Density Estimation

Data: $D = \{D_1, D_2, \ldots, D_n\}$

$D_i = \mathbf{x}_i$ a vector of attribute values

Attributes:

- modeled by random variables $\mathbf{X} = \{X_1, X_2, \ldots, X_d\}$ with:
  - **Continuous values**
  - **Discrete values**

  E.g. ***blood pressure*** with numerical values

  or ***chest pain*** with discrete values

  [no-pain, mild, moderate, strong]

Underlying true probability distribution:

$$p(\mathbf{X})$$

Parametric

- the distribution is modeled using a set of parameters Θ

$$p(\mathbf{X} \mid \Theta)$$

- **Example:** mean and covariances of a multivariate normal
- **Estimation:** find parameters Θ describing data D

Binomial distribution.

Data: D a set of order-independent outcomes with two possible values – 0 or 1 (head or tail)

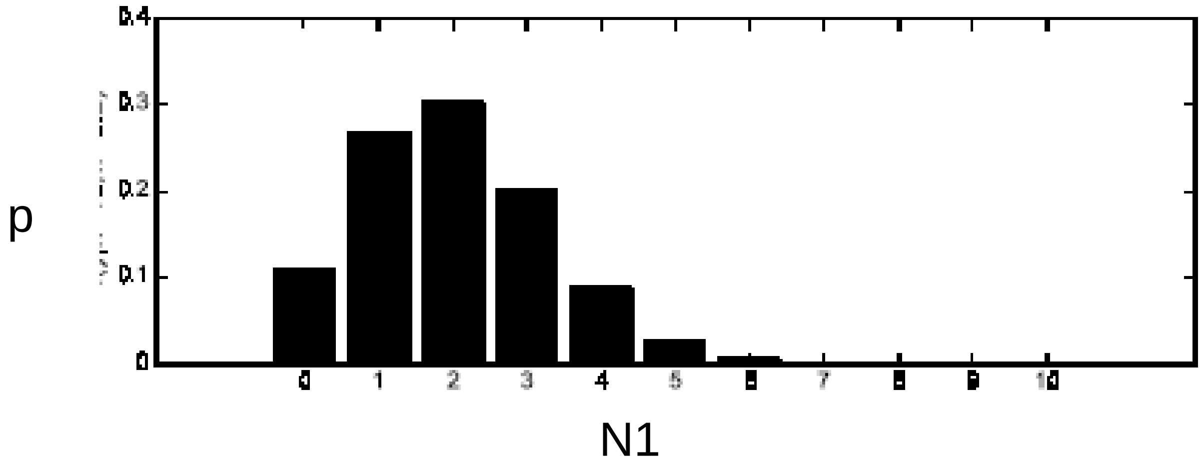
N_1 - number of heads seen N_2 - number of tails seen

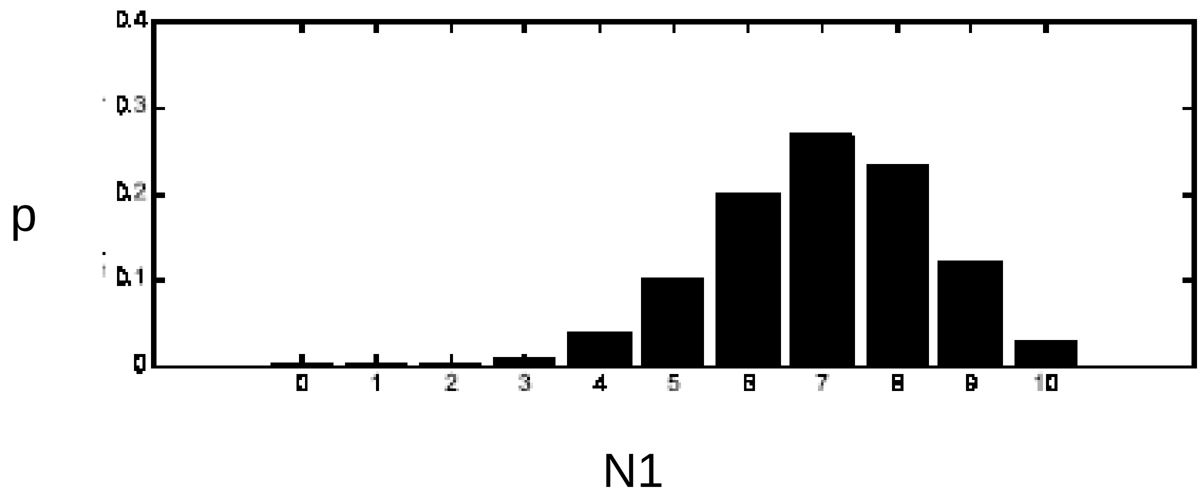
We treat *D* as a multi-set !!!

