Neuroinformatics Lecture (L9)

Prof. Dr. Gordon Pipa

Institute of Cognitive Science University of Osnabrück

Poisson Distribution: probability of k events in given that λ events are expected

$$p(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

With k=0,1,2, …. (discrete)   and λ >0 (parameter is continuous)

E(k) = λ
var(k) = λ

Example: p(k) spikes in an interval between 0 and T if λ spikes are expected

Y = poisspdf(X,lambda)

computes the Poisson pdf at each of the values in X using the corresponding parameters in lambda. The parameters in lambda must all be positive.

The Gaussian Distribution

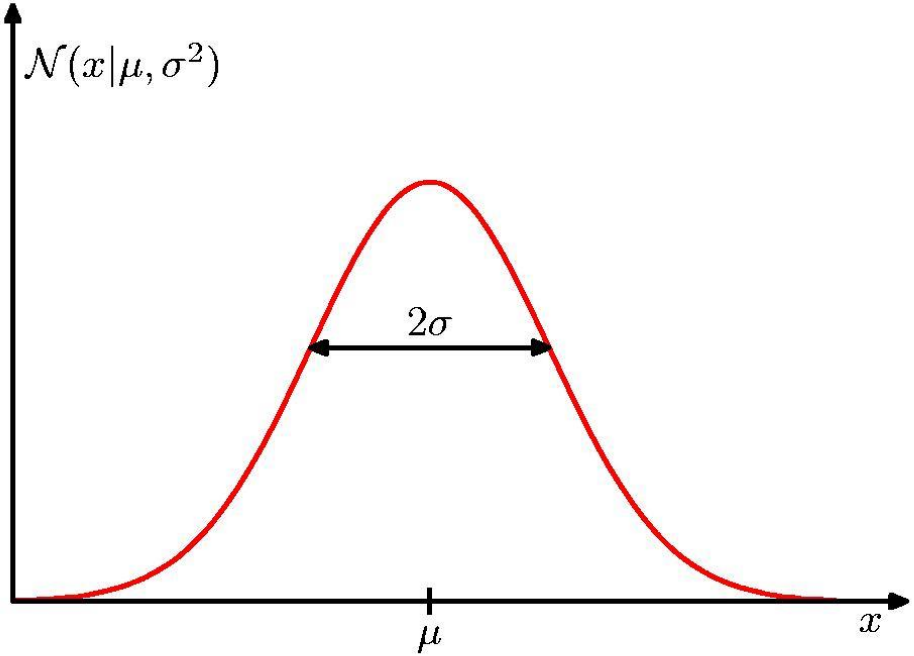
$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

With x continuous; also parameter μ, σ are continuous

With

$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1$$



Y = normpdf(X,mu,sigma)

computes the pdf at each of the values in X using the normal distribution with mean mu and standard deviation sigma. The parameters in sigma must be positive.

Gamma Distribution

$$p(x) = c \cdot x^{\alpha-1}\, e^{-\frac{x}{\beta}}$$

with x,α,β > 0

c is used to nomalize such that

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

Y = gampdf(X,α,β)

computes the gamma pdf at each of the values in X using the corresponding parameters in α and β. The parameters in α and β must all be positive, and the values in X must lie on the interval [0 ∞

The chi^2 distribution is a special case of the gamma distribution.

chi^2 distribution k degrees of freedom :

$$p(x) = c \cdot x^{\alpha-1}\, e^{-\frac{x}{\beta}}$$

with $\alpha=k/2$, $\beta=2$

c is a norm

Y = gampdf(X,k/2,2)

computes the chi^2 pdf at each of the values in X using the corresponding degree of freedom k. The parameters k must all be positive, and the values in X must lie on the interval [0 ∞

Central Limit Theorem

Rehearsal

R

Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample of size n with expected values μ and variances σ^2 (independent and identically distributed random variables)

Suppose we are interested in the sample average of these random variables:

$$S_N = \frac{1}{N}\sum_{n=1}^{N} X_n$$

Then the central limit theorem asserts that for large N's, the distribution of S_n is approximately normal with

distribution of S_n ~ Normal with

mean: μ

variance: $\frac{1}{N}\sigma^2$

The true strength of the theorem is that S_n approaches normality regardless of the shapes of the distributions of individual X_i's.

Suppose we have two independent random variables X and Y.

To get the probability distribution of p(Z=X+Y), we use a convolution. That is the integral of the joint probability of p(x,y), given that x+y=z (line integral)

$$
\begin{aligned}
p(z) &= \int_{-\infty}^{\infty} p(x, y \mid x + y = z)\,dx\,dy \\
&= \int_{-\infty}^{\infty} p(x)p(z-x)\,dx
\end{aligned}
$$

For a discrete distribution, we get:

$$p(z) = \sum p(x)p(z-x)$$

Neuroinformatics Lecture

Topics:

