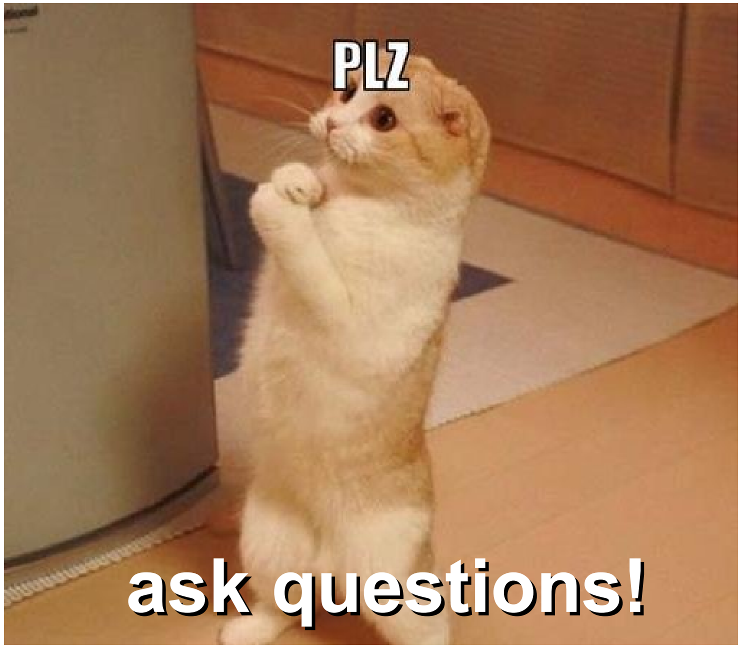
# Data Mining & Machine Learning

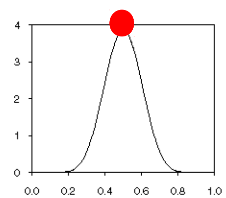
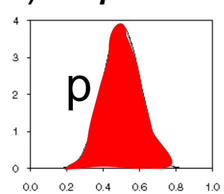
Dipl.-Inf. Christoph Carl Kling

People and Knowledge Networks




**DM@C-Kling.de**

# Parameter Estimation

| **Methode** | **Eigenschaften** |
| --- | --- |
| Maximum likelihood estimation (MLE)<br>$max(p(c\|p))$ | - Einfach<br>- Standard<br>- Bei wenigen Beobachtungen naiv!  |
| Maximum a posteriori estimation (MAP)<br>$max(p(p\|c)) = max(p(c\|p) \cdot p(p))$<br> | - Erlaubt Prior<br>- Auch bei wenigen Beobachtungen gut |
| Bayesian Inference<br>$E(\hat{p}) = \int_0^1 p \cdot p(c\|p) \cdot p(p)\ dp$<br> | - Berücksichtigt ganze Verteilung, nicht nur das Maximum |

# Beta distribution

$$Beta(\alpha, \beta) = \frac{p^{\alpha-1} \cdot (1-p)^{\beta-1}}{Z}$$

$$Z = \int_0^1 p^{\alpha-1} \cdot (1-p)^{\beta-1} \, dp = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

# Expected value for p

$$Beta(\alpha, \beta)$$

$$E(\hat{p}) = \int_0^1 p \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}\ dp = \frac{\alpha}{\alpha+\beta}$$

# Posterior probability

WeST
People and Knowledge Networks

$$posterior = \frac{likelihood \cdot prior}{evidence}$$

**For binomial distributed observations:**

$$p(p|\mathbf{c}) \propto p(\mathbf{c}|p) \cdot p(p)$$

$$= p^{n_1} \cdot (1-p)^{n_0} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$= p^{n_1+\alpha-1} \cdot (1-p)^{n_0+\beta-1}$$

$$= Beta(n_1+\alpha, n_0+\beta)$$

# Expected value for p

$$Beta(n_1 + \alpha, n_0 + \beta)$$

$$\Rightarrow E(\hat{p}) = \frac{n_1 + \alpha}{n_1 + \alpha + n_0 + \beta}$$

Das ist die Parameter-Schätzung für p mit **Bayesian inference**!

