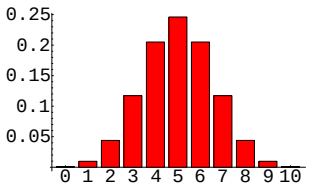
A simple example: A coin

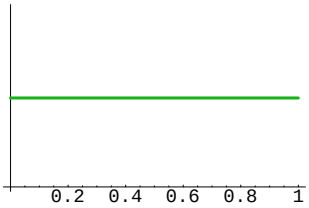
$h = p(H)$ probability to find "head" H
(h represents the hypothesis)
k denotes the number of "heads" found in n trials
(k represents the data D).

Likelihood:
$p(D|h) = p(k|n,h) = \binom{n}{k} h^k (1-h)^{n-k}$



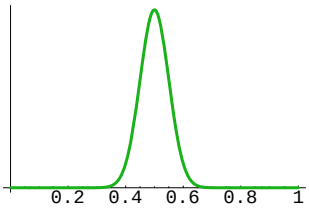
Some possible priors:
uniform prior (every value of h in [0,1] is equally likely)

$p(h) = \Theta(1-h)\Theta(h)$



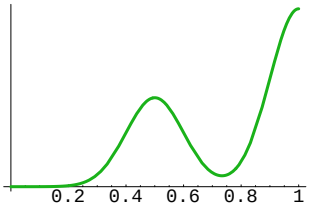
"expert" prior (h near 0.5)

$p(h) \approx \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(h-0.5)^2}{2\sigma^2}} \Theta(1-h)\Theta(h)$



"Las Vegas" prior (h near 0.5 OR near 1)

$$p(h) \approx \frac{1}{2\sqrt{2\pi}\sigma} \left(e^{-\frac{(h-0.5)^2}{2\sigma^2}} + 2e^{-\frac{(h-1)^2}{2\sigma^2}} \right) \Theta(1-h)\Theta(h)$$



A simple example: Maximum A Posteriori Approximation

Choose as "best" estimate for h the maximum of $p(h|D)$, i.e., the minimum of the "(posterior) energy" $E(h|D) = -\ln p(h|D)$. To find the MAP solution h^* solve
$0 = \frac{d}{dh} p(h|D)$ or $0 = -\frac{d}{dh} \ln p(h|D)$

For binomial model
$-\ln p(D|h) = -k \ln h - (n-k) \ln(1-h) - \ln \binom{n}{k}$

For uniform prior

$$0 = \frac{d}{dh} \ln p(D|h) = \frac{k}{h} - \frac{n-k}{1-h}$$

$$\Rightarrow \boxed{h^* = \frac{k}{n}}$$

That means the MAP solution for uniform prior is the observed relative frequency.

For Gaussian prior $-\ln p(h) = \frac{(h-h_0)^2}{2\sigma^2} + \frac{1}{2} \ln(2\pi\sigma^2)$

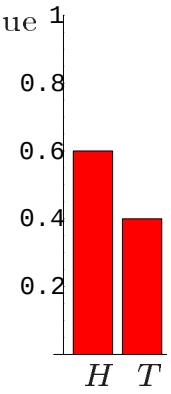
$$0 = \frac{d}{dh} \ln p(D|h) = \frac{k}{h} - \frac{n-k}{1-h} - \frac{h-h_0}{\sigma^2}$$

$\Rightarrow h^* = \frac{k}{n} - \frac{1}{n\sigma^2}(h-h_0)h(1-h)$ (nonlinear)

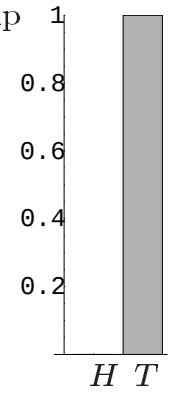
A simple example
$n = 1$

$h = p_{\text{true}}(H)$

true value (unknown)



sampling $\longrightarrow$

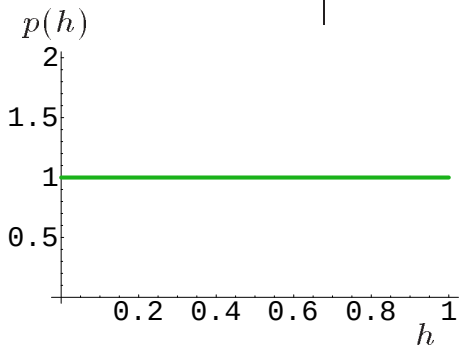


observed data
$n = 1$
$k = 0$

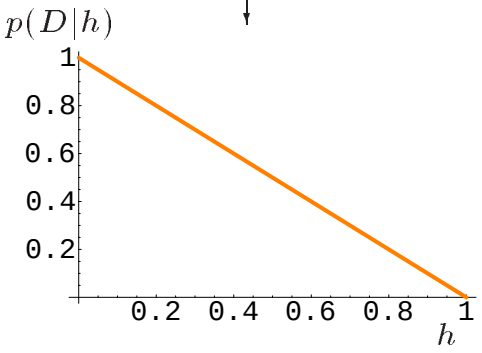
$p_{\text{emp}}(H) = \frac{k}{n}$

prior model $\uparrow$

likelihood model $\downarrow$



$\times$

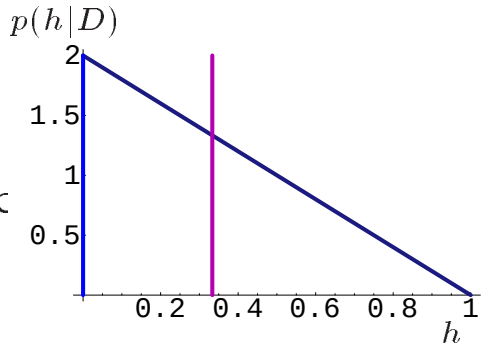


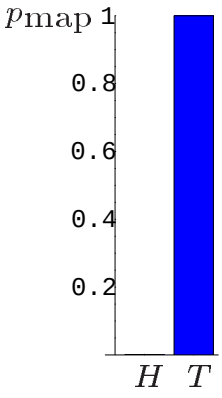
$p(h) = \text{uniform}$

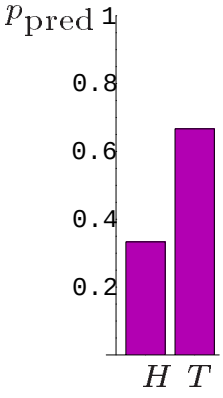
$p(k|h,n) = \binom{n}{k} h^k (1-h)^{n-k}$

prior $\times$ likelihood

$\propto$







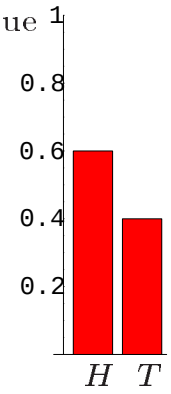
$\propto$ posterior $\Rightarrow$ MAP

mean of posterior

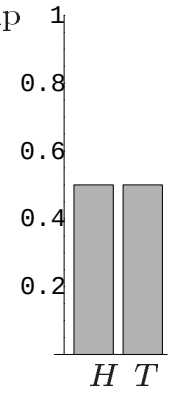
A simple example
$n = 2$

$h = p_{\text{true}}(H)$

true value (unknown)