Maximum Likelihood

Bayes' Theorem:

posterior $\propto$ likelihood × prior

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Now use:

y=D (observed data)

x=$\vec{w}$ (set model parameter)

$$p(\vec{w}|D) = \frac{p(D|\vec{w})p(\vec{w})}{p(D)}$$

$$p(D) = \int P(D|\vec{w})p(\vec{w})d\vec{w}$$

Normalisation

Now $p(\vec{w}|D)$ probability of a model with set of parameters given the data

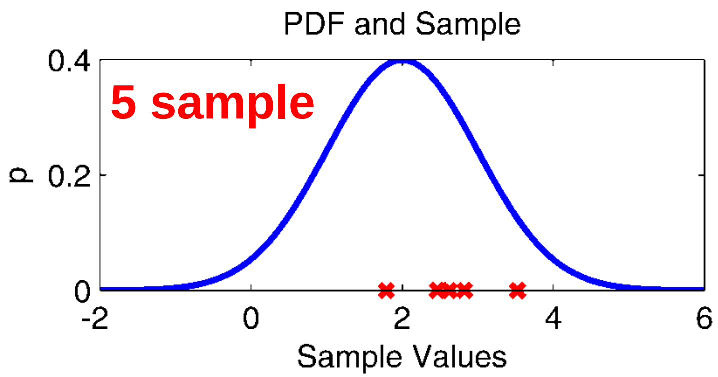
Bayes' Theorem:

Posterior:
probability of a model with set of parameters given the data

$$p(\vec{w}|D)$$

Likelihood:
probability of the data given with set of parameters of a model

$$p(D|\vec{w})$$



Likelihood: $p(D|\vec{w})$

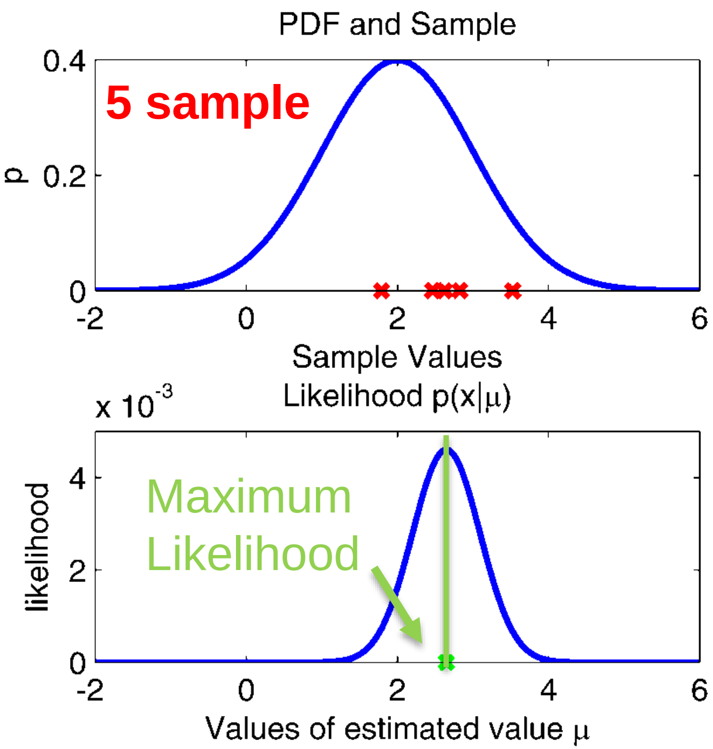
Likelihood function joint probability for all x_n conditioned on the parameters! Here we assume a Gaussian PDF.

$$p(\mathbf{x}|\mu,\sigma^2) = \prod_{i=1}^{N} \mathcal{N}\left(x_i|\mu,\sigma^2\right)$$

$$p(\vec{x}|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

$$p(\vec{x}|\mu,\sigma^2) = f(\mu,\vec{x},\sigma^2) = L(\mu)$$

Matlab code to generate these figures is available on STUDIP: Likelihood_Gauss.m



Likelihood: $p(D|\vec{w})$

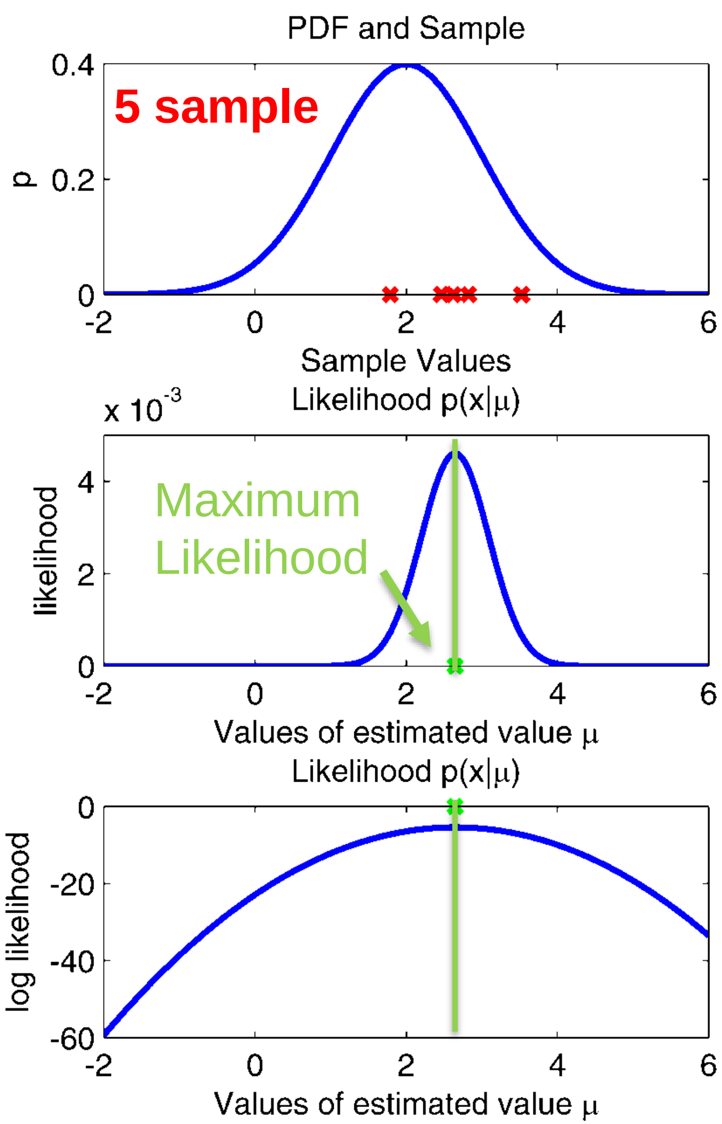
Likelihood function joint probability for all x_n conditioned on the parameters! Here we assume a Gaussian PDF.

