Bayesian book

Ben Lambert

December 24, 2014

$$P(\theta|data) = \frac{P(data|\theta) \times P(\theta)}{P(data)}$$

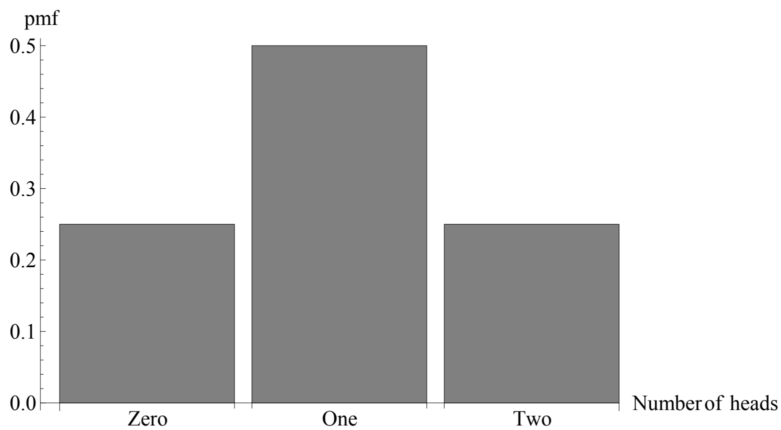
θ

$$P(\theta|data) = \frac{P(data|\theta) \times P(\theta)}{P(data)}$$

$P(data|\theta)$

$$P(data|\theta) = Probability(data|\theta, Model\ Choice)$$

$|\theta\theta$



$\theta = \frac{1}{2}\frac{1}{2}$

$$
\begin{aligned}
P(HH|\theta, Model) &= P(H|\theta, Model) \times P(H|\theta, Model) \\
&= \theta \times \theta = \theta^2 \\
&= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}
\end{aligned}
$$

$theta = \frac{1}{2} theta$

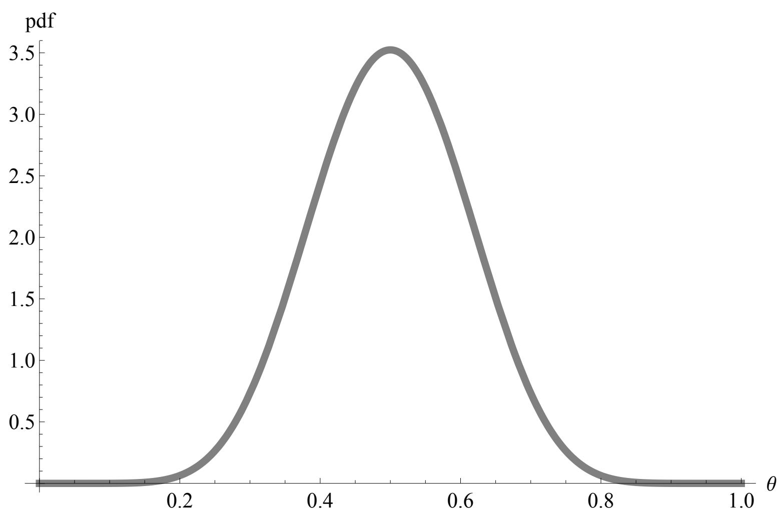
-
-

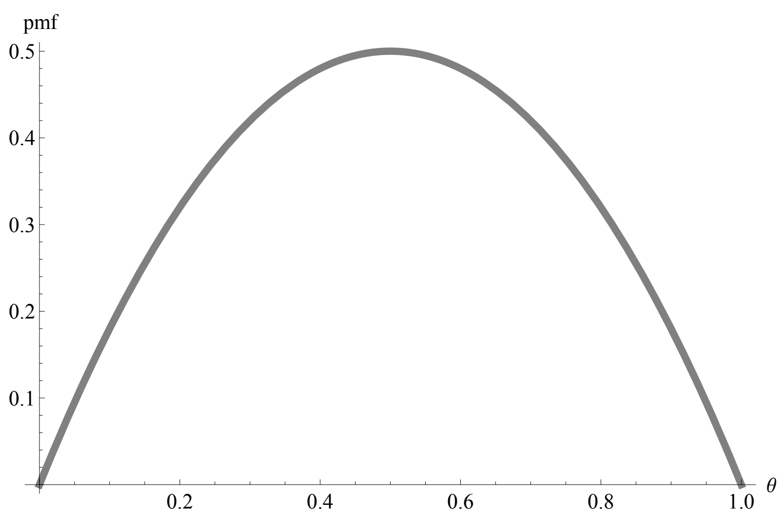
$theta P(data|\theta)$

$\theta = \frac{1}{2}$

$\theta P(\theta|data) P(data|\theta) theta \theta$

$\theta P(data|\theta) P(data|\theta) \theta\theta$


pdf
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.2
0.4
0.6
0.8
1.0
θ


pmf
0.5
0.4
0.3
0.2
0.1
0.2
0.4
0.6
0.8
1.0
θ

$\frac{1}{3}$

$$\mathcal{L}(\theta|data) = P(data|\theta)$$

$\theta \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\} X$

$$\begin{aligned}
P(X = 0|\theta) &= P(TT|\theta) = P(T|\theta) \times P(T|\theta) = (1 - \theta)^2 \\
P(X = 1|\theta) &= P(HT|\theta) + P(TH|\theta) = 2 \times P(T|\theta) \times P(H|\theta) = 2\theta(1 - \theta) \\
P(X = 2|\theta) &= P(HH|\theta) = P(H|\theta) \times P(H|\theta) = \theta^2
\end{aligned}$$

$(1 - \theta)(1 - \theta)\{\}$

$\theta\theta\theta\theta$

θ

$\{\}$

-
-

$\theta\theta$

-
-
-
-
-
-

X

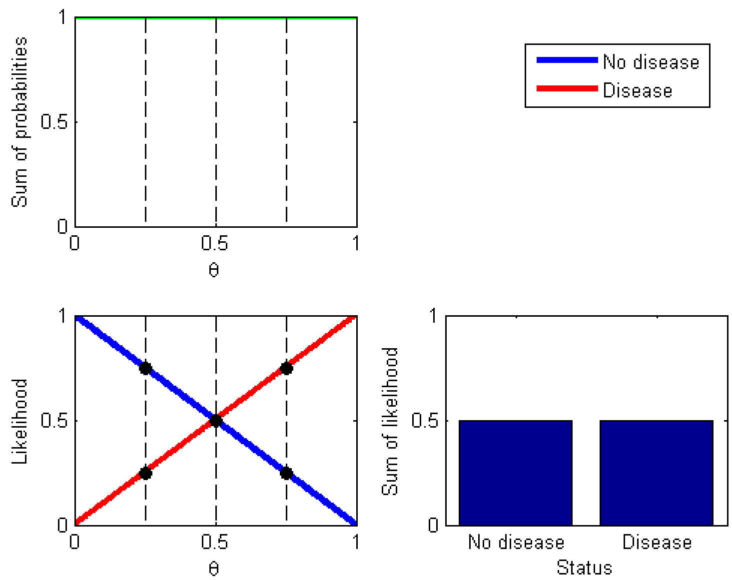
$$X = \begin{cases} O, No\ disease \\ 1\ , Positive\ diagnosis \end{cases}$$

