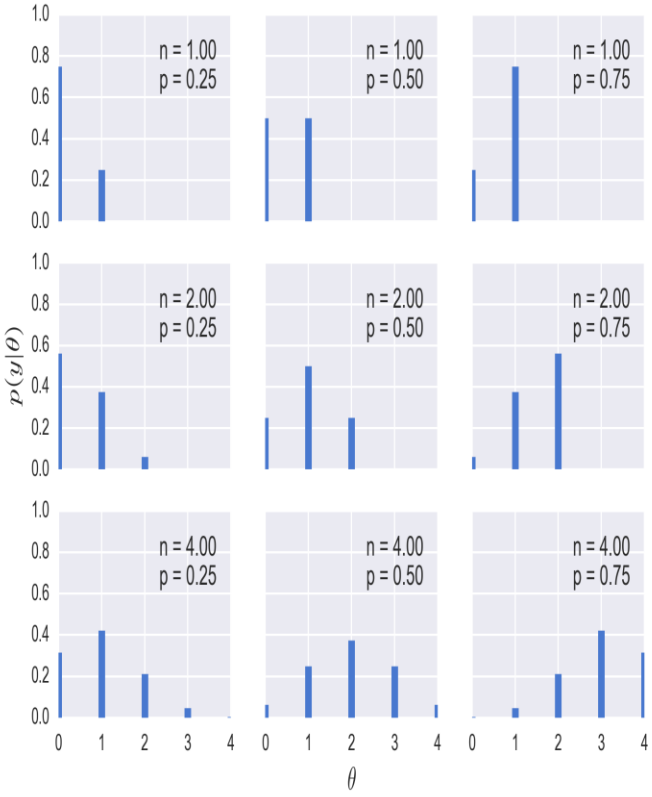
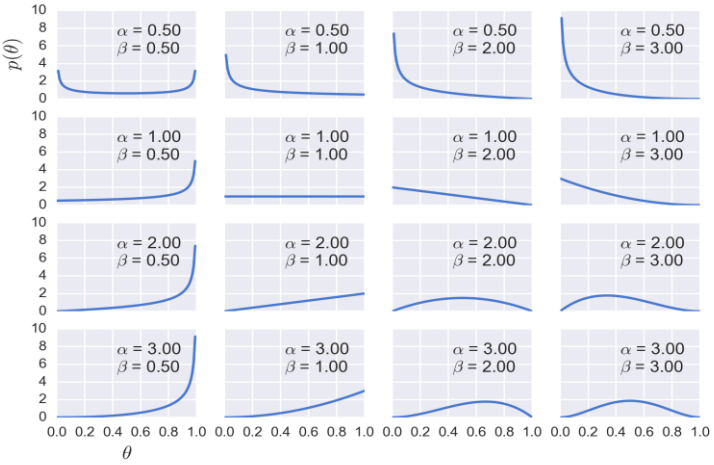
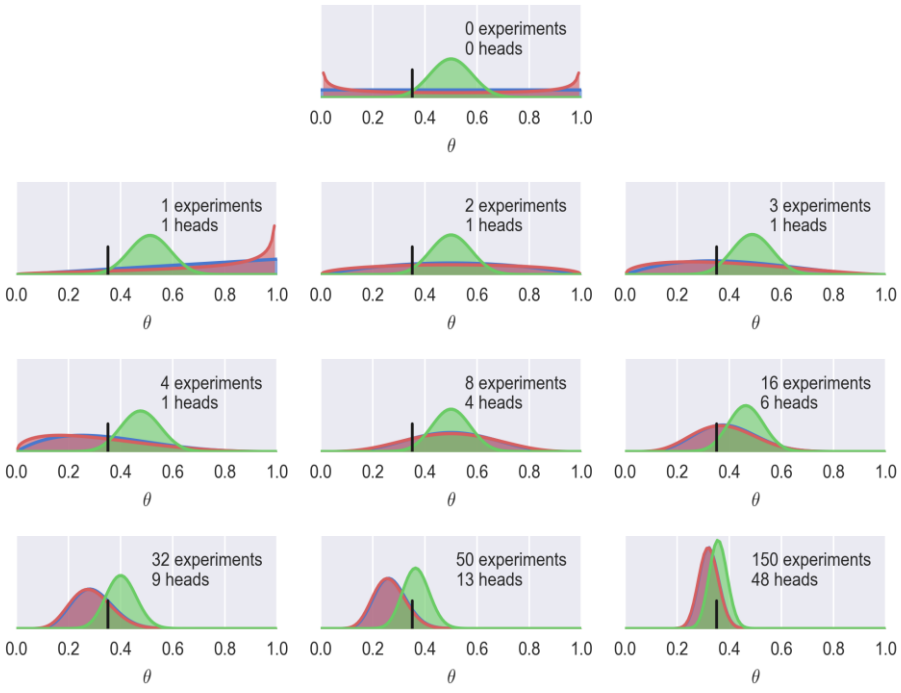
Chapter 1 – Thinking Probabilistically – A Bayesian Inference Primer

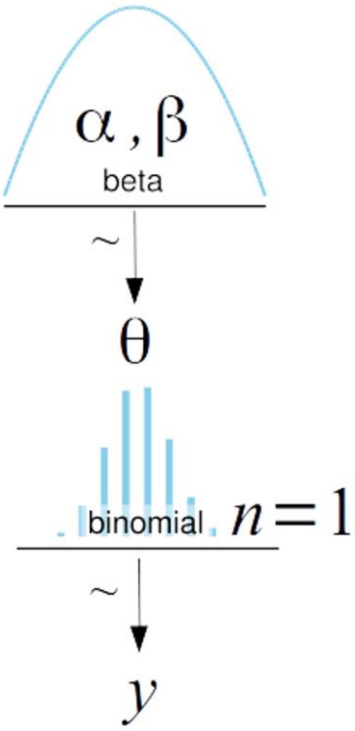


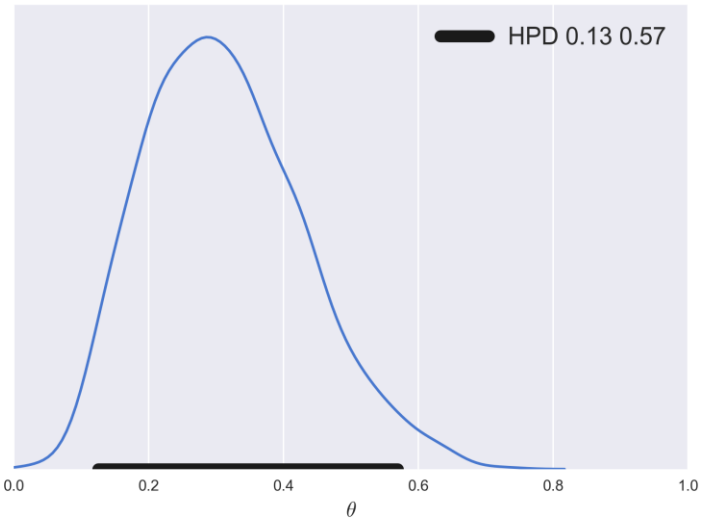


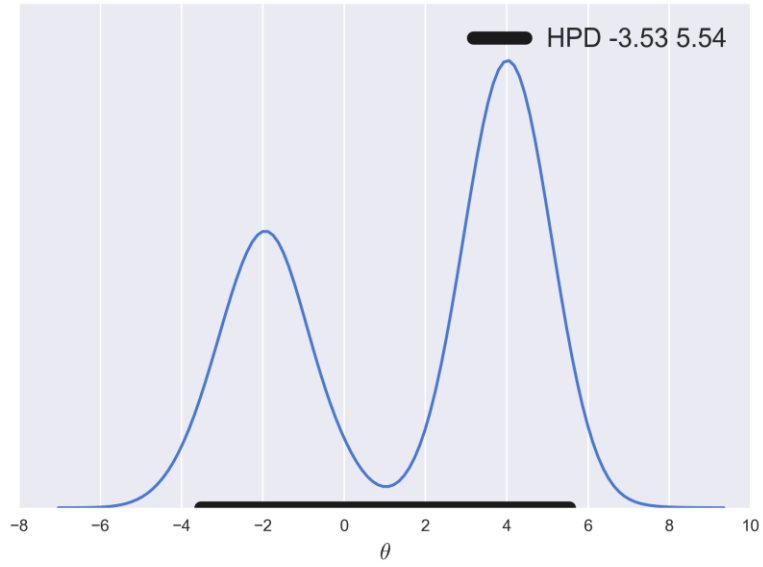

n = 1.00
p = 0.25
n = 1.00
p = 0.50
n = 1.00
p = 0.75
n = 2.00
p = 0.25
n = 2.00
p = 0.50
n = 2.00
p = 0.75
n = 4.00
p = 0.25
n = 4.00
p = 0.50
n = 4.00
p = 0.75
$p(y|\theta)$
θ
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4

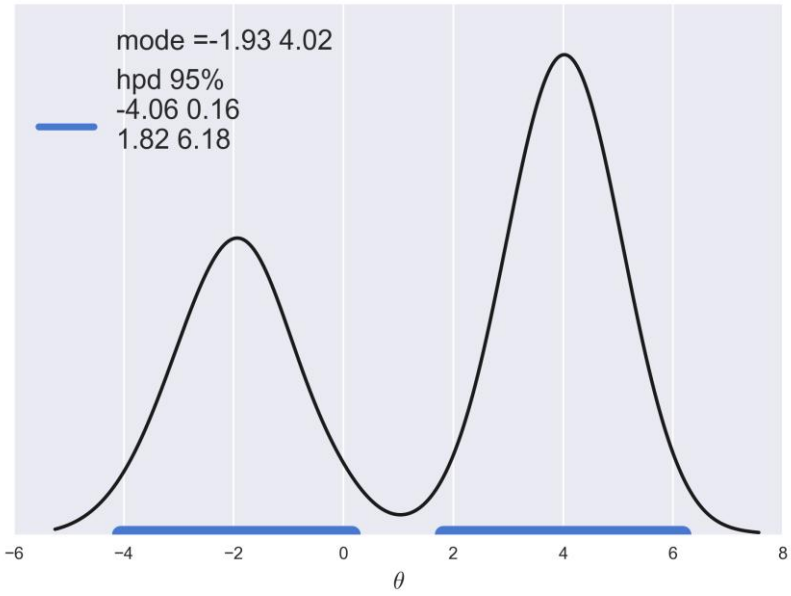

$p(\theta)$
α = 0.50 β = 0.50
α = 0.50 β = 1.00
α = 0.50 β = 2.00
α = 0.50 β = 3.00
α = 1.00 β = 0.50
α = 1.00 β = 1.00
α = 1.00 β = 2.00
α = 1.00 β = 3.00
α = 2.00 β = 0.50
α = 2.00 β = 1.00
α = 2.00 β = 2.00
α = 2.00 β = 3.00
α = 3.00 β = 0.50
α = 3.00 β = 1.00
α = 3.00 β = 2.00
α = 3.00 β = 3.00
θ


0 experiments
0 heads
1 experiments
1 heads
2 experiments
1 heads
3 experiments
1 heads
4 experiments
1 heads
8 experiments
4 heads
16 experiments
6 heads
32 experiments
9 heads
50 experiments
13 heads
150 experiments
48 heads
θ

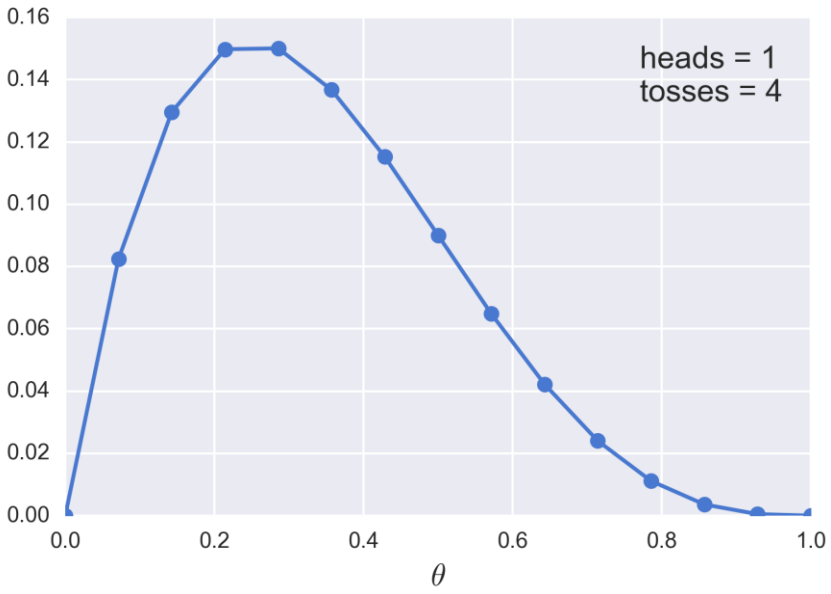

α , β
beta
~
θ
binomial
n = 1
~
y

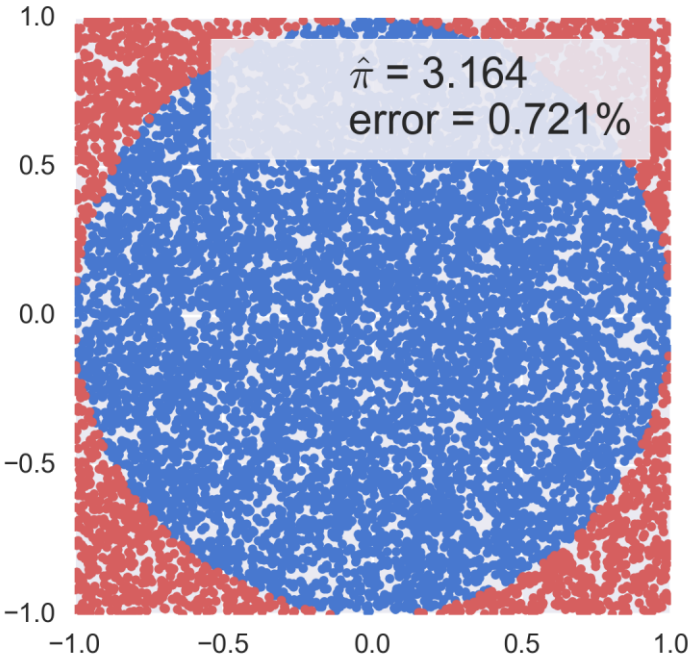

HPD 0.13 0.57
0.0
0.2
0.4
0.6
0.8
1.0
θ

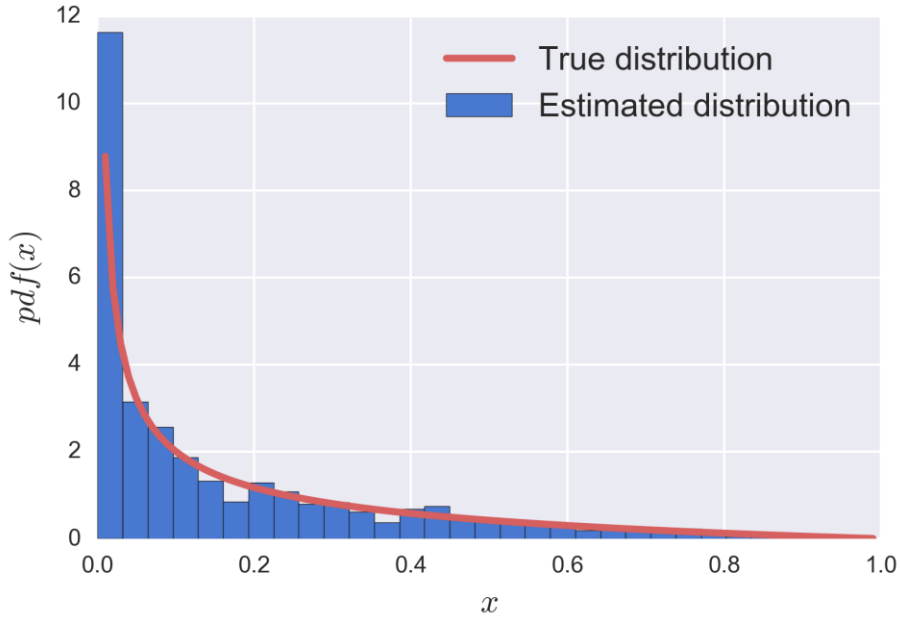

HPD -3.53 5.54
−8
−6
−4
−2
0
2
4
6
8
10
θ

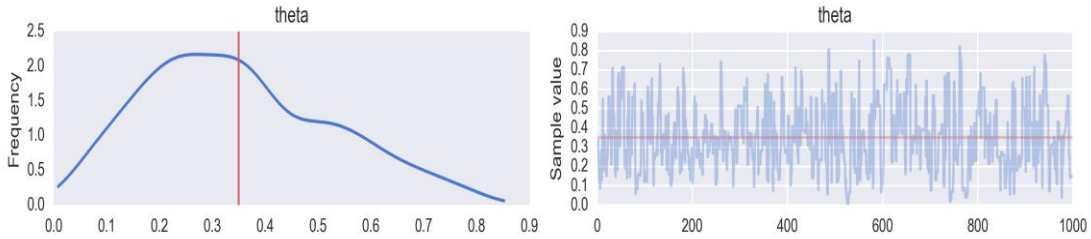

mode =-1.93 4.02
hpd 95%
-4.06 0.16
1.82 6.18
−6
−4
−2
0
2
4
6
8
θ

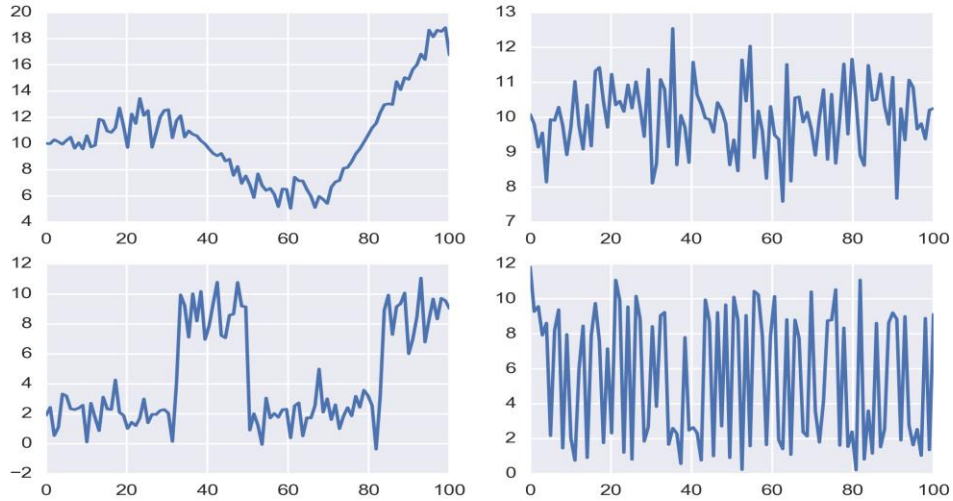
Chapter 2 – Programming Probabilistically – A PyMC3 Primer

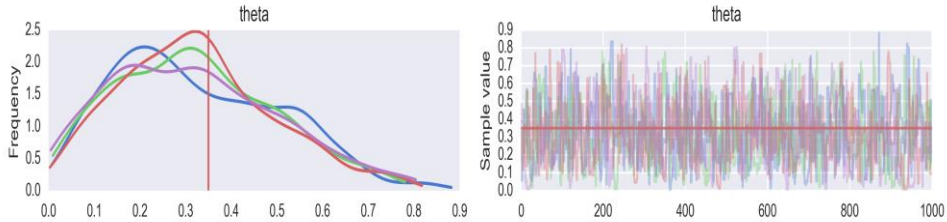


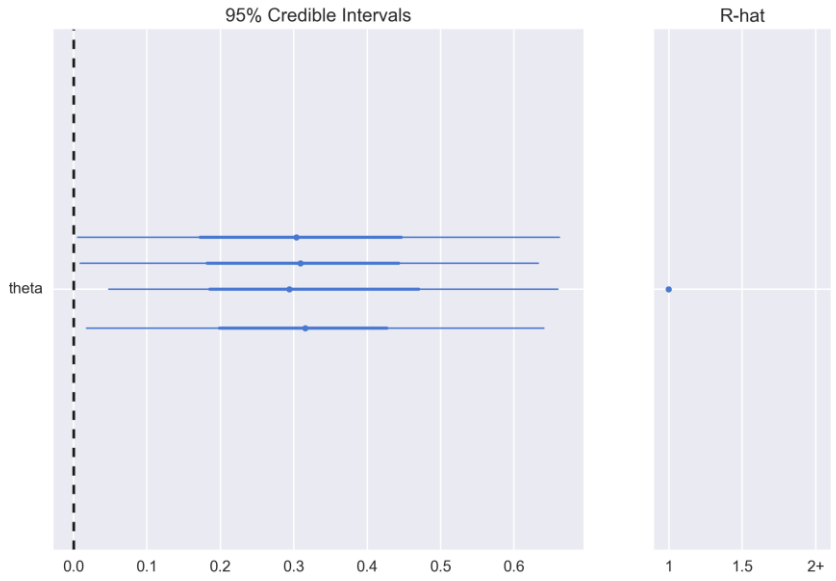

$\hat{\pi}$ = 3.164
error = 0.721%
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

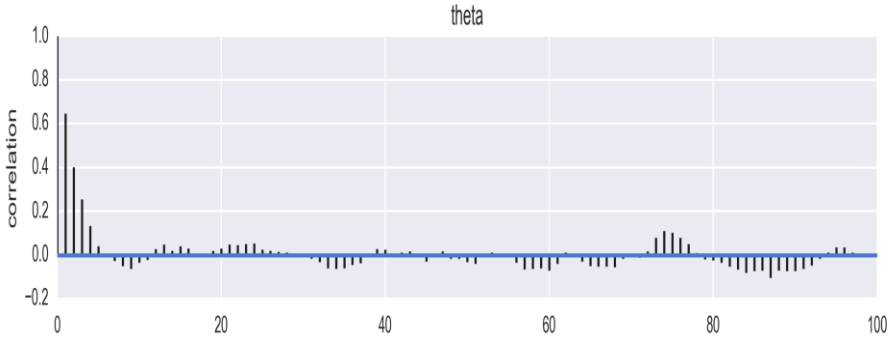

True distribution
Estimated distribution
$pdf(x)$
x
12
10
8
6
4
2
0
0.0
0.2
0.4
0.6
0.8
1.0

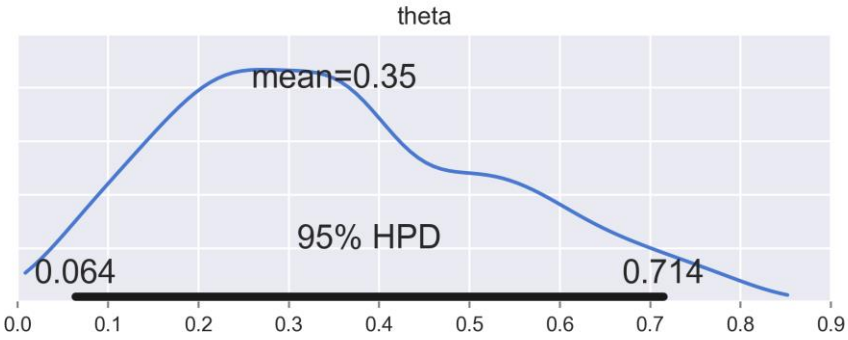

theta
Frequency
theta
Sample value

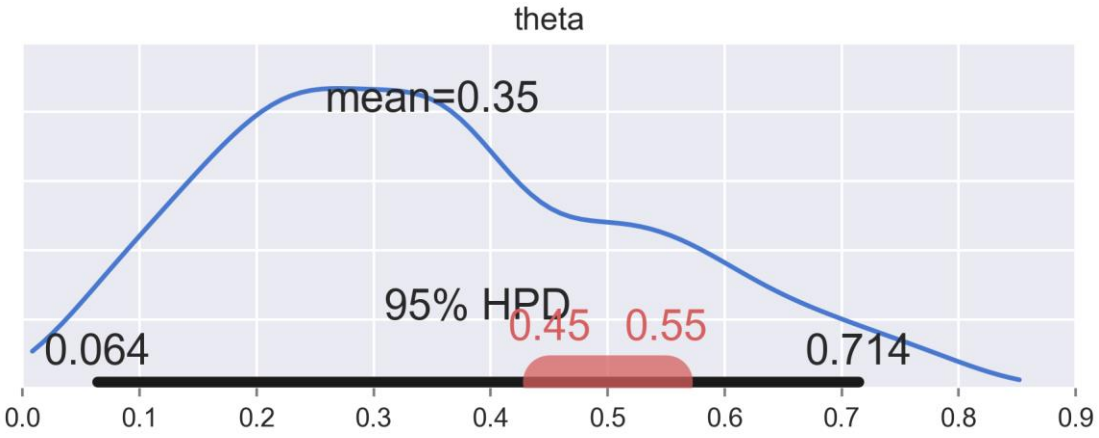


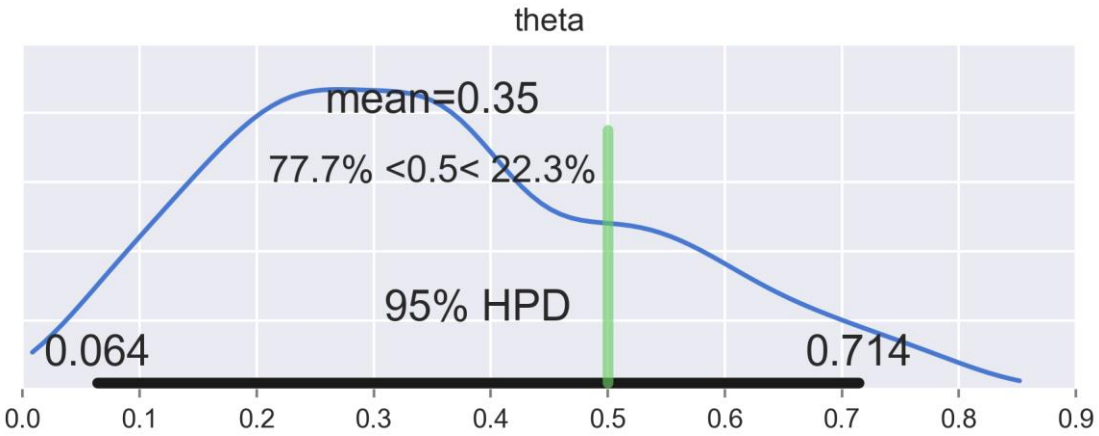

theta
Frequency
theta
Sample value


95% Credible Intervals
R-hat
theta
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1
1.5
2+