$$p(\mathbf{x}|\mu,\sigma^2) = \prod_{i=1}^{N} \mathcal{N}\left(x_i|\mu,\sigma^2\right)$$

$$p(\vec{x}|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

$$p(\vec{x}|\mu,\sigma^2) = f(\mu,\vec{x},\sigma^2) = L(\mu)$$

Matlab code to generate these figures is available on STUDIP: Likelihood_Gauss.m



Likelihood: $p(D|\vec{w})$

Likelihood function joint probability for all x_n conditioned on the parameters! Here we assume a Gaussian PDF.

$$p(\mathbf{x}|\mu,\sigma^2) = \prod_{i=1}^{N} \mathcal{N}\left(x_i|\mu,\sigma^2\right)$$

$$p(\vec{x}|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

$$p(\vec{x}|\mu,\sigma^2) = f(\mu,\vec{x},\sigma^2) = L(\mu)$$

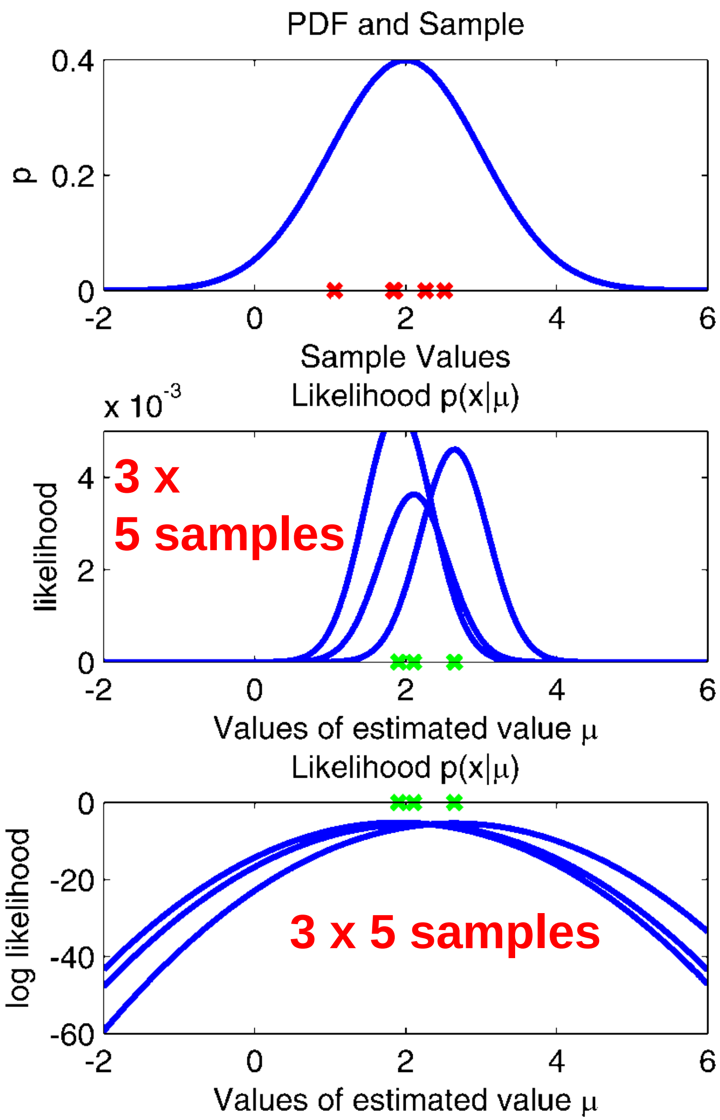
Matlab code to generate these figures is available on STUDIP: Likelihood_Gauss.m

So far:

- We sampled 5 i.i.d. samples from a given distribution and derived the maximum likelihood estimation for the expected value μ.
- We decided that the best estimator is the parameter that gives us the largest probability (highest likelihood) to explain the observed data.
- The maximum likelihood estimation is call μ_{ML}

Question:

- What happens if we repeat the same experiment more times.



Likelihood: $p(D|\vec{w})$

Likelihood function Joint probability for all x_n conditioned on the parameters!

The likelihood is a function of the data. Therefore the maximum depends of the data as well.

Interpretation:

The best guess about the real unknown parameter is data driven and therefore it self distributed.

Can we derive an analyticcal expression ?

