Tag 5: Bayes-Statistik und Wahlprognosen

Philip Biegel Tanja Bien Maximilian Wiesmann

Sommerakademie der Studienstiftung Olang - Arbeitsgruppe 4: Empirische Wahlforschung und Wahlprognosen

6. September 2019

Inhalt

Bayessche Inferenz

„*At the heart of statistics lie the ideas of statistical inference. Methods of statistical inference enable the investigator to argue from the particular observations in a sample to the general case. In contrast to logical deductions made from the general case to the specific case, a statistical inference can sometimes be incorrect. Nevertheless, one of the great intellectual advances of the twentieth century is the realization that strong scientific evidence can be developed on the basis of many, highly variable, observations.*“ (aus: Encyclopedia of Mathematics)

Bayessche Inferenz

„*At the heart of statistics lie the ideas of statistical inference. Methods of statistical inference enable the investigator to argue from the particular observations in a sample to the general case. In contrast to logical deductions made from the general case to the specific case, a statistical inference can sometimes be incorrect. Nevertheless, one of the great intellectual advances of the twentieth century is the realization that strong scientific evidence can be developed on the basis of many, highly variable, observations.*“ (aus: Encyclopedia of Mathematics)

Satz (Bayes)

$$Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)} = \frac{Pr(B|A)Pr(A)}{\sum Pr(B|A)Pr(A)}$$

Bayesscher Wasserfall

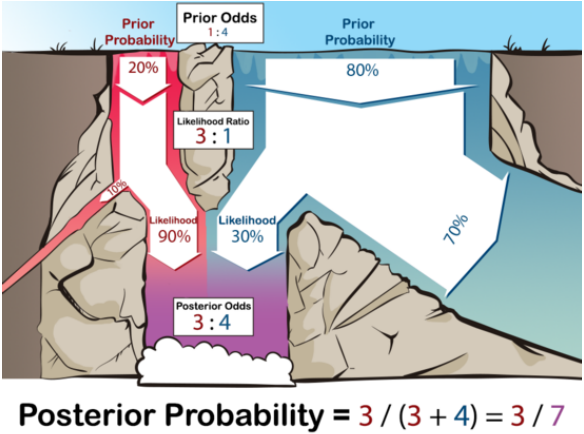


Abbildung: arbital.com

Bsp.: Diagnosetest

- $D+$: Person ist krank; $D-$: Person nicht krank
- $T+$: Test positiv; $T-$: Test negativ
- $Pr(T+|D+) = Pr(T-|D-) = 90\%$
- $Pr(D+) = 1\%$

Bsp.: Diagnosetest

- $D+$: Person ist krank; $D-$: Person nicht krank
- $T+$: Test positiv; $T-$: Test negativ
- $Pr(T+|D+) = Pr(T-|D-) = 90\%$
- $Pr(D+) = 1\%$

$$
\begin{aligned}
\Rightarrow Pr(T+) &= Pr(T+|D+)Pr(D+) + Pr(T+|D-)Pr(D-)\\
&= Pr(T+|D+)Pr(D+) + (1 - Pr(T+|D+))(1 - Pr(D-))\\
&= 0,108
\end{aligned}
$$

Bsp.: Diagnosetest

- $D+$: Person ist krank; $D-$: Person nicht krank
- $T+$: Test positiv; $T-$: Test negativ
- $Pr(T+|D+) = Pr(T-|D-) = 90\%$
- $Pr(D+) = 1\%$

$$
\begin{aligned}
\Rightarrow Pr(T+) &= Pr(T+|D+)Pr(D+) + Pr(T+|D-)Pr(D-) \\
&= Pr(T+|D+)Pr(D+) + (1 - Pr(T+|D+))(1 - Pr(D-)) \\
&= 0,108
\end{aligned}
$$

$$
\overset{\text{Bayes}}{\Rightarrow} Pr(D+|T+) = \frac{Pr(T+|D+)Pr(D+)}{Pr(T+)} \approx 0,083
$$