θ

$$P(X = 0|\theta) = (1 - \theta)$$
$$P(X = 1|\theta) = \theta$$

$X = 0 X = 1$

$$P(X = \alpha|\theta) = \theta^{\alpha}(1 - \theta)^{1-\alpha}$$



$\alpha \in \{0, 1\} X$

α

$$P(X = 0|\theta) = \theta^0 (1 - \theta)^1 = (1 - \theta)$$
$$P(X = 1|\theta) = \theta^1 (1 - \theta)^0 = \theta$$

$X \theta theta \{0, 1\}$

NZ

ZN

$$P(X = \alpha|\theta) = \theta^\alpha (1 - \theta)^{1-\alpha}$$

$\alpha \in \{0,1\} X X1 X2$

$$\begin{aligned} P(X_1 = \alpha_1, X_2 = \alpha_2 | \theta_1, \theta_2) &= P(X_1 = \alpha_1 | \theta_1) \ \times \ P(X_2 = \alpha_2 | \theta_2) \\ &= \theta_1^{\alpha_1} (1 - \theta_1)^{1-\alpha_1} \times \theta_2^{\alpha_2} (1 - \theta_2)^{1-\alpha_2} \end{aligned}$$

$\theta_1 \theta_2$

$\theta_1 = \theta_2$

$$\begin{aligned} P(X_1 = \alpha_1, X_2 = \alpha_2 | \theta) &= \theta^{\alpha_1} (1 - \theta)^{1-\alpha_1} \times \theta^{\alpha_2} (1 - \theta)^{1-\alpha_2} \\ &= \theta^{\alpha_1 + \alpha_2} (1 - \theta)^{2-\alpha_1-\alpha_2} \end{aligned}$$

$a^b \times a^c = a^{b+c} \theta (1 - \theta)$

$Z X_1 X_2 Z$

$$Z = X_1 + X_2$$

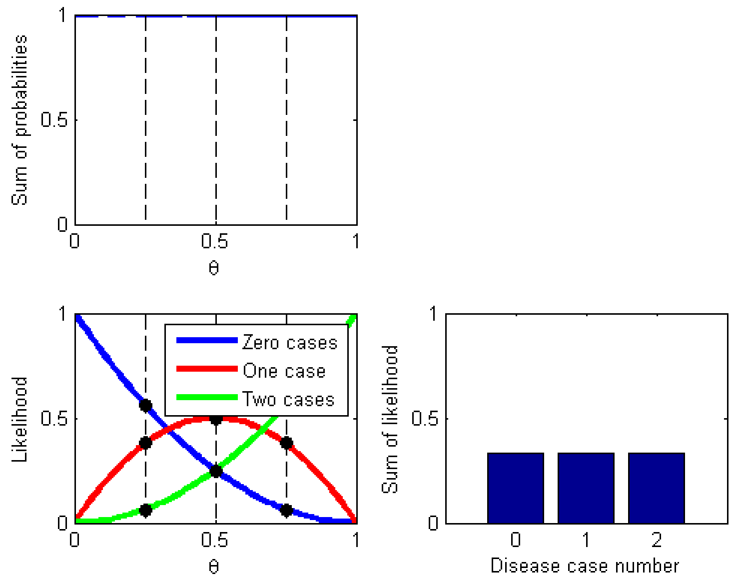
$$\begin{aligned} P(Z = 0|\theta) &= P(X_1 = 0, X_2 = 0|\theta) = \theta^{0+0}(1-\theta)^{2-0-0} = (1-\theta)^2 \\ P(Z = 1|\theta) &= P(X_1 = 1, X_2 = 0|\theta) + P(X_1 = 0, X_2 = 1|\theta) = 2\theta(1-\theta) \\ P(Z = 2|\theta) &= P(X_1 = 1, X_2 = 1|\theta) = \theta^{1+1}(1-\theta)^{2-1-1} = \theta^2 \end{aligned}$$

Z

$$\begin{aligned} P(Z = 0|\theta) &= \ \theta^0 (1-\theta)^2 \\ P(Z = 1|\theta) &= 2\theta^1 (1-\theta)^1 \\ P(Z = 2|\theta) &= \ \theta^2 (1-\theta)^0 \end{aligned}$$

$\theta^\beta (1 - \theta)^{2-\beta} \beta \in \{0, 1, 2\}$

$$P(Z = \beta|\theta) \sim \theta^\beta (1 - \theta)^{2-\beta}$$



$$(x+1)^2 = x^2 + 2x + 1$$

$\{1, 2, 1\}\{x^2, x^1, x^0\}^n C_r$

$$\binom{2}{\beta} = \frac{2!}{(2-\beta)!\beta!}$$

$!\beta \in \{0, 1, 2\} Z$

$$P(Z = \beta|\theta) = \binom{2}{\beta}\theta^{\beta}(1-\theta)^{2-\beta}$$

N

$$
\begin{aligned}
P(Z=0|\theta) &= \ P(X_1=0|\theta)P(X_2=0|\theta)P(X_3=0|\theta) \\
P(Z=1|\theta) &= 3P(X_1=1|\theta)P(X_2=0|\theta)P(X_3=0|\theta) \\
P(Z=2|\theta) &= 3P(X_1=1|\theta)P(X_2=1|\theta)P(X_3=0|\theta) \\
P(Z=3|\theta) &= \ P(X_1=1|\theta)P(X_2=1|\theta)P(X_3=1|\theta)
\end{aligned}
$$

$\{1,3,3,1\}(x+1)^3$

$$P(Z=\beta|\theta) = \binom{3}{\beta}\theta^{\beta}(1-\theta)^{3-\beta}$$

$$P(Z=\beta|\theta) = \binom{N}{\beta}\theta^{\beta}(1-\theta)^{N-\beta}$$

$\theta = 1\%$

$$P(Z \geq 10|\theta = 0.01) = \sum_{Z=10}^{100}\binom{100}{Z}0.01^{Z}(1-0.01)^{100-Z} = 7.63 \times 10^{-8}$$

$\mu = 70\sigma^2 = 81X$

$$P(X=\alpha|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(\alpha-\mu)^2}{2\sigma^2}}$$

$$\mu = 70\sigma^2 = 81\mu\sigma$$

$$P(X \geq 90|\mu = 70, \sigma^2 = 81) = \int_{90}^{\infty} \frac{1}{\sqrt{2\pi \times 10}} e^{-\frac{(\alpha-70)^2}{2\times 10}} \mathrm{d}\alpha$$
$$= 1 - \Phi\left(\frac{90-70}{9}\right) \approx 0.0131$$

Φ

$$P(X_1 = \alpha_1, X_2 = \alpha_2, ..., X_N = \alpha_N|\mu, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\alpha_i-\mu)^2}{2\sigma^2}}$$

θ

$$L(\theta|data) = \binom{100}{10} \theta^{10} (1-\theta)^{100-10}$$

$\theta\theta\theta\theta$

$\mu\sigma^2$

$$l(\theta|data) = LogL(\theta|data) = log\binom{100}{10} + 10log(\theta) + 90log(1-\theta)$$

$$log(ab) = log(a) + log(b)$$
$$log(a^b) = blog(a)$$

$l(\theta|data)$

$$\frac{\partial l}{\partial \theta} = \frac{10}{\hat{\theta}} - \frac{90}{1-\hat{\theta}} = 0$$