Maximum Likelihood Approach for model fitting

- We observe Data *D* composed of samples x_i : $D = \{x_1, x_2, \dots, x_N\} = \vec{x}$
- *We assume the data is sampled form a single probability distribution characterized by some parameters :* $\vec{w} = (w_1, w_2, \dots .)$
- *In case of the normal distribution these parameters correspond to the expected value μ and the standard deviation:* $\vec{w} = (\mu, \sigma)$

Approach, we derive a function of the samples that defines the Maximum likelihood estimator of the parameter.

$$w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$$

For a Gaussian : $\mu_{ML} = f(x_1, x_2, \dots, x_N)$

Maximum Likelihood

- Data *D* composed of samples x_i: $D = \{x_1, x_2, \dots, x_N\} = \vec{x}$
- *single probability distribution with parameters :* $\vec{w} = (w_1, w_2, \dots .)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$

Step 1: Write down likelihood: $p(\vec{x}|w_1, w_2, \dots)$

Step 2: Assume sample are independent and from the same distribution (i.i.d):

$$p(\vec{x}|w_1, w_2, \dots) = \prod_{i=1}^{N} p(x_i|w_1, w_2, \dots)$$

Maximum Likelihood

- Data D composed of samples x_i: $D = \{x_1, x_2, \dots, x_N\} = \vec{x}$
- *single probability distribution with parameters* : $\vec{w} = (w_1, w_2, \dots.)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$

Step 1: Write down likelihood: $p(\vec{x}|w_1, w_2, \dots)$

Step 2: Assume sample are independent and from the same distribution (i.i.d):

$$p(\vec{x}|w_1, w_2, \dots) = \prod_{i=1}^{N} p(x_i|w_1, w_2, \dots)$$

Step 3: Assume a certain type of a distribution for example Gaussian

$$p(\vec{x}|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

Maximum Likelihood

- Data D composed of samples x_i: $D = \{x_1, x_2, \ldots, x_N\} = \vec{x}$
- *single probability distribution with parameters :* $\vec{w} = (w_1, w_2, \ldots .)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \ldots, x_N | w_2, \ldots)$

Step 1: Write down likelihood: $p(\vec{x}|w_1, w_2, \ldots)$

Step 2: Assume sample are independent and from the same distribution (i.i.d):

$$p(\vec{x}|w_1, w_2, \ldots) = \prod_{i=1}^{N} p(x_i|w_1, w_2, \ldots)$$

Step 3: Assume a certain type of a distribution for example Gaussian

$$p(\vec{x}|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

Step 4: Maximize this function in respect the wanted parameter

$$\underset{\mu}{\arg\max} \quad p(\vec{x}|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

Maximum Likelihood

- Data D composed of samples x_i: $D = \{x_1, x_2, \ldots, x_N\} = \vec{x}$
- *single probability distribution with parameters :* $\vec{w} = (w_1, w_2, \ldots.)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \ldots, x_N | w_2, \ldots)$

Step 4: Maximize this function in respect the wanted parameter

$$\arg\max_{\mu} \quad p(\vec{x}|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

argument of the maximum (abbreviated **arg max or argmax**) is the set of points of the given argument for which the given function attains its maximum value

Trick: Instead of maximizing the Likelihood we maximize the log of the likelihood (log likelihood)

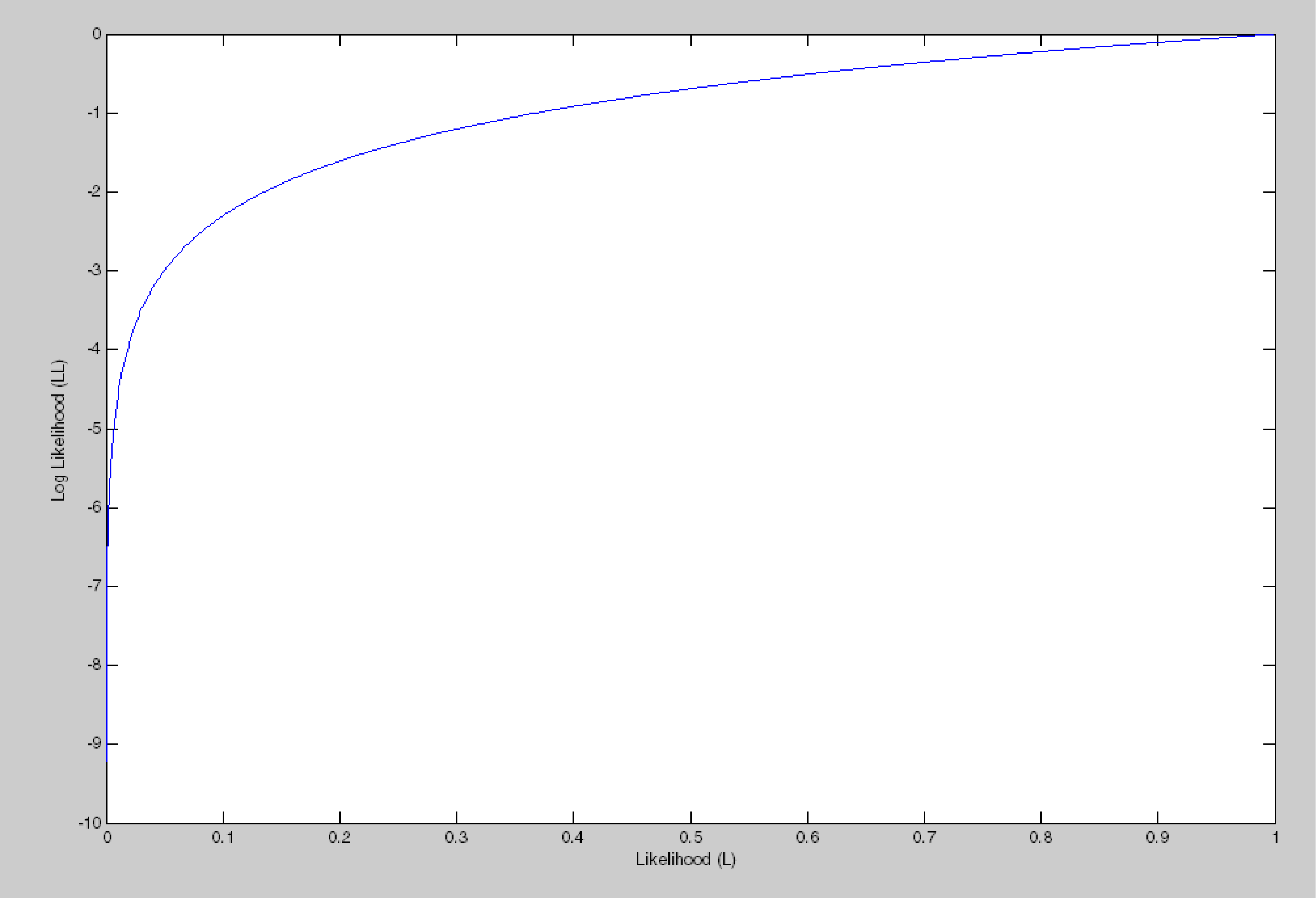
$$\log(a \cdot b \cdot c) = \log a + \log b + \log c$$