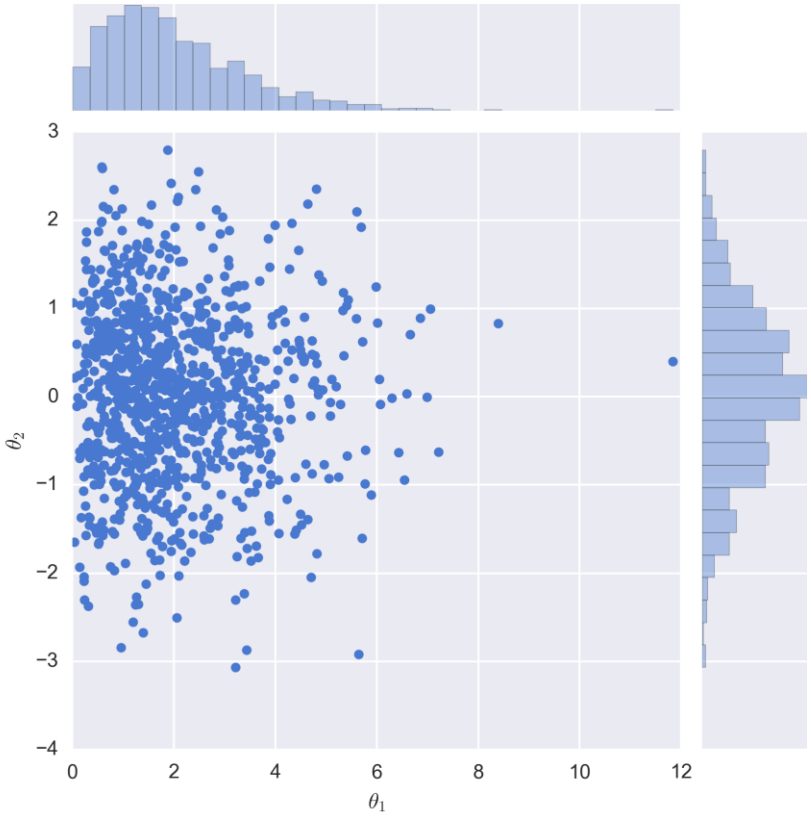

theta
correlation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
20
40
60
80
100

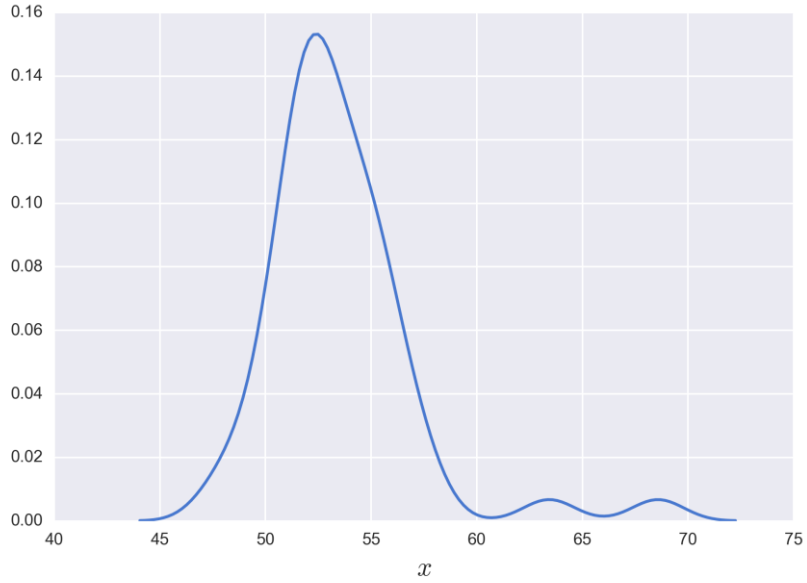

theta
mean=0.35
95% HPD
0.064
0.714
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

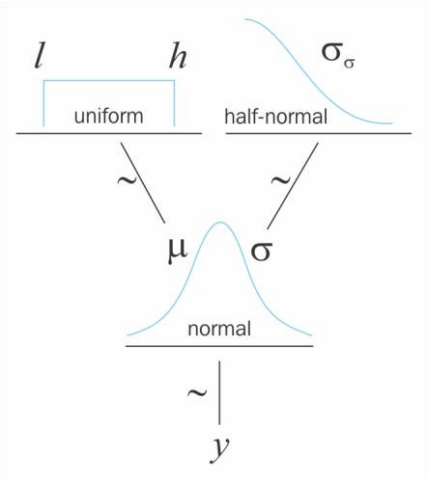

theta
mean=0.35
95% HPD
0.45
0.55
0.064
0.714
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

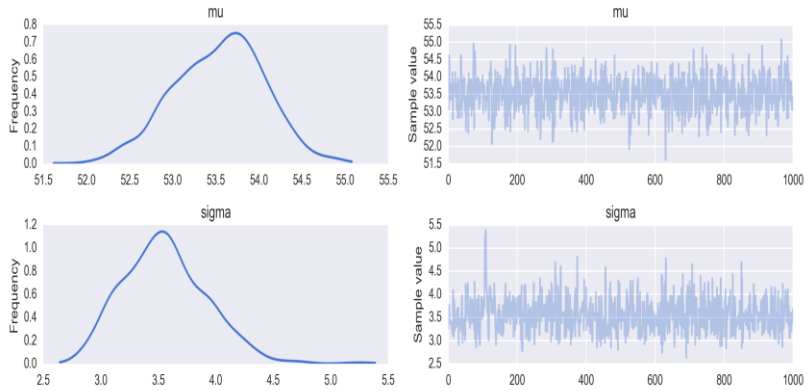

theta
mean=0.35
77.7% <0.5< 22.3%
95% HPD
0.064
0.714
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

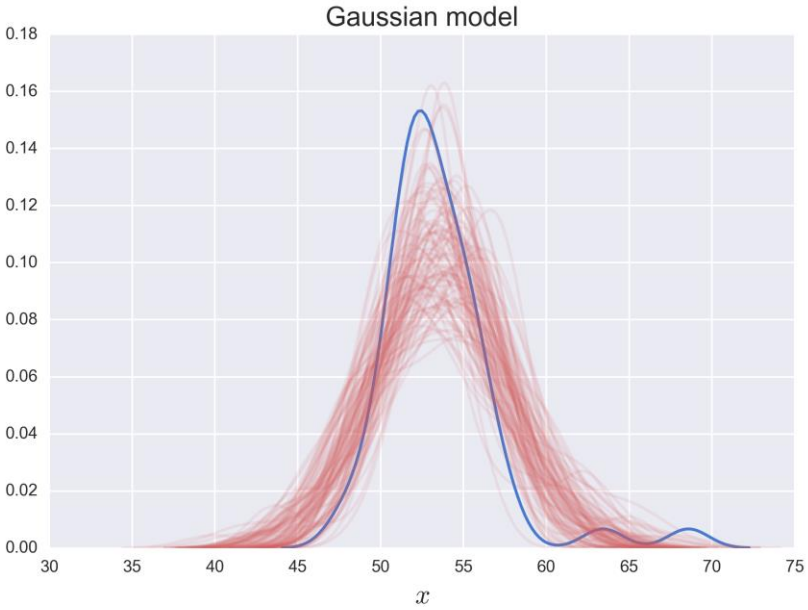
Chapter 3 – Juggling with Multi-Parametric and Hierarchical Models

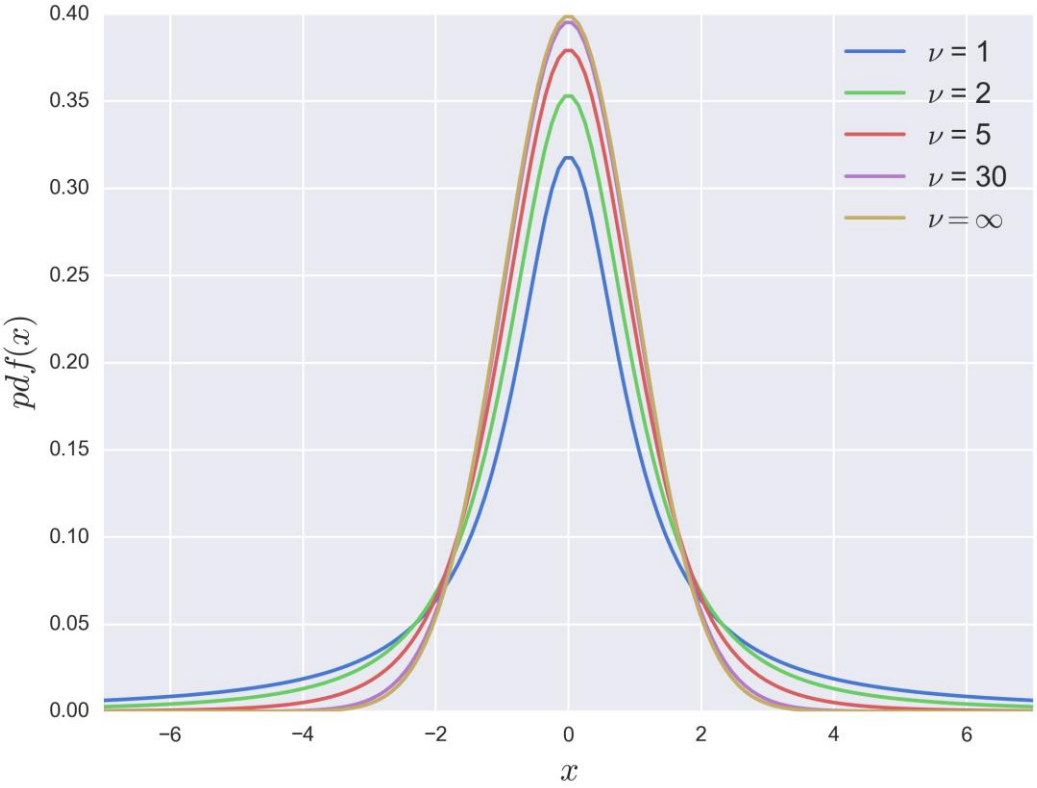


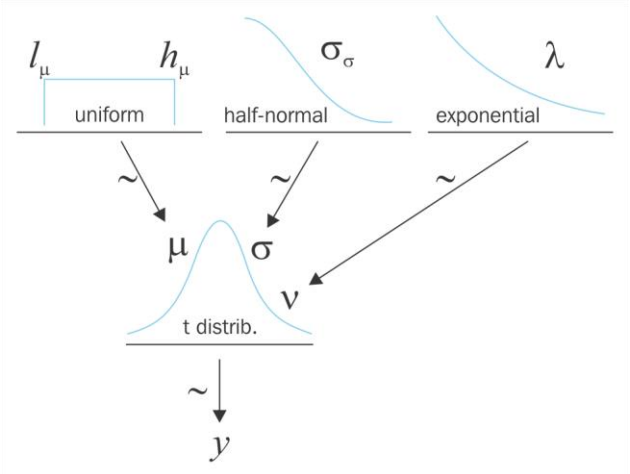

0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
40
45
50
55
60
65
70
75
x

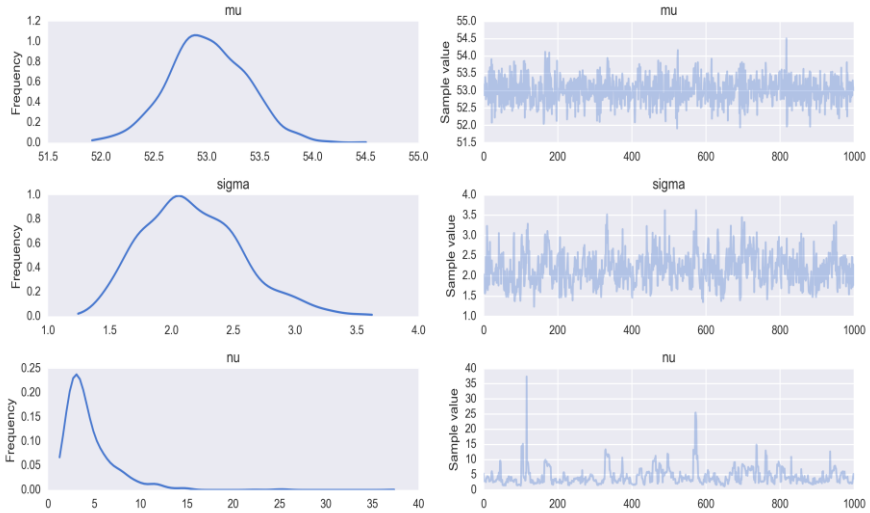

l
h
σ_σ
uniform
half-normal
$\sim$
$\sim$
μ
σ
normal
$\sim$
y

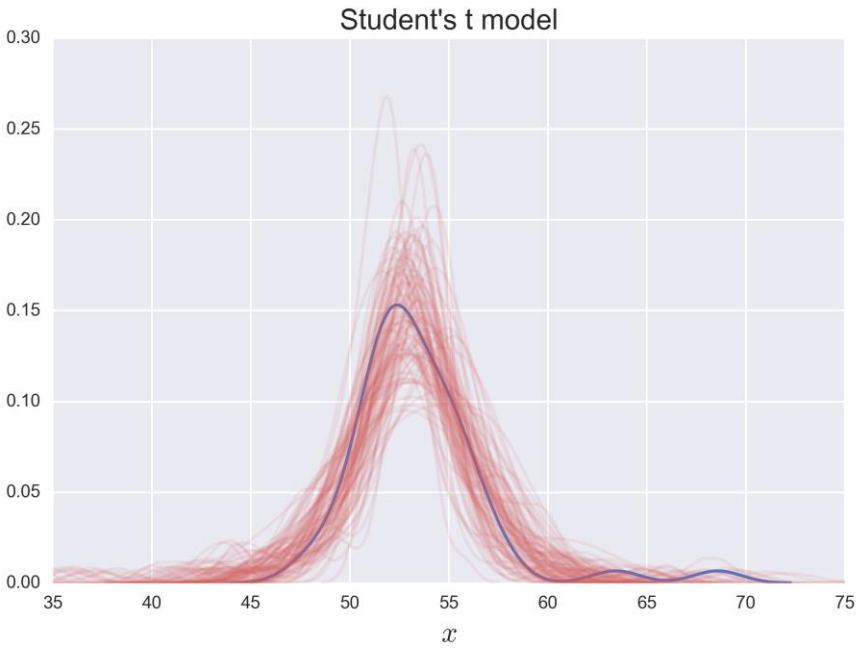

mu
Frequency
mu
Sample value
sigma
Frequency
sigma
Sample value

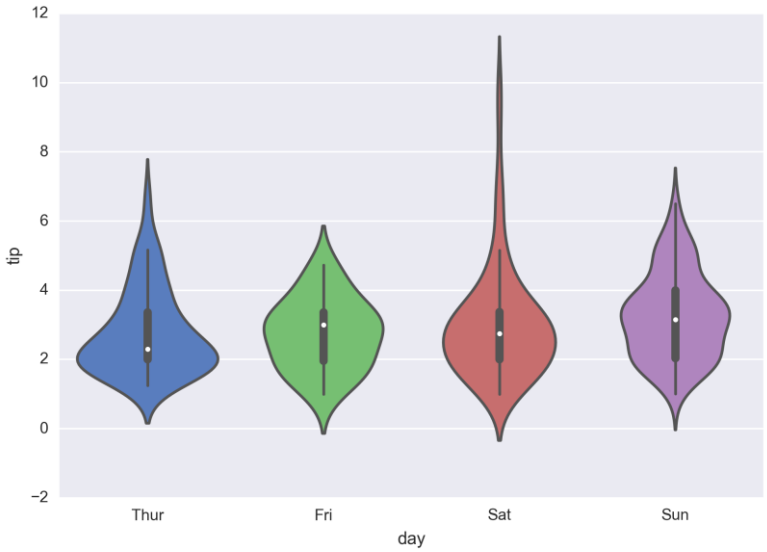

Gaussian model
x

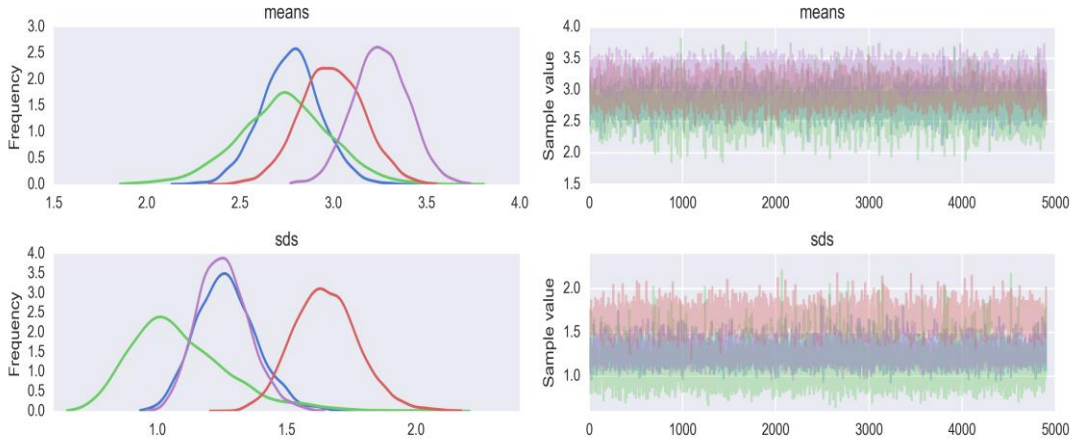

0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$pdf(x)$
−6
−4
−2
0
2
4
6
x
$\nu = 1$
$\nu = 2$
$\nu = 5$
$\nu = 30$
$\nu = \infty$

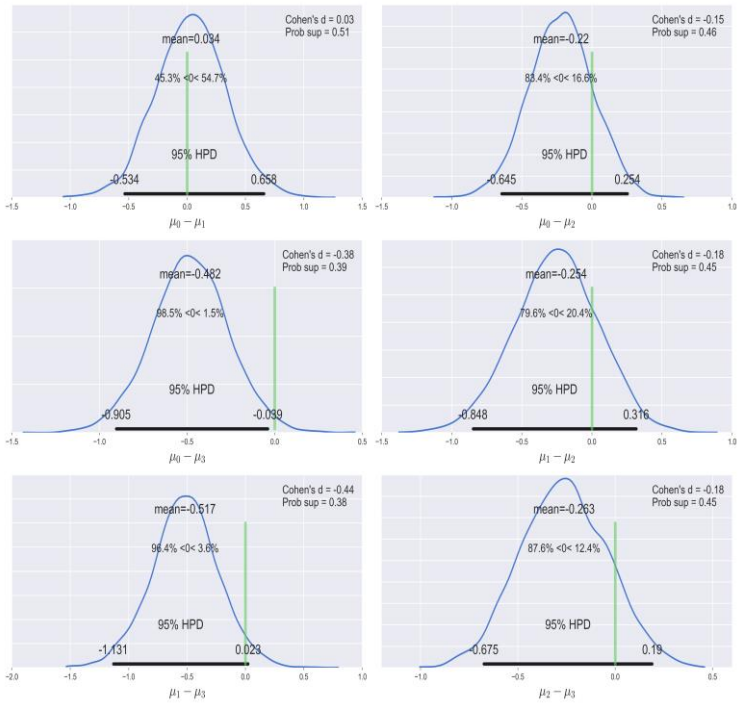

l_μ
h_μ
σ_σ
λ
uniform
half-normal
exponential
~
~
~
μ
σ
ν
t distrib.
~
y

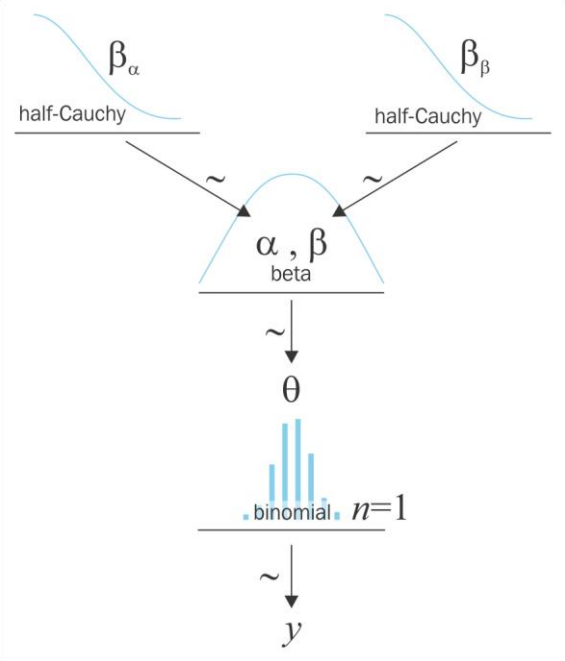

mu
Frequency
mu
Sample value
sigma
Frequency
sigma
Sample value
nu
Frequency
nu
Sample value

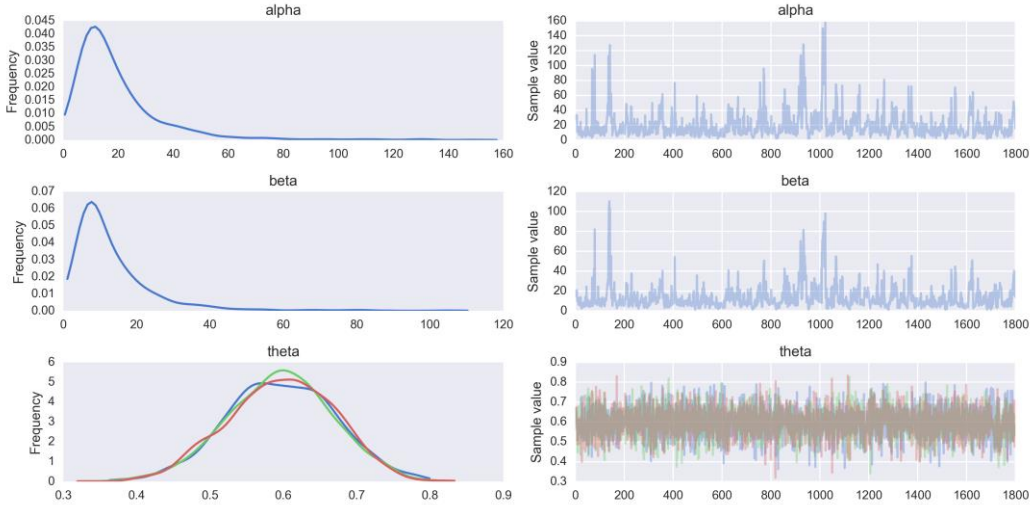

Student's t model
x

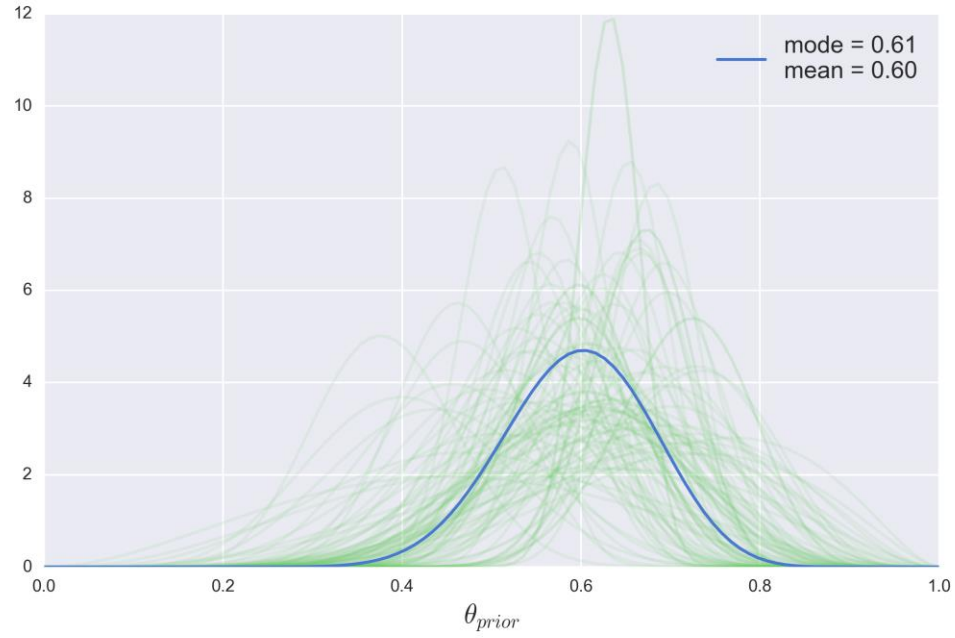

12
10
8
6
4
2
0
−2
tip
Thur
Fri
Sat
Sun
day


means
Frequency
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.5
2.0
2.5
3.0
3.5
4.0
means
Sample value
4.0
3.5
3.0
2.5
2.0
1.5
0
1000
2000
3000
4000
5000
sds
Frequency
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
1.5
2.0
sds
Sample value
2.0
1.5
1.0
0
1000
2000
3000
4000
5000