sampling $\longrightarrow$

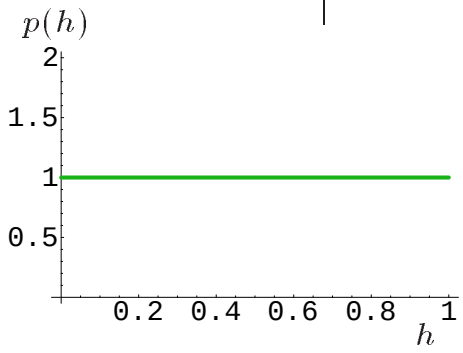


observed data
$n = 2$
$k = 1$

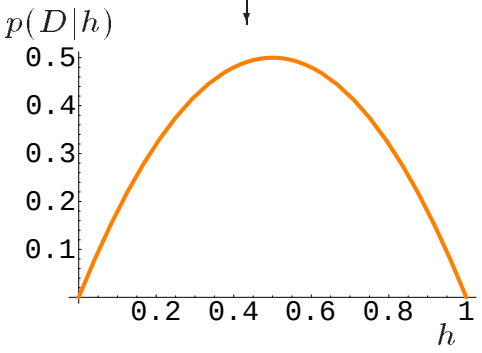
$p_{\text{emp}}(H) = \frac{k}{n}$

prior model $\uparrow$

likelihood model $\downarrow$



$\times$

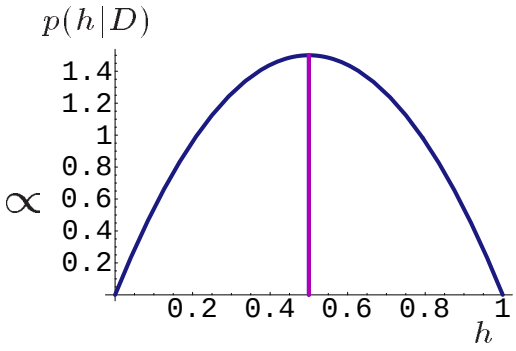


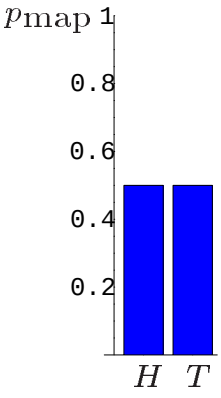
$p(h) = \text{uniform}$

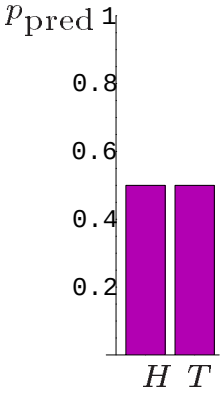
$p(k|h,n) = \binom{n}{k} h^k (1-h)^{n-k}$

prior $\times$ likelihood

$\propto$







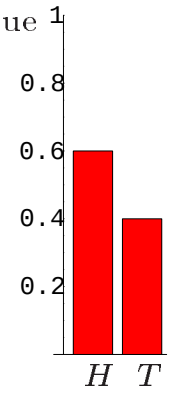
$\propto$ posterior $\Rightarrow$ MAP

mean of posterior

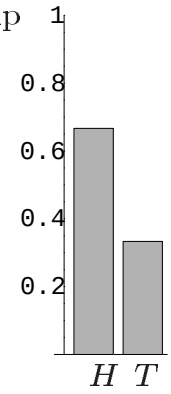
A simple example
$n = 3$

$h = p_{\text{true}}(H)$

true value (unknown)



sampling $\longrightarrow$

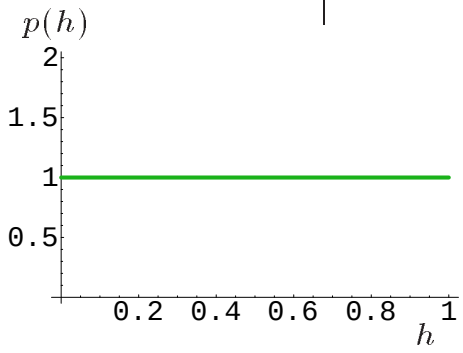


observed data
$n = 3$
$k = 2$

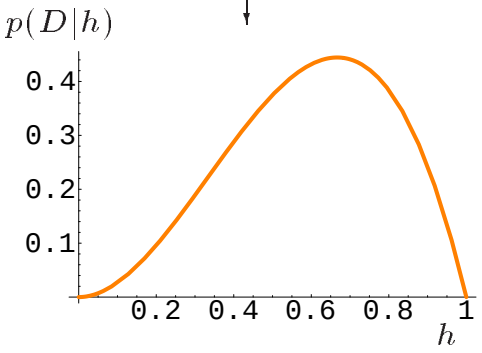
$p_{\text{emp}}(H) = \frac{k}{n}$

prior model $\uparrow$

likelihood model $\downarrow$



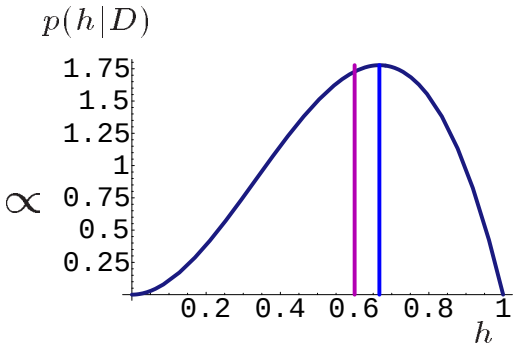
$\times$



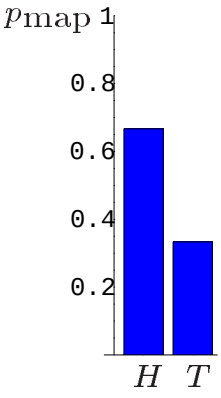
$p(h) = \text{uniform}$

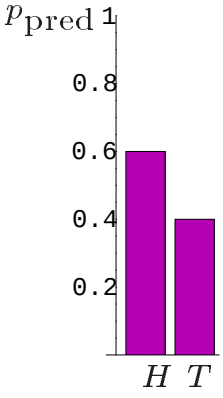
$p(k|h,n) = \binom{n}{k} h^k (1-h)^{n-k}$

prior $\times$ likelihood



$\propto$





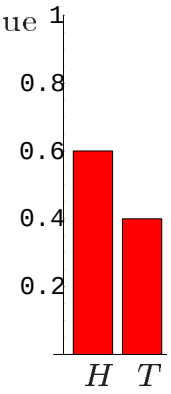
$\propto$ posterior $\Rightarrow$ MAP

mean of posterior

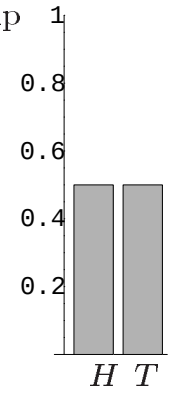
A simple example
$n = 10$

$h = p_{\text{true}}(H)$

true value (unknown)



