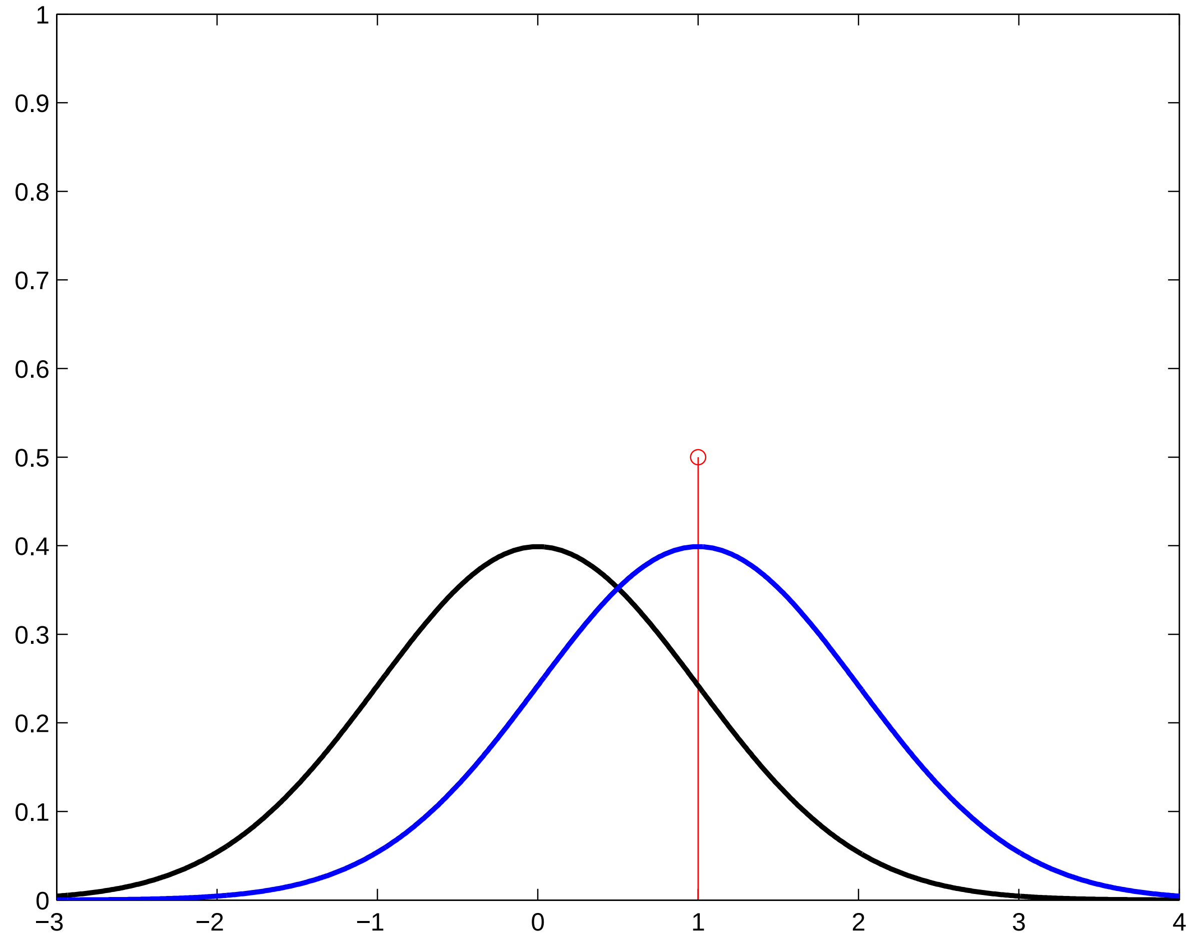