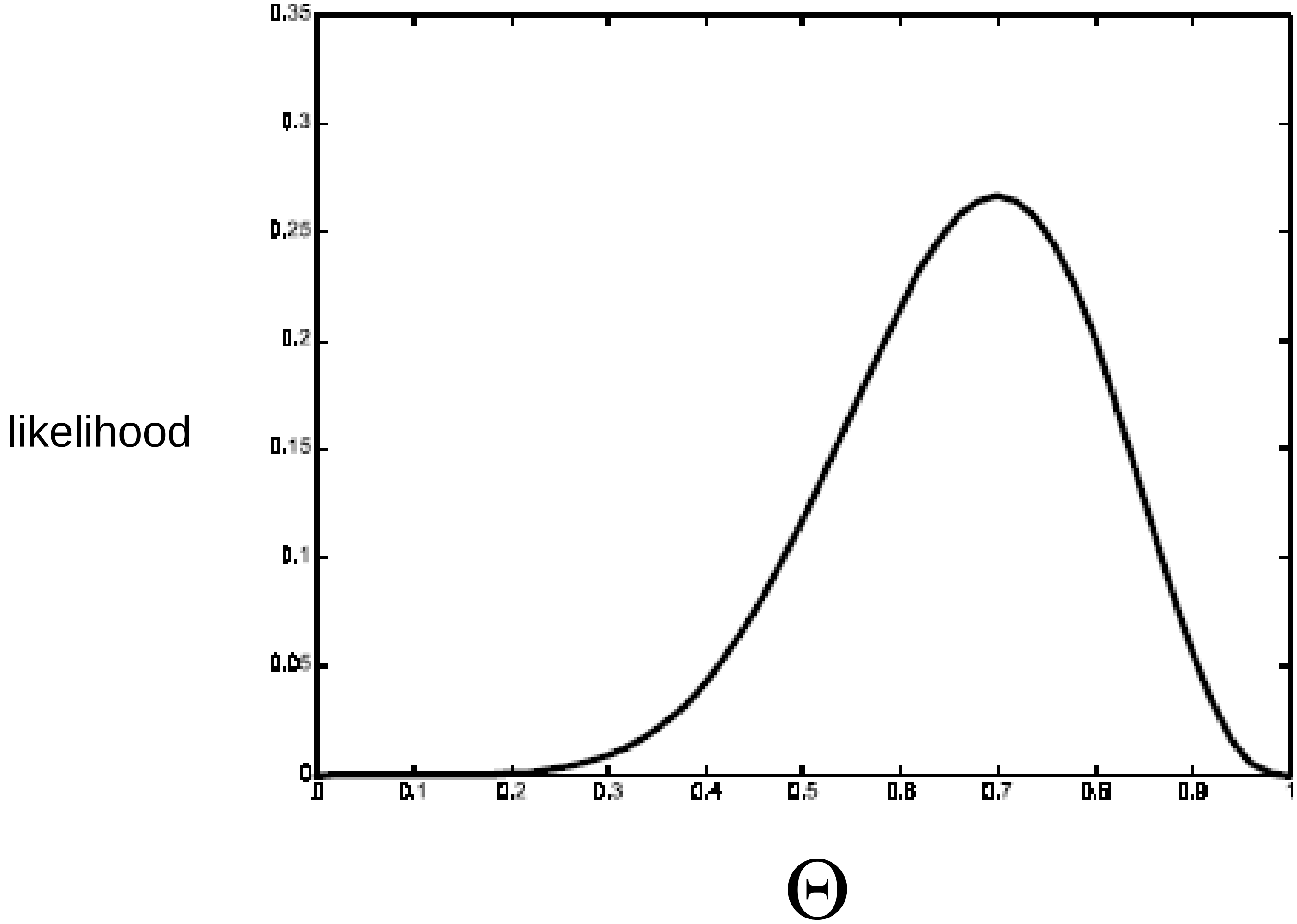
Model: probability of a 1 θ
probability of a 0 $(1-\theta)$

Probability of an outcome

$$P(D \mid \theta) = \binom{N}{N_1} \theta^{N_1} (1-\theta)^{N_2}$$ **Binomial distribution**


p
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10
N1


p
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10
N1


likelihood
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Θ

Parameter learning

What is the best set of parameters?