sampling $\longrightarrow$

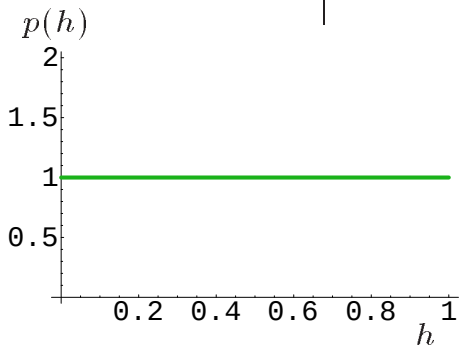


observed data
$n = 10$
$k = 5$

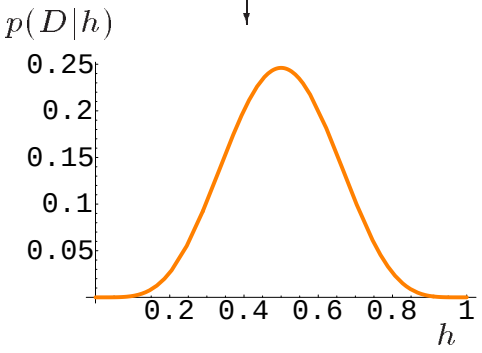
$p_{\text{emp}}(H) = \frac{k}{n}$

prior model $\uparrow$

likelihood model $\downarrow$



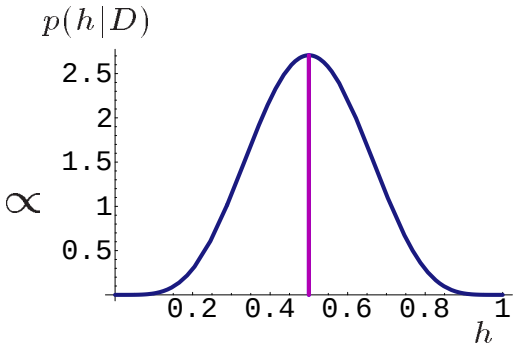
$\times$



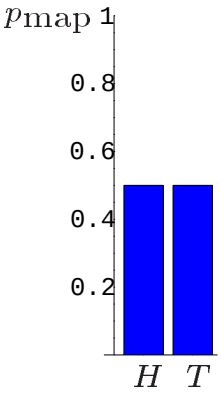
$p(h) = \text{uniform}$

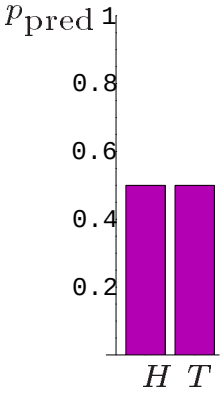
$p(k|h,n) = \binom{n}{k} h^k (1-h)^{n-k}$

prior $\times$ likelihood



$\propto$





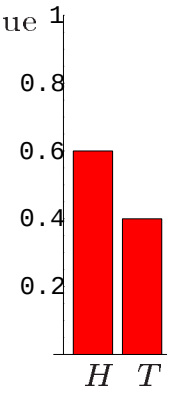
$\propto$ posterior $\Rightarrow$ MAP

mean of posterior

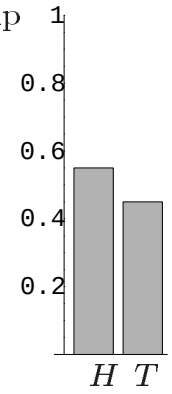
A simple example
$n = 100$

$h = p_{\text{true}}(H)$

true value (unknown)



sampling $\longrightarrow$

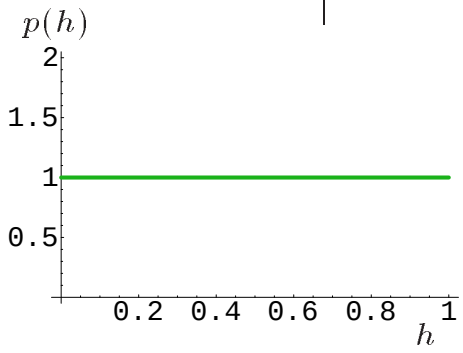


observed data
$n = 100$
$k = 55$

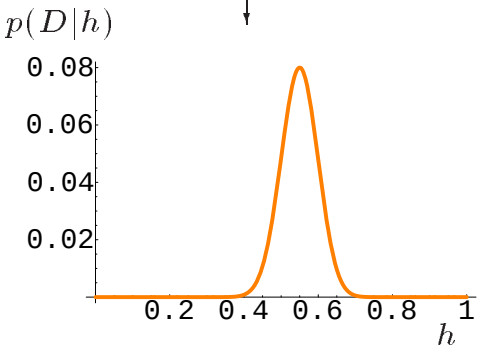
$p_{\text{emp}}(H) = \frac{k}{n}$

prior model $\uparrow$

likelihood model $\downarrow$



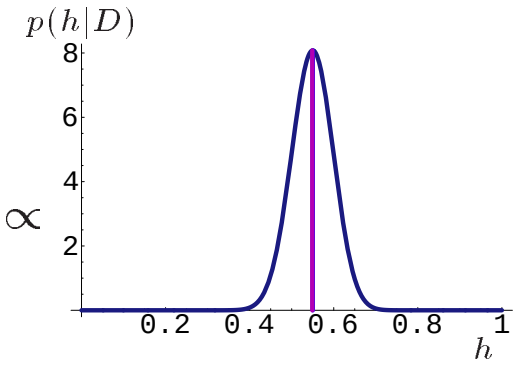
$\times$



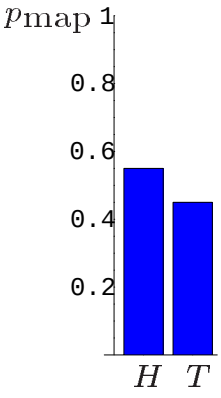
$p(h) = \text{uniform}$

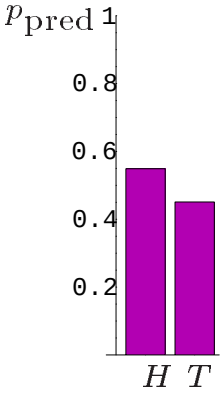
$p(k|h,n) = \binom{n}{k} h^k (1-h)^{n-k}$

prior $\times$ likelihood



$\propto$





$\propto$ posterior $\Rightarrow$ MAP

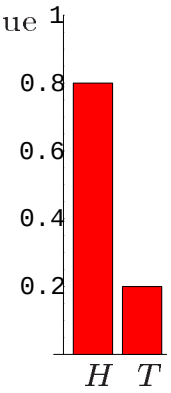
mean of posterior

A simple example with "expert" prior

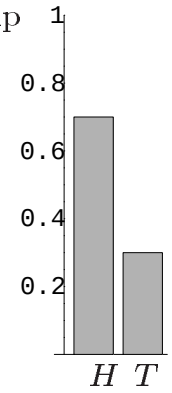
$n = 10$

$h = p_{\text{true}}(H)$

true value (unknown)

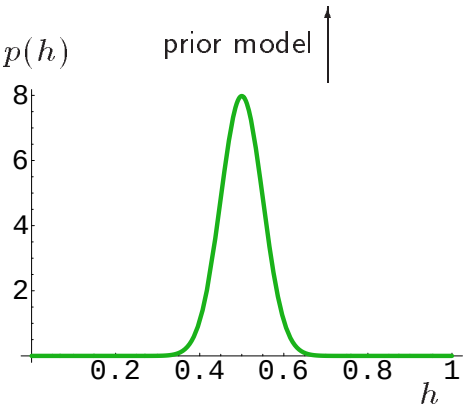


sampling $\longrightarrow$

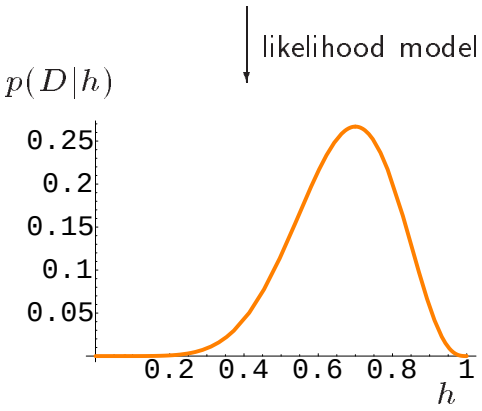


observed data
$n = 10$
$k = 7$

$p_{\text{emp}}(H) = \frac{k}{n}$



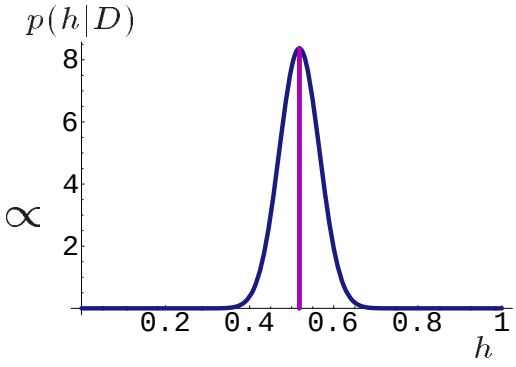
$\times$



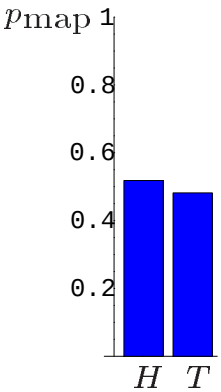
$p(h) =$ Gaussian

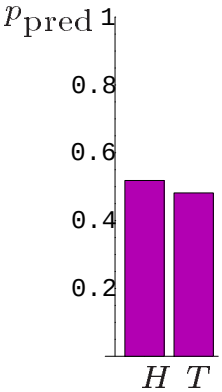
$p(k|h,n) = \binom{n}{k} h^k (1-h)^{n-k}$