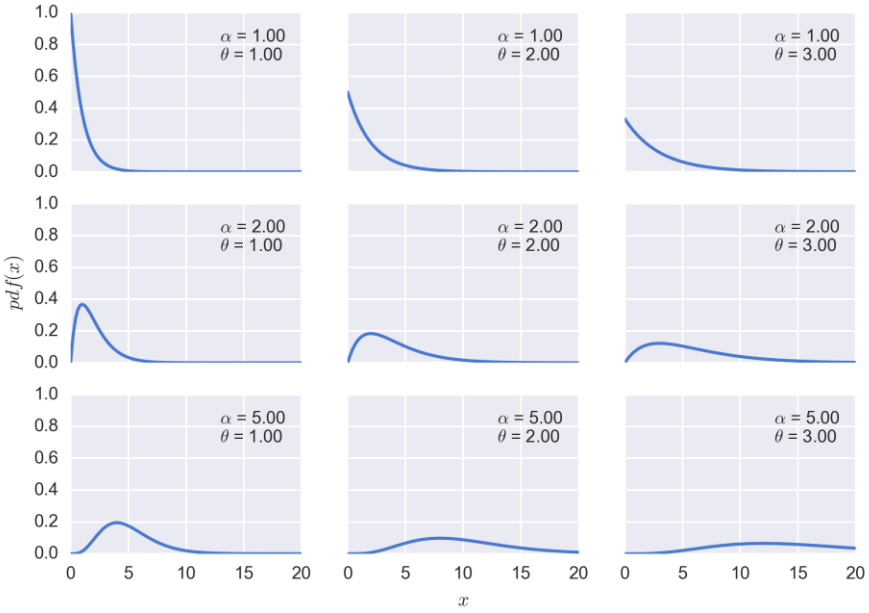

Cohen's d = 0.03
Prob sup = 0.51
mean=0.034
45.3% <0< 54.7%
95% HPD
-0.534
0.658
$\mu_0 - \mu_1$
Cohen's d = -0.15
Prob sup = 0.46
mean=-0.22
83.4% <0< 16.6%
95% HPD
-0.645
0.254
$\mu_0 - \mu_2$
Cohen's d = -0.38
Prob sup = 0.39
mean=-0.482
98.5% <0< 1.5%
95% HPD
-0.905
-0.039
$\mu_0 - \mu_3$
Cohen's d = -0.18
Prob sup = 0.45
mean=-0.254
79.6% <0< 20.4%
95% HPD
-0.848
0.316
$\mu_1 - \mu_2$
Cohen's d = -0.44
Prob sup = 0.38
mean=-0.517
96.4% <0< 3.6%
95% HPD
-1.131
0.023
$\mu_1 - \mu_3$
Cohen's d = -0.18
Prob sup = 0.45
mean=-0.263
87.6% <0< 12.4%
95% HPD
-0.675
0.19
$\mu_2 - \mu_3$

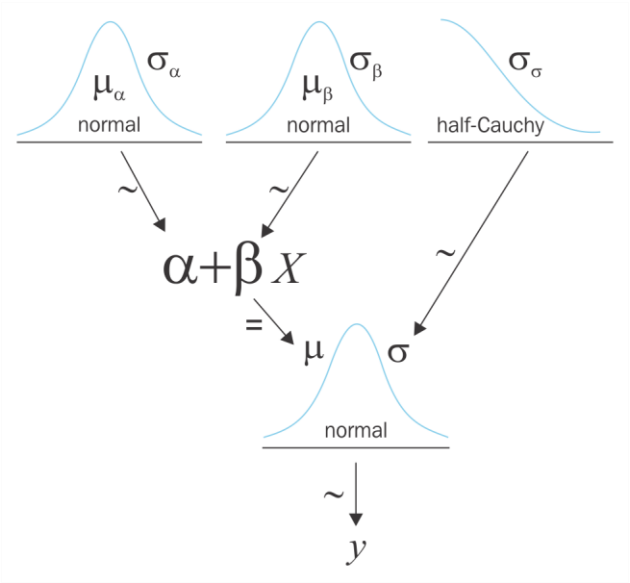

β_α
half-Cauchy
β_β
half-Cauchy
~
~
α , β
beta
~
θ
binomial
n=1
~
y

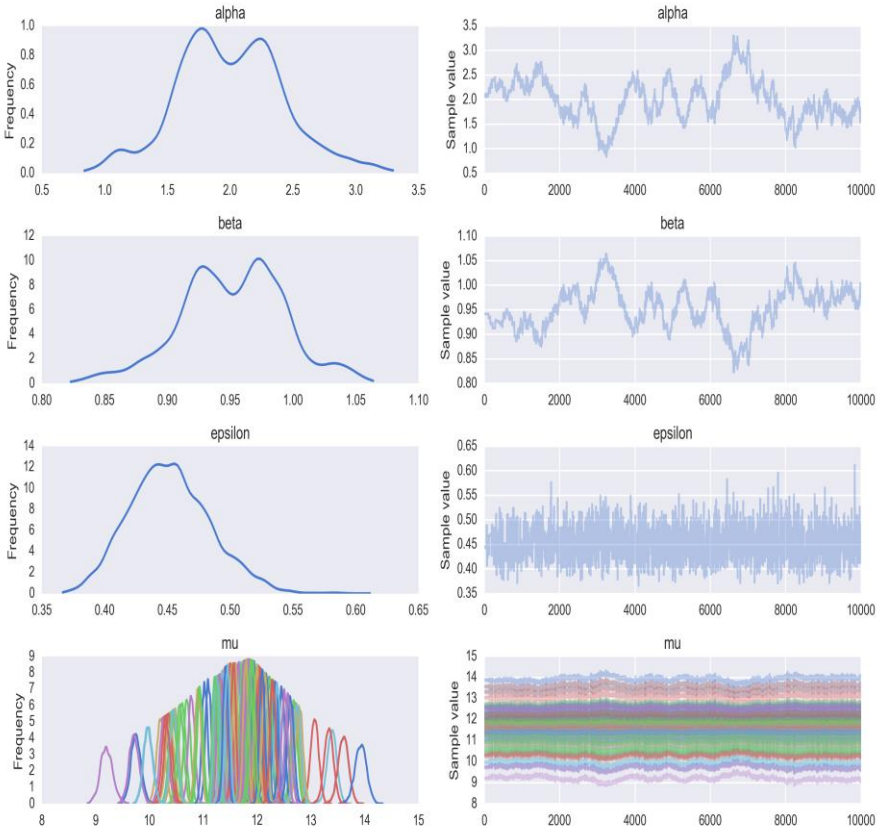

alpha
Frequency
alpha
Sample value
beta
Frequency
beta
Sample value
theta
Frequency
theta
Sample value

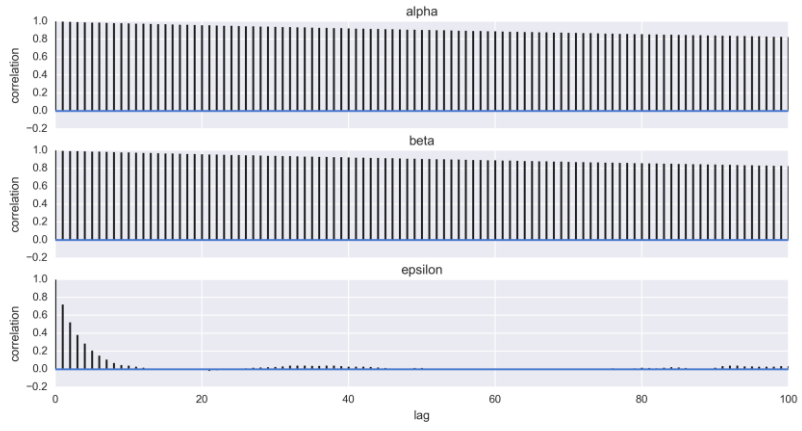

mode = 0.61
mean = 0.60
θ_{prior}

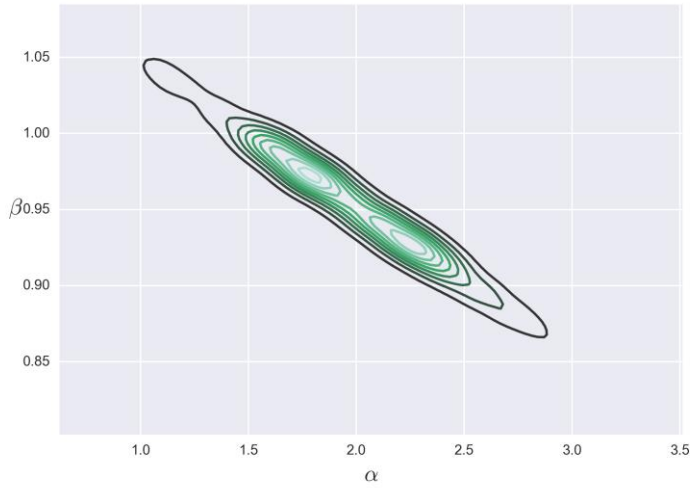
Chapter 4: Understanding and Predicting Data with Linear Regression Models

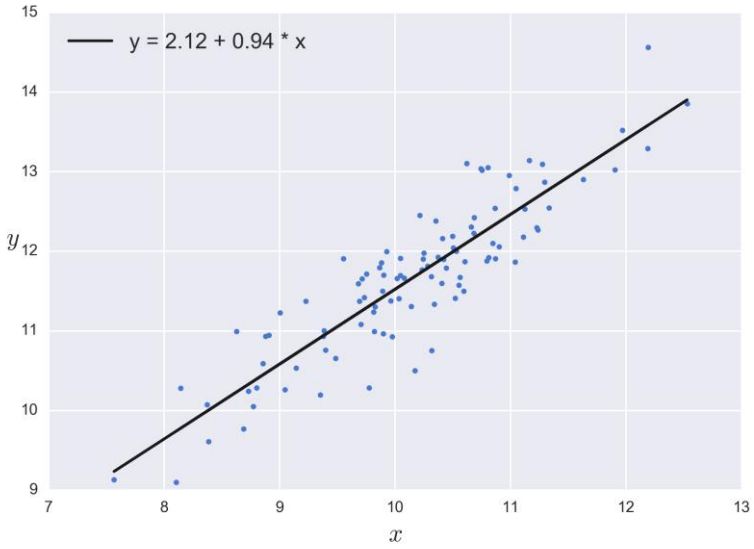


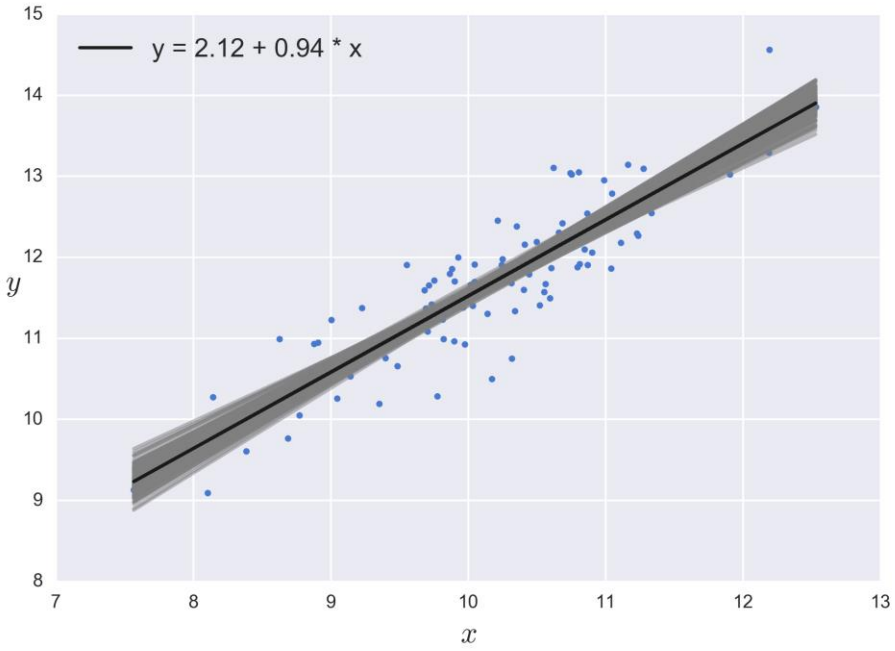

σ_α
μ_α
normal
σ_β
μ_β
normal
σ_σ
half-Cauchy
~
~
~
$\alpha+\beta X$
=
μ
σ
normal
~
y

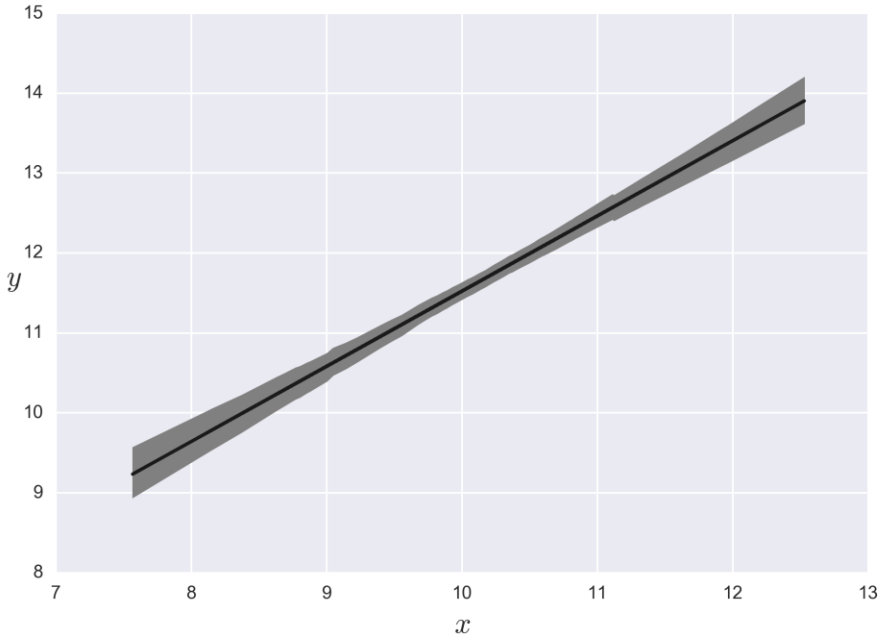

alpha
Frequency
alpha
Sample value
beta
Frequency
beta
Sample value
epsilon
Frequency
epsilon
Sample value
mu
Frequency
mu
Sample value

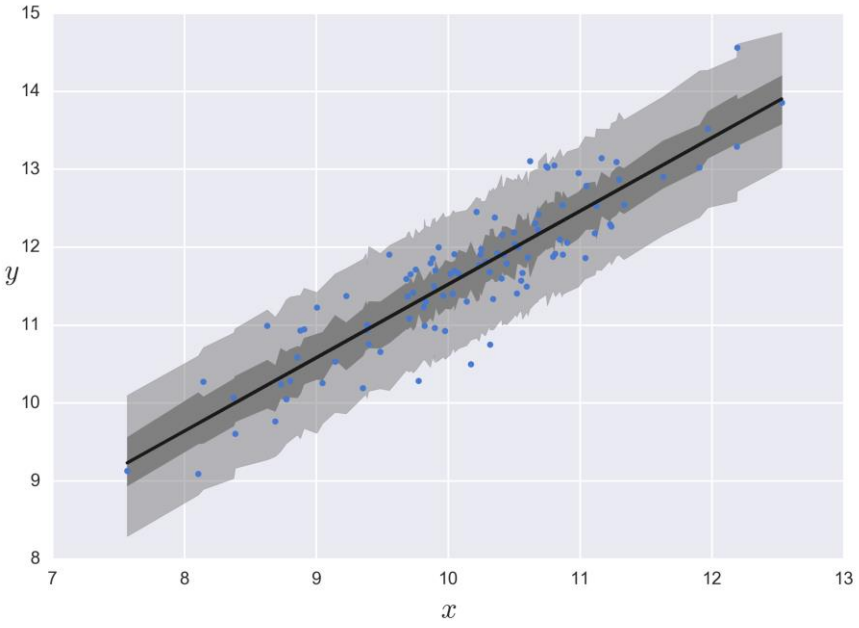

alpha
correlation
beta
correlation
epsilon
correlation
lag

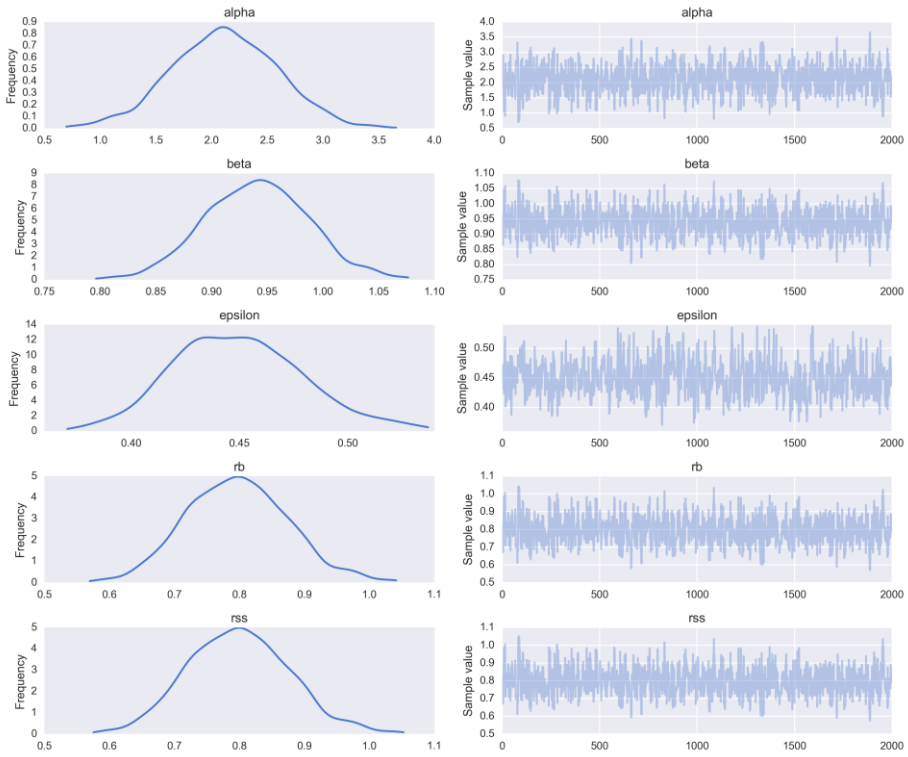

1.05
1.00
β0.95
0.90
0.85
1.0
1.5
2.0
2.5
3.0
3.5
α

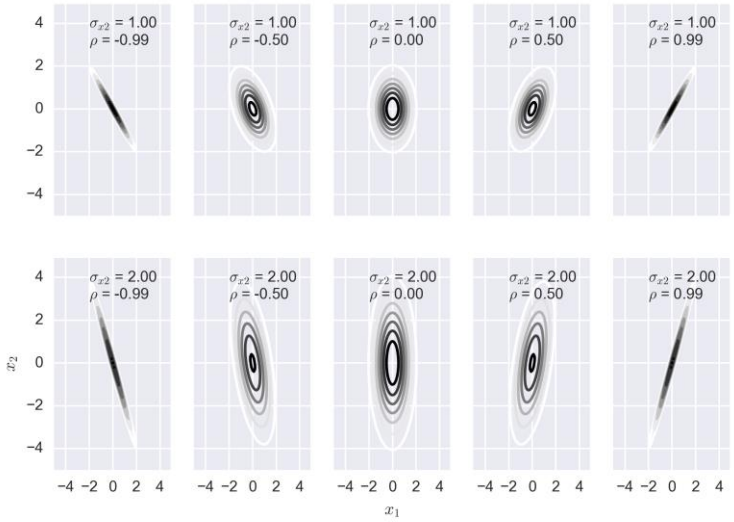

y = 2.12 + 0.94 * x
15
14
13
y12
11
10
9
7
8
9
10
11
12
13
x

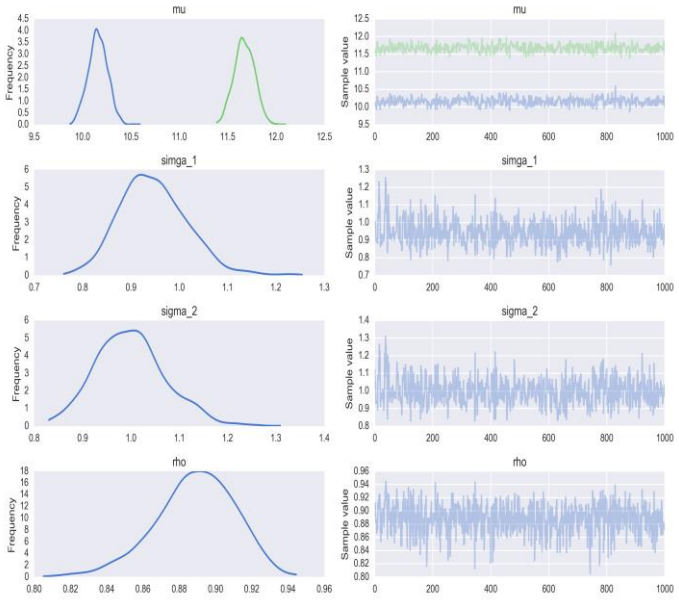

y = 2.12 + 0.94 * x
15
14
13
12
11
10
9
8
7
8
9
10
11
12
13
y
x

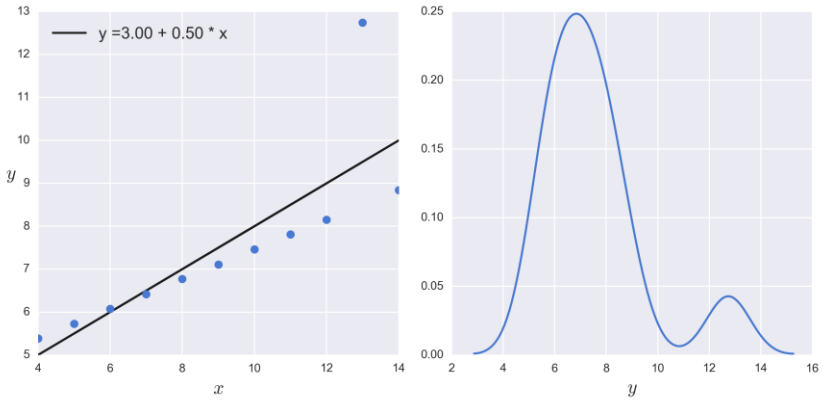

15
14
13
12
11
10
9
8
y
7
8
9
10
11
12
13
x

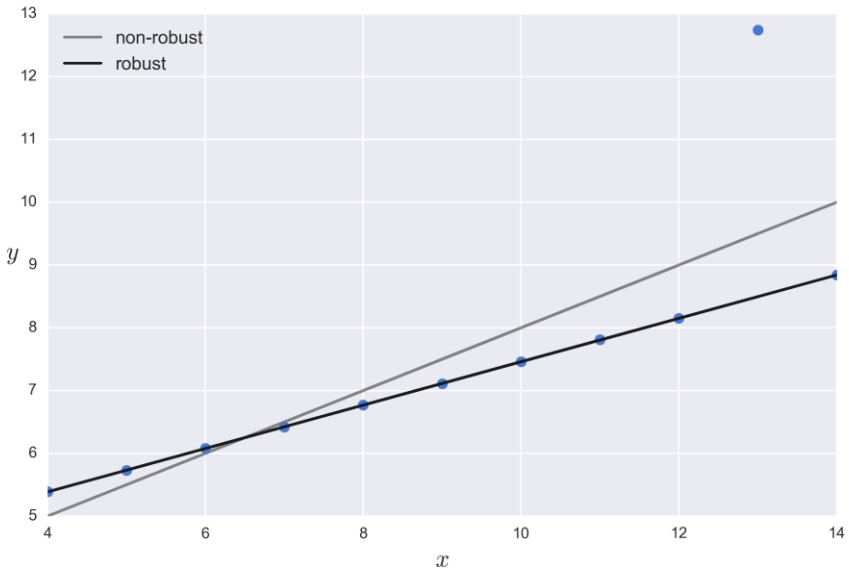

15
14
13
12
11
10
9
8
y
7
8
9
10
11
12
13
x

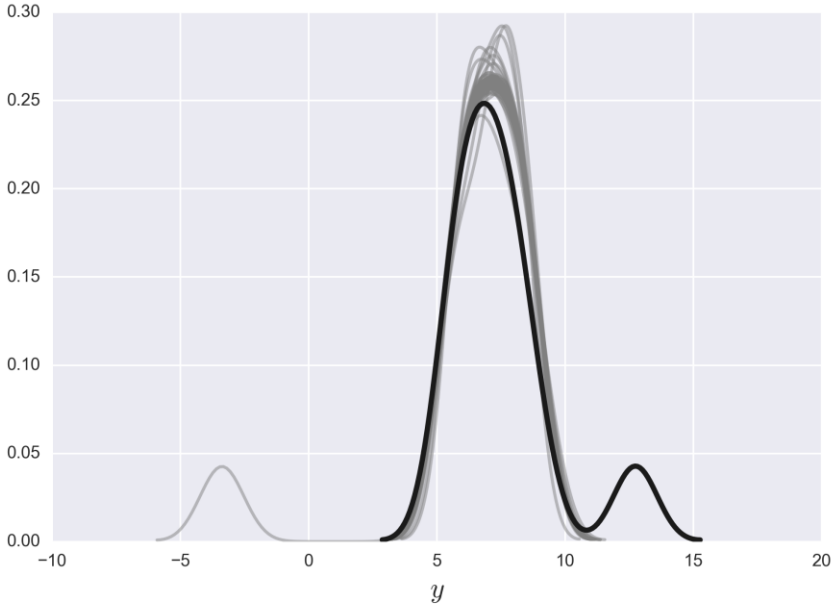

alpha
Frequency
alpha
Sample value
beta
Frequency
beta
Sample value
epsilon
Frequency
epsilon
Sample value
rb
Frequency
rb
Sample value
rss
Frequency
rss
Sample value