n1: Anzahl Würfe Kopf
n0: Anzahl Würfe Zahl

## Maximum likelihood estimation

$$\max_p p(p(\mathbf{c}|p)) = \frac{n_1}{n_1 + n_0}$$

## Maximum a posteriori estimation

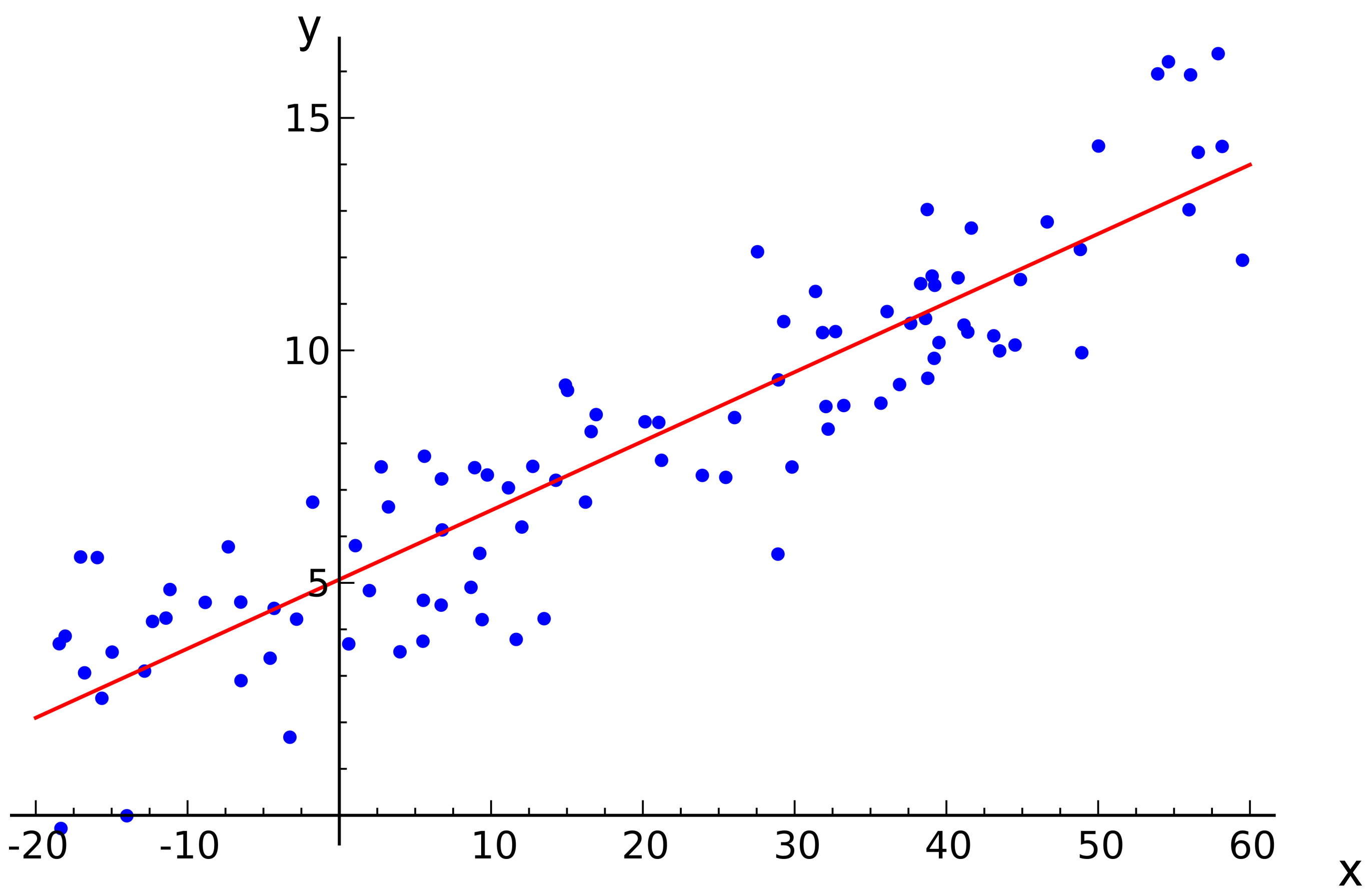
$$\max_p p(p(p|\mathbf{c})) = \frac{n_1 + \alpha - 1}{n_1 + \alpha - 1 + n_0 + \beta - 1}$$