prior $\times$ likelihood



$\propto$





$\propto$ posterior $\Rightarrow$ MAP

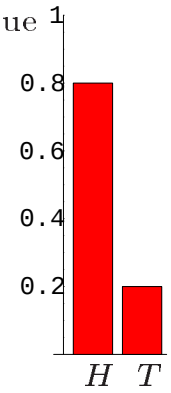
mean of posterior

A simple example with "expert" prior

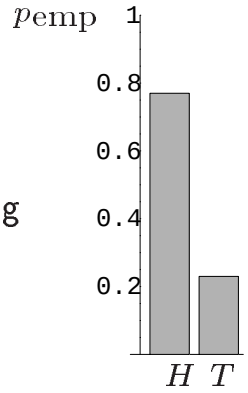
$n = 100$

$h = p_{\text{true}}(H)$

true value (unknown)

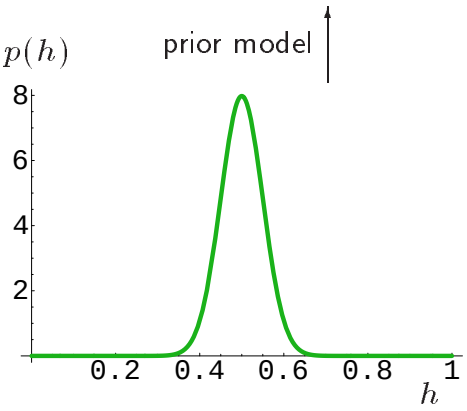


sampling $\longrightarrow$

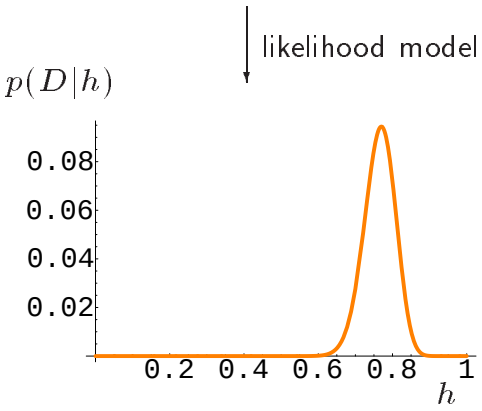


observed data
$n = 100$
$k = 77$

$p_{\text{emp}}(H) = \frac{k}{n}$



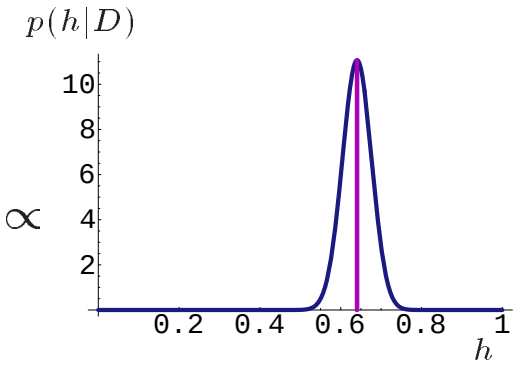
$\times$



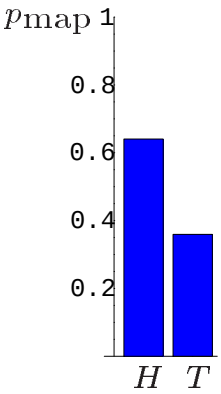
$p(h) = \text{Gaussian}$

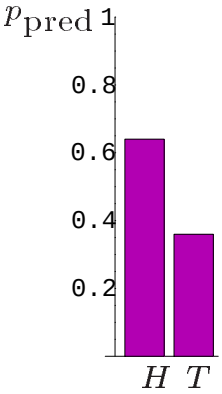
$p(k|h,n) = \binom{n}{k} h^k (1-h)^{n-k}$

prior $\times$ likelihood



$\propto$





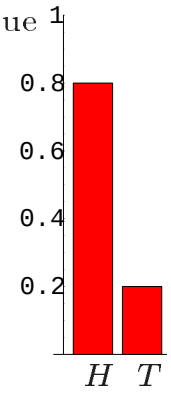
$\propto$ posterior $\Rightarrow$ MAP

mean of posterior

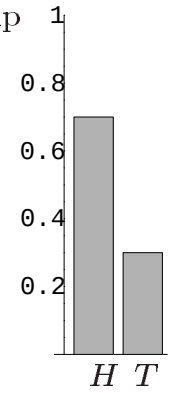
A simple example:
A fair coin in Las Vegas?
$n = 10$

$h = p_{\text{true}}(H)$

true value (unknown)

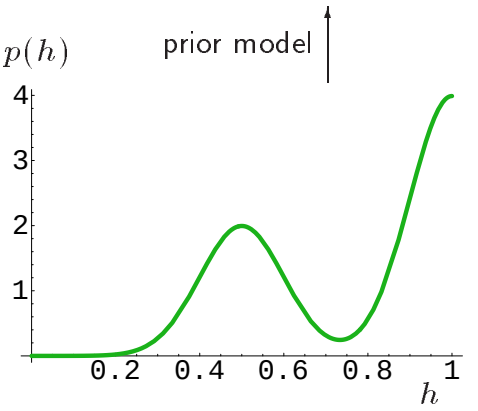


sampling $\longrightarrow$

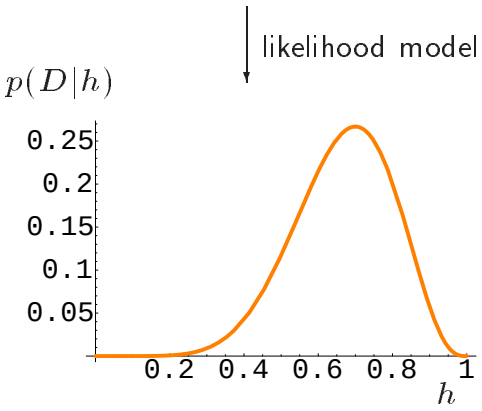


observed data
$n = 10$
$k = 7$

$p_{\text{emp}}(H) = \frac{k}{n}$



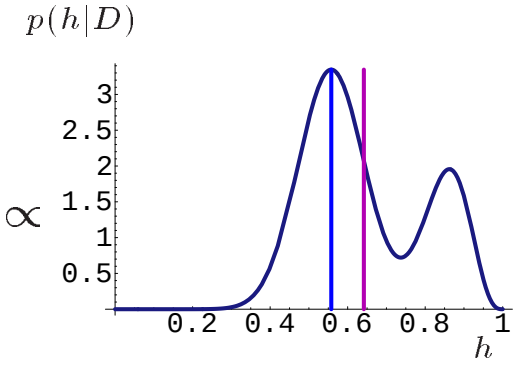
$\times$



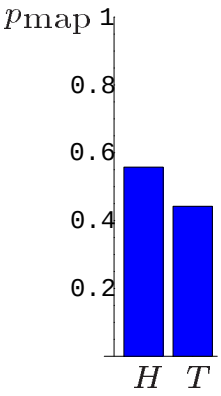
$p(h) = \text{Gaussian}$

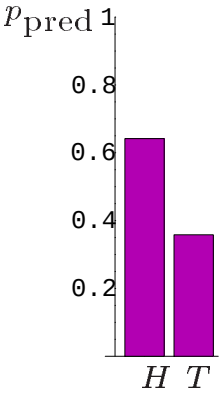
$p(k|h,n) = \binom{n}{k} h^k (1-h)^{n-k}$

prior $\times$ likelihood



$\propto$





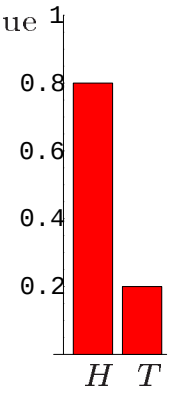
$\propto$ posterior $\Rightarrow$ MAP

mean of posterior

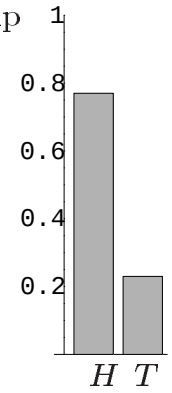
A simple example:
A fair coin in Las Vegas?
$n = 100$