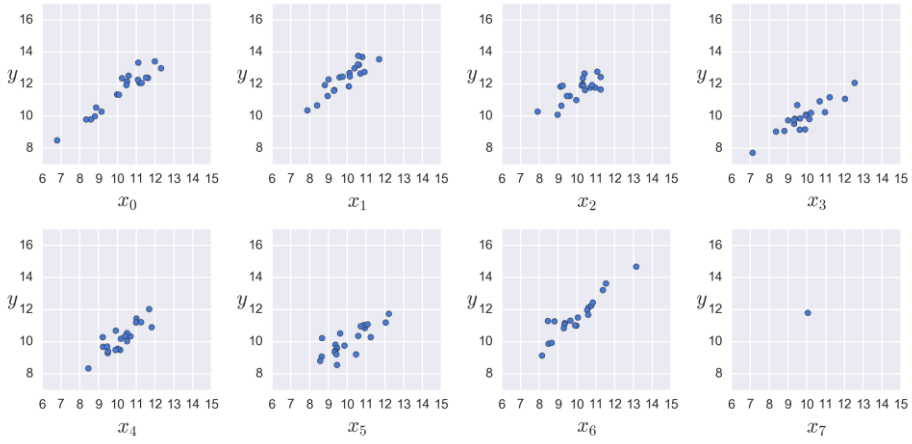

σ_{x2} = 1.00
ρ = -0.99
σ_{x2} = 1.00
ρ = -0.50
σ_{x2} = 1.00
ρ = 0.00
σ_{x2} = 1.00
ρ = 0.50
σ_{x2} = 1.00
ρ = 0.99
σ_{x2} = 2.00
ρ = -0.99
σ_{x2} = 2.00
ρ = -0.50
σ_{x2} = 2.00
ρ = 0.00
σ_{x2} = 2.00
ρ = 0.50
σ_{x2} = 2.00
ρ = 0.99
x_2
x_1

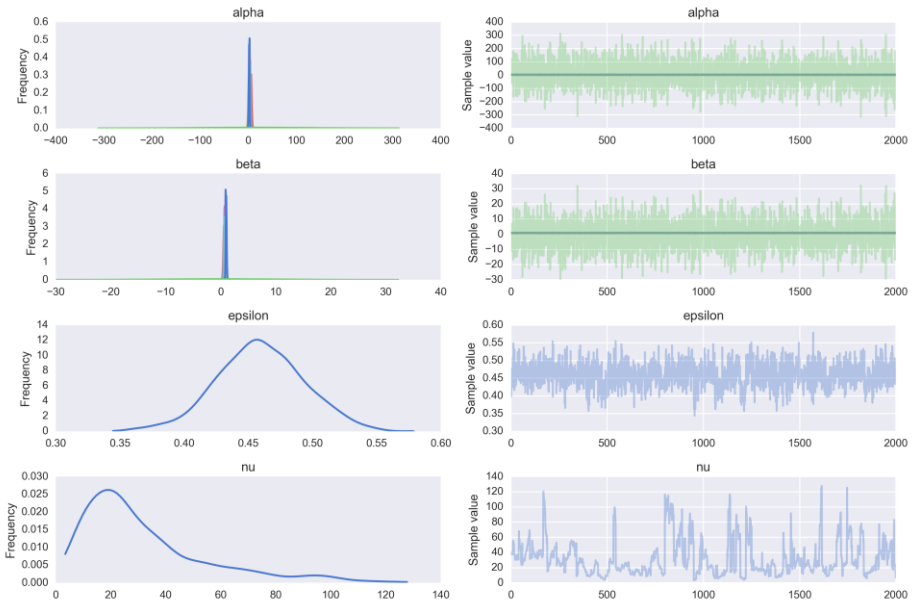

mu
Frequency
Sample value
simga_1
sigma_2
rho

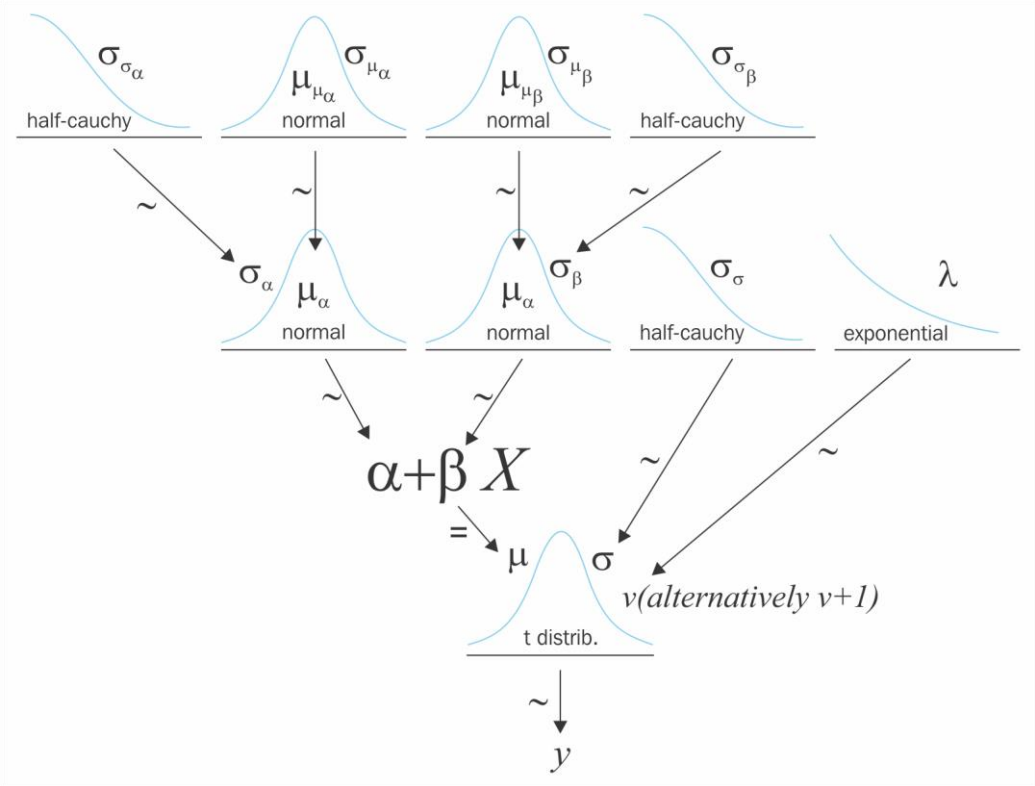

y =3.00 + 0.50 * x
y
x

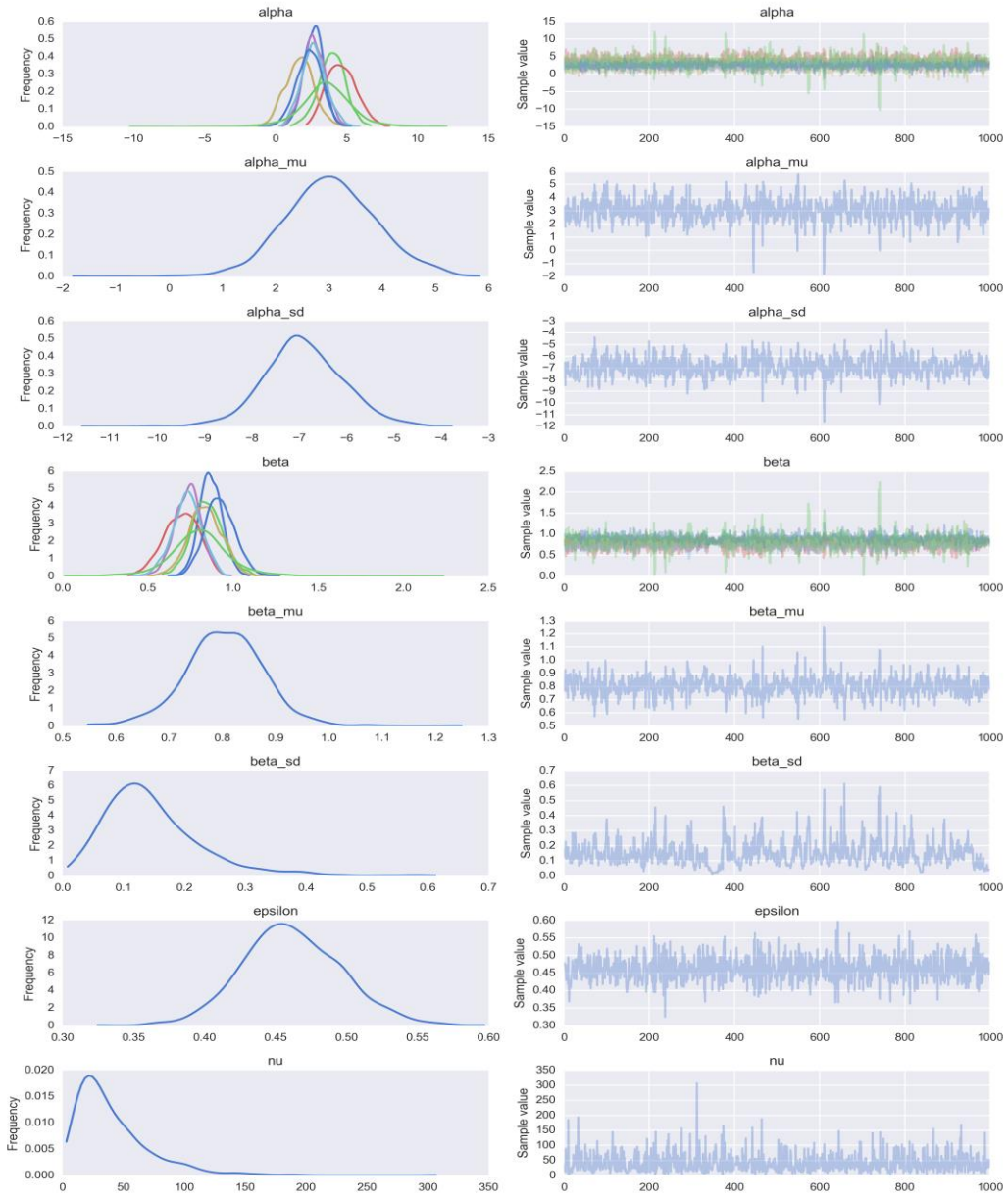

non-robust
robust
13
12
11
10
9
8
7
6
5
y
4
6
8
10
12
14
x

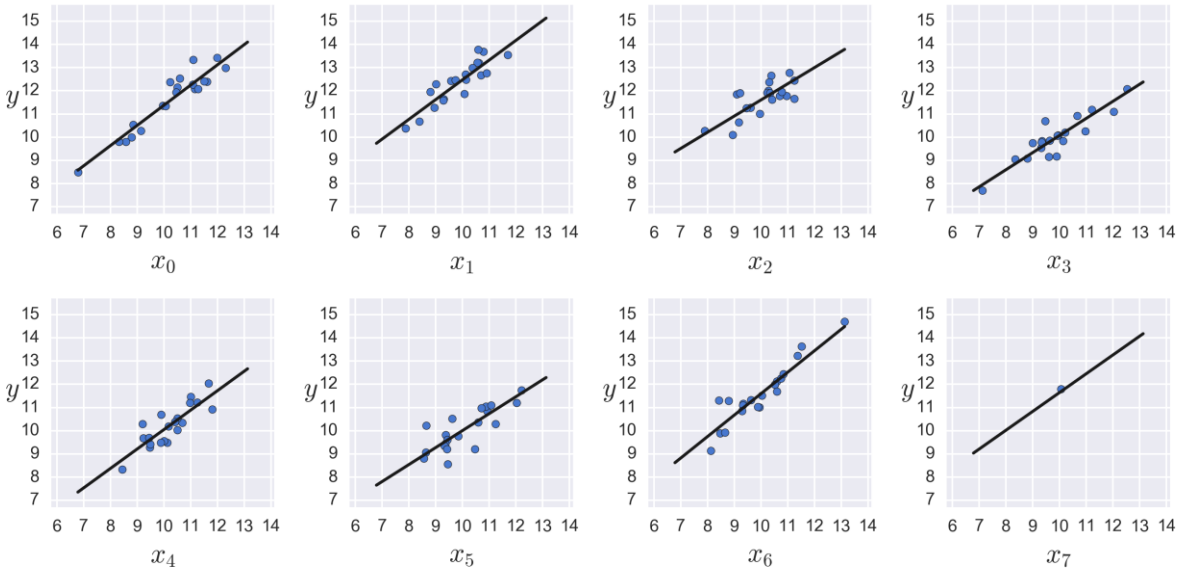

0.30
0.25
0.20
0.15
0.10
0.05
0.00
−10
−5
0
5
10
15
20
y

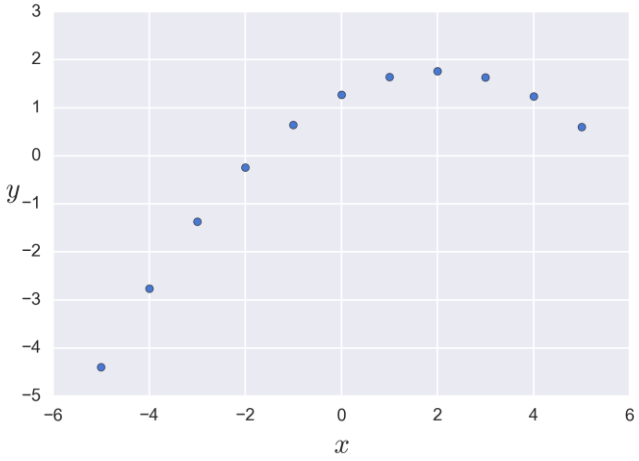

y
x_0
x_1
x_2
x_3
x_4
x_5
x_6
x_7

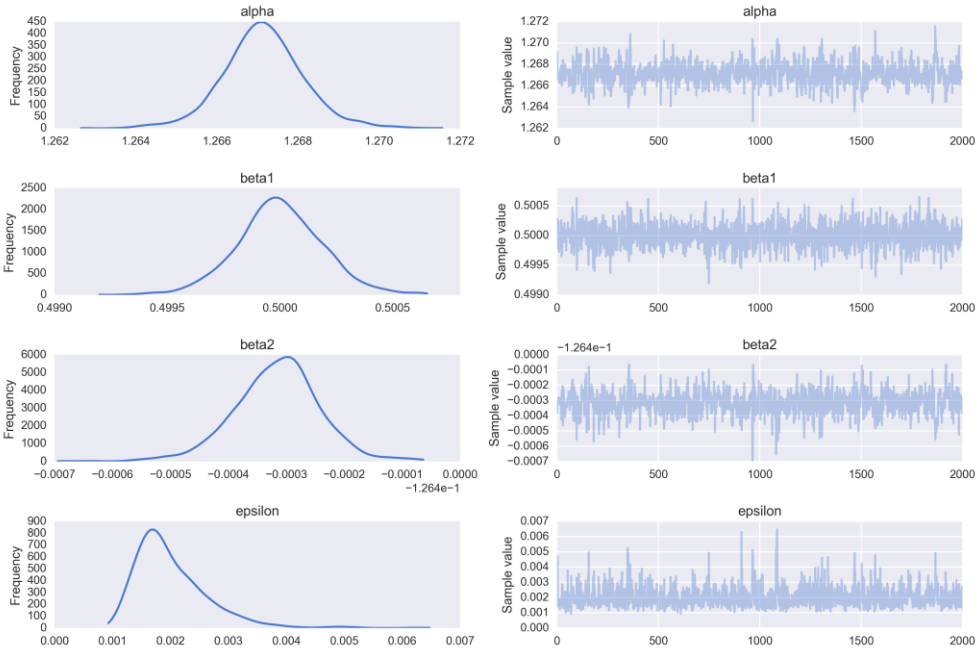

alpha
beta
epsilon
nu
Frequency
Sample value

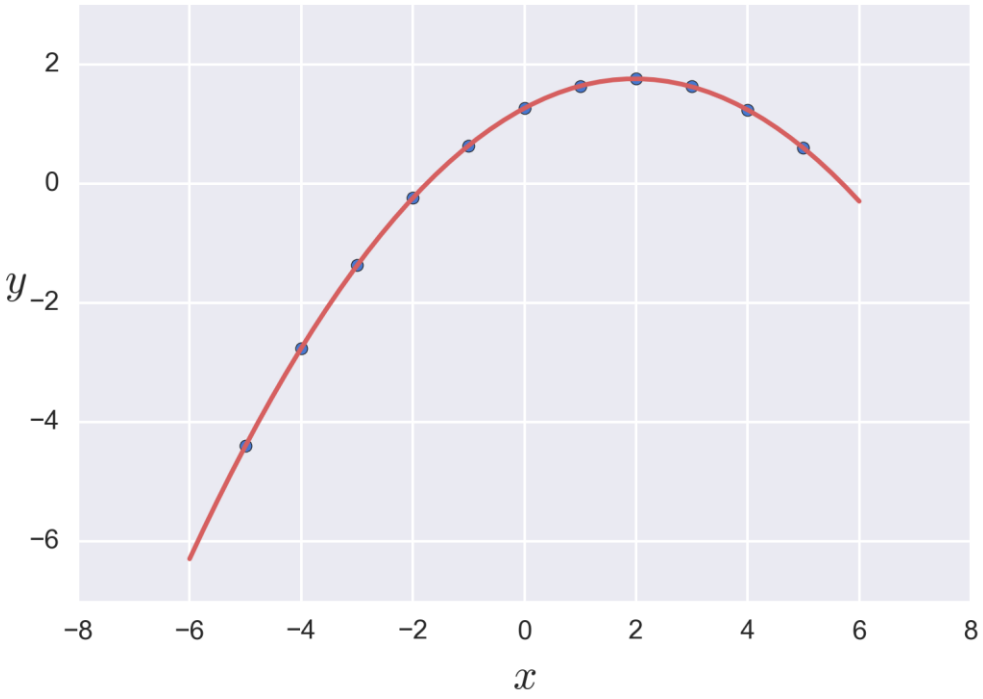

σσα
half-cauchy
μμα
σμα
normal
μμβ
σμβ
normal
σσβ
half-cauchy
~
~
~
~
σα
μα
normal
μα
σβ
normal
σσ
half-cauchy
λ
exponential
~
~
~
~
α+β X
=
μ
σ
v(alternatively v+1)
t distrib.
~
y

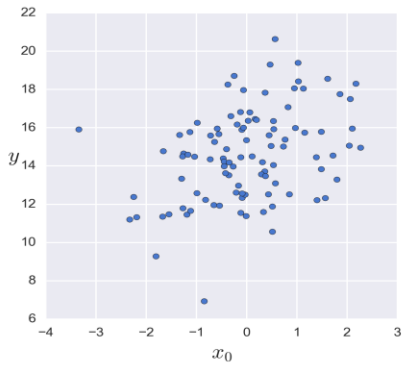

alpha
Frequency
Sample value
alpha_mu
alpha_sd
beta
beta_mu
beta_sd
epsilon
nu

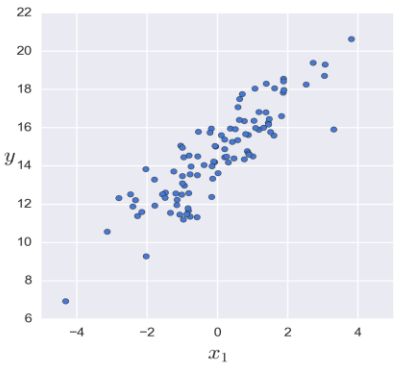

y
x_0
x_1
x_2
x_3
x_4
x_5
x_6
x_7

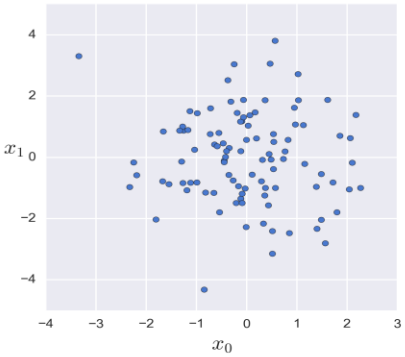

y
x

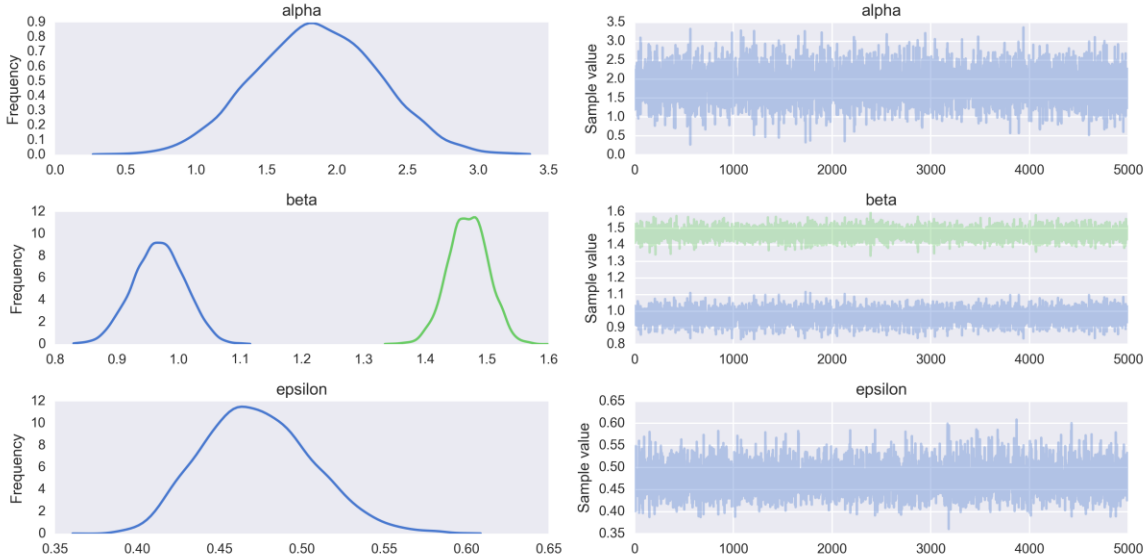

alpha
Frequency
alpha
Sample value
beta1
Frequency
beta1
Sample value
beta2
Frequency
beta2
Sample value
−1.264e−1
epsilon
Frequency
epsilon
Sample value