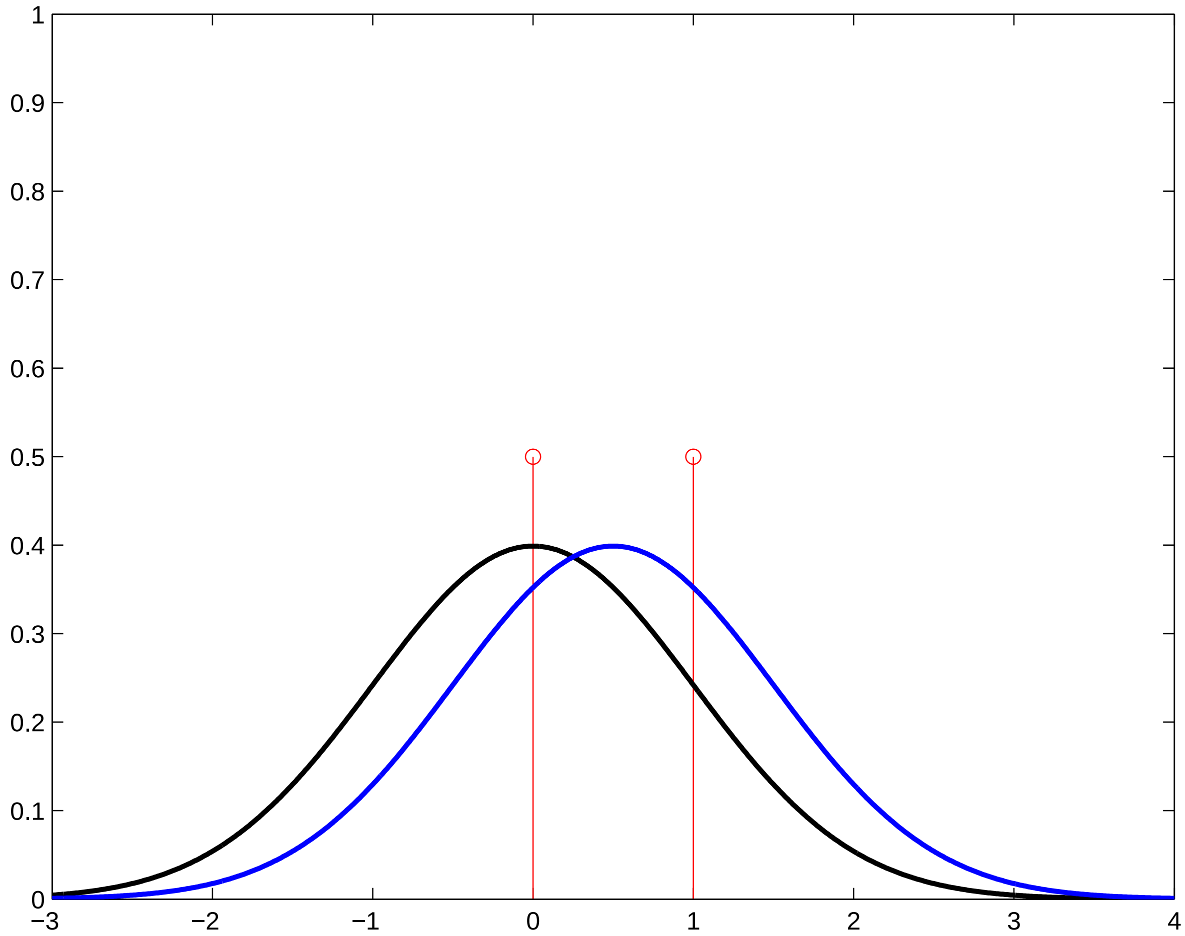
Maximum Likelihood Estimation