$\hat{\theta} = \frac{1}{10}$

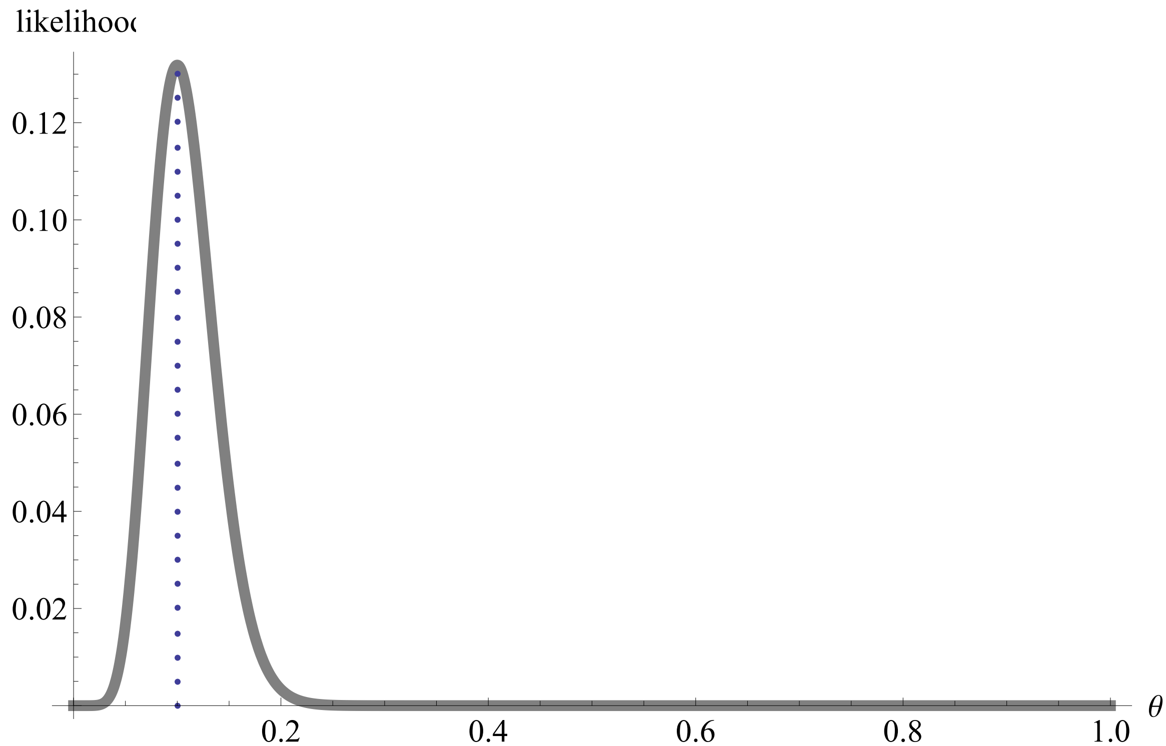
βN

$$\hat{\theta} = \frac{\beta}{N}$$

$\{75, 71\}$

$$L(\mu, \sigma^2|X_1 = 75, X_2 = 71) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(75-\mu)^2}{2\sigma^2}} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(71-\mu)^2}{2\sigma^2}}$$

$$l(\mu, \sigma^2|X_1 = 75, X_2 = 71) = 2log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(75-\mu)^2}{2\sigma^2} - \frac{(71-\mu)^2}{2\sigma^2}$$

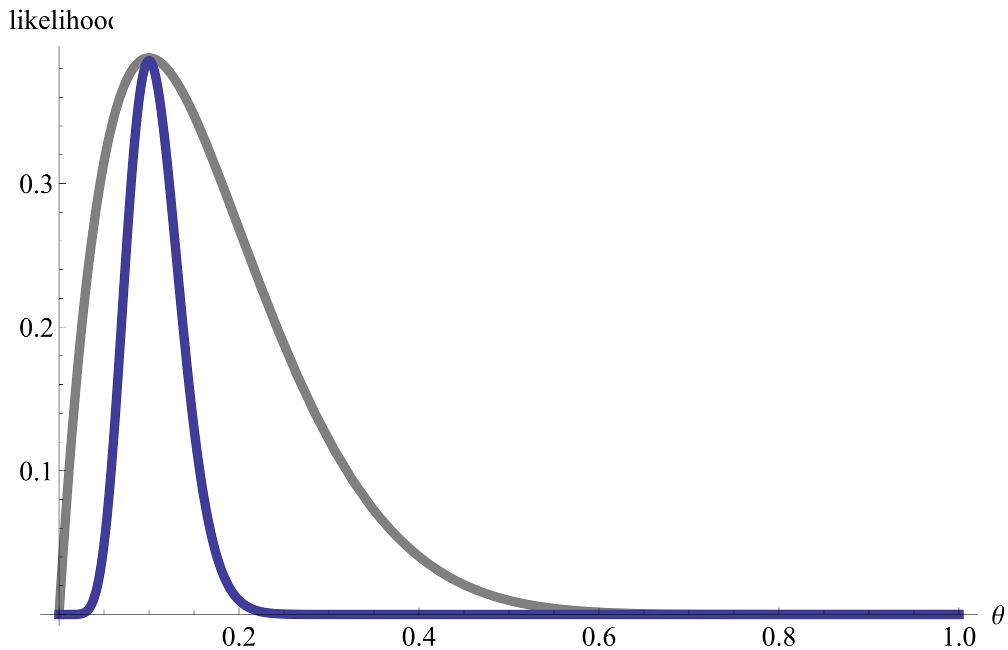

likelihood
0.12
0.10
0.08
0.06
0.04
0.02
0.2
0.4
0.6
0.8
1.0
θ

$$\frac{\partial l}{\partial \mu} = \frac{(75-\hat{\mu})}{\hat{\sigma^2}} + \frac{(71-\hat{\mu})}{\hat{\sigma^2}} = 0$$
$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{\hat{\sigma^2}} + \frac{(75-\hat{\mu})^2 + (71-\hat{\mu})^2}{2\hat{\sigma^4}} = 0$$

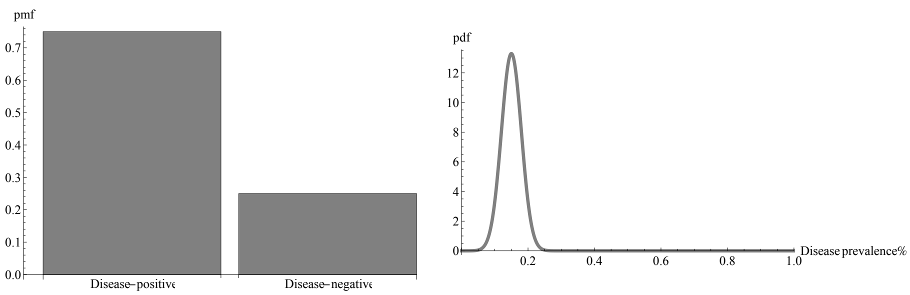
$\hat{\mu} = \frac{71+75}{2} = 73$

$$\hat{\sigma^2} = \frac{1}{2}\left[(75-73)^2 + (71-73)^2)\right] = 4$$

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i = \bar{X}$$
$$\hat{\sigma^2} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 = s^2$$


likelihood
0.3
0.2
0.1
0.2
0.4
0.6
0.8
1.0
θ

$$P(\theta|data) = \frac{P(data|\theta) \times P(\theta)}{P(data)}$$



θ

$\theta P(\theta) = 1$

$$P(\theta|data) = \frac{P(data|\theta)}{P(data)}$$

$P(\theta)P(\theta) = 1 \int\limits_{-\infty}^{+\infty} P(\theta)\mathrm{d}\theta \to \infty P(\theta)$