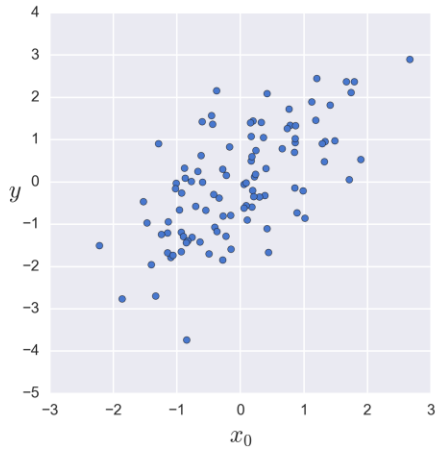

2
0
−2
−4
−6
y
−8
−6
−4
−2
0
2
4
6
8
x

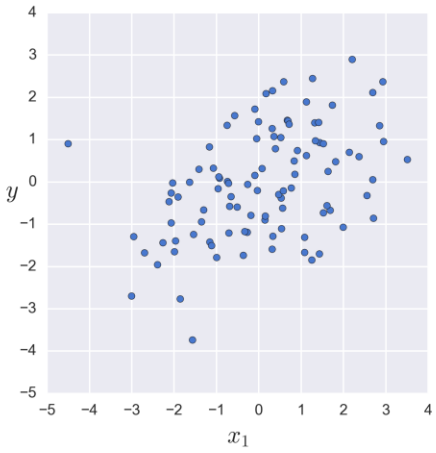

22
20
18
16
14
12
10
8
6
y
−4
−3
−2
−1
0
1
2
3
x_0

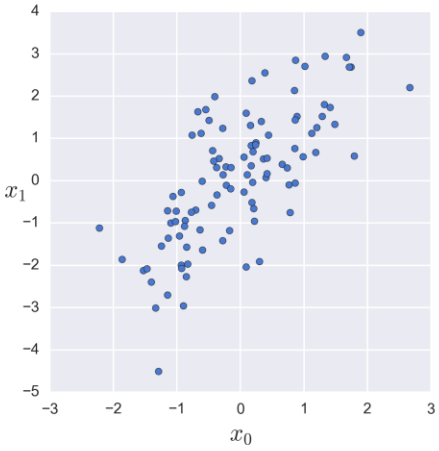

22
20
18
16
14
12
10
8
6
y
−4
−2
0
2
4
x_1

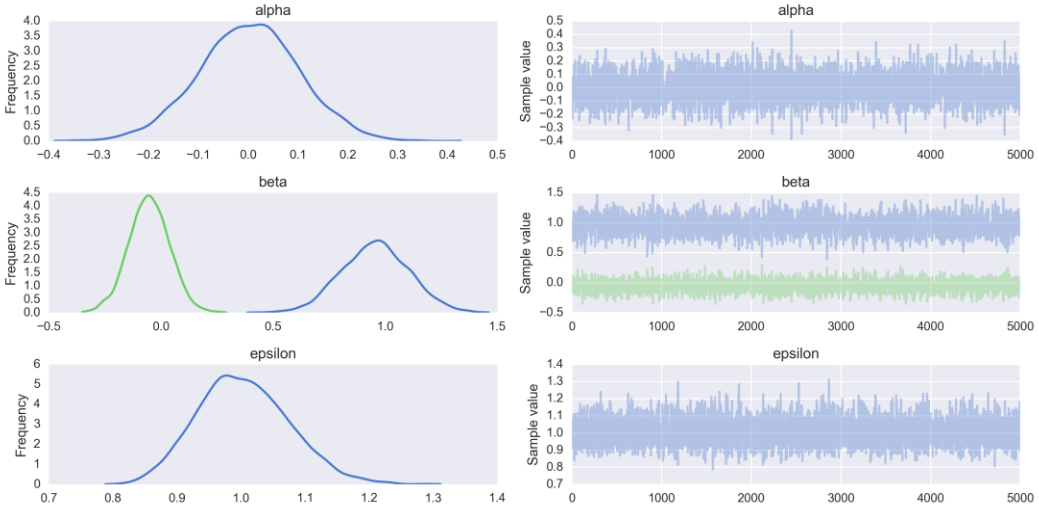

4
2
0
−2
−4
x_1
−4
−3
−2
−1
0
1
2
3
x_0

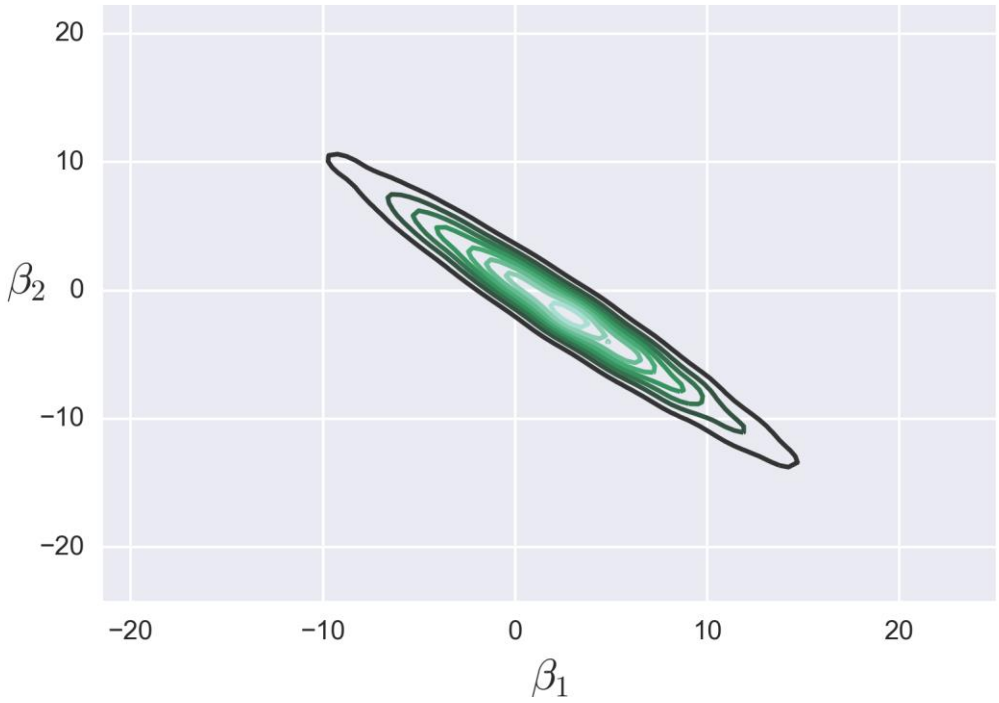

alpha
Frequency
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
alpha
Sample value
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
0 1000 2000 3000 4000 5000
beta
Frequency
0 2 4 6 8 10 12
0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6
beta
Sample value
0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6
0 1000 2000 3000 4000 5000
epsilon
Frequency
0 2 4 6 8 10 12
0.35 0.40 0.45 0.50 0.55 0.60 0.65
epsilon
Sample value
0.35 0.40 0.45 0.50 0.55 0.60 0.65
0 1000 2000 3000 4000 5000

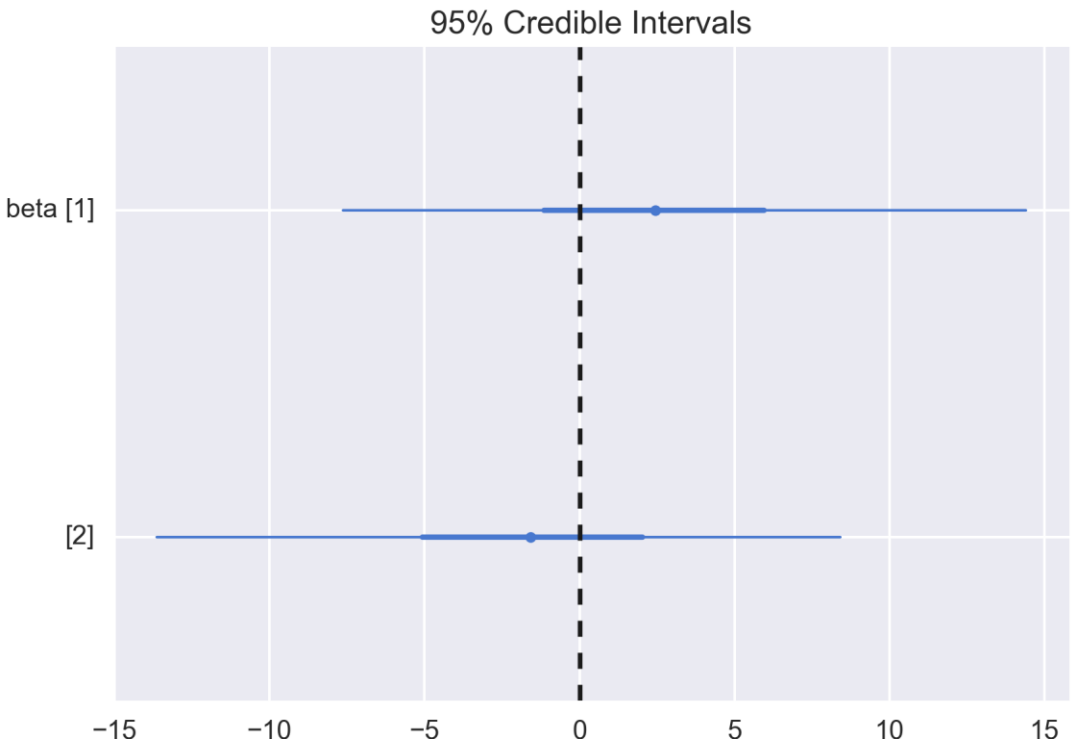

y
x0

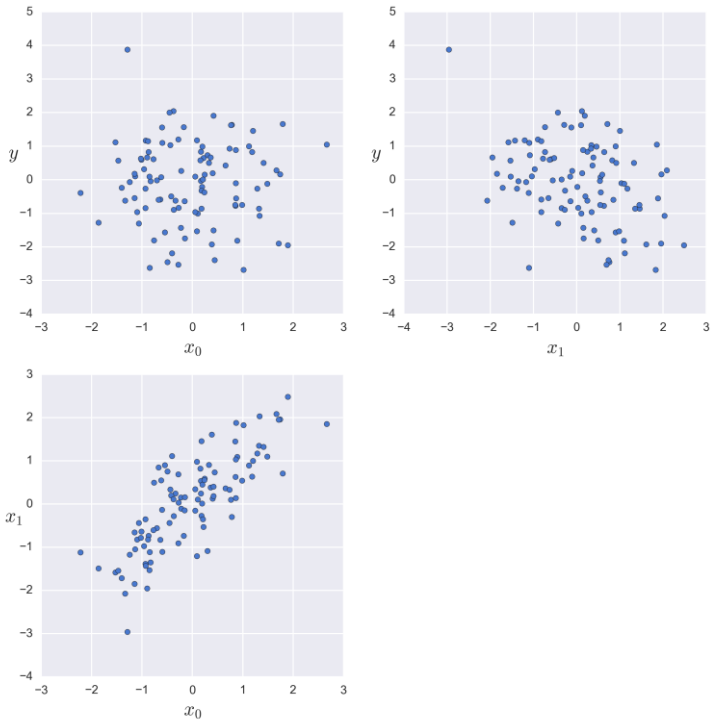

y
x1

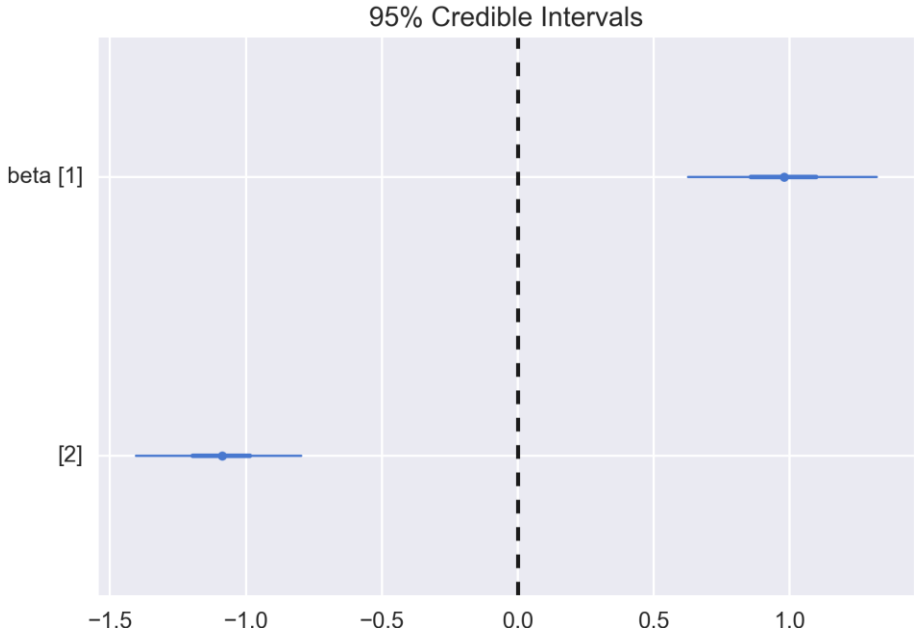

x1
x0


alpha
Frequency
alpha
Sample value
beta
Frequency
beta
Sample value
epsilon
Frequency
epsilon
Sample value

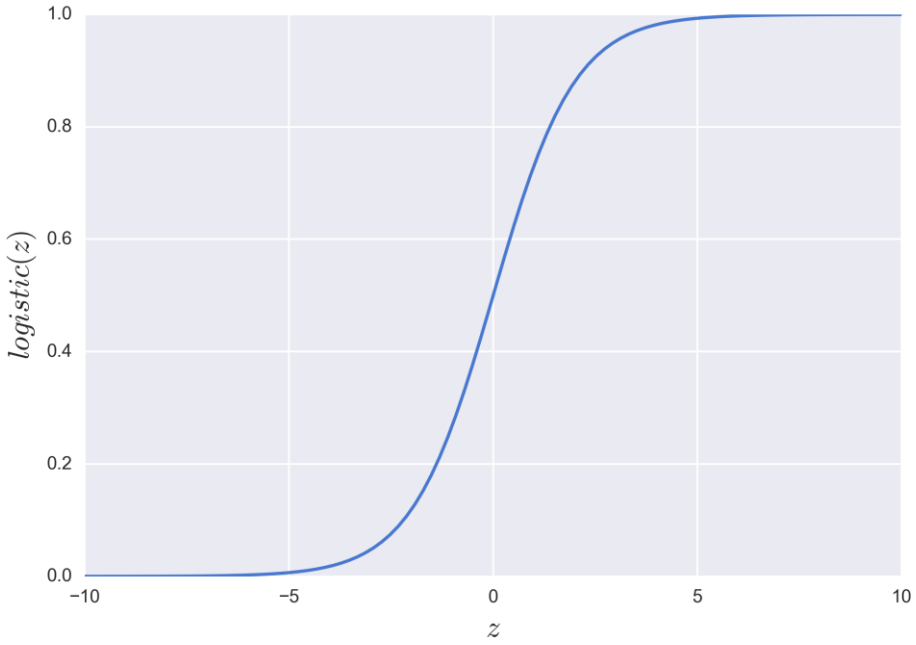

β_2
β_1

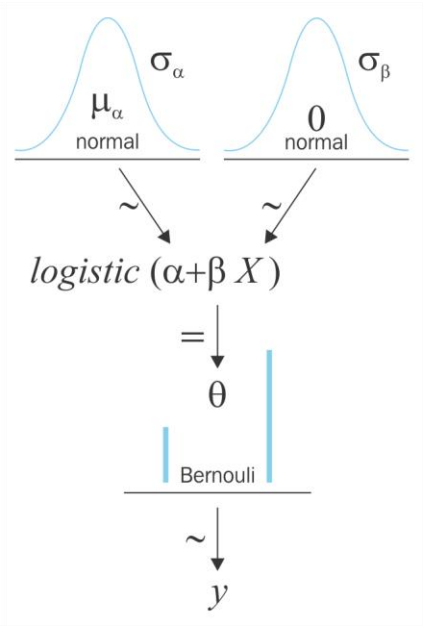

95% Credible Intervals
beta [1]
[2]
−15
−10
−5
0
5
10
15

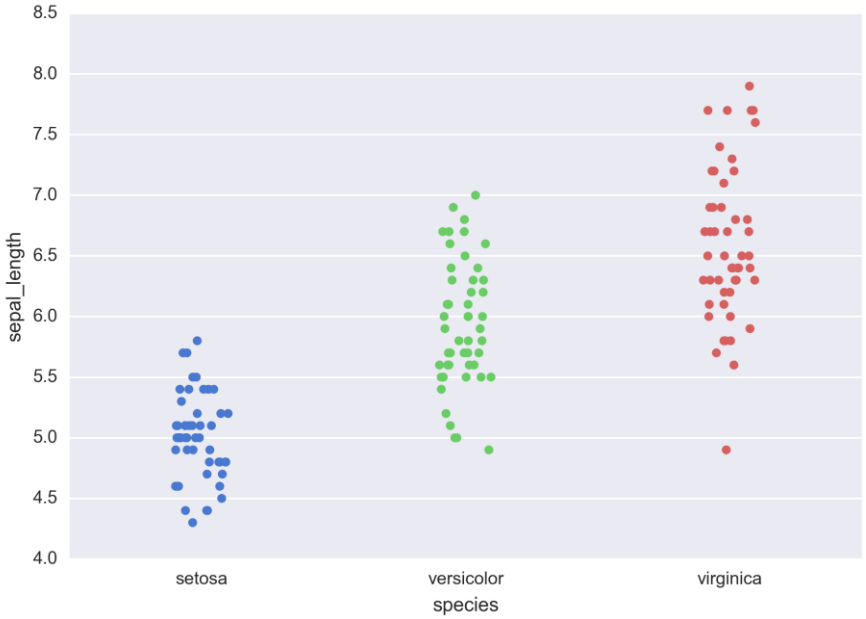

y
x_0
y
x_1
x_1
x_0

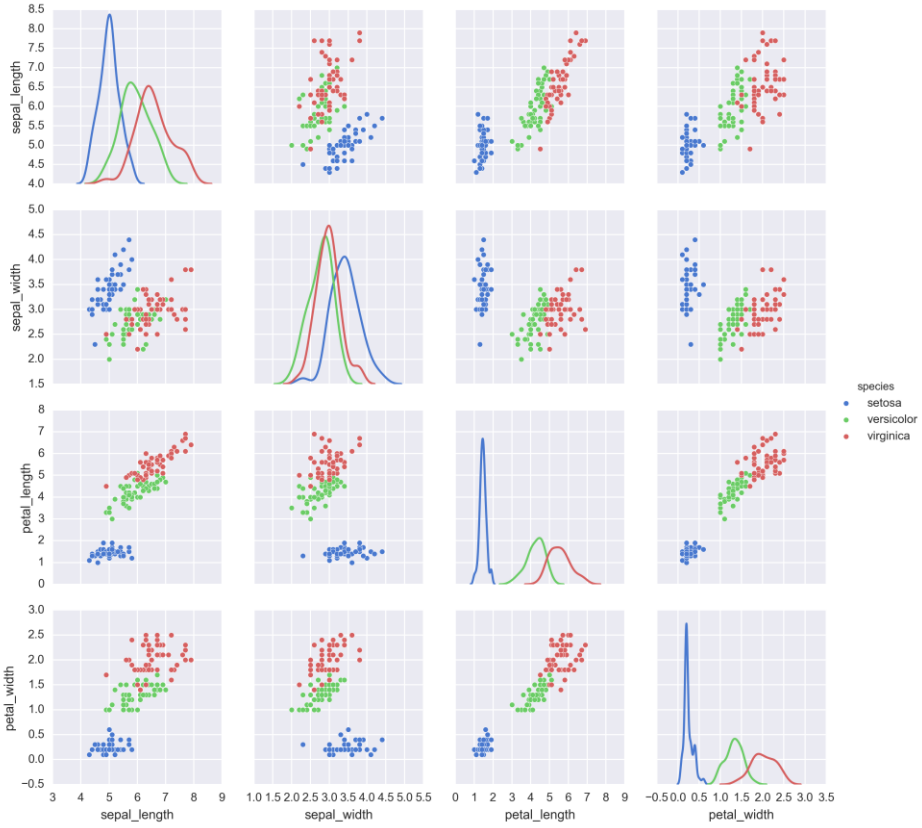

95% Credible Intervals
beta [1]
[2]
−1.5
−1.0
−0.5
0.0
0.5
1.0

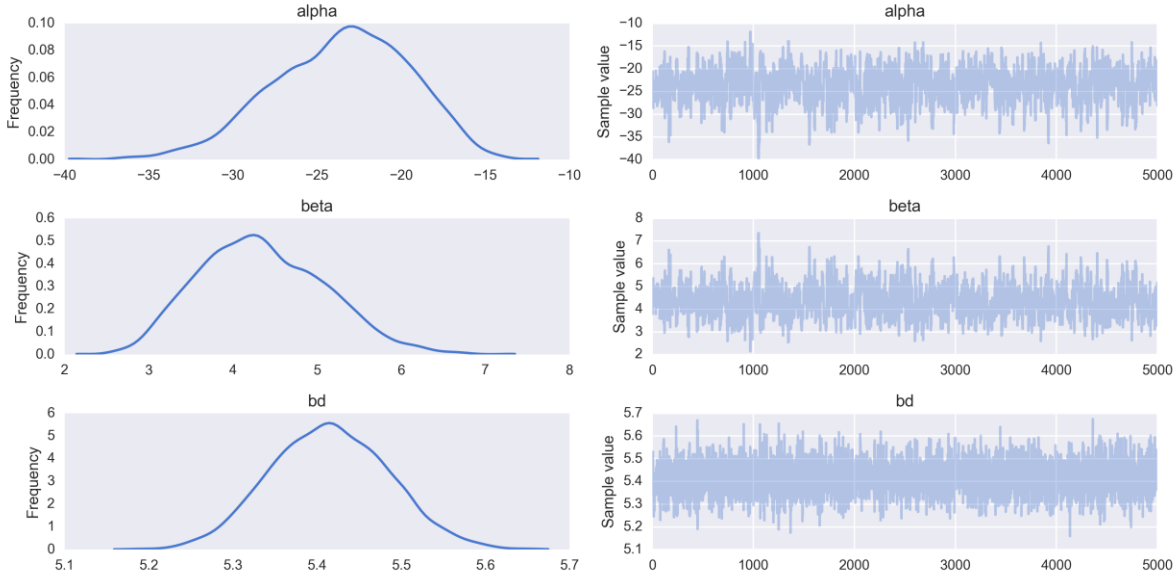
Chapter 5: Classifying Outcomes with Logistic Regression

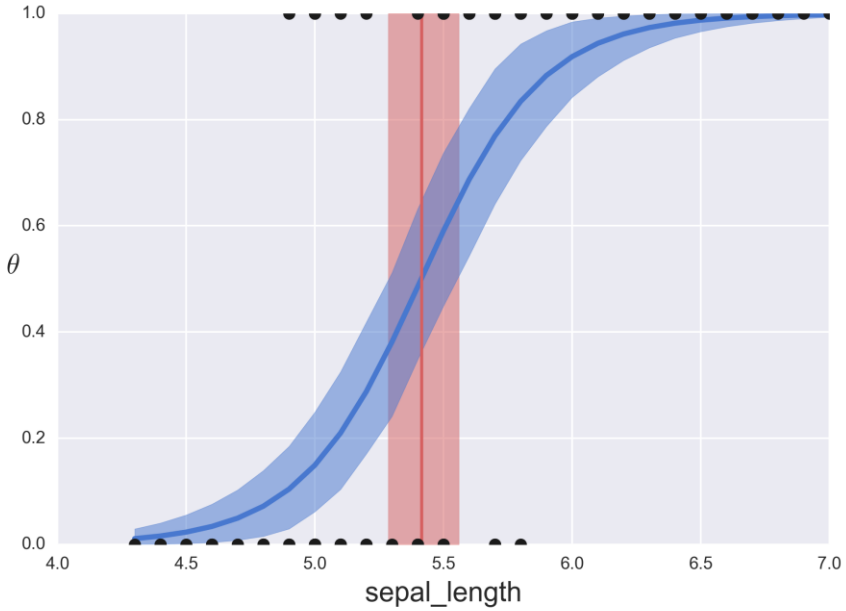


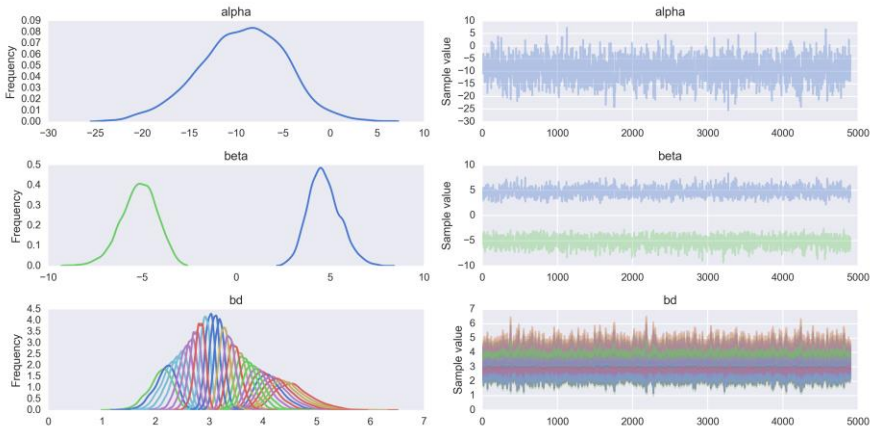

σα
μα
normal
σβ
0
normal
~
~
logistic (α+β X)
=
θ
Bernouli
~
y

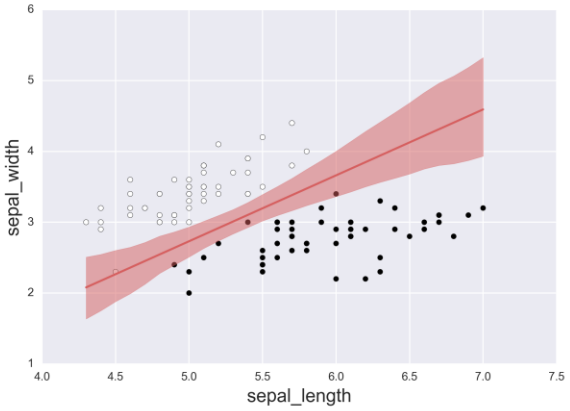

8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
sepal_length
setosa
versicolor
virginica
species