- **Maximum likelihood (ML) estimates**

maximize $p(D \mid \Theta, \xi)$

ξ - represents prior (background) knowledge

Bernoulli distribution.

Outcomes: x_i with values 0 or 1 (head or tail)
Data: D a sequence of outcomes x_i

Model: probability of an outcome 1 θ
probability of 0 $(1-\theta)$

$$P(x_i \mid \theta) = \theta^{x_i}(1-\theta)^{(1-x_i)}$$ **Bernoulli distribution**

Maximum likelihood (ML) estimate.

Optimize log-likelihood

$$l(D,\theta) = N_1 \log\theta + N_2 \log(1-\theta)$$

Set derivative to zero

$$\frac{\partial l(D,\theta)}{\partial\theta} = \frac{N_1}{\theta} - \frac{N_2}{(1-\theta)} = 0$$

Solving

$$\theta = \frac{N_1}{N_1 + N_2}$$

ML Solution:

$$\boxed{\theta_{ML} = \frac{N_1}{N} = \frac{N_1}{N_1 + N_2}}$$

A General MLE strategy

Suppose **q** = (*q*1, *q*2, …, *q*n)T is a vector of parameters.

Task: Find MLE θ assuming known form for p(Data| θ,stuff)

1. Write LL = log P(Data| θ,stuff)
2. Work out ∂LL/∂θ
3. Solve the set of simultaneous equations

$$\frac{\partial LL}{\partial \theta_1} = 0$$
$$\frac{\partial LL}{\partial \theta_2} = 0$$
$$\vdots$$
$$\frac{\partial LL}{\partial \theta_n} = 0$$

4. Check that you're at a maximum

MLE for univariate Gaussian

- Suppose you have $x_1, x_2, \dots x_R$ ~(i.i.d) $N(\mu,\sigma^2)$
- But you don't know μ or σ^2
- MLE: For which $\theta =(\mu,\sigma^2)$ is $x_1, x_2,\dots x_R$ most likely?

$$\log p(x_1, x_2, \dots x_R \mid \mu, \sigma^2) = -R(\log \pi + \frac{1}{2}\log \sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{R}(x_i - \mu)^2$$

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

$$\frac{\partial LL}{\partial \sigma^2} = -\frac{R}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{R}(x_i - \mu)^2$$

MLE for univariate Gaussian

- Suppose you have $x_1, x_2, \dots x_R$ ~(i.i.d) $N(\mu,\sigma^2)$
- But you don't know μ or σ^2
- MLE: For which $\theta = (\mu,\sigma^2)$ is $x_1, x_2, \dots x_R$ most likely?

$$\log p(x_1, x_2, \dots x_R \mid \mu, \sigma^2) = -R(\log \pi + \frac{1}{2}\log \sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{R}(x_i - \mu)^2$$

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

$$0 = -\frac{R}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{R}(x_i - \mu)^2$$

MLE for univariate Gaussian

- Suppose you have $x_1, x_2, \ldots x_R$ ~(i.i.d) $N(\mu,\sigma^2)$
- But you don't know μ or σ^2
- MLE: For which $\theta = (\mu,\sigma^2)$ is $x_1, x_2, \ldots x_R$ most likely?

$$\log p(x_1, x_2, \ldots x_R \mid \mu, \sigma^2) = -R(\log \pi + \frac{1}{2}\log \sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{R}(x_i - \mu)^2$$

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

$$0 = -\frac{R}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{R}(x_i - \mu)^2$$

MLE for univariate Gaussian

- Suppose you have $x_1, x_2, \ldots x_R$ ~(i.i.d) $N(\mu,\sigma^2)$
- But you don't know μ or σ^2
- MLE: For which $\theta = (\mu,\sigma^2)$ is $x_1, x_2, \ldots x_R$ most likely?

$$\mu^{mle} = \frac{1}{R}\sum_{i=1}^{R} x_i$$

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