Bayes für kontinuierliche Verteilungen

Satz (Bayes)

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta} \propto f(x|\theta)f(\theta)$$

Bayes für kontinuierliche Verteilungen

Satz (Bayes)

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta} \propto f(x|\theta)f(\theta)$$

- $f(\theta|x)$: posterior distribution
- $f(\theta)$: prior distribution
- $f(x|\theta)$: likelihood function (auch bez. mit $L(\theta)$)
- $f(x)$: marginal likelihood

Bayesian Point Estimates

- posterior mean $E(\theta|x) = \int \theta f(\theta|x)\,\mathrm{d}\theta$

Bayesian Point Estimates

- posterior mean $E(\theta|x) = \int \theta f(\theta|x)\,\mathrm{d}\theta$
- posterior mode $Mod(\theta|x) = \arg\max_{\theta} f(\theta|x)$

Bayesian Point Estimates

- posterior mean $E(\theta|x) = \int \theta f(\theta|x)\,\mathrm{d}\theta$
- posterior mode $Mod(\theta|x) = \arg\max_{\theta} f(\theta|x)$
- posterior median $Med(\theta|x) = a$, sd. $\int_{-\infty}^{a} f(\theta|x)\,\mathrm{d}\theta = \frac{1}{2} = \int_{a}^{\infty} f(\theta|x)\,\mathrm{d}\theta$

Credible Intervals

- Das Intervall $I = [t_l, t_u]$ heißt $\gamma \cdot 100\%$ credible interval (Glaubwürdigkeitsintervall), falls gilt:

$$\int_{t_l}^{t_u} f(\theta|x)\,\mathrm{d}\theta = \gamma$$

Credible Intervals

- Das Intervall $I = [t_l, t_u]$ heißt $\gamma \cdot 100\%$ credible interval (Glaubwürdigkeitsintervall), falls gilt:

$$\int_{t_l}^{t_u} f(\theta|x)\,\mathrm{d}\theta = \gamma$$

- equi-tailed:

$$\int_{-\infty}^{t_l} f(\theta|x)\,\mathrm{d}\theta = \int_{t_u}^{\infty} f(\theta|x)\,\mathrm{d}\theta = \frac{1-\gamma}{2}$$

Credible Intervals

- Das Intervall $I = [t_l, t_u]$ heißt $\gamma \cdot 100\%$ credible interval (Glaubwürdigkeitsintervall), falls gilt:

$$\int_{t_l}^{t_u} f(\theta|x)\,\mathrm{d}\theta = \gamma$$

- equi-tailed:

$$\int_{-\infty}^{t_l} f(\theta|x)\,\mathrm{d}\theta = \int_{t_u}^{\infty} f(\theta|x)\,\mathrm{d}\theta = \frac{1-\gamma}{2}$$

- highest posterior density (HPD):

$$f(\theta|x) \geq f(\tilde{\theta}|x)\ \forall \theta \in I, \tilde{\theta} \notin I$$

in „schönen“ Fällen gilt: $f(t_l|x) = f(t_u|x)$

Beta-Verteilung

- $\theta \sim Be(\alpha, \beta)$

Beta-Verteilung

- $\theta \sim Be(\alpha, \beta)$
- Dichtefunktion

$$f(\theta) = \begin{cases} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} & \text{für } \theta \in [0,1] \\ 0 & \text{sonst} \end{cases}$$

mit Beta-Funktion $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, \mathrm{d}t = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

Beta-Verteilung

- $\theta \sim Be(\alpha, \beta)$
- Dichtefunktion

$$f(\theta) = \begin{cases} \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & \text{für } \theta \in [0,1] \\ 0 & \text{sonst} \end{cases}$$

mit Beta-Funktion $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

- $$E(\theta) = \frac{\alpha}{\alpha+\beta} \quad , \quad Mod(\theta) = \frac{\alpha-1}{\alpha+\beta-2} \quad , \quad Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Beta-Verteilung