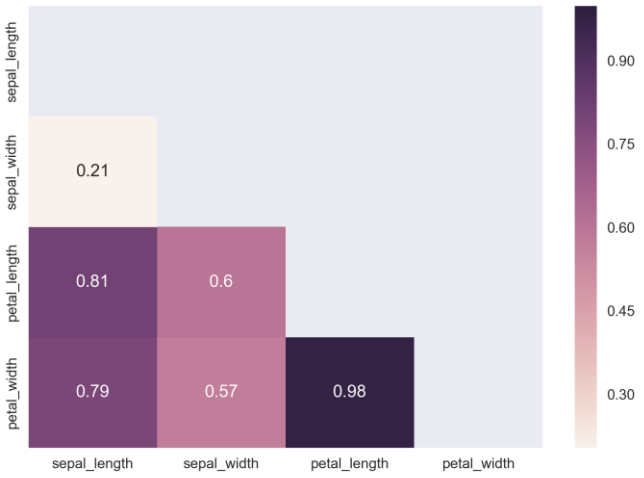

sepal_length
sepal_width
petal_length
petal_width
species
setosa
versicolor
virginica

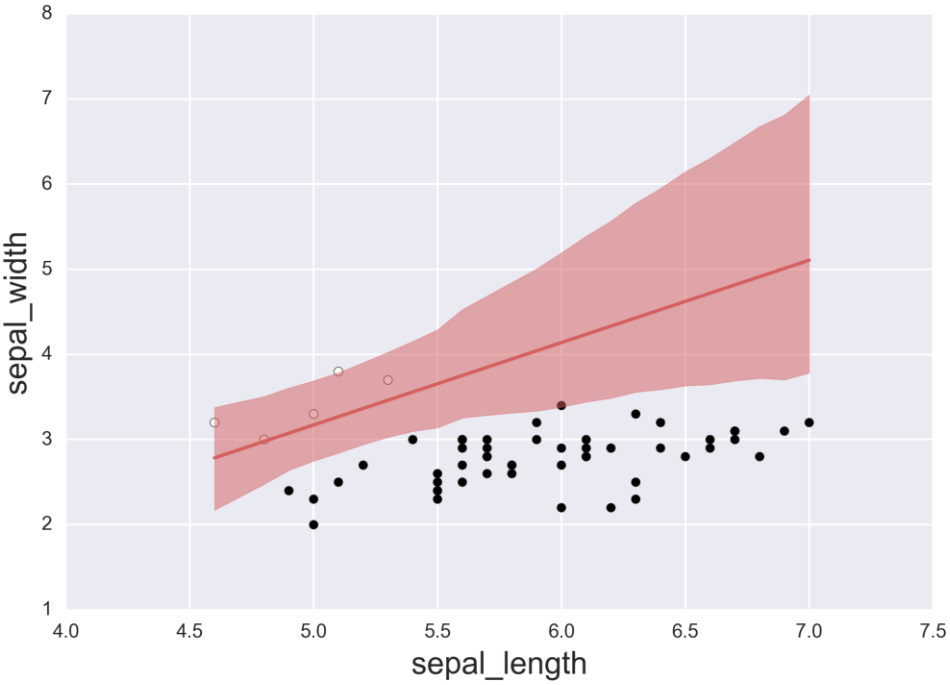

alpha
Frequency
alpha
Sample value
beta
Frequency
beta
Sample value
bd
Frequency
bd
Sample value

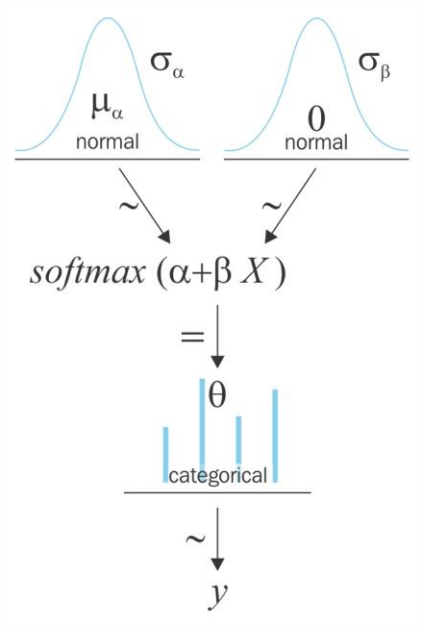

θ
sepal_length

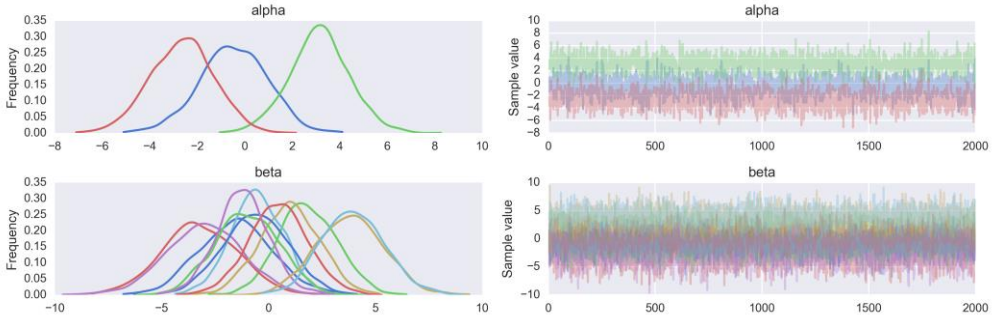

alpha
Frequency
alpha
Sample value
beta
Frequency
beta
Sample value
bd
Frequency
bd
Sample value

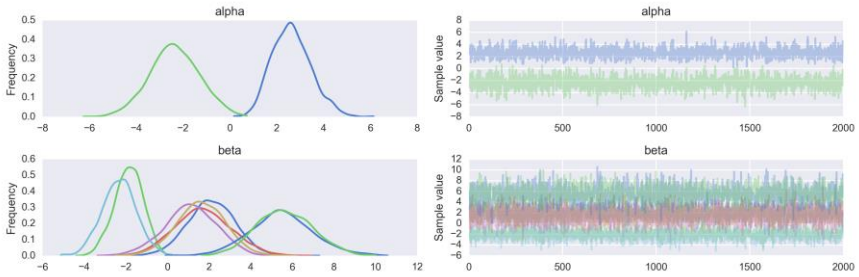

sepal_width
sepal_length

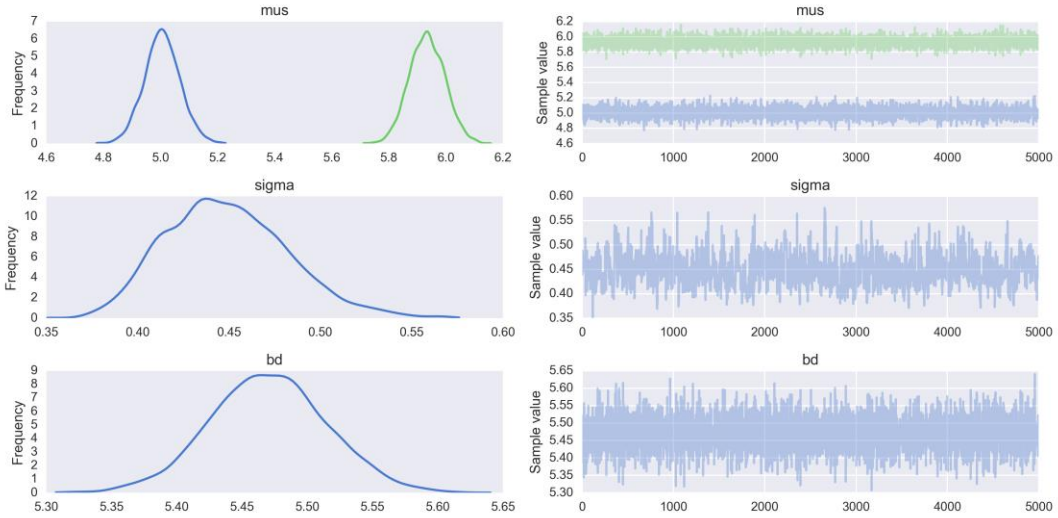

sepal_length
sepal_width
petal_length
petal_width
0.21
0.81
0.6
0.79
0.57
0.98
0.90
0.75
0.60
0.45
0.30
sepal_length
sepal_width
petal_length
petal_width

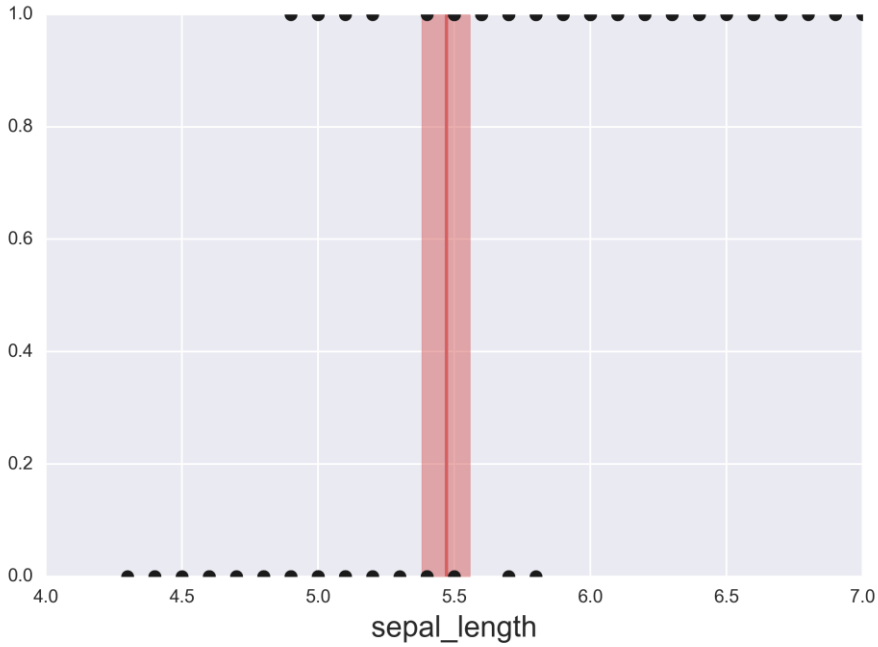

8
7
6
5
4
3
2
1
sepal_width
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
sepal_length


σα
μα
normal
σβ
0
normal
~
~
softmax (α+β X)
=
θ
categorical
~
y

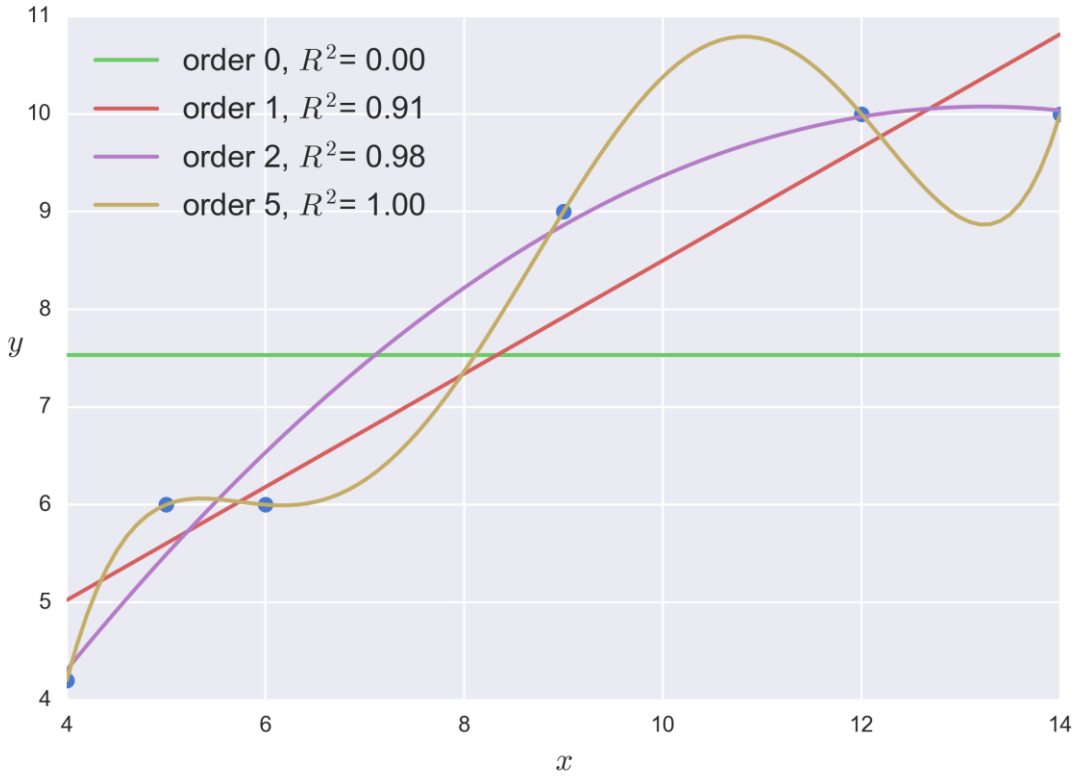

alpha
Frequency
alpha
Sample value
beta
Frequency
beta
Sample value

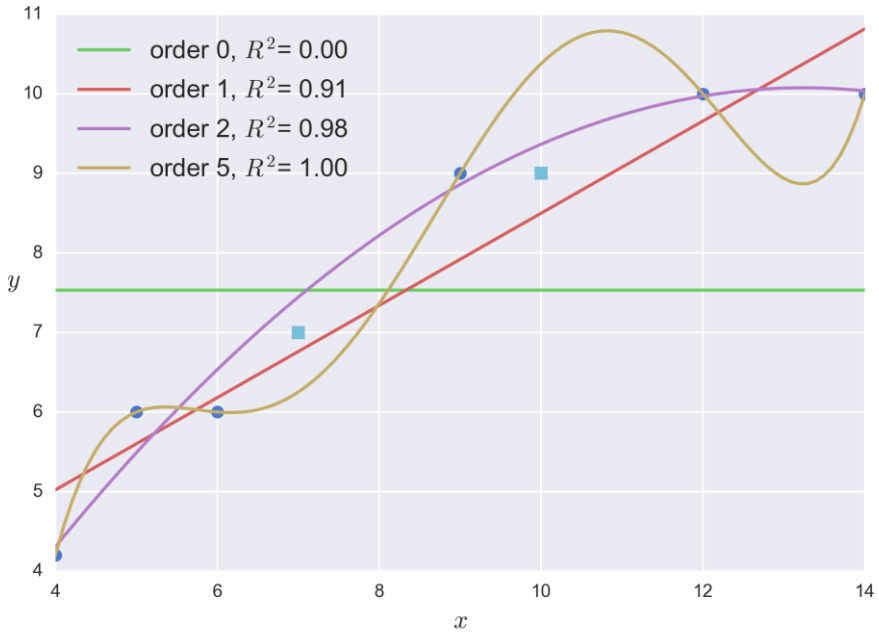

alpha
Frequency
alpha
Sample value
beta
Frequency
beta
Sample value

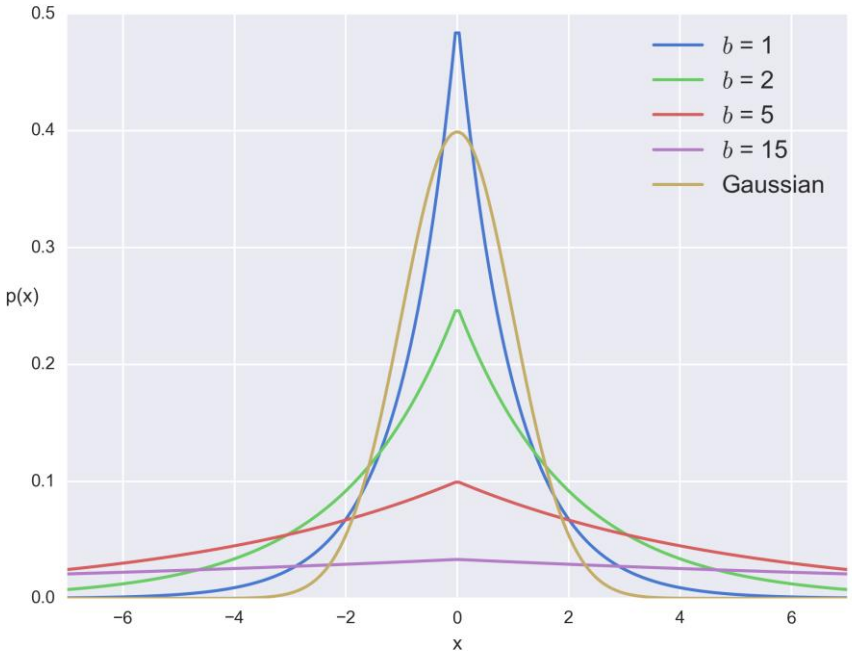

mus
Frequency
mus
Sample value
sigma
Frequency
sigma
Sample value
bd
Frequency
bd
Sample value

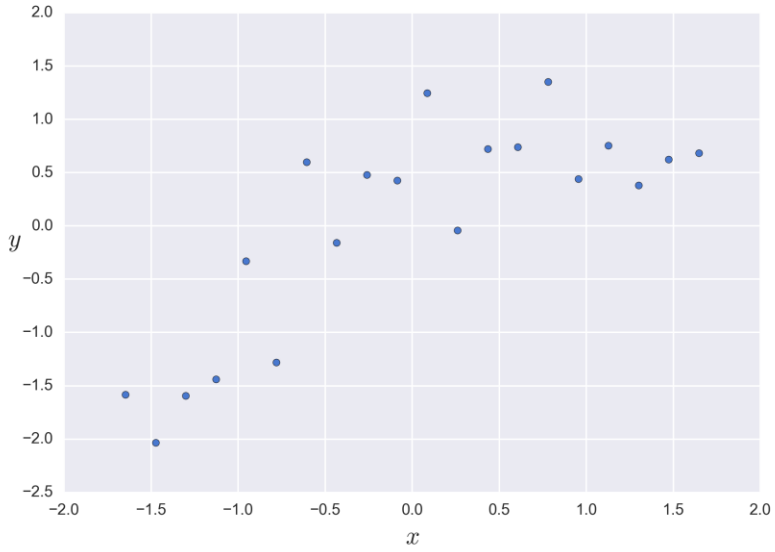

sepal_length

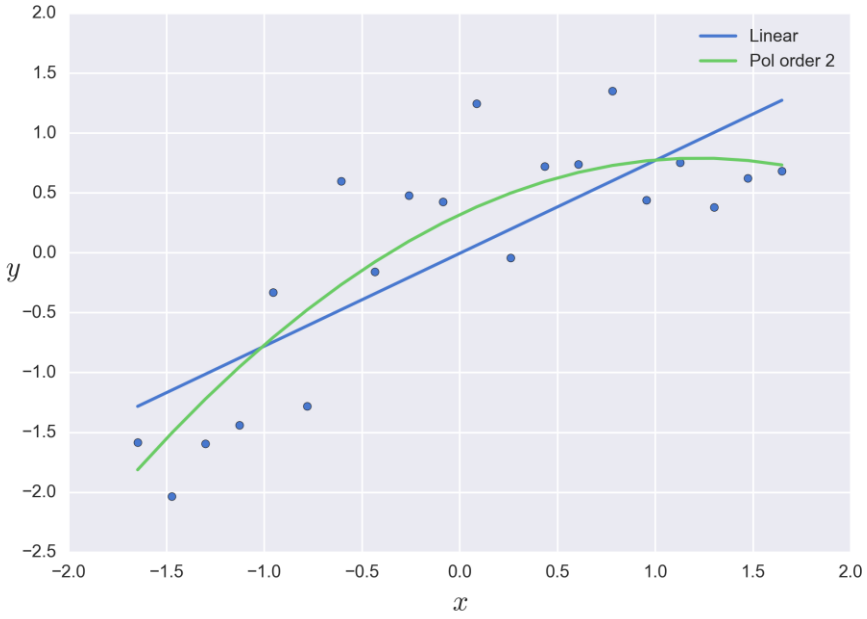
Chapter 6: Model Comparison

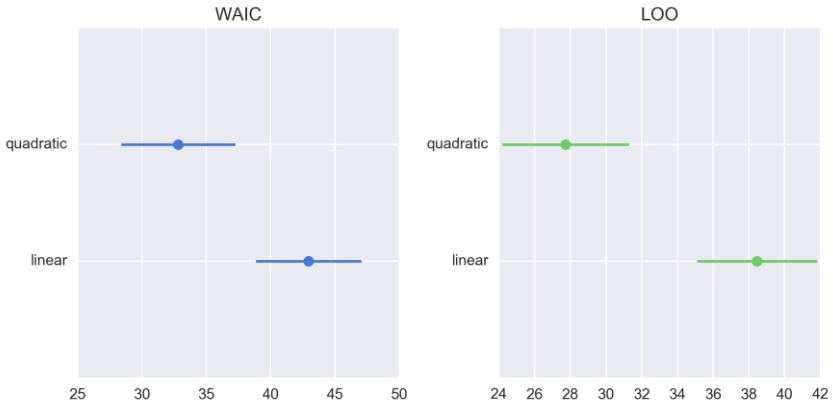


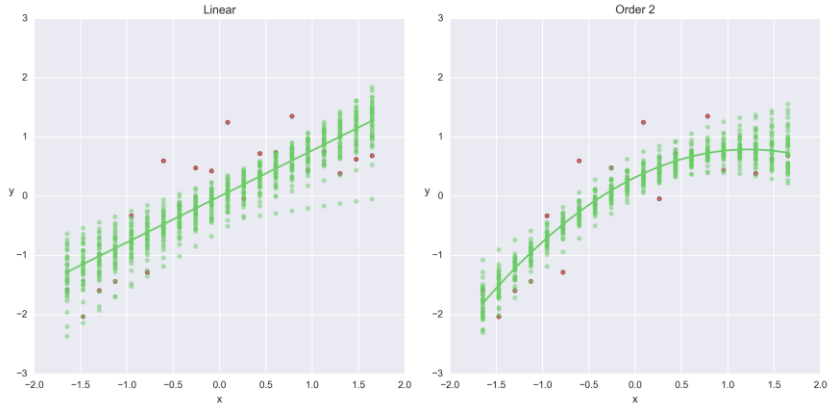

order 0, R^2= 0.00
order 1, R^2= 0.91
order 2, R^2= 0.98
order 5, R^2= 1.00
11
10
9
8
7
6
5
4
y
4
6
8
10
12
14
x

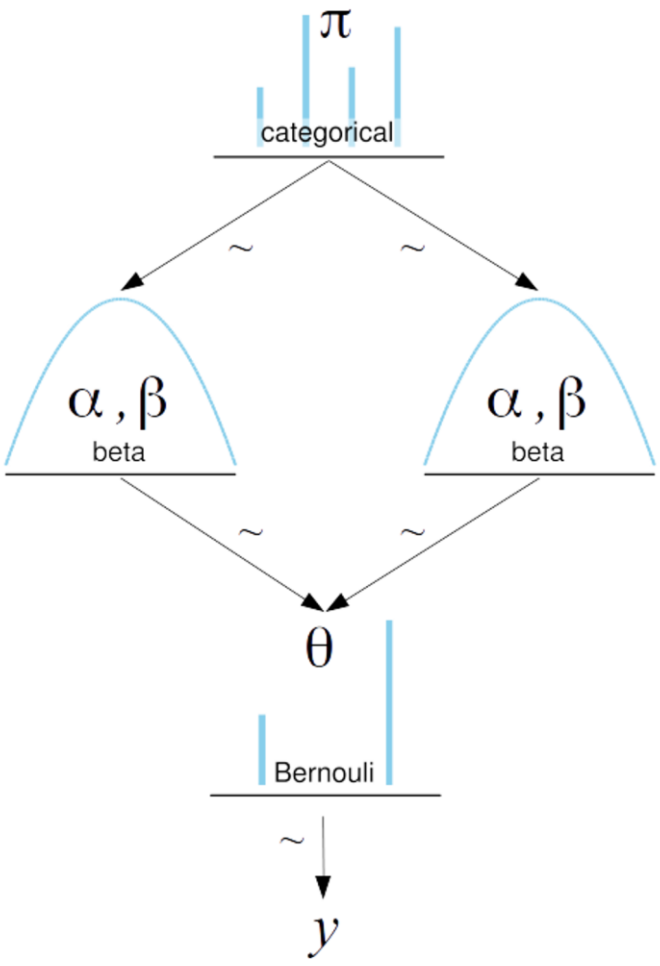

b = 1
b = 2
b = 5
b = 15
Gaussian
0.5
0.4
0.3
0.2
0.1
0.0
p(x)
−6
−4
−2
0
2
4
6
x