Find the distribution most likely to have generated the data



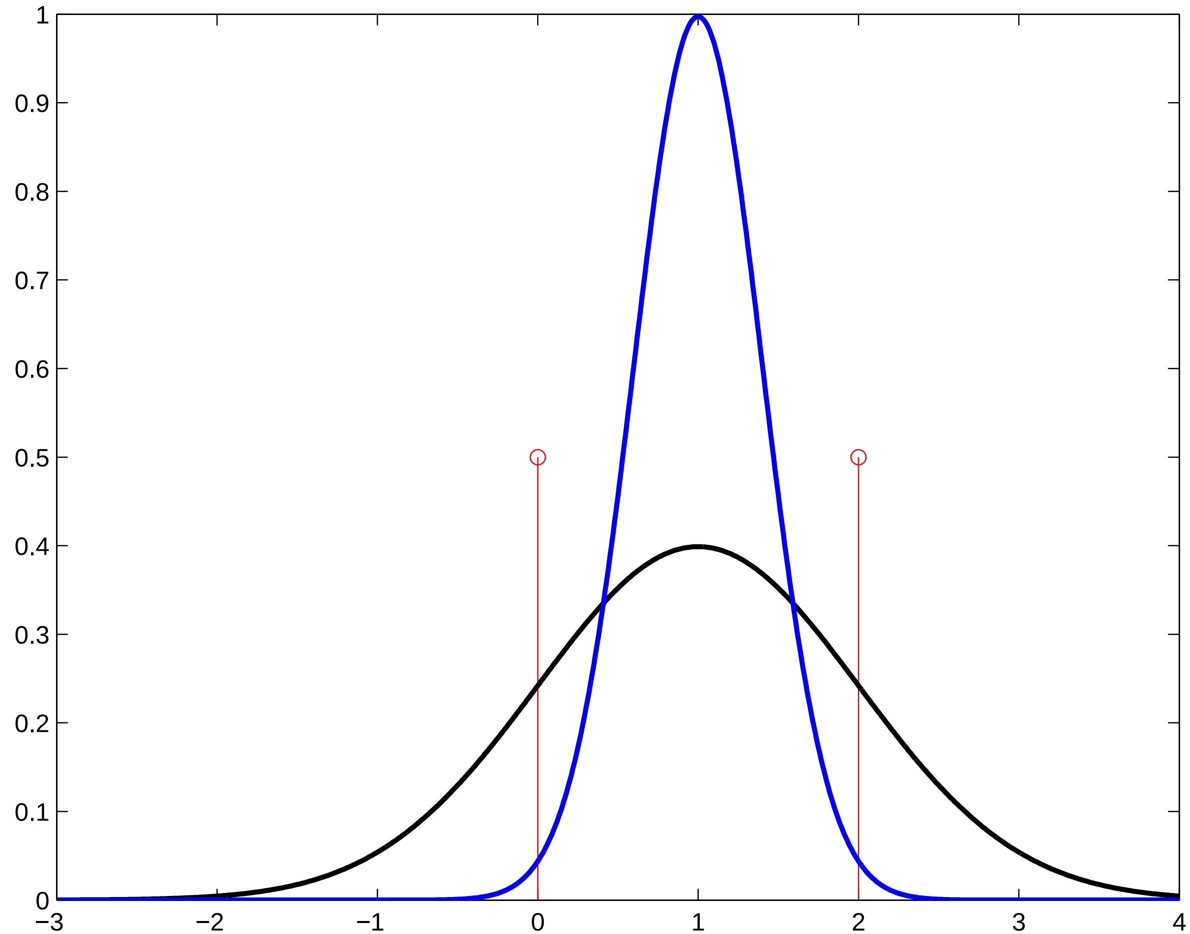
Maximum Likelihood Estimation

Find the distribution most likely to have generated the data



Maximum Likelihood Estimation

Find the distribution most likely to have generated the data



Maximum Likelihood Principle

- Suppose we have a dataset $x_1, x_2, \ldots, x_n$
- Suppose we model each datum as Gaussian with density function $f(x; \mu, \sigma)$ and independent
- For $X_1 = x_1$ and $X_2 = x_2$, the likelihood is

$$L(X_1 = x_1, X_2 = x_2; \mu, \sigma) = f(x_1; \mu, \sigma) f(\mu, \sigma)$$

- The likelihood of the dataset is

$$\begin{aligned} L(x_1, \ldots, x_n; \theta) &= f(x_1; \theta) \times f(x_2; \theta) \times \cdots \times f(x_n; \theta) \\ &= \prod_{i=1}^{n} f(x_i; \theta) \end{aligned}$$

- Choose $\theta = \mu, \sigma$ to maximize likelihood

A little notation

Normally, we denote a pdf by $f(x)$
We might remind ourselves that the density depends on the parameters θ
$f(x) = f(x;\theta)$

In maximum likelihood estimation, we turn the problem around

Instead of thinking of the likelihood as a function of the data X with given parameters θ