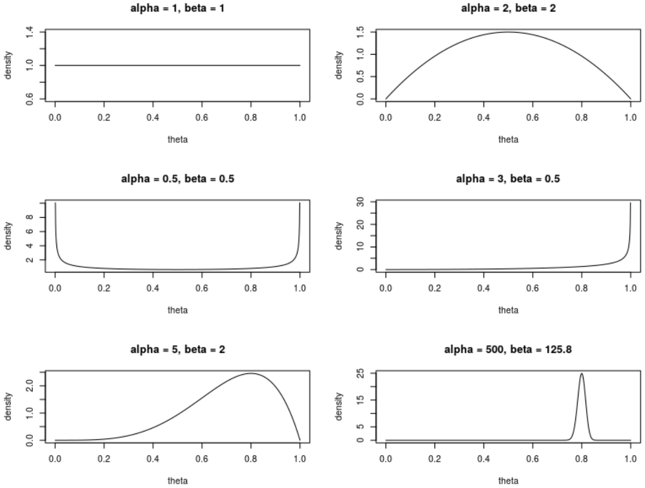
- $\theta \sim Be(\alpha, \beta)$
- Dichtefunktion

$$f(\theta) = \begin{cases} \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & \text{für } \theta \in [0,1] \\ 0 & \text{sonst} \end{cases}$$

mit Beta-Funktion $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

- $$E(\theta) = \frac{\alpha}{\alpha+\beta} \quad , \quad Mod(\theta) = \frac{\alpha-1}{\alpha+\beta-2} \quad , \quad Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
- für $\alpha = \beta = 1$ erhält man die Gleichverteilung auf $[0, 1]$

Beta-Verteilung



Beta-Verteilung

- $X \sim Bin(n, \theta)\ , \quad \theta \sim Be(\alpha, \beta)$

Beta-Verteilung

- $X \sim Bin(n, \theta)$, $\theta \sim Be(\alpha, \beta)$
- $\Rightarrow f(\theta|x) \propto f(x|\theta) f(\theta) \propto \theta^x (1-\theta)^{n-x} \theta^{\alpha-1} (1-\theta)^{\beta-1}$

Beta-Verteilung

- $X \sim Bin(n, \theta)$, $\quad \theta \sim Be(\alpha, \beta)$
- $\Rightarrow f(\theta|x) \propto f(x|\theta)f(\theta) \propto \theta^x(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}$
- $\Rightarrow \theta|x \sim Be(\alpha + x, \beta + n - x)$

Beta-Verteilung

- $X \sim Bin(n, \theta)$, $\quad \theta \sim Be(\alpha, \beta)$
- $\Rightarrow f(\theta|x) \propto f(x|\theta)f(\theta) \propto \theta^x(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}$
- $\Rightarrow \theta|x \sim Be(\alpha + x, \beta + n - x)$
- Beta-Verteilung *konjugiert* zur Binomialverteilung

Beta-Verteilung Beispiel

- $\theta \sim Be(3, 2)$, $\quad X \sim Bin(n, \theta)$

Beta-Verteilung Beispiel

- $\theta \sim Be(3,2)$, $\quad X \sim Bin(n,\theta)$
- Ergebnis der Stichprobe mit Umfang $n = 10$: $x = 8$

Beta-Verteilung Beispiel

- $\theta \sim Be(3,2)$, $\quad X \sim Bin(n,\theta)$
- Ergebnis der Stichprobe mit Umfang $n = 10$: $x = 8$
- $\Rightarrow \theta|x \sim Be(11,4)$

Beta-Verteilung Beispiel

- $\theta \sim Be(3,2)$, $\quad X \sim Bin(n,\theta)$
- Ergebnis der Stichprobe mit Umfang $n = 10$: $x = 8$
- $\Rightarrow \theta|x \sim Be(11,4)$