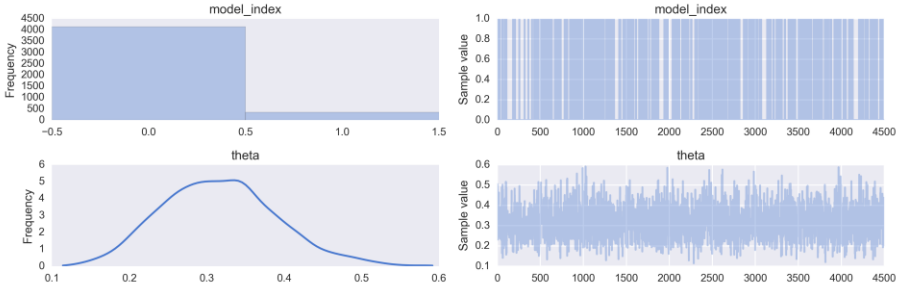

2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
y
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
x

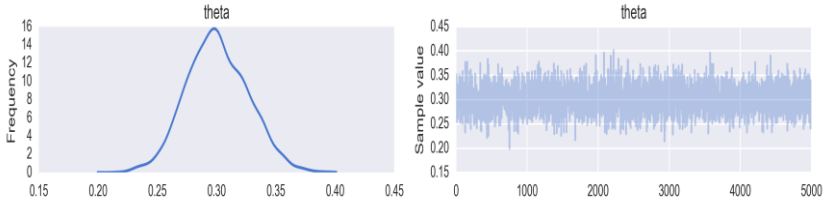

Linear
Pol order 2
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
y
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
x

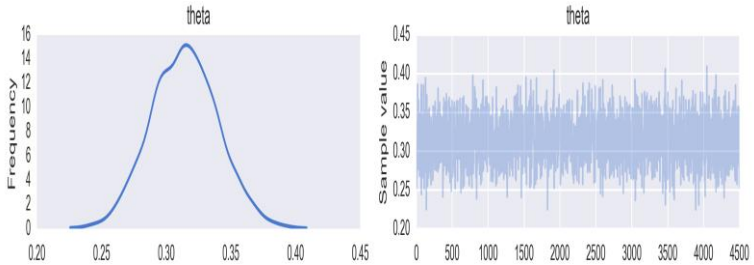

WAIC
quadratic
linear
25
30
35
40
45
50
LOO
quadratic
linear
24
26
28
30
32
34
36
38
40
42

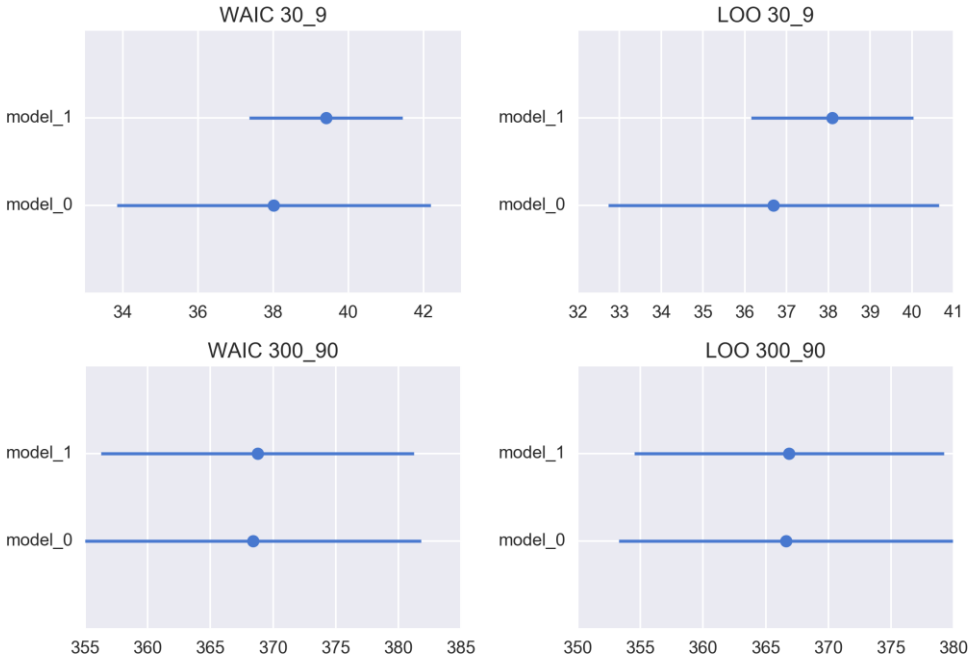

Linear
Order 2
y
x
3
2
1
0
−1
−2
−3
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0


π
categorical
~
~
α , β
beta
α , β
beta
~
~
θ
Bernouli
~
y

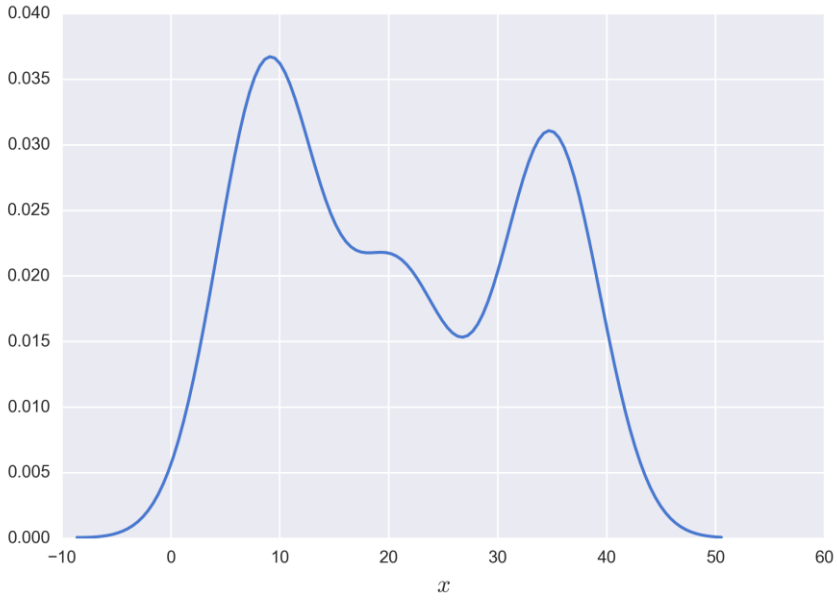

model_index
Frequency
model_index
Sample value
theta
Frequency
theta
Sample value

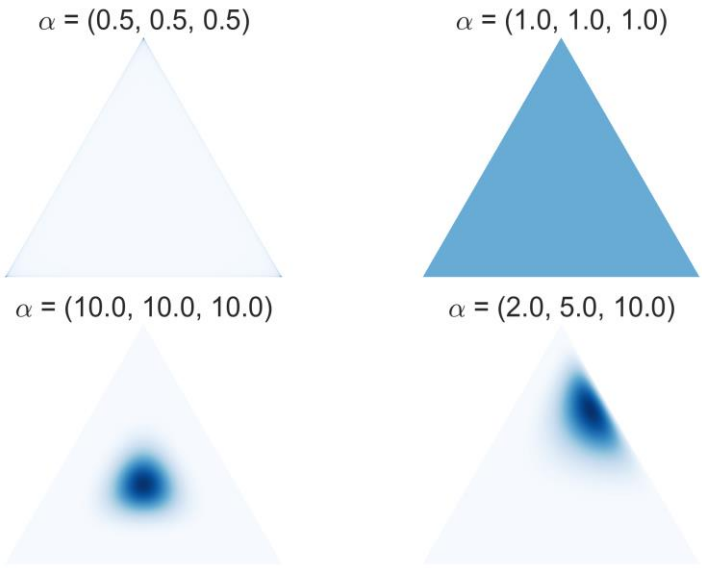

theta
Frequency
Sample value
theta

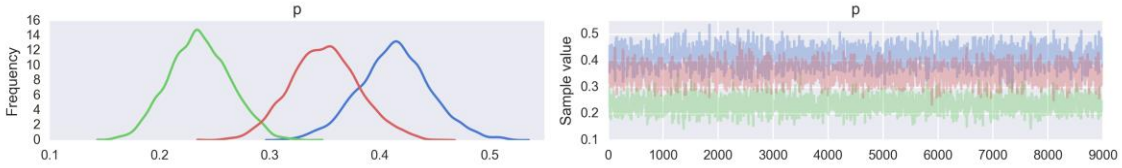

theta
Frequency
Sample value
theta

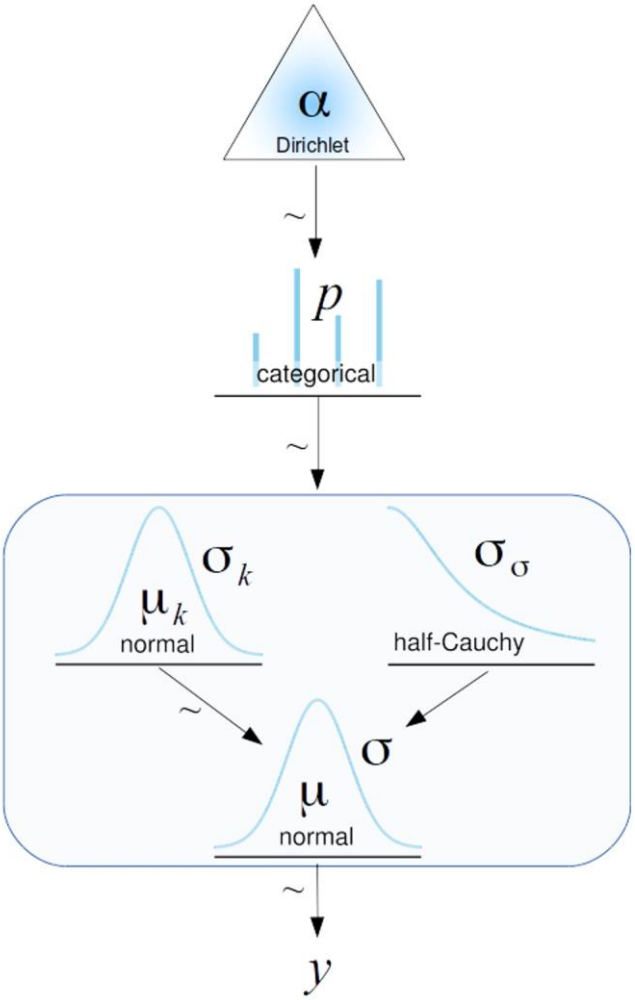

WAIC 30_9
LOO 30_9
WAIC 300_90
LOO 300_90
model_1
model_0

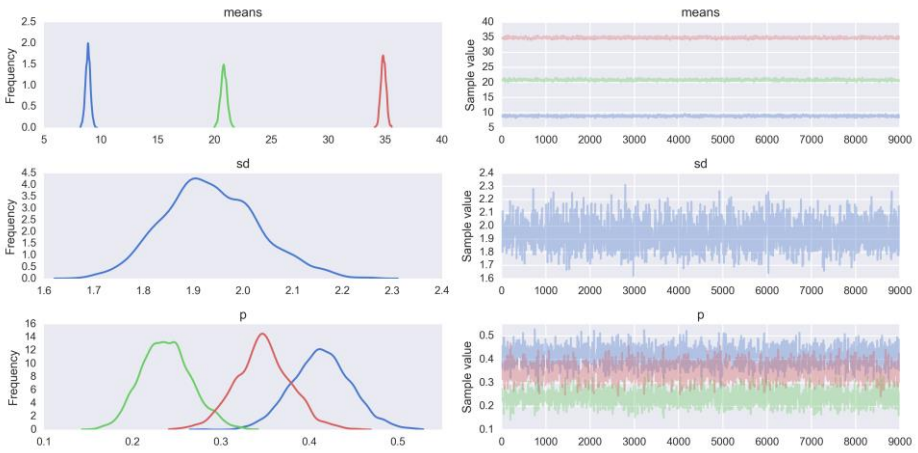
Chapter 7: Mixture Models

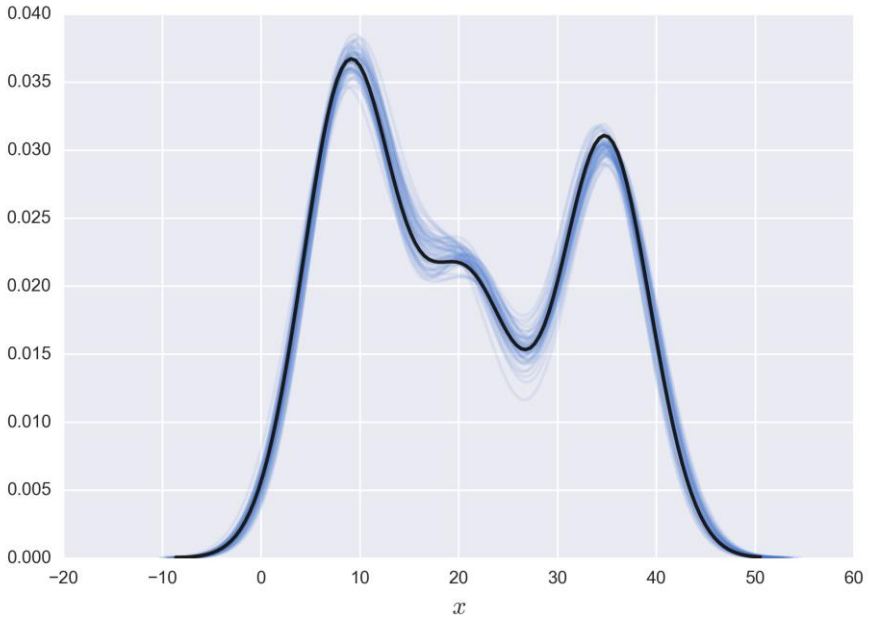


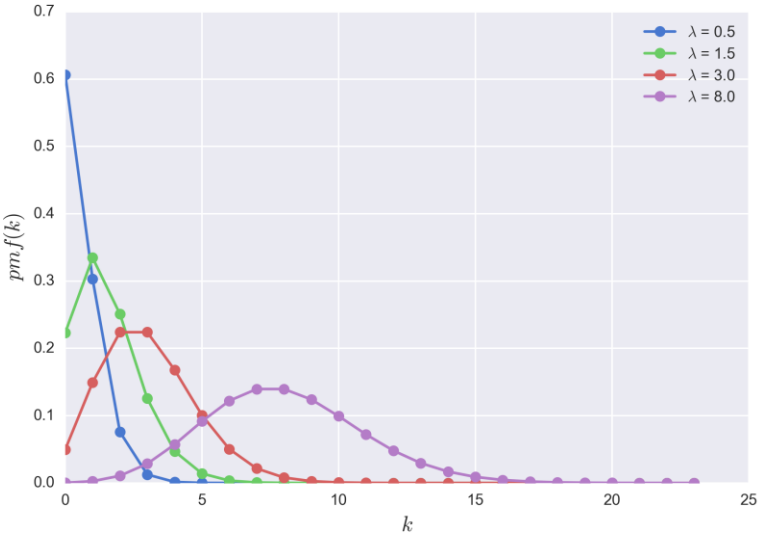

α = (0.5, 0.5, 0.5)
α = (1.0, 1.0, 1.0)
α = (10.0, 10.0, 10.0)
α = (2.0, 5.0, 10.0)

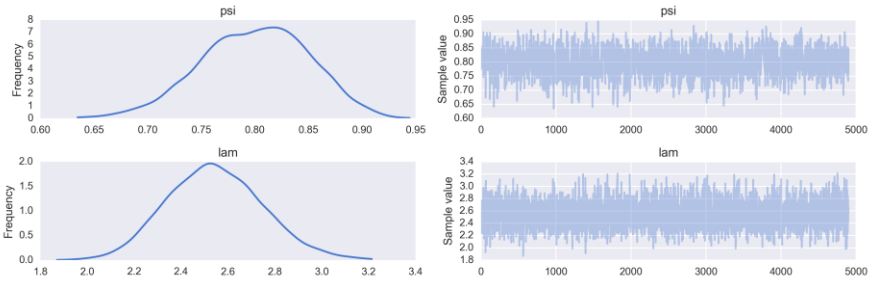

p
Frequency
16
14
12
10
8
6
4
2
0
0.1
0.2
0.3
0.4
0.5
p
Sample value
0.5
0.4
0.3
0.2
0.1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

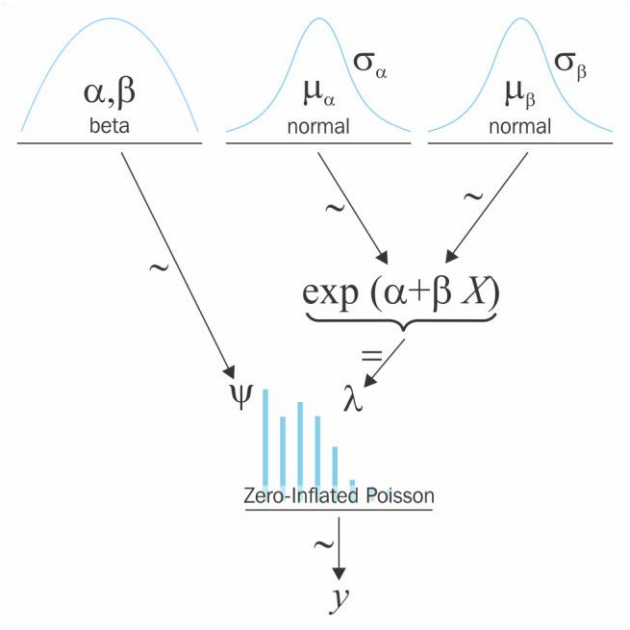

α
Dirichlet
~
p
categorical
~
σ_k
μ_k
normal
σ_σ
half-Cauchy
~
σ
μ
normal
~
y

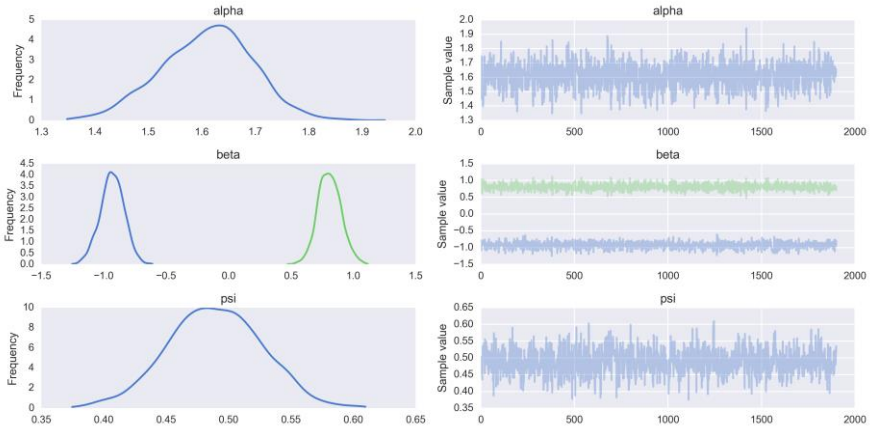

means
Frequency
means
Sample value
sd
Frequency
sd
Sample value
p
Frequency
p
Sample value

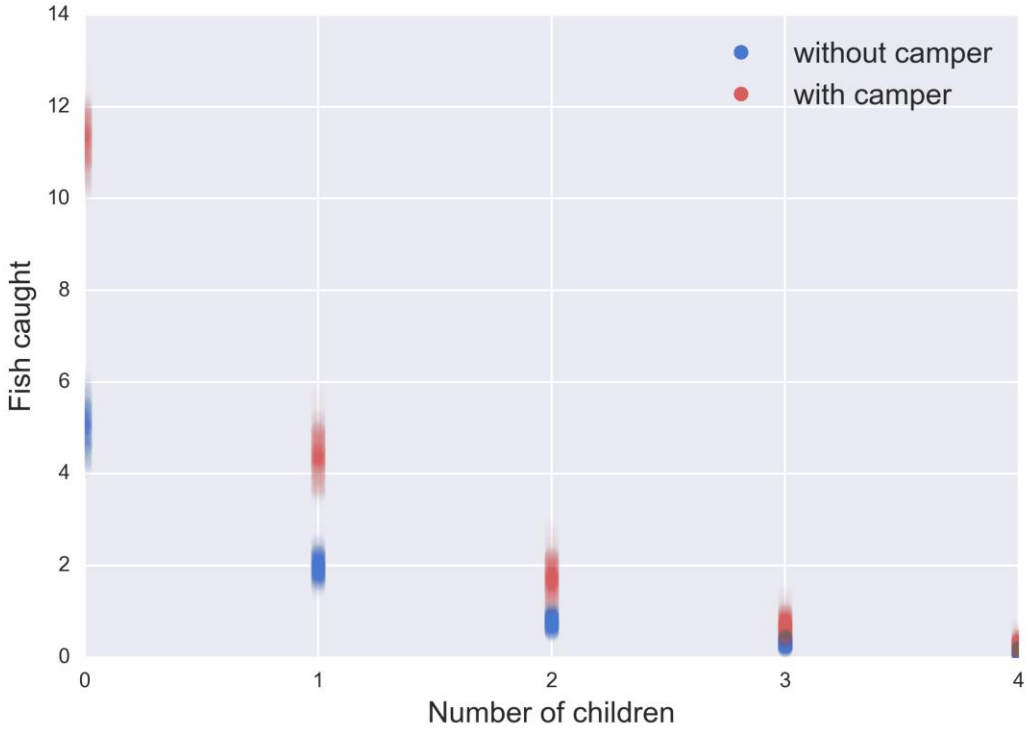

x

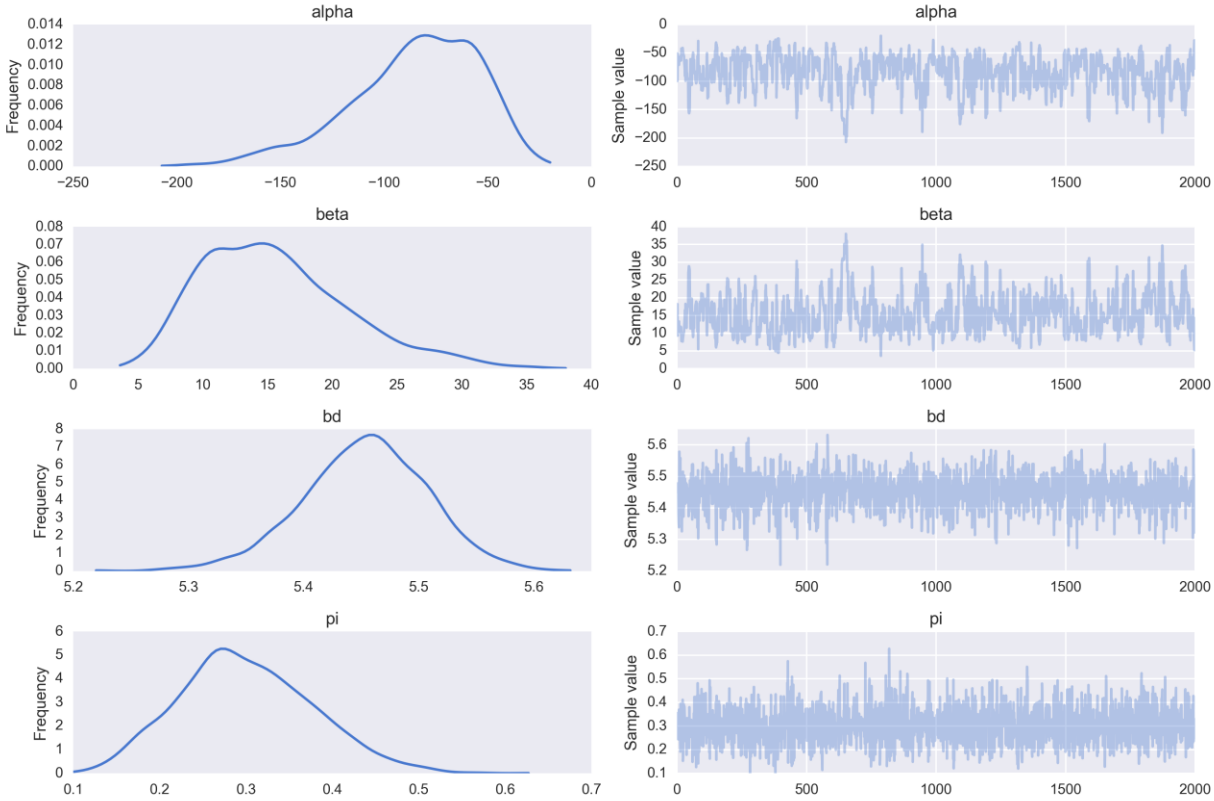

λ = 0.5
λ = 1.5
λ = 3.0
λ = 8.0
pmf(k)
k

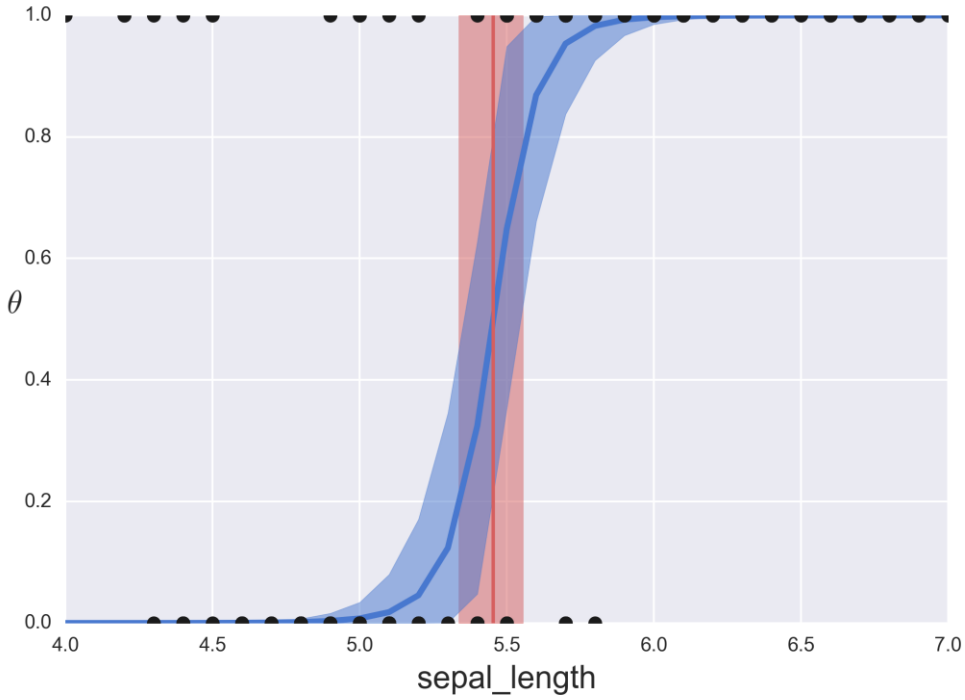

psi
Frequency
psi
Sample value
lam
Frequency
lam
Sample value


α,β
beta
μα
σα
normal
μβ
σβ
normal
~
~
~
exp (α+β X)
=
ψ
λ
Zero-Inflated Poisson
~
y