$h = p_{\text{true}}(H)$

true value (unknown)

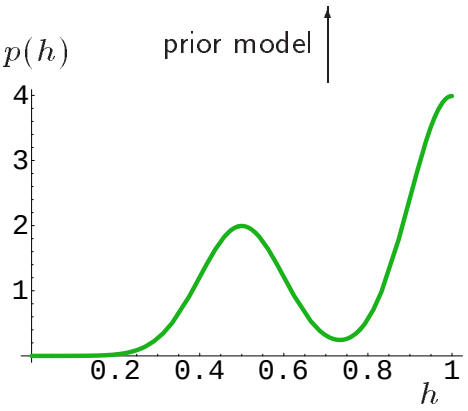


sampling $\longrightarrow$

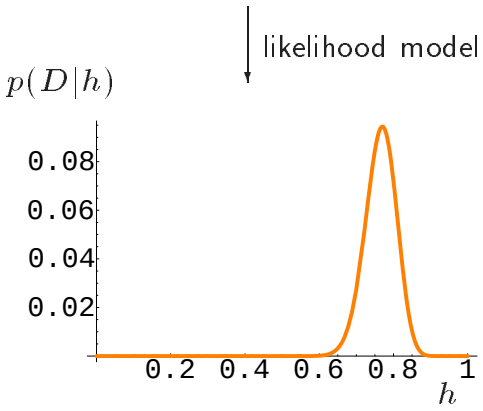


$p_{\text{emp}}(H) = \frac{k}{n}$

observed data
$n = 100$
$k = 77$



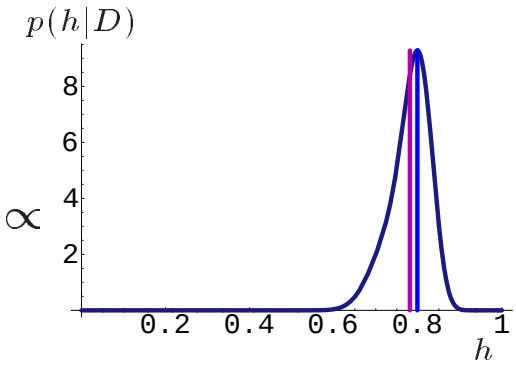
$\times$



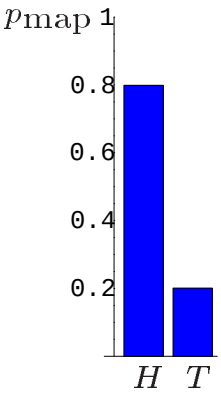
$p(h) =$ Gaussian

$p(k|h,n) = \binom{n}{k} h^k (1-h)^{n-k}$

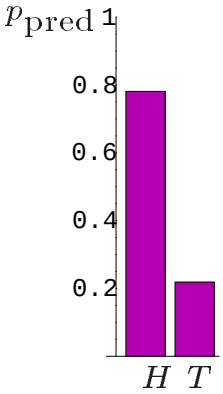
prior $\times$ likelihood



$\propto$ posterior $\Rightarrow$



MAP

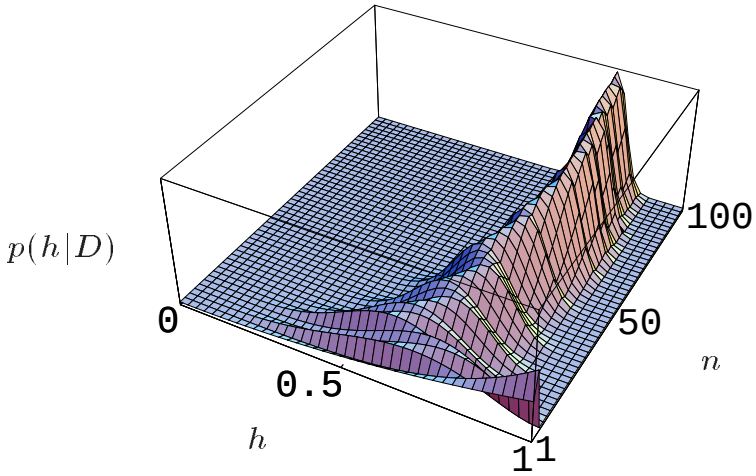


mean of posterior

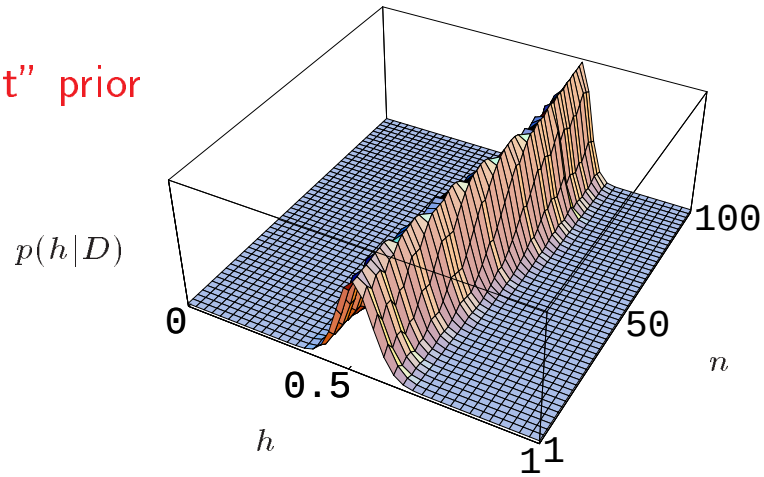
A simple example: Three posteriors

(Sampling $n = 100$ data points for true $h = 0.8$)

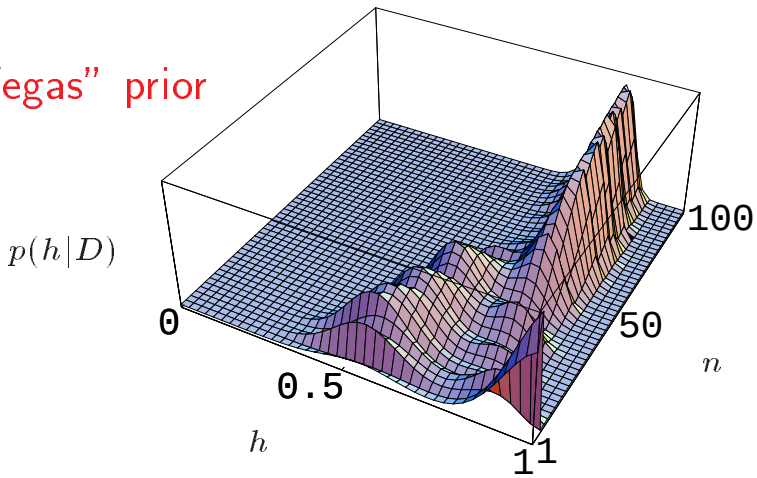
uniform prior



"expert" prior



"Las Vegas" prior



A fair coin?