$$\log p(\vec{x}|\mu,\sigma) = \log \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)} = \sum_{i=1}^{N} log \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

Log Likelihood

Matlab Code:

```
L= 0.0001:0.001:1;
LL=log(L);
plot(L,LL)
xlabel('Likelihood (L) ');  ylabel('Log Likelihood (LL)')
```



Log is strictly monotonic

Maps L between [0,1] to LL from –inf to 0.

Maximum Likelihood

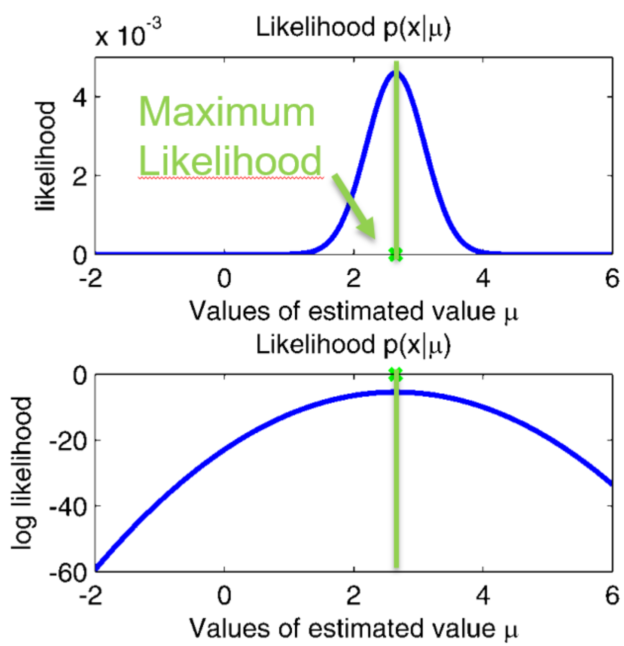
- Data *D* composed of samples x_i: $D = \{x_1, x_2, \dots, x_N\} = \vec{x}$
- *single probability distribution with parameters :* $\vec{w} = (w_1, w_2, \dots.)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$

Step 4: Maximize the likelihood in respect to the wanted parameter

$$\arg\max_{\mu} \; p(\vec{x}|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

$$\arg\max_{\mu} \; \log p(\vec{x}|\mu,\sigma) = \sum_{i=1}^{N} \log \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

The resulting parameter μ^{ML} is the same since log is a strictly monotonic and increasing function



Maximum Likelihood

- Data D composed of samples x_i: $D = \{x_1, x_2, \dots, x_N\} = \vec{x}$
- *single probability distribution with parameters :* $\vec{w} = (w_1, w_2, \dots.)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$

Step 4: Maximize the log likelihood in respect to the wanted parameter

$$\arg\max_{\mu} \log p(\vec{x}|\mu, \sigma) = \sum_{i=1}^{N} \log \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

With $\log(a \cdot b \cdot c) = \log a + \log b + \log c$

$$\arg\max_{\mu} \log p(\vec{x}|\mu, \sigma) = \sum_{i=1}^{N} \left[\log \frac{1}{\sigma\sqrt{2\pi}} + \log e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}\right]$$

$$\arg\max_{\mu} \log p(\vec{x}|\mu, \sigma) = \sum_{i=1}^{N} \left[\log \frac{1}{\sigma\sqrt{2\pi}} - \left(\frac{(x_i - \mu)^2}{2\sigma^2}\right)\right]$$