$\{H, H\} theta = 1 theta = 0$

$P(data|\theta)\frac{1}{4}$

$\frac{1}{6}YYX \in \{0, 1\}$

$$P(X = 1|Y = \alpha) = \frac{\alpha}{5}$$

$\alpha \in \{0, 1, 2, 3, 4, 5\}$

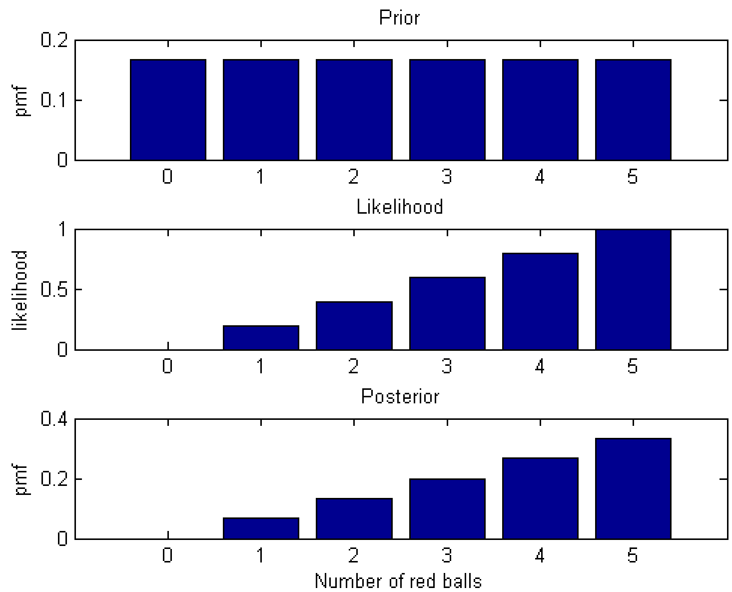
$P(data) = \frac{1}{2}$

$= \frac{Prior \times Likelihood}{P(data)}$

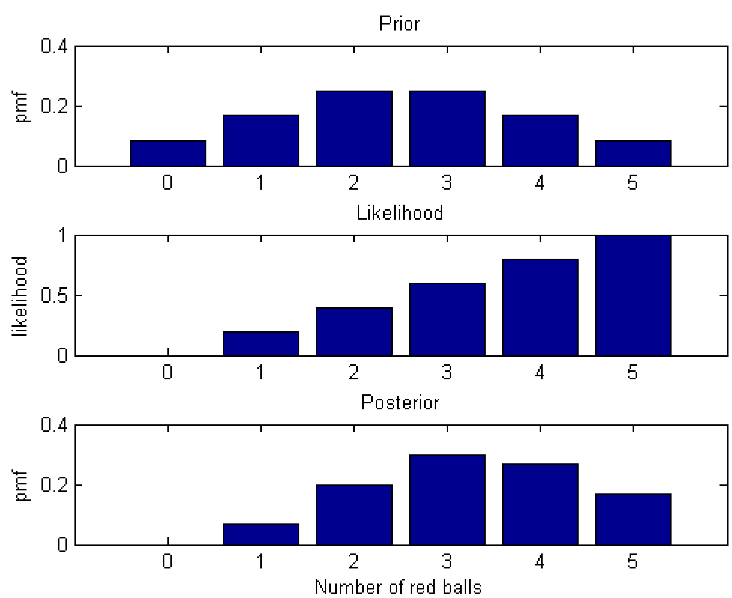
$P(data) = 1/2$

$= \frac{Prior \times Likelihood}{P(data)}$

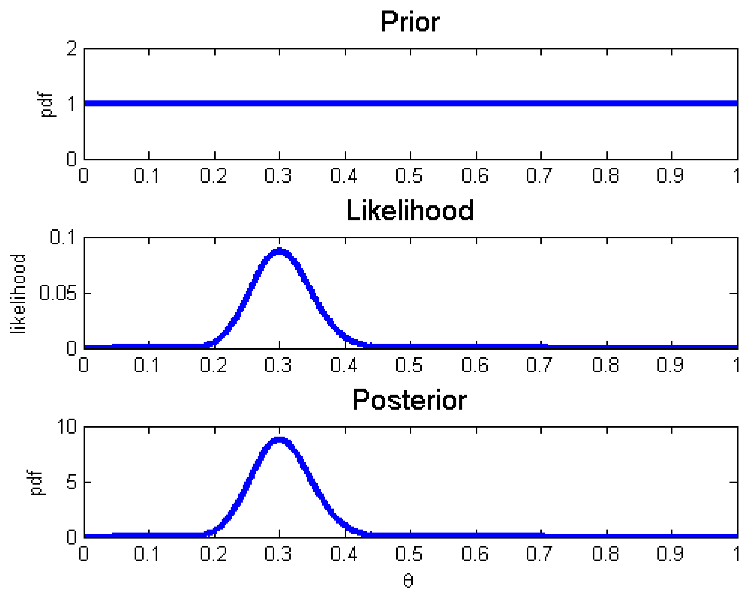
$\theta\theta$



Prior
pmf
0.2
0.1
0
0 1 2 3 4 5
Likelihood
likelihood
1
0.5
0
0 1 2 3 4 5
Posterior
pmf
0.4
0.2
0
0 1 2 3 4 5
Number of red balls



Prior
pmf
0.4
0.2
0
0 1 2 3 4 5
Likelihood
likelihood
1
0.5
0
0 1 2 3 4 5
Posterior
pmf
0.4
0.2
0
0 1 2 3 4 5
Number of red balls



θ

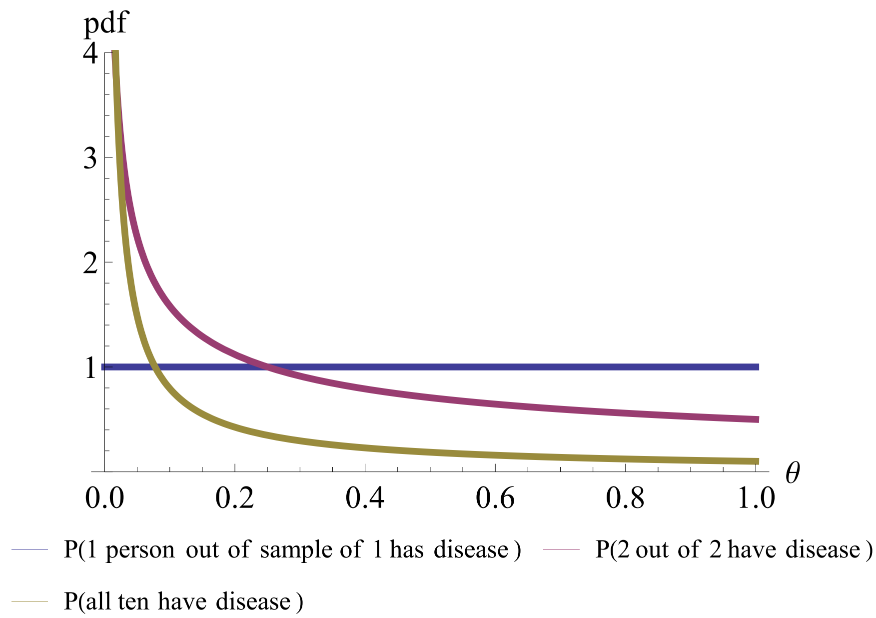
$$P(Z=3|\theta) = \binom{100}{3}\theta^{\beta}(1-\theta)^{100-3}$$