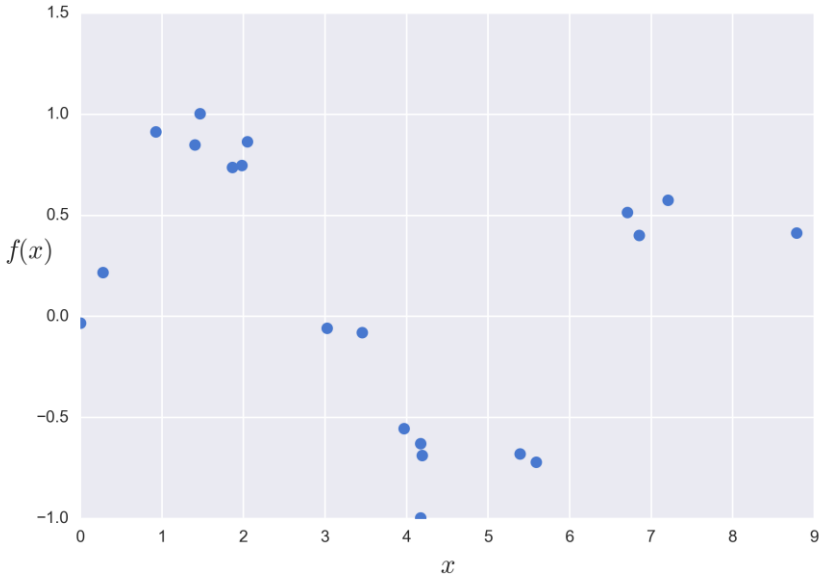

alpha
Frequency
alpha
Sample value
beta
Frequency
beta
Sample value
psi
Frequency
psi
Sample value

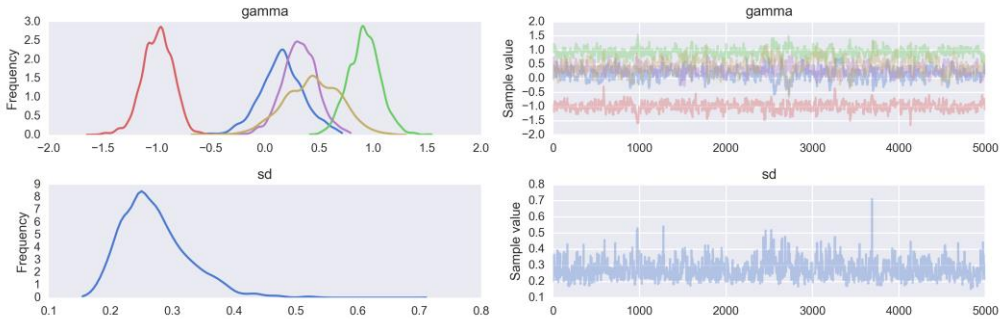

without camper
with camper
Fish caught
Number of children
0
2
4
6
8
10
12
14
0
1
2
3
4

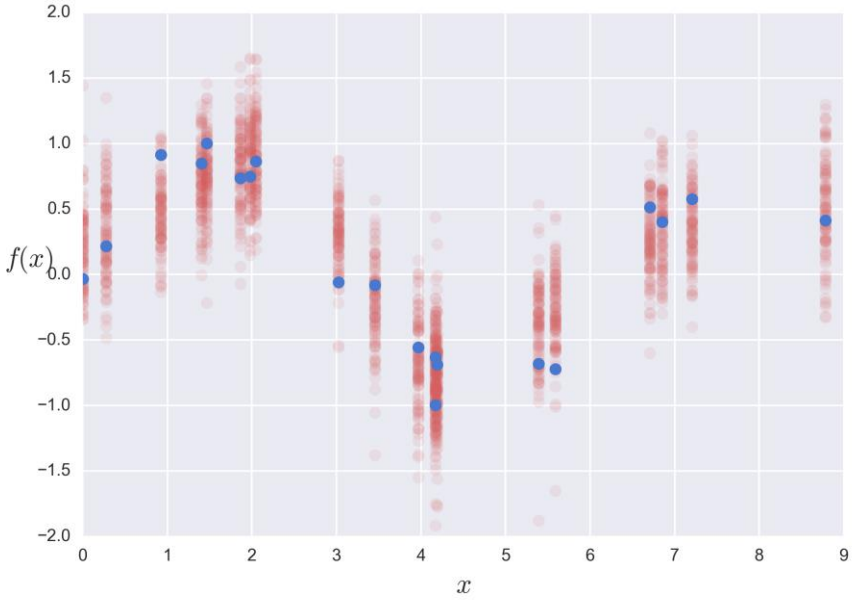

alpha
Frequency
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
−250
−200
−150
−100
−50
0
alpha
Sample value
0
−50
−100
−150
−200
−250
0
500
1000
1500
2000
beta
Frequency
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
25
30
35
40
beta
Sample value
40
35
30
25
20
15
10
5
0
0
500
1000
1500
2000
bd
Frequency
8
7
6
5
4
3
2
1
0
5.2
5.3
5.4
5.5
5.6
bd
Sample value
5.6
5.5
5.4
5.3
5.2
0
500
1000
1500
2000
pi
Frequency
6
5
4
3
2
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
pi
Sample value
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
500
1000
1500
2000

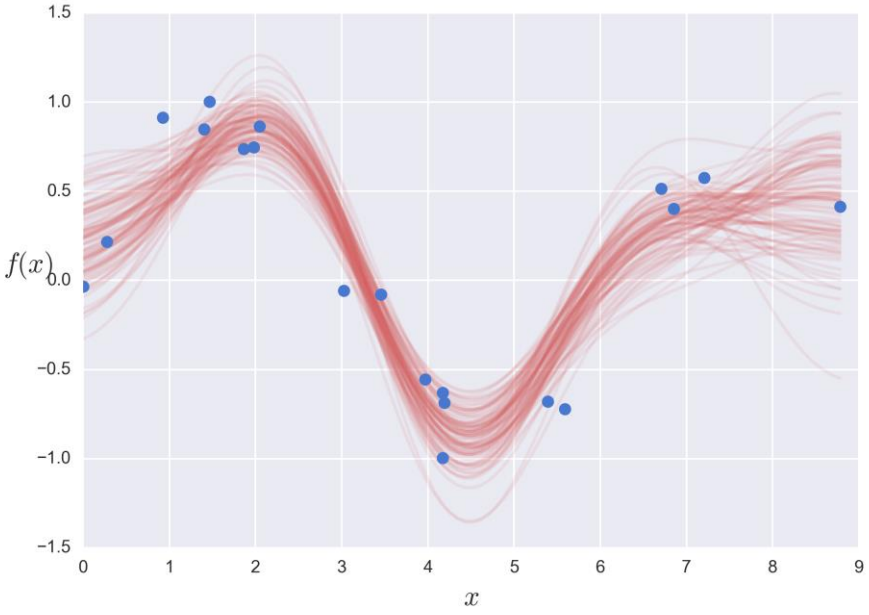

1.0
0.8
0.6
θ
0.4
0.2
0.0
4.0
4.5
5.0
5.5
6.0
6.5
7.0
sepal_length

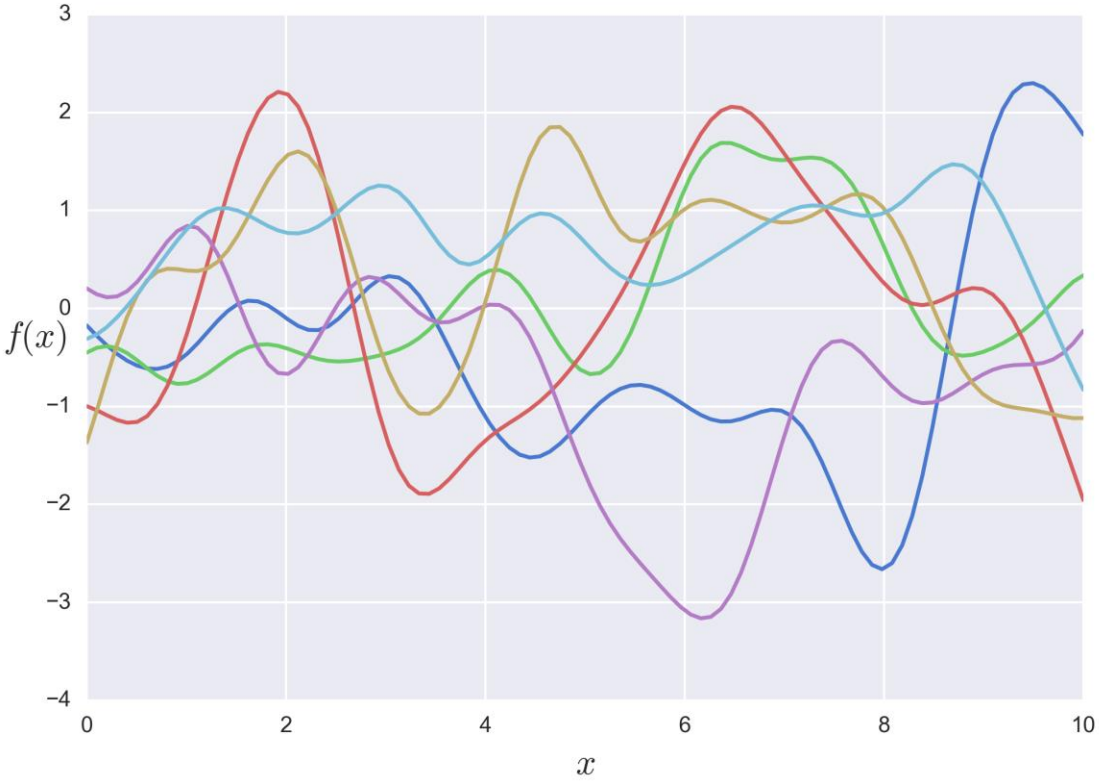
Chapter 8: Gaussian Processes

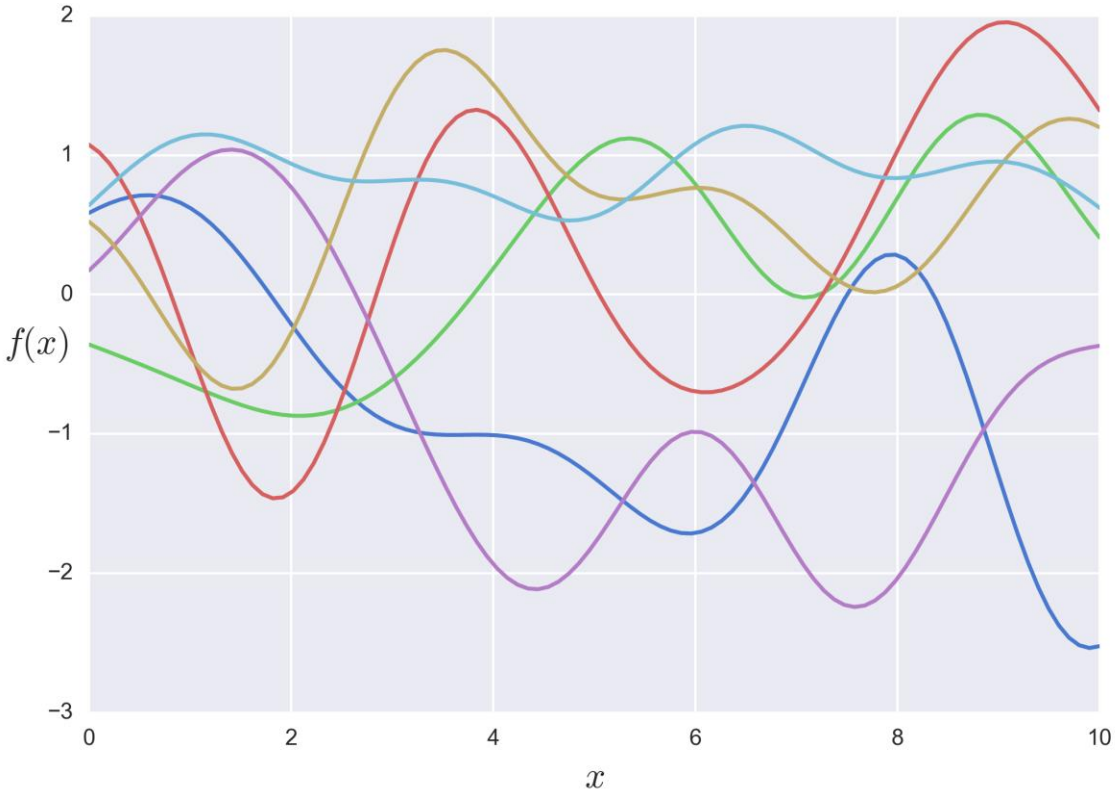


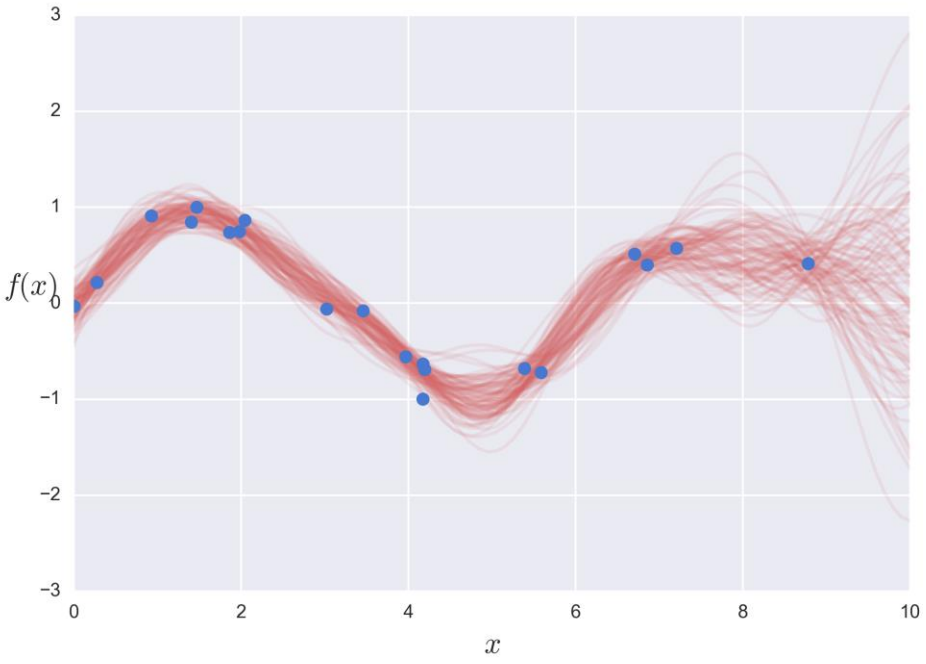


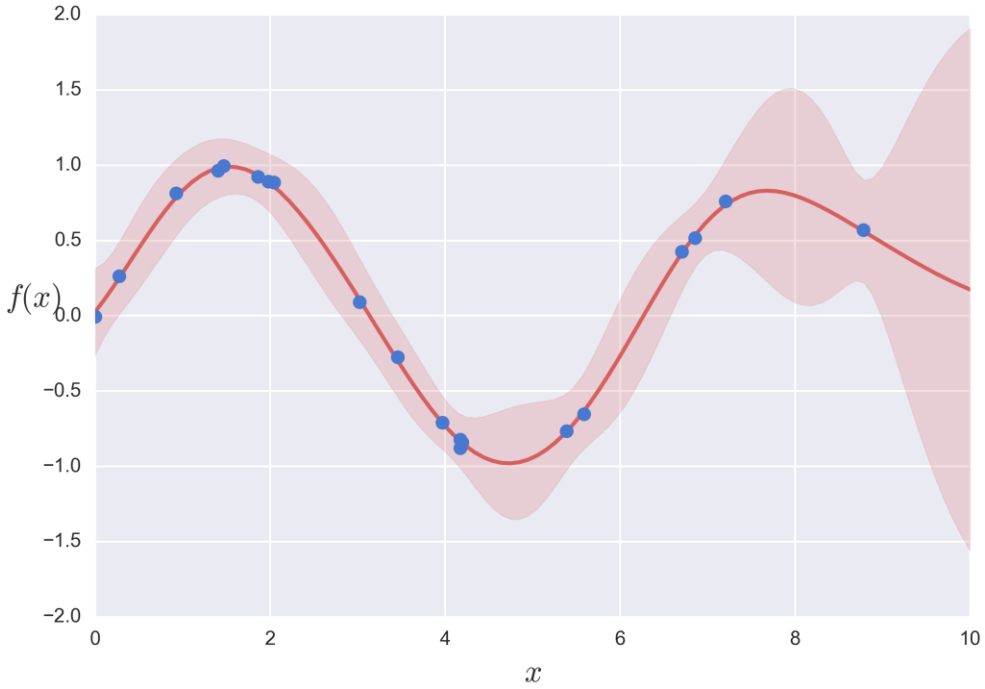

2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
$f(x)$
0
1
2
3
4
5
6
7
8
9
x

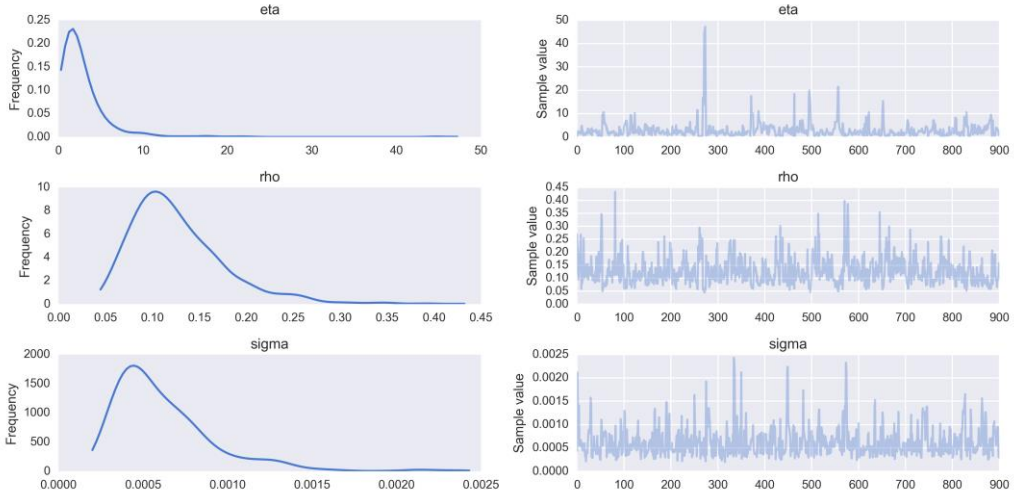

1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
$f(x)$
0
1
2
3
4
5
6
7
8
9
x

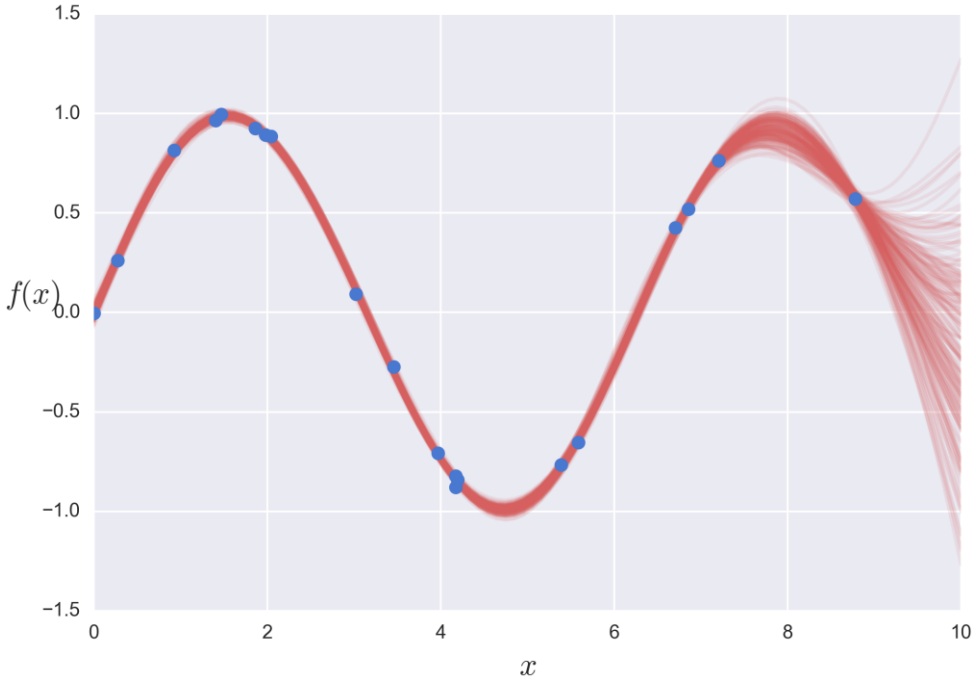

3
2
1
0
−1
−2
−3
−4
$f(x)$
0
2
4
6
8
10
x

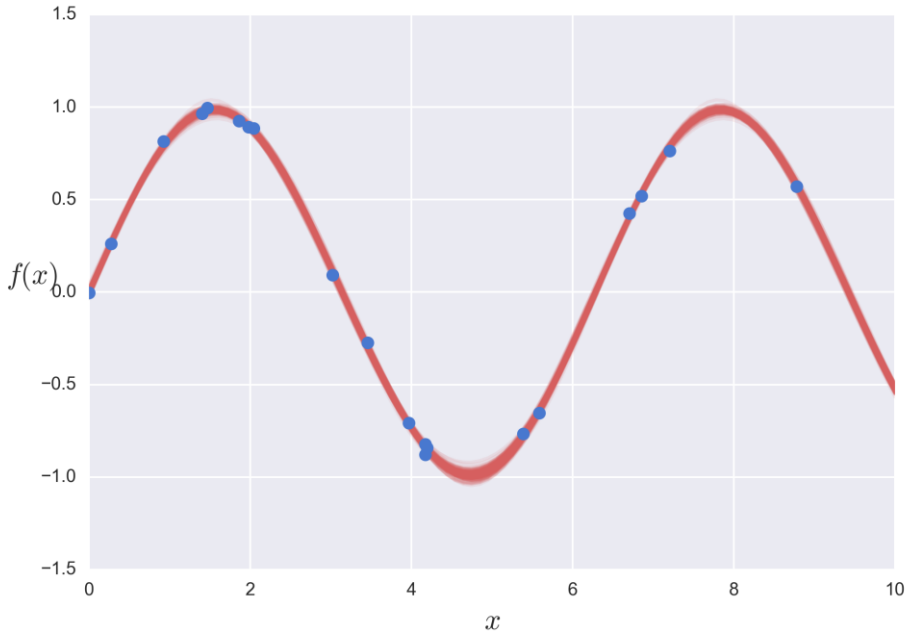

2
1
0
−1
−2
−3
$f(x)$
0
2
4
6
8
10
x


3
2
1
0
−1
−2
−3
$f(x)$
0
2
4
6
8
10
x


2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
$f(x)$
0
2
4
6
8
10
x


eta
Frequency
rho
Frequency
sigma
Frequency
eta
Sample value
rho
Sample value
sigma
Sample value


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
$f(x)$
0
2
4
6
8
10
x


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
$f(x)$
0
2
4
6
8
10
x