Maximum Likelihood

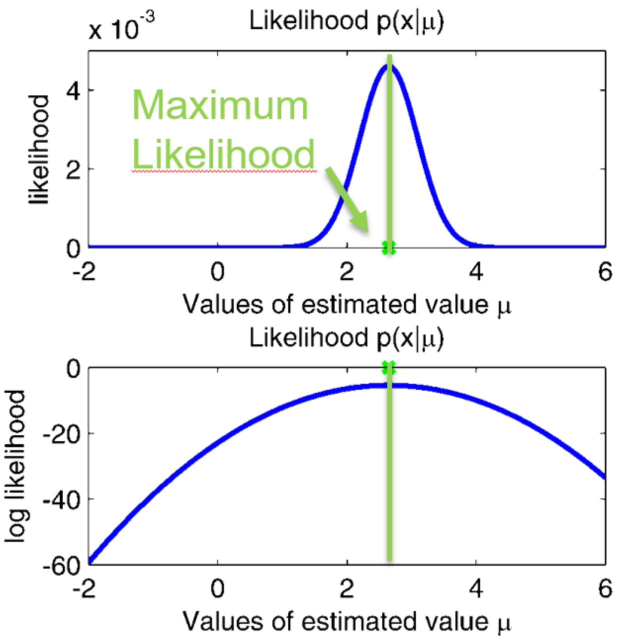
- Data *D* composed of samples x_i: $D = \{x_1, x_2, \dots, x_N\} = \vec{x}$
- *single probability distribution with parameters :* $\vec{w} = (w_1, w_2, \dots.)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$

Step 5: To derive μ that maximizes, we compute the derivative and set this to zero (the point of the likelihood with slope zero)

$$\arg\max_{\mu} \log p(\vec{x}|\mu,\sigma) = \sum_{i=1}^{N}\left[\log\frac{1}{\sigma\sqrt{2\pi}} - \left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right]$$

$$\Rightarrow \frac{\partial}{\partial\mu} \log p(\vec{x}|\mu,\sigma) = 0$$

$$\frac{\partial}{\partial\mu}\sum_{i=1}^{N}\left[\log\frac{1}{\sigma\sqrt{2\pi}} - \left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right] = 0$$



Compute the derivative in respect to just μ

Maximum Likelihood

- Data D composed of samples x_i: $D = \{x_1, x_2, \dots, x_N\} = \vec{x}$
- *single probability distribution with parameters :* $\vec{w} = (w_1, w_2, \dots.)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$

Step 6: Execute the derivative

$$\frac{\partial}{\partial \mu} \sum_{i=1}^{N} \left[\log \frac{1}{\sigma\sqrt{2\pi}} - \left(\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right] = 0 - \frac{\partial}{\partial \mu} \sum_{i=1}^{N} \left(\frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

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

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

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

Maximum Likelihood

- Data D composed of samples x_i: $D = \{x_1, x_2, \dots, x_N\} = \vec{x}$
- *single probability distribution with parameters :* $\vec{w} = (w_1, w_2, \dots.)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$

Step 7: set derivative to zero and resort term to get a function $f(x_1, x_2, \dots, x_N | w_2, \dots)$

$$\frac{\partial}{\partial \mu} \sum_{i=1}^{N} \left[\log \frac{1}{\sigma\sqrt{2\pi}} - \left(\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right] = +\frac{2}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu) \quad = 0$$

Maximum = slope flat

$$0 = +\frac{2}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)$$

Lösung 1: σ infinitely large(irrelevant for us)

Lösung 2: $0 = \sum_{i=1}^{N} (x_i - \mu)$

Lösung 2: $$0 = \sum_{i=1}^{N} (x_i - \mu^{ML}) \Leftrightarrow 0 = -N\mu^{ML} + \sum_{i=1}^{N} x_i$$

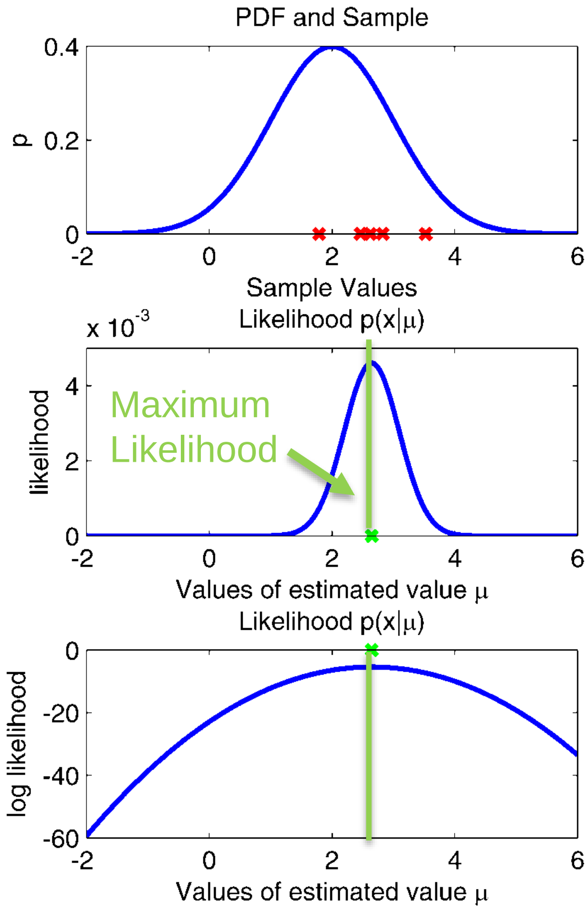
$$\boxed{\mu^{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i}$$