$\binom{100}{3} = 161,700$

$\theta\theta$

$P(\theta) = constant$

$$\begin{aligned} P(\theta|data) &= \frac{P(\theta) \times P(data|\theta)}{P(data)} \\ &\propto P(\theta) \times P(data|\theta) \\ &\propto P(data|\theta) \end{aligned}$$



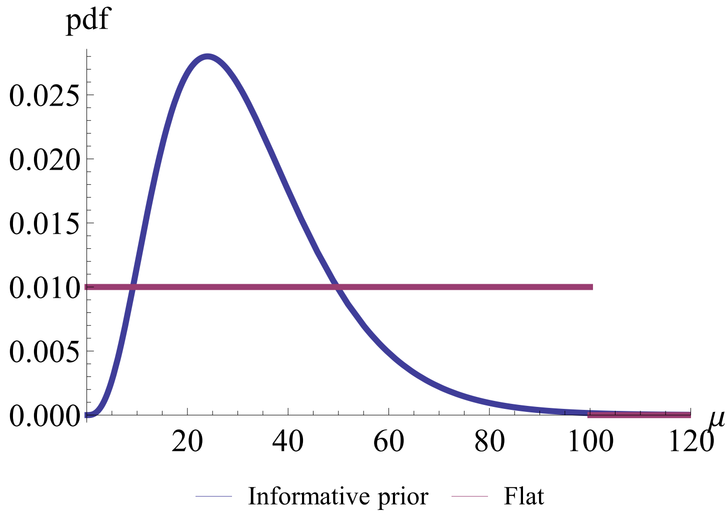
$\theta\theta^2\theta\theta^2\theta$

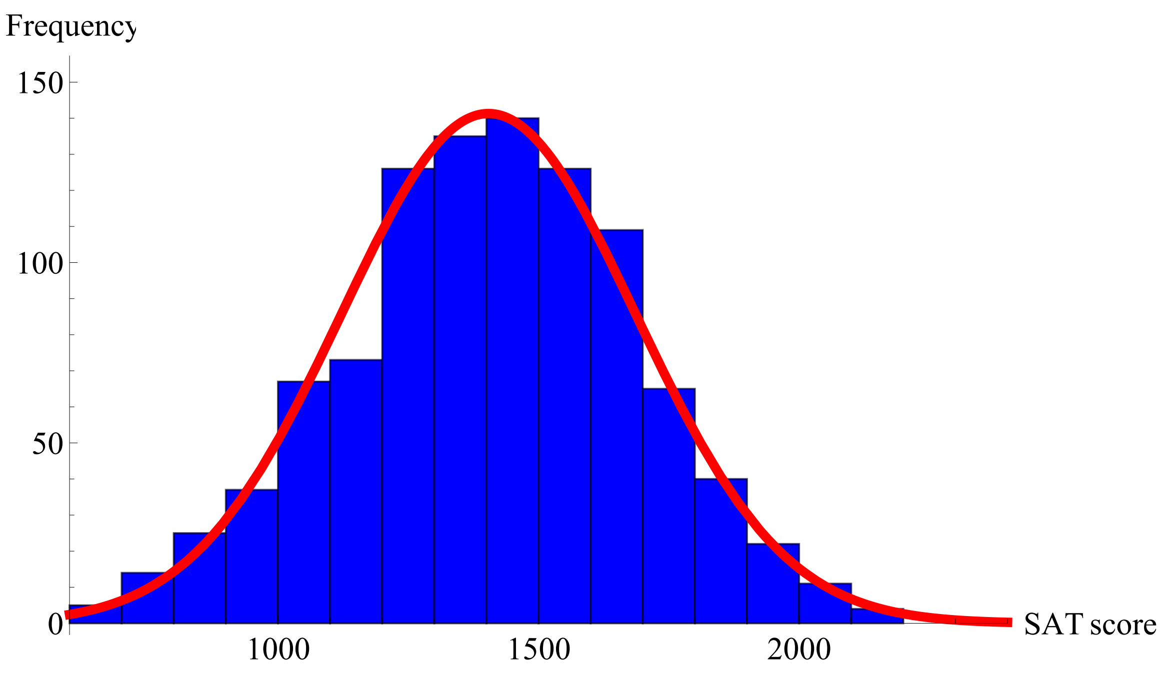
$\theta \in \{0,1\}\mu\mu\mu \sim Unif(0,\infty)\mu \sim Unif(0,\infty)\infty\mu\mu\mu\mu$

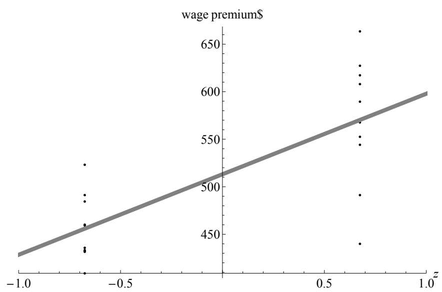
$\mu\sigma^2\mu = 1404\sigma^2 = 79,716$

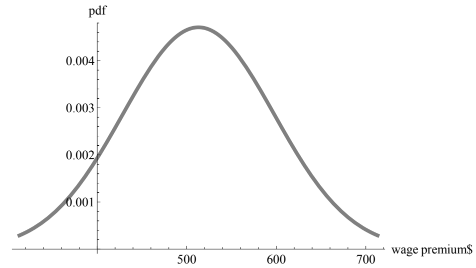
$P(\theta)P(data|\theta)$

$$P(\theta|data) = \frac{P(\theta) \times P(data|\theta)}{P(data)} \propto P(\theta) \times P(data|\theta)$$


pdf
0.025
0.020
0.015
0.010
0.005
0.000
20
40
60
80
100
120
μ
Informative prior
Flat