## Bayesian inference

$$E(\hat{p}) = \frac{n_1 + \alpha}{n_1 + \alpha + n_0 + \beta}$$

# Lineare Regression

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots$$




$$\hat{y} = \beta_0 + \beta_1 x_1$$

**Annahmen:**

-Linearer Zusammenhang zwischen x und y

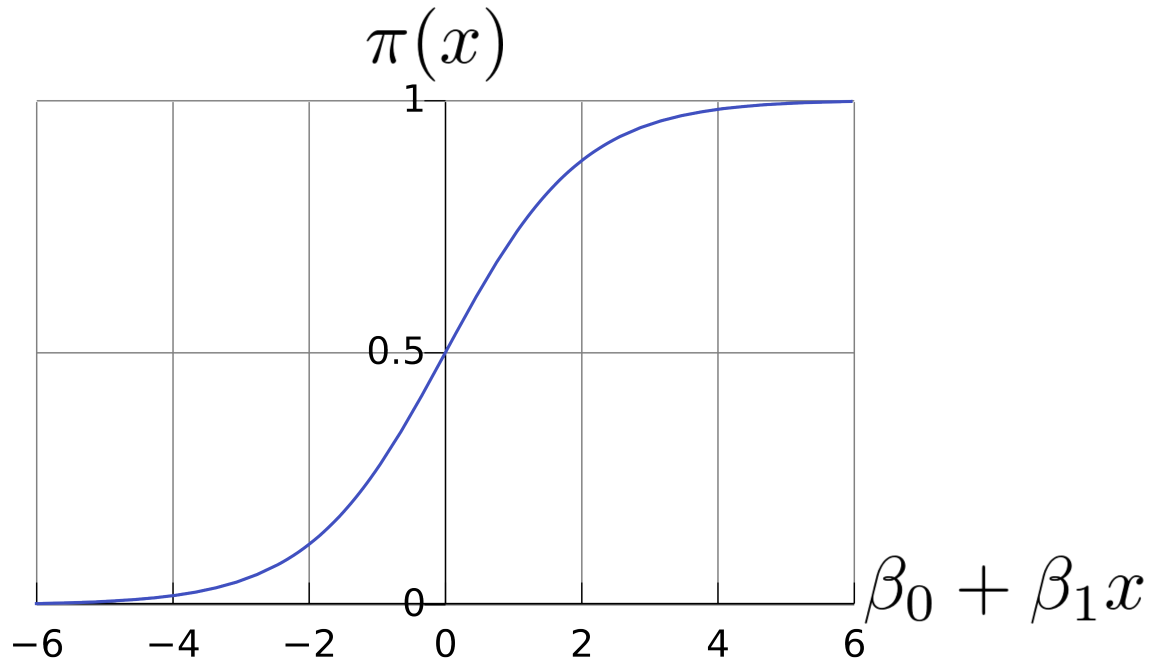
-Varianz konstant

-Keine starke Korrelation zwischen zwei Variablen

# Lineare Regression

| y = Größe | x1 = Geschlecht | x2 = Gewicht |
|---|---|---|
| 168 | 1 | 65 |
| 172 | 0 | 80 |
| 164 | 1 | 52 |
| 187 | 0 | 120 |
| 194 | 0 | 90 |

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 x)}}{e^{(\beta_0+\beta_1 x)} + 1} = \frac{1}{e^{-(\beta_0+\beta_1 x)} + 1}$$



# Logistische Regression

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 x)}}{e^{(\beta_0+\beta_1 x)}+1} = \frac{1}{e^{-(\beta_0+\beta_1 x)}+1}$$

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x$$

# Logistische Regression


Münze mit Gewicht x1 auf Rückseite

| | | | | | |
|---|---|---|---|---|---|
| Kopf | 6 | 4 | 3 | 3 | 2 |
| Zahl | 4 | 6 | 7 | 7 | 8 |
| x1 | 0g | 1g | 2g | 3g | 4g |

Web Science & Technologies

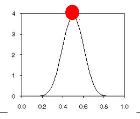
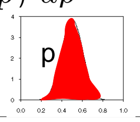
University of Koblenz ▪ Landau, Germany

# Data Mining & Machine Learning

## Dipl.-Inf. Christoph Carl Kling


PLZ
ask questions!
DM@C-Kling.de

## Parameter Estimation

| **Methode** | **Eigenschaften** |
| --- | --- |
| Maximum likelihood estimation (MLE)<br>$max(p(c\|p))$ | - Einfach<br>- Standard<br>- Bei wenigen Beobachtungen naiv!  |
| Maximum a posteriori estimation (MAP)<br>$max(p(p\|c)) = max(p(c\|p) \cdot p(p))$<br> | - Erlaubt Prior<br>- Auch bei wenigen Beobachtungen gut |
| Bayesian Inference<br>$E(\hat{p}) = \int_0^1 p \cdot p(c\|p) \cdot p(p)\ dp$<br> | - Berücksichtigt ganze Verteilung, nicht nur das Maximum |

## Beta distribution

$$Beta(\alpha, \beta) = \frac{p^{\alpha-1} \cdot (1-p)^{\beta-1}}{Z}$$

$$Z = \int_0^1 p^{\alpha-1} \cdot (1-p)^{\beta-1}\ dp = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

## Expected value for p

$$Beta(\alpha, \beta)$$

$$E(\hat{p}) = \int_0^1 p \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}\, dp = \frac{\alpha}{\alpha + \beta}$$

# Posterior probability

$$posterior = \frac{likelihood \cdot prior}{evidence}$$

**For binomial distributed observations:**

$$p(p|\mathbf{c}) \propto \quad p(\mathbf{c}|p) \cdot p(p)$$

$$= p^{n_1} \cdot (1-p)^{n_0} \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$= p^{n_1+\alpha-1} \cdot (1-p)^{n_0+\beta-1}$$

$$= Beta(n_1+\alpha, n_0+\beta)$$

## Expected value for p

$$Beta(n_1 + \alpha, n_0 + \beta)$$

$$\Rightarrow E(\hat{p}) = \frac{n_1 + \alpha}{n_1 + \alpha + n_0 + \beta}$$

Das ist die Parameter-Schätzung für p mit **Bayesian inference**!