Maximum Likelihood

- Data D composed of samples x_i : $D = \{x_1, x_2, \dots, x_N\} = \vec{x}$
- *single probability distribution with parameters :* $\vec{w} = (w_1, w_2, \dots.)$
- *Derive* $w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$

Maximize the log likelihood in respect the wanted parameter

$$\arg\max_{\mu} \log p(\vec{x}|\mu, \sigma) = \sum_{i=1}^{N} \log \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$



Solution $w_1^{ML} = f(x_1, x_2, \dots, x_N | w_2, \dots)$
for the Maximum likelihood estimator

$$\mu^{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

The arithmetic mean is the best guess of the expected value

Q

1. Derive Maximum likelihood estimator of μ for a Gaussian distribution.

$$p(\vec{x}|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

Timer (5 min): Start Stop

Maximum Likelihood

Step 1: Write down likelihood: $p(\vec{x}|w_1, w_2, \ldots)$

Step 2: Assume sample are independent and from the same distribution (i.i.d):

$$p(\vec{x}|w_1, w_2, \ldots) = \prod_{i=1}^{N} p(x_i|w_1, w_2, \ldots)$$

Step 3: Assume a certain type of a distribution for example Gaussian

$$p(\vec{x}|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

Step 4: Maximize this function in respect the wanted parameter

$$\arg\max_{\mu} \quad p(\vec{x}|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}$$

$$\arg\max_{\mu} \ \log p(\vec{x}|\mu, \sigma) = \sum_{i=1}^{N} \left[\log \frac{1}{\sigma\sqrt{2\pi}} - \left(\frac{(x_i - \mu)^2}{2\sigma^2}\right)\right]$$

Step 5: Compute the derivative and set this to zero

$$\frac{\partial}{\partial\mu} \sum_{i=1}^{N} \left[\log \frac{1}{\sigma\sqrt{2\pi}} - \left(\frac{(x_i - \mu)^2}{2\sigma^2}\right)\right] = 0 - \frac{\partial}{\partial\mu} \sum_{i=1}^{N} \left(\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

- Step 7: set derivative to zero and resort term to get a function $f(x_1, x_2, \dots, x_N | w_2, \dots)$

Maximum = slope flat

$$\frac{\partial}{\partial \mu} \sum_{i=1}^{N} \left[\log \frac{1}{\sigma\sqrt{2\pi}} - \left(\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right] = + \frac{2}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu) \quad = 0$$

$$0 = + \frac{2}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)$$

Lösung 1: σ infinitely large (irrelevant for us)
Lösung 2: $0 = \sum_{i=1}^{N} (x_i - \mu)$

Lösung 2: $$0 = \sum_{i=1}^{N} (x_i - \mu^{ML}) \quad \Leftrightarrow 0 = -N\mu^{ML} + \sum_{i=1}^{N} x_i$$

$$\boxed{\mu^{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i}$$

Next Topic: ML for the expected value of a Bernoulli distribution

Summary: Bernoulli Distribution

Distribution: $p(x|\mu) = \mu^x(1-\mu)^{1-x}$

Expected value: $E[x] = \mu$

Variance: $var[x] = \mu(1-\mu)$

Y = binopdf(x,n,p)

For n=1, computes the Bernoulli pdf at each of the values in X using the corresponding probability P. The values in P must lie on the interval [0, 1].

ML for Bernoulli

Given: $\mathcal{D} = \{x_1, \ldots, x_N\}$, m heads (1), $N - m$ tails (0)

likelihood:

$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n}$$

$$\ln p(\mathcal{D}|\mu) = \sum_{n=1}^{N} \ln p(x_n|\mu) = \sum_{n=1}^{N} \{x_n \ln \mu + (1-x_n)\ln(1-\mu)\}$$

Partial derivative after μ (for ML set to zero):

$$\frac{\partial}{\partial \mu} \ln p(\vec{x}\,|\mu) = \sum_{n=1}^{N} \left(x_n \frac{\partial}{\partial \mu} \ln \mu + (1-x_n) \frac{\partial}{\partial \mu} \ln(1-\mu) \right)$$

ML for Bernoulli

Given: $\mathcal{D} = \{x_1, \ldots, x_N\}$, m heads (1), $N - m$ tails (0)

Partial derivative after μ (for ML set to zero):

$$\frac{\partial}{\partial \mu} \ln p(\vec{x}\,|\mu) = \sum_{n=1}^{N} \left(x_n \frac{\partial}{\partial \mu} \ln \mu + (1 - x_n) \frac{\partial}{\partial \mu} \ln(1 - \mu) \right)$$

$$0 = \sum_{n=1}^{N} \left(x_n \frac{1}{\mu} + (1 - x_n) \frac{-1}{1 - \mu} \right)$$

$$0 = \sum_{n=1}^{N} (x_n (1 - \mu) - (1 - x_n) \mu)$$

$$0 = \sum_{n=1}^{N} (x_n - \mathrm{x}_n \mu - \mu + x_n \mu)$$

ML for Bernoulli

Given: $\mathcal{D} = \{x_1, \ldots, x_N\}$, m heads (1), $N - m$ tails (0)