Frequency
150
100
50
0
1000
1500
2000
SAT score





$\theta\theta\theta\propto$ $P(\theta) \times P(data|\theta)$

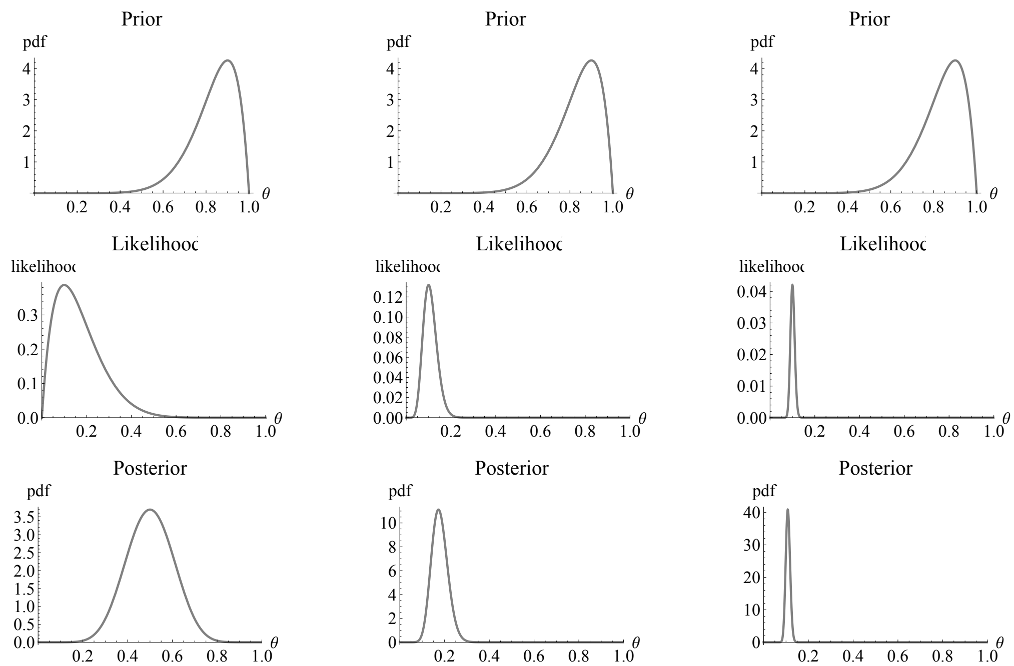
$wage_{25} wage_{75}$

$$z_{25} = \frac{wage_{25} - \mu}{\sigma}$$
$$z_{75} = \frac{wage_{75} - \mu}{\sigma}$$

$z_{25} z_{50}$

$$wage_{25} = \mu + \sigma z_{25}$$
$$wage_{75} = \mu + \sigma z_{75}$$

$y = mx + cc = \mu m = \sigma\mu\sigma$


Prior
pdf
θ
Prior
pdf
θ
Prior
pdf
θ
Likelihooc
likelihooc
θ
Likelihooc
likelihooc
θ
Likelihooc
likelihooc
θ
Posterior
pdf
θ
Posterior
pdf
θ
Posterior
pdf
θ

$$
\begin{aligned}
P(Y = \alpha | X = 1) &= \frac{P(X = 1 | Y = \alpha) \times P(Y = \alpha)}{P(X = 1)} \\
&= \frac{P(X = 1 | Y = \alpha) \times P(Y = \alpha)}{\sum\limits_{\alpha=0}^{5} P(X = 1 | Y = \alpha) \times P(Y = \alpha)} \\
&= \frac{\frac{\alpha}{5} \times \frac{1}{6}}{\sum\limits_{\alpha=0}^{5} \frac{\alpha}{5} \times \frac{1}{6}}
\end{aligned}
$$

$\theta P(\theta) = 1 P(Y) = P(\theta^2)$

$$P(Y) = P(\theta(Y)) \times |\theta'(Y)|$$

$\theta(Y) = Y^{-\frac{1}{2}} Y = \theta^2 \theta' Y$

$$P(Y) = \frac{1}{2\sqrt{Y}}$$

$P(data)$

$$P(\theta|data) = \frac{P(data|\theta) \times P(\theta)}{P(data)}$$

$P(data)$

$P(data)$

$P(data)\theta P(data)\theta P(\theta|data)\theta$

θ

$$P(data) = \sum_{All\ \theta} P(data|\theta) \times P(\theta)$$

8

$$P(data) = \int\limits_{All\ \theta} P(data|\theta) \times P(\theta)\mathrm{d}\theta$$

θ

$$\theta = \begin{cases} O, Disease\ negative \\ 1\ , Disease\ positive \end{cases}$$

$\frac{1}{4}$

$$P(test\ positive|\theta) = \begin{cases} \frac{1}{10}, \theta = 0 \\ \frac{4}{5}\ , \theta = 1 \end{cases}$$

$P(positive|\theta = 0)$

$$\begin{aligned} P(test\ positive) &= \sum_{\theta=0}^{1} P(test\ positive|\theta) \times P(\theta) \\ &= P(test\ positive|\theta = 0) \times P(\theta = 0) + P(test\ positive|\theta = 1) \times P(\theta = 1) \\ &= \frac{1}{10} \times \frac{3}{4} + \frac{4}{5} \times \frac{1}{4} = \frac{11}{40} \end{aligned}$$

$P(test\ negative) = 1 - P(test\ positive) = \frac{29}{40} P(test\ negative)$

θ

$$
\begin{aligned}
P(\theta = 1|test\ positive) &= \frac{P(test\ positive|\theta = 1) \times P(\theta = 1)}{P(test\ positive)} \\
&= \frac{\frac{4}{5} \times \frac{1}{4}}{\frac{1}{10} \times \frac{3}{4} + \frac{4}{5} \times \frac{1}{4}} \\
&= \frac{8}{11}
\end{aligned}
$$

$\frac{1}{4}$

θ

$$
P(Z = \beta|\theta) = \binom{100}{\beta} \theta^{\beta} (1-\theta)^{100-\beta}
$$

$\beta \in [0, 100] \theta = 40\%$

$\theta \in \{0, 1\}$

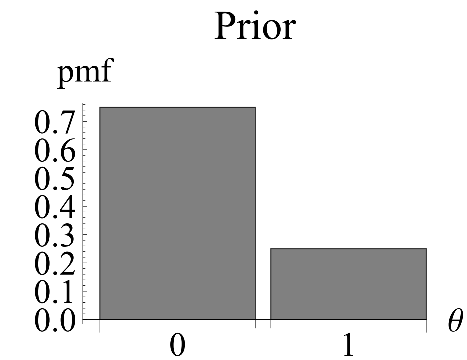
$$
\begin{aligned}
P(Z = 40) &= \int_0^1 P(Z = 40|\theta) \times P(\theta) \mathrm{d}\theta \\
&= \int_0^{0.45} \binom{100}{40} \theta^{40} (1-\theta)^{60} \times \frac{20}{9} \mathrm{d}\theta \\
&\approx 0.018
\end{aligned}
$$

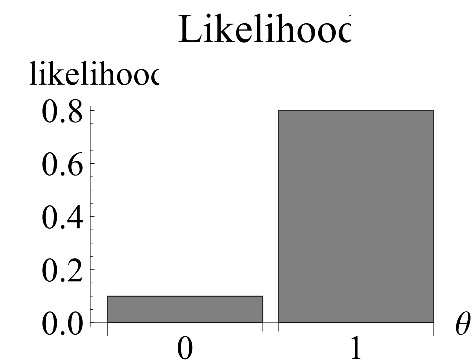
$P(\theta) = 0 \theta > 0.45 \theta \approx 0.018$