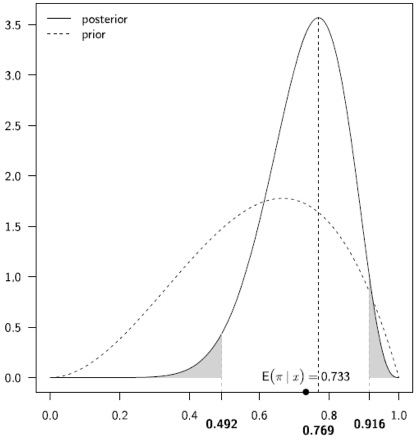


Abbildung: Held, Bové: Applied Statistical Inference, S. 173

Capture-Recapture-Method

- Möchte Anzahl N von Fischen in einem See bestimmen

Capture-Recapture-Method

- Möchte Anzahl N von Fischen in einem See bestimmen
- Entnehme M Fische, markiere diese und gebe sie wieder zurück

Capture-Recapture-Method

- Möchte Anzahl N von Fischen in einem See bestimmen
- Entnehme M Fische, markiere diese und gebe sie wieder zurück
- Entnehme n Fische und bestimme Anzahl x der markierten Fische

Capture-Recapture-Method

- Möchte Anzahl N von Fischen in einem See bestimmen
- Entnehme M Fische, markiere diese und gebe sie wieder zurück
- Entnehme n Fische und bestimme Anzahl x der markierten Fische
- Möglicher Ansatz: $\frac{M}{N} \approx \frac{x}{n}$

Capture-Recapture-Method

- Ansatz mittels Bayesscher Inferenz:

Capture-Recapture-Method

- Ansatz mittels Bayesscher Inferenz:
- Likelihood:

$$f(x|N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Capture-Recapture-Method

- Ansatz mittels Bayesscher Inferenz:
- Likelihood:

$$f(x|N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

- Geometrische Verteilung als Prior: $f(N) \propto \pi(1-\pi)^{N-1}$

Capture-Recapture-Method

- Ansatz mittels Bayesscher Inferenz:
- Likelihood:

$$f(x|N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

- Geometrische Verteilung als Prior: $f(N) \propto \pi(1-\pi)^{N-1}$
- Bsp.: $M = 26\,,\ n = 63\,,\ x = 5\,,\ \pi = 0,0011$

Capture-Recapture-Method

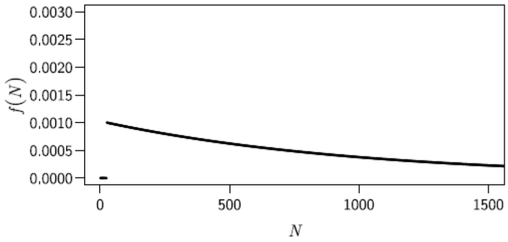


Abbildung: Held, Bové: Applied Statistical Inference, S. 179

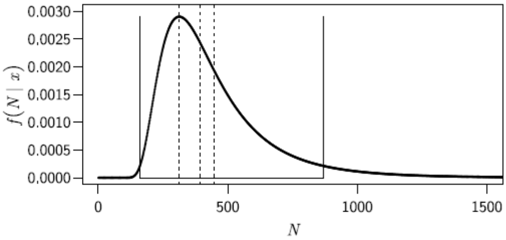


Abbildung: ebd.

- $Mod(N|x) = 313\,,\ \ Med(N|x) = 392\,,\ \ E(N|x) = 446.5$

Capture-Recapture-Method

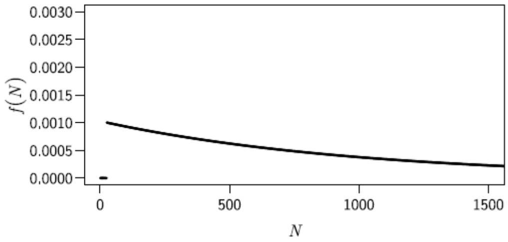


Abbildung: Held, Bové: Applied Statistical Inference, S. 179

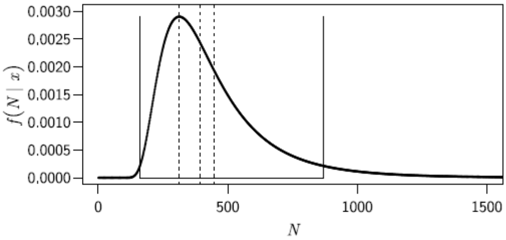


Abbildung: ebd.