Partial derivative after μ (for ML set to zero):

$$\frac{\partial}{\partial \mu} \ln p(\vec{x} \,|\mu) = \sum_{n=1}^{N} \left(x_n \frac{\partial}{\partial \mu} \ln \mu + (1 - x_n) \frac{\partial}{\partial \mu} \ln(1 - \mu) \right)$$

$$0 = \sum_{n=1}^{N} (x_n - \mathrm{x}_n \mu - \mu + x_n \mu)$$

$$0 = \sum_{n=1}^{N} (x_n - \mu) = -N\mu + \sum_{n=1}^{N} x_n$$

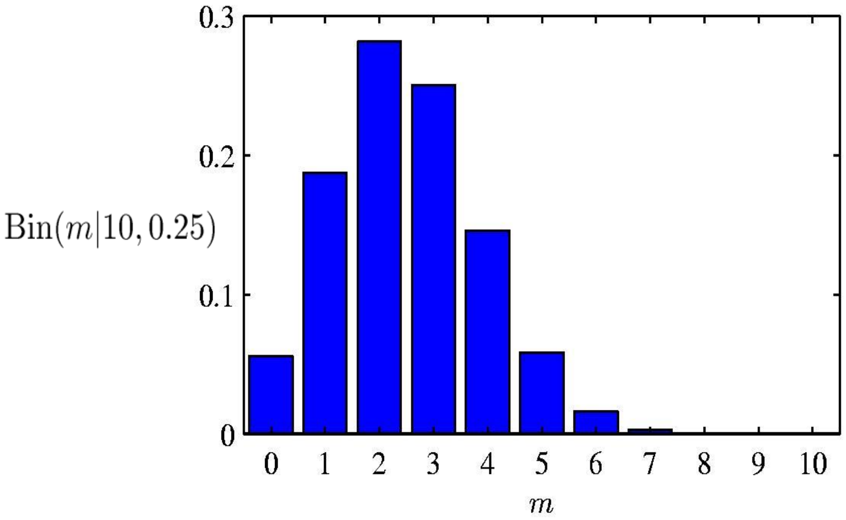
$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

ML estimator of $p(x_n=1)$, with $p(x_n=0)=1-p(x_n=1)$

Next Topic: ML for the expected value of a Binominal distribution

Binomial Distribution

$$\mathrm{Bin}(m|N,\mu) = \binom{N}{m}\mu^m(1-\mu)^{N-m}$$



$$\mathbb{E}[m] \equiv \sum_{m=0}^{N} m\mathrm{Bin}(m|N,\mu) \quad = \quad N\mu$$

$$\mathrm{var}[m] \equiv \sum_{m=0}^{N} (m - \mathbb{E}[m])^2\,\mathrm{Bin}(m|N,\mu) \quad = \quad N\mu(1-\mu)$$

Y = binopdf(x,n,p)

computes the binomial pdf at each of the values in X using the corresponding parameters in N and P. The parameters in N must be positive integers, and the values in P must lie on the interval [0, 1].

ML for Binominal

Given: $\mathcal{D} = \{x_1, \ldots, x_N\}$, m heads (1), $N - m$ tails (0)

likelihood: $p(m|\mu, N) = \binom{N}{m} \cdot \mu^m (1-\mu)^{N-m}$

Log likelihood: $\log p(m|\mu, N) = \log\left(\binom{N}{m} \cdot \mu^m (1-\mu)^{N-m}\right)$

$$\log p(m|\mu, N) = \log \binom{N}{m} + \mathrm{m} \log \mu + (N-m)\log(1-\mu)$$

Partial derivative after μ (for ML set to zero):

$$\frac{\partial}{\partial \mu} \ln p(m\,|\mu, N) = \frac{m}{\mu} - \frac{N-m}{1-\mu} = 0$$

$$m(1-\mu) - \mu(N-m) = m - m\mu - \mu N + m\mu = 0$$

$$\mu_{ML} = \frac{m}{N}$$

ML estimator of $p(x_n=1)$

Q

ML for Binominal

Given: $\mathcal{D} = \{x_1, \ldots, x_N\}$, m heads (1), $N - m$ tails (0)

likelihood: $p(m|\mu, N) = \binom{N}{m} \cdot \mu^m (1-\mu)^{N-m}$

Derive the log-likelihood and the ML estimator of μ

Rules you may want to use

$$ln \prod x = \sum \ln(x) \qquad \ln(x \cdot y \cdot z) = \ln(x) + \ln(y) + \ln(z) \qquad \ln(a^x) = x\ln(a)$$

Timer (8min): Start Stop

ML for Binominal

Given: $\mathcal{D} = \{x_1, \ldots, x_N\}$, m heads (1), $N - m$ tails (0)

likelihood: $p(m|\mu, N) = \binom{N}{m} \cdot \mu^m (1-\mu)^{N-m}$

Log likelihood: $\log p(m|\mu, N) = \log\left(\binom{N}{m} \cdot \mu^m (1-\mu)^{N-m}\right)$

$$\log p(m|\mu, N) = \log \binom{N}{m} + \mathrm{m} \log \mu + (N-m)\log(1-\mu)$$

Partial derivative after μ (for ML set to zero):

$$\frac{\partial}{\partial \mu} \ln p(m\,|\mu, N) = \frac{m}{\mu} - \frac{N-m}{1-\mu} = 0$$

$$m(1-\mu) - \mu(N-m) = m - m\mu - \mu N + m\mu = 0$$

$$\mu_{ML} = \frac{m}{N}$$

ML estimator of $p(x_n=1)$