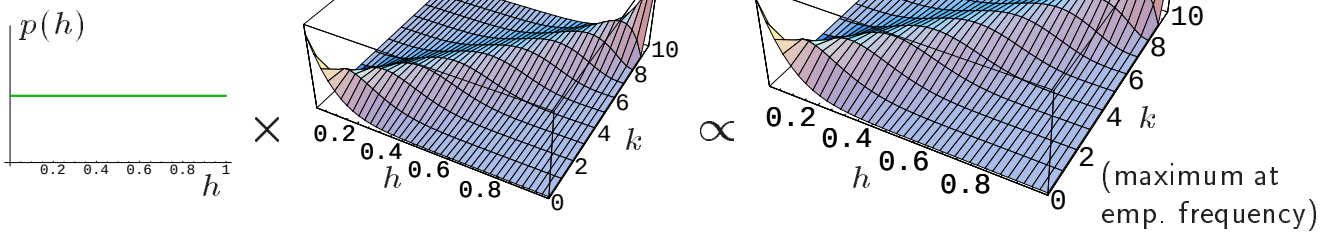
$h = p(K)$ probability to find "head"
k denotes the number of "heads" found in $n = 10$ trials (corresponds to data D).

Likelihood: $p(D|h) = p(k|h) = \binom{n}{k} h^k (1-h)^{(n-k)}$

prior $p(h)$ $\times$ likelihood $p(D|h)$ $\propto$ posterior $p(h|D)$

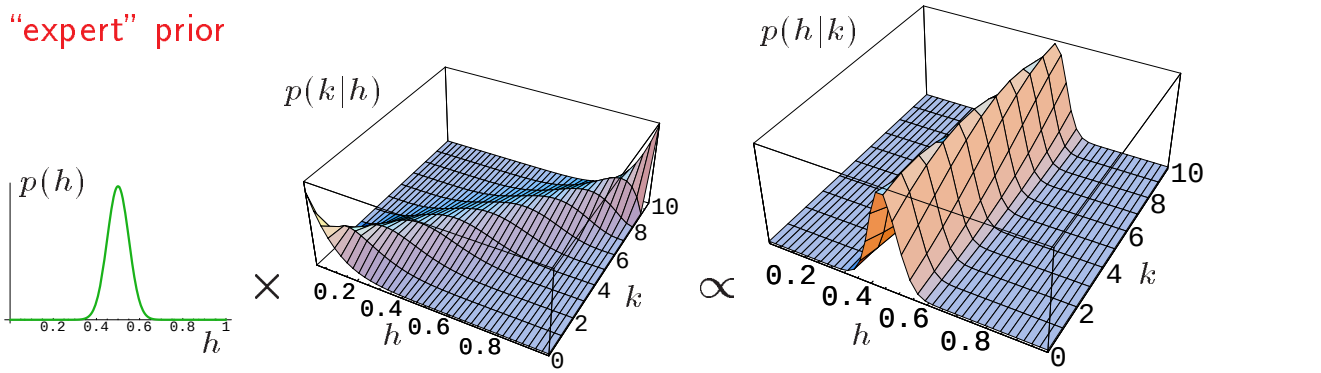
"extraterrestrian" prior

(uniform)

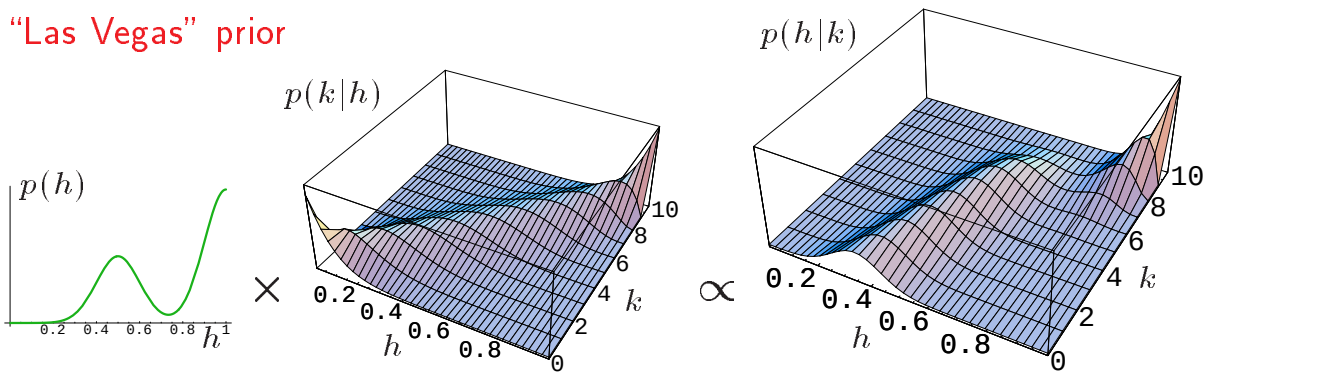


(maximum at emp. frequency)

"expert" prior



"Las Vegas" prior



A simple example: Bayesian error bars

Maximum a posteriori approximation (saddle point approximation, Laplace approximation):

$$e^{-f(x)} \approx e^{-f(x_0)-\frac{(x-x^*)^2}{2}\frac{d^2}{dx^2}f\big|_{x=x^*}} = e^{-f(x_0)}e^{-\frac{(x-x^*)^2}{2\sigma_{\rm map}^2}}$$

where $\sigma_{\rm map}$ gives the MAP error bars.

For binomial model with likelihood energy

$$-\ln p(D|h) = -k\ln h - (n-k)\ln(1-h) - \ln\binom{n}{k}$$

follows

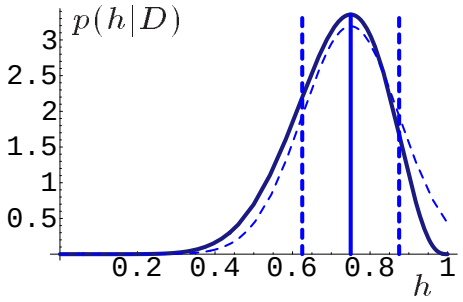
$$0 = -\frac{d^2}{dh^2}\ln p(D|h) = \left(-\frac{k}{h^2} + \frac{n-k}{(1-h)^2}\right),$$

i.e., at h^*,

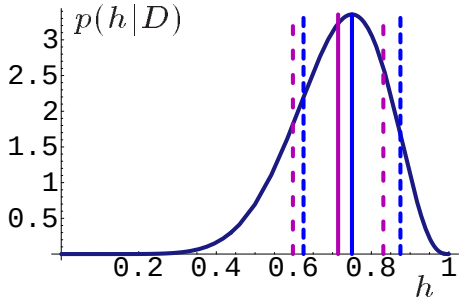
$$0 = -\frac{n^2}{k} + \frac{n-k}{(1-\frac{k}{n})^2} = \frac{n}{h^*(1-h^*)}$$