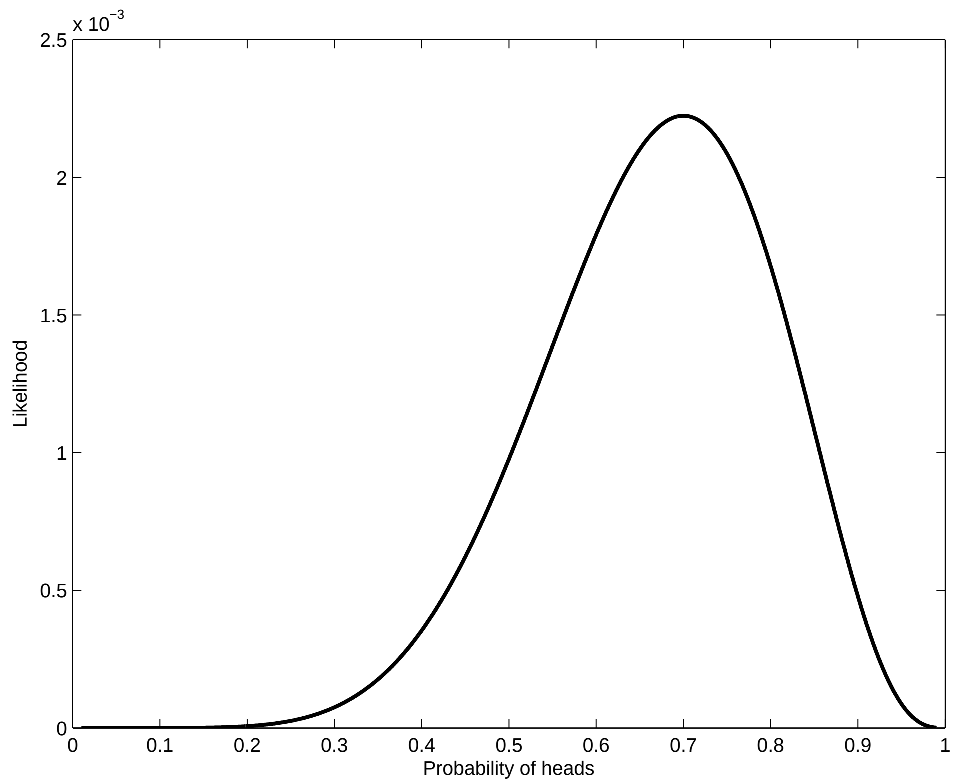
Think of the likelihood as a function of the parameters θ with given data X
$L(x;\theta) = L(\theta;x)$

Simple Binary Example

A coin was tossed 10 times, 7 heads, 3 tails

$Prob(X = \text{'heads'}) = p,\ Prob(X = \text{'tail'}) = 1 - p$

$L(p; x) = ppppppp(1-p)(1-p)(1-p) = p^7(1-p)^3$



Solution

$$
\begin{aligned}
L(p;x) &= p^7(1-p)^3 \\
\frac{dL(p;x)}{dp} &= 7p^6(1-p)^3 + p^7(3)(1-p)^2(-1) = 0 \\
7p^6(1-p)^3 &= 3p^7(1-p)^2 \\
7(1-p) &= 3p \\
p &= 7/10
\end{aligned}
$$

The Maximum Likelihood estimate of p is .7

Benefits and Costs

- No other estimate has lower asymptotic variance than ML
- Possible to exactly determine variance of estimates
  The variance is
  $(-E($second derivative of loglikelihood$))^{-1}$
- Sometimes difficult to solve without computer

Log Likilihood

Many densities have exponentials
log likelihood often more tractable

$$\begin{aligned}
L(p;x) &= p^7(1-p)^3 \\
\log L(p) &= 7\log(p) + 3\log(1-p) \\
d\log L(p)/dp &= 7/p - 3/(1-p) = 0 \\
3p &= 7(1-p) \\
p &= 7/10
\end{aligned}$$

Estimating the normal distribution

$$f(x; \mu\sigma) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

$$\log f(x; \mu, \sigma) = -\log(\sigma) - \log(\sqrt{2\pi}) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

$$\log L(\mu, \sigma; x) = \log(f(x_1)) + \ldots + \log(f(x_n))$$

$$= -n\log(\sigma) - n\log(\sqrt{2\pi}) - \frac{1}{2\sigma^2}\left((x_1 - \mu)^2 + \ldots + (x_n - \mu)^2\right)$$

$$\frac{d \log L(\mu,\sigma)}{d\mu} = \frac{1}{\sigma^2}\left((x_1 - \mu) + \ldots + (x_n - \mu)\right) = \frac{n}{\sigma^2}(\bar{x}_n - \mu)$$

Setting the derivative to zero:

$\hat{\mu} = \bar{x}_n$ The sample mean is the most likely value for μ

Estimating normal variance

$$-n\log(\sigma) - n\log(\sqrt{2\pi}) - \frac{1}{2\sigma^2}\left((x_1-\mu)^2 + \ldots + (x_n-\mu)^2\right)$$

$$\begin{aligned}\frac{d\log L(\mu,\sigma)}{d\sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left((x_1-\mu)^2 + \ldots + (x_n-\mu)^2\right)\\ &= -\frac{n}{\sigma^3}\left(\sigma^2 - \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right)\end{aligned}$$

Set the derivative to 0

$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{n}$

Substituting in the ML estimate of μ:

$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x}_n)^2}{n}$

The ML variance estimate is biased, (but not asymptotically)