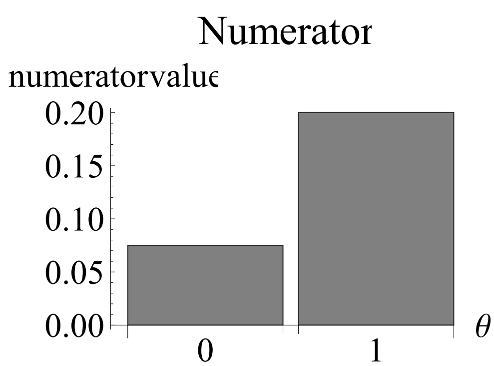
$\theta = 0.45$

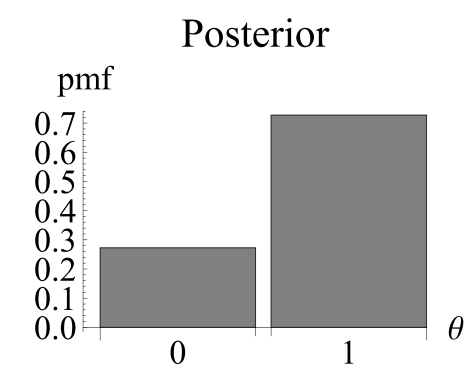
$P(data, \theta)$

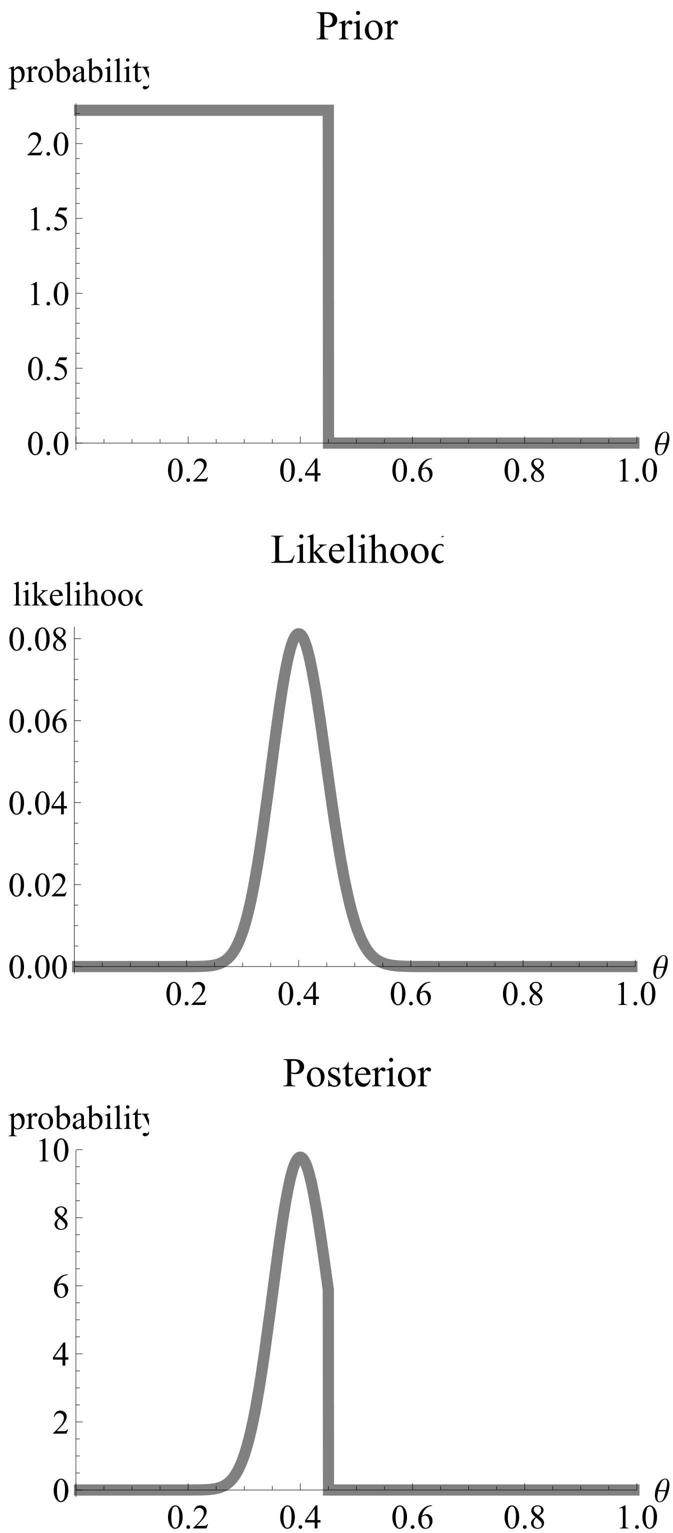
$\frac{20}{9} \theta \leq 0.45$


Prior
pmf
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
θ


likelihood
0.8
0.6
0.4
0.2
0.0
0
1
θ


0.20
0.15
0.10
0.05
0.00
0
1
θ


Posterior
pmf
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
θ


Prior
probability
2.0
1.5
1.0
0.5
0.0
θ
0.2
0.4
0.6
0.8
1.0
Likelihood
likelihood
0.08
0.06
0.04
0.02
0.00
θ
0.2
0.4
0.6
0.8
1.0
Posterior
probability
10
8
6
4
2
0
θ
0.2
0.4
0.6
0.8
1.0

$$
\begin{aligned}
p(data) &= \int\limits_{All\ \theta} p(data|\theta) \times p(\theta)\mathrm{d}\theta \\
&= \int\limits_{All\ \theta} p(data, \theta)\mathrm{d}\theta
\end{aligned}
$$

$$
p(data|\theta) = \frac{p(data, \theta)}{p(\theta)}
$$

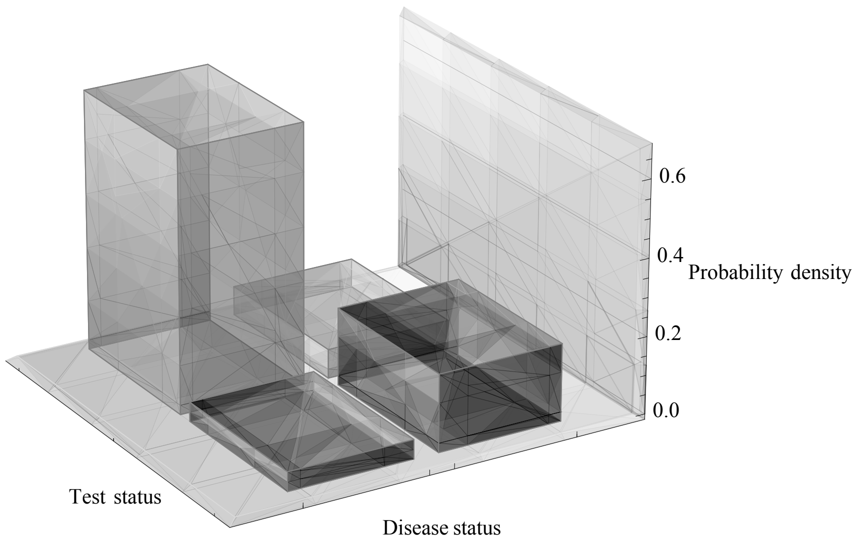
θ

	$\times \quad \times$
	$p(data)$

$p(data)p(data)$

$$
p(data) = E_\theta\left[p(data|\theta)\right] = \int\limits_{All\ \theta} p(data|\theta) \times p(\theta)\mathrm{d}\theta
$$