$$\Rightarrow \sigma_{\rm map} = \sqrt{\frac{h^*(1-h^*)}{n}}$$

MAP with error bars

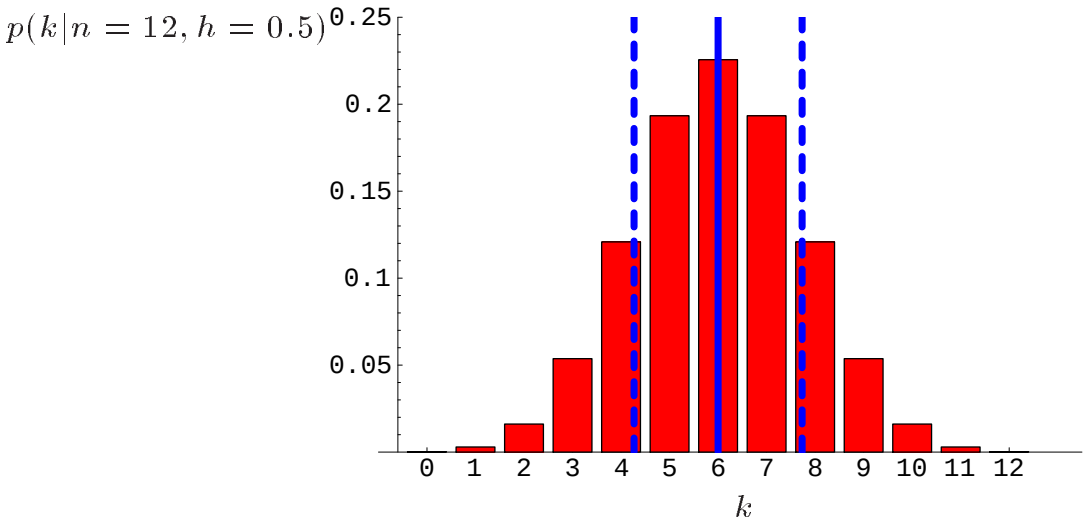


MAP vs. exact error bars

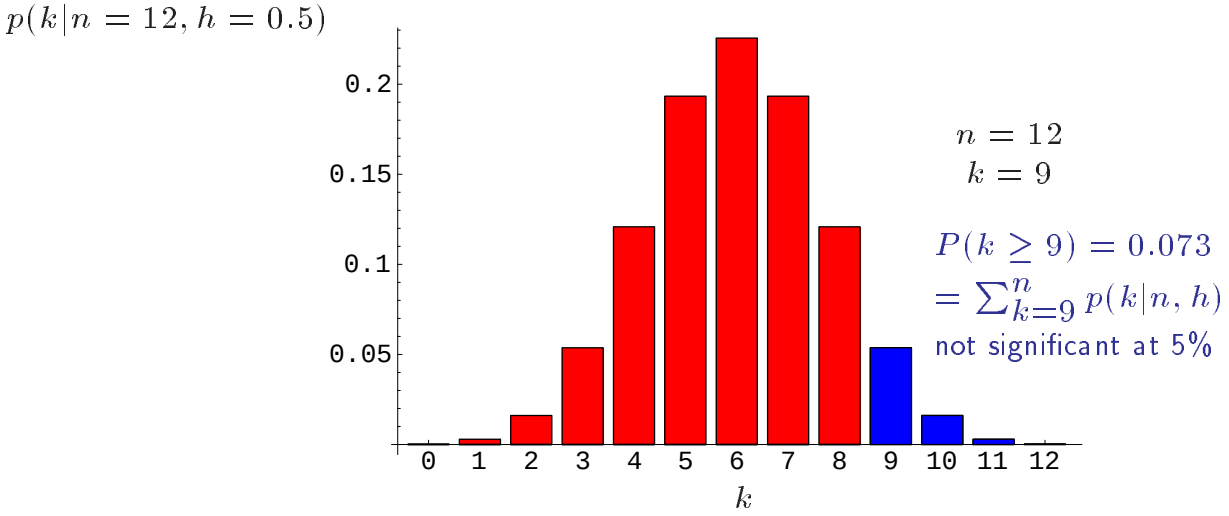


A simple example: Frequentist error bars and P-values

Choose a null hypothesis, say $h_0 = 0.5$. Conditioned on h_0 the sampling distribution (or likelihood) is a binomial distribution $p(k|n, h_0) = \binom{n}{k} h^k (1-h)^{n-k}$ with mean $< k >= nh$ and variance $< k - < k >>= nh(1-h)$

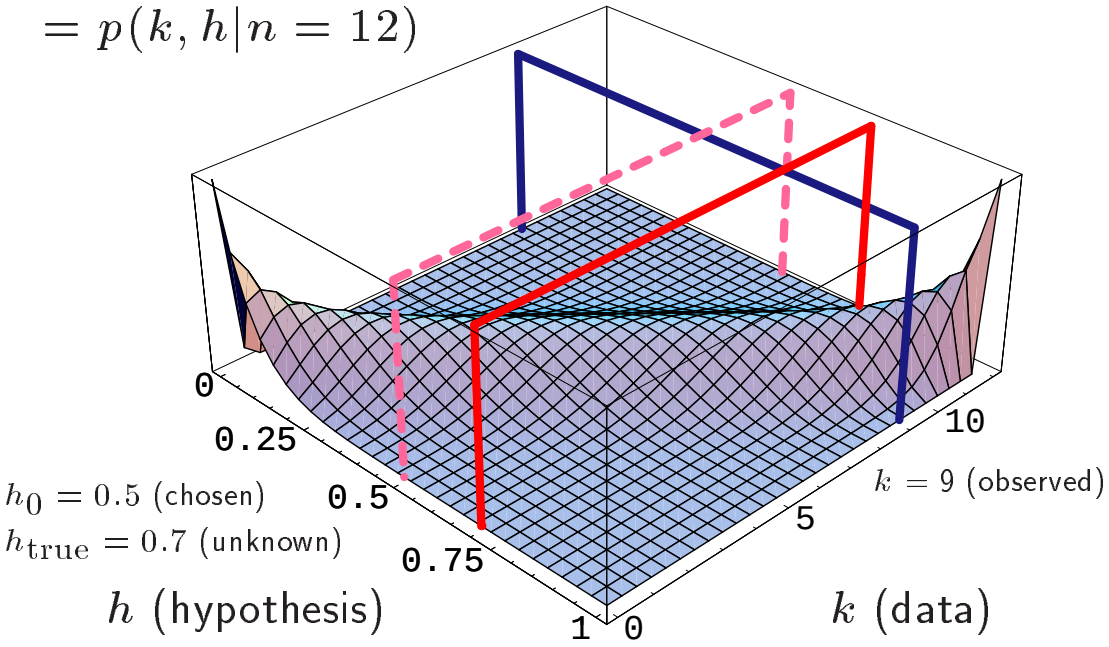


The (one–sided) P-value $P(k \geq x) = \sum_{k=x}^{n} p(x|n, h)$ gives the approximate relative frequency of getting the same or a more extreme result if you repeat the experiment very often and the null hypothesis is true



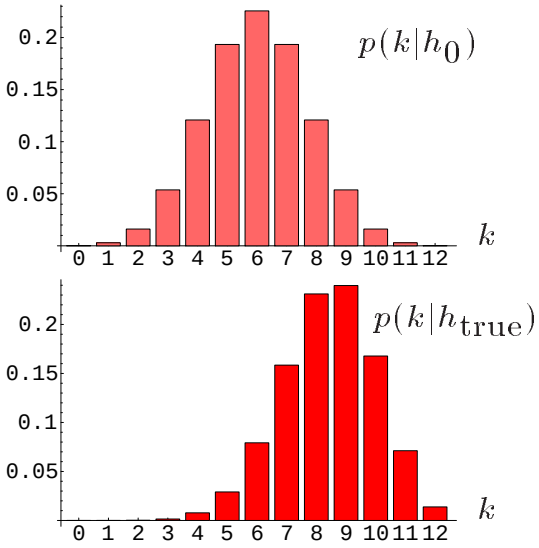
Bayesian vs. Frequentist statistics

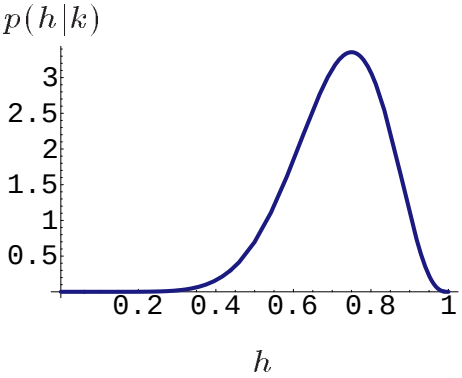
joint probability $p(D, h)$
$= p(k, h|n = 12)$