## Estimation of p

n1: Anzahl Würfe Kopf
n0: Anzahl Würfe Zahl

### Maximum likelihood estimation

$$\max_p p(p(\mathbf{c}|p)) = \frac{n_1}{n_1 + n_0}$$

### Maximum a posteriori estimation

$$\max_p p(p(p|\mathbf{c})) = \frac{n_1 + \alpha - 1}{n_1 + \alpha - 1 + n_0 + \beta - 1}$$

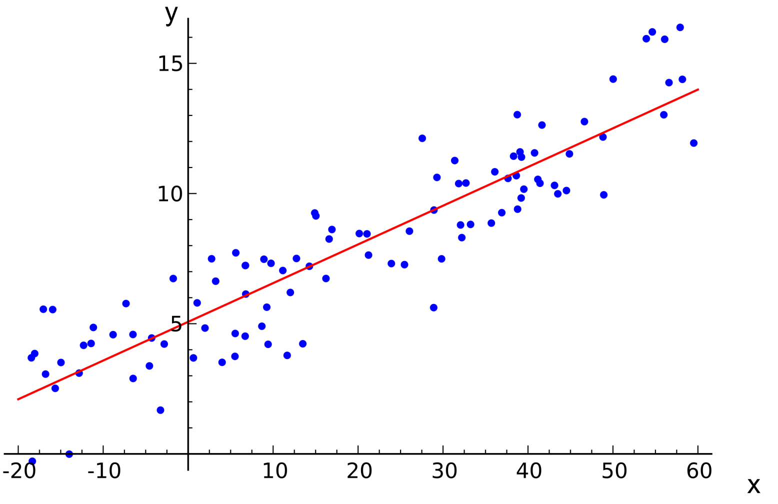
### Bayesian inference

$$E(\hat{p}) = \frac{n_1 + \alpha}{n_1 + \alpha + n_0 + \beta}$$

## Lineare Regression

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots$$

# Lineare Regression




$$\hat{y} = \beta_0 + \beta_1 x_1$$

**Annahmen:**

-Linearer Zusammenhang zwischen x und y

-Varianz konstant

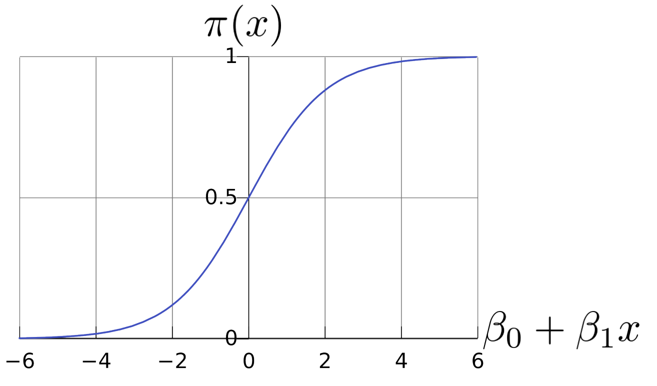
-Keine starke Korrelation zwischen zwei Variablen

# Lineare Regression

| y = Größe | x1 = Geschlecht | x2 = Gewicht |
|---|---|---|
| 168 | 1 | 65 |
| 172 | 0 | 80 |
| 164 | 1 | 52 |
| 187 | 0 | 120 |
| 194 | 0 | 90 |

## Logistische Regression

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 x)}}{e^{(\beta_0+\beta_1 x)}+1} = \frac{1}{e^{-(\beta_0+\beta_1 x)}+1}$$



## Logistische Regression

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 x)}}{e^{(\beta_0+\beta_1 x)}+1} = \frac{1}{e^{-(\beta_0+\beta_1 x)}+1}$$

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x$$

# Logistische Regression

Münze mit Gewicht x1 auf Rückseite

| | | | | | |
|---|---|---|---|---|---|
| Kopf | 6 | 4 | 3 | 3 | 2 |
| Zahl | 4 | 6 | 7 | 7 | 8 |
| x1 | 0g | 1g | 2g | 3g | 4g |