- $Mod(N|x) = 313\,,\;\; Med(N|x) = 392\,,\;\; E(N|x) = 446.5$
- 95% HPD Intervall: $[61; 869]$

Capture-Recapture-Method

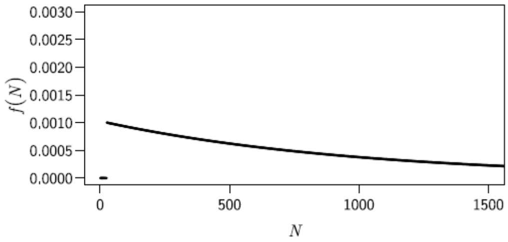


Abbildung: Held, Bové: Applied Statistical Inference, S. 179

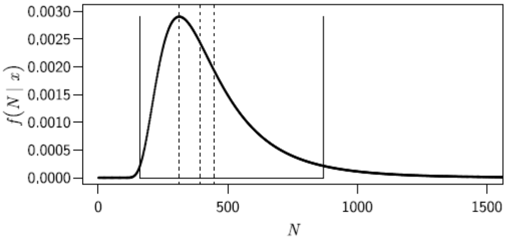


Abbildung: ebd.

- $Mod(N|x) = 313\,,\;\; Med(N|x) = 392\,,\;\; E(N|x) = 446.5$
- 95% HPD Intervall: $[61; 869]$
- Erster Ansatz liefert: $N \approx 327,6$

Variablentransformation

- $Y = g(X)$
- $$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} \right|$$

Jeffrey's prior

- Ziel: Priori Verteilung invariant unter Variablentransformation

Jeffrey's prior

- Ziel: Priori Verteilung invariant unter Variablentransformation
- Fisher Information:

$$I(\theta) = -\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta^2}$$

mit log-likelihood function $l(\theta) = ln(L(\theta))$

Jeffrey's prior

- Ziel: Priori Verteilung invariant unter Variablentransformation
- Fisher Information:

$$I(\theta) = -\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta^2}$$

  mit log-likelihood function $l(\theta) = ln(L(\theta))$
- Erwartete Fisher Information:

$$J(\theta) = E(I(\theta; X))$$

  wobei der Erwartungswert bezüglich X gebildet wird

Jeffrey's prior

- Ziel: Priori Verteilung invariant unter Variablentransformation
- Fisher Information:

$$I(\theta) = -\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta^2}$$

  mit log-likelihood function $l(\theta) = ln(L(\theta))$
- Erwartete Fisher Information:

$$J(\theta) = E(I(\theta; X))$$

  wobei der Erwartungswert bezüglich X gebildet wird
- Jeffrey's prior

$$f(\theta) \propto \sqrt{J(\theta)}$$

Jeffrey's prior

- Ziel: Priori Verteilung invariant unter Variablentransformation
- Fisher Information:

$$I(\theta) = -\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta^2}$$

  mit log-likelihood function $l(\theta) = ln(L(\theta))$
- Erwartete Fisher Information:

$$J(\theta) = E(I(\theta; X))$$

  wobei der Erwartungswert bezüglich X gebildet wird
- Jeffrey's prior

$$f(\theta) \propto \sqrt{J(\theta)}$$

- drückt Uninformiertheit aus

Jeffrey's prior

- Für multivariate Verteilungen: Erwartete Fisher Information Matrix

$$[J(\boldsymbol{\theta})]_{ij} = E(-\partial_i \partial_j l(\boldsymbol{\theta}))$$

Jeffrey's prior

- Für multivariate Verteilungen: Erwartete Fisher Information Matrix

$$[J(\boldsymbol{\theta})]_{ij} = E(-\partial_