Frequentists condition on h_0 and work with $p(D|h_0)$

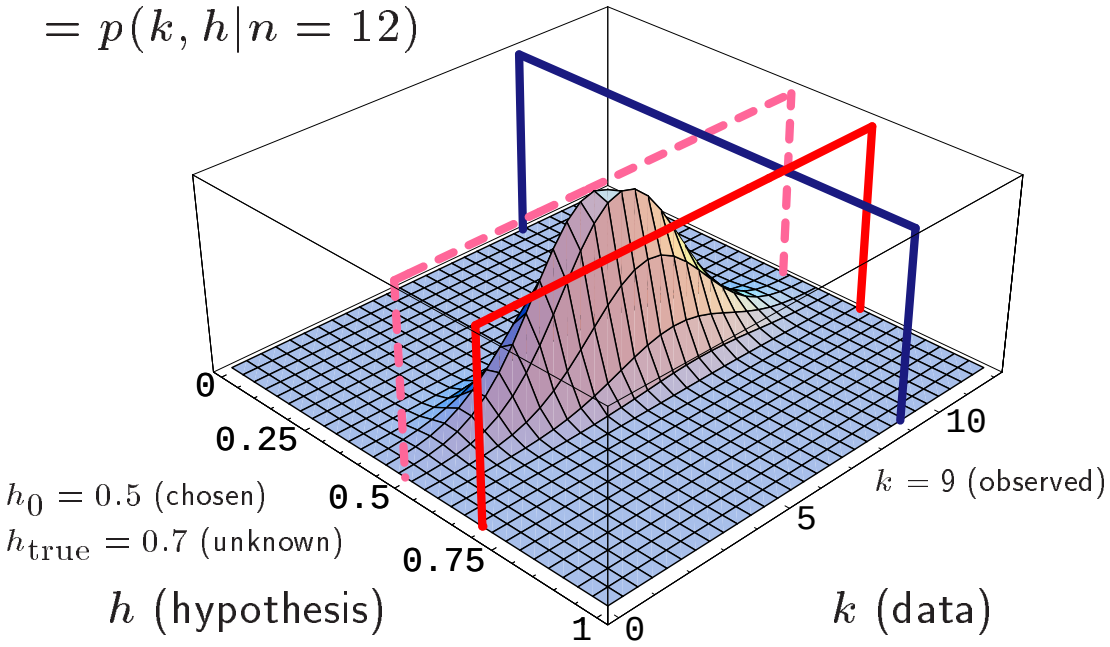
Bayesians condition on data and work with the posterior $p(h|D)$





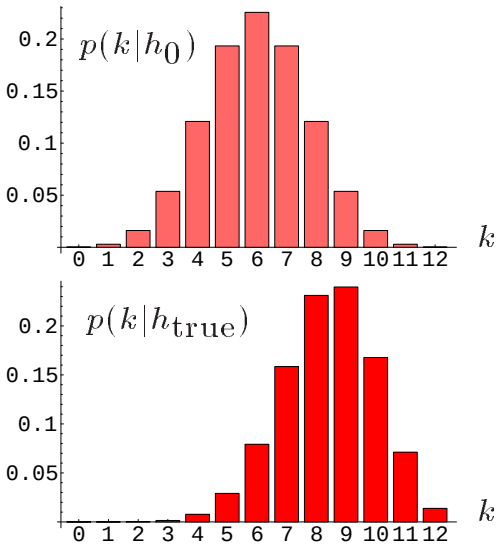
Bayesian vs. Frequentist statistics (Gaussian prior centered at $h = 0.5$)

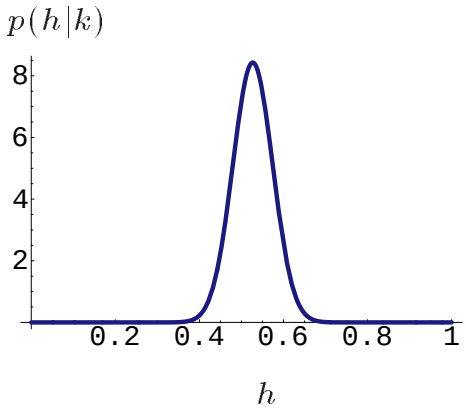
joint probability $p(D, h)$
$= p(k, h|n = 12)$



Frequentists condition on h_0 and work with $p(D|h_0)$

Bayesians condition on data and work with the posterior $p(h|D)$





Stopping rules

$n = 12$

$k = 9$

$h = 0.5$

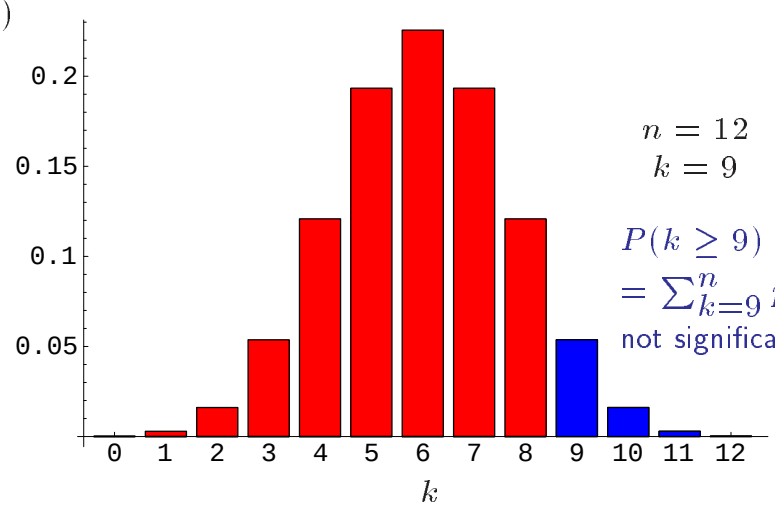
Stop at $n = 12$: Binomial distribution

$$p(k|n,h) = \binom{n}{k} h^k (1-h)^{n-k}$$

$p(k|n = 12, h = 0.5)$

(0111)
(1011)
(1101)
(1110)

$\frac{n!}{k!(n-k)!}$



$n = 12$
$k = 9$

$P(k \geq 9) = 0.073$
$= \sum_{k=9}^{n} p(k|n,h)$
not significant at 5%

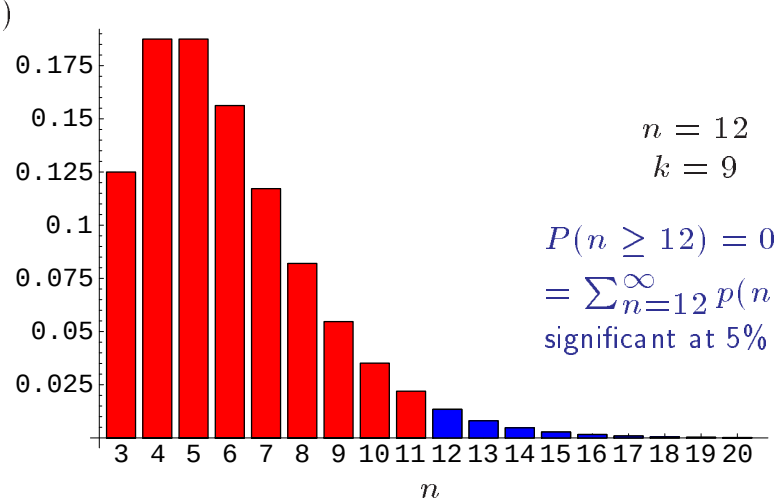
Stop at $n - k = 3$: Negative Binomial distribution

$$p(n|k,h) = \binom{n-1}{n-k-1} h^k (1-h)^{n-k}$$

$p(n|k = 9, h = 0.5)$

(000)
(1000)
(0100)
(0010)
(11000)
$\cdots$

$\frac{(n-1)!}{(n-k-1)!k!}$



$n = 12$
$k = 9$

$P(n \geq 12) = 0.033$
$= \sum_{n=12}^{\infty} p(n|k,h)$
significant at 5%