$\frac{p(data|model_1)}{p(data|model_2)}p(model|data)p(data|model)$



$$p(model|data) = \frac{p(data|model) \times p(model)}{p(data)}$$

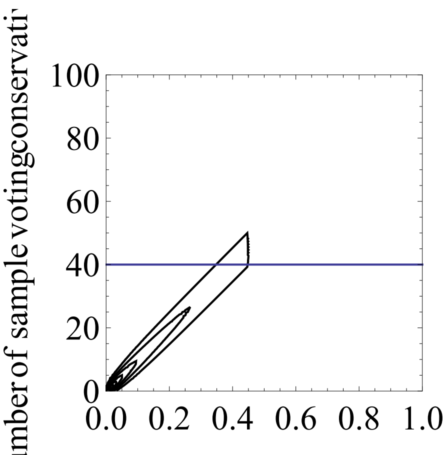
$p(model)$

$$\frac{p(model_1|data)}{p(model_2|data)} = \frac{p(data|model_1)}{p(data|model_2)} \times \frac{p(model_1)}{p(model_2)}$$

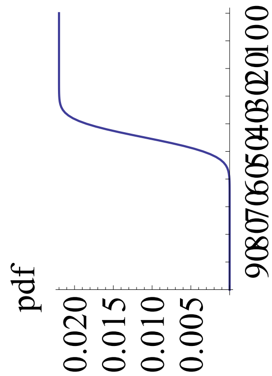
$p(model_1) = p(model_2)$

$$Bayes\ factor(model_1, model_2) = \frac{p(data|model_1)}{p(data|model_2)}$$

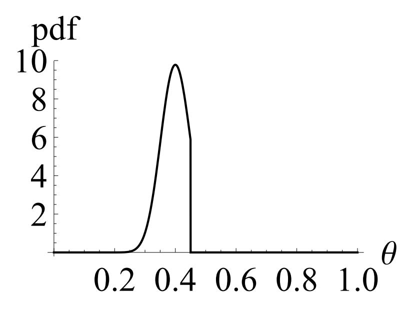
$p(data, \theta)$



mber of sample voting conservati
100
80
60
40
20
0
0.0 0.2 0.4 0.6 0.8 1.0



pdf
0.020 0.015 0.010 0.005



pdf
10
8
6
4
2
0.2 0.4 0.6 0.8 1.0
θ

$p(data)\theta$

$$p(data) = \sum_{All\ \theta_1} \sum_{All\ \theta_2} p(data, \theta_1, \theta_2)$$

$$p(data) = \int_{All\ \theta_1} \int_{All\ \theta_2} p(data, \theta_1, \theta_2) \mathrm{d}\theta_1 \mathrm{d}\theta_2$$

$p(data)$

$D \in \{0, 1\} A \in \{0, 1\} p(D = 1, A = 1) = 0.6$

$$\begin{aligned} p(A = 1|D = 1) &= \frac{p(A = 1, D = 1)}{p(D = 1)} \\ &= \frac{0.2}{0.35} \\ &= \frac{4}{7} \approx 0.57 \end{aligned}$$

$p(A = 1) = 0.25$

$X \in \{0, 1\} (D = 1, A = 1)$

$X = 1 \boldsymbol{\theta} = (D, A)$

$$\begin{aligned} p(\boldsymbol{\theta}|X = 1) &= \frac{p(X = 1|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})}{p(X = 1)} \\ &= \frac{p(X = 1|A, D) \times p(A, D)}{p(X = 1)} \end{aligned}$$

$\boldsymbol{\theta} = (D, A)$

$\boldsymbol{IQ} = \{100, 50, 150\} IQ$

	$p(D)$
$p(A)$	
	$p(D\|X=1)$
$p(A\|X=1)$	

$$p(IQ_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(IQ_i-\mu)^2}{2\sigma^2}\right)$$

$IQ \in [0, 300] \mu\sigma$

$$p(\mu,\sigma^2) = p(\mu) \times p(\sigma^2)$$

$\sigma^2 \geq 0 \sigma^2 \sim Unif(0,\infty) log(\sigma^2) \sigma log(\sigma^2) \sigma$

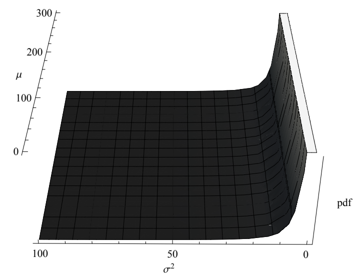
$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

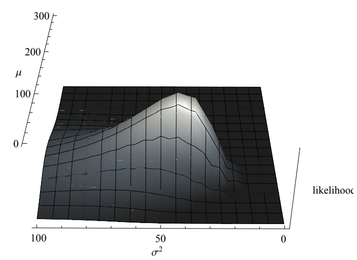
$(\mu,\sigma^2) \int\limits_0^\infty \frac{1}{\sigma^2} \mathrm{d}\sigma^2 \to \infty$

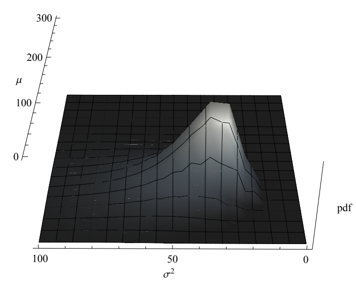
$(\mu,\sigma^2) = \boldsymbol{\theta}$

$$\begin{aligned}
p(\boldsymbol{\theta}|\boldsymbol{IQ}) &= \frac{p(\boldsymbol{IQ}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})}{p(\boldsymbol{IQ})} \\
&= \frac{p(\boldsymbol{IQ}|\mu,\sigma^2) \times p(\mu,\sigma^2)}{p(\boldsymbol{IQ})} \\
&= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{\sum\limits_{i=1}^{3}(IQ_i-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma^2}}{\int\limits_0^{300}\int\limits_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{\sum\limits_{i=1}^{3}(IQ_i-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma^2} \mathrm{d}\sigma^2 \mathrm{d}\mu} \\
&\propto \sigma^{-3} exp\left(-\frac{\sum\limits_{i=1}^{3}(IQ_i-\mu)^2}{2\sigma^2}\right)
\end{aligned}$$

$\boldsymbol{\theta} = (\mu,\sigma^2)$


300
200
μ
100
0
pdf
100
50
0
σ^2


300
200
μ
100
0
likelihooc
100
50
0
σ^2


300
200
μ
100
0
pdf
100
50
0
σ^2

$$p(\mu|\boldsymbol{IQ}) = \int_0^{\infty} p(\mu, \sigma^2|\boldsymbol{IQ})\mathrm{d}\sigma^2$$

$$p(\sigma^2|\boldsymbol{IQ}) = \int_0^{300} p(\mu, \sigma^2|\boldsymbol{IQ})\mathrm{d}\mu$$

-
-

$$\begin{aligned} p(\theta|data) &= \frac{p(data|\theta) \times p(\theta)}{p(data)} \\ &\propto p(data|\theta) \times p(\theta) \end{aligned}$$

$p(data)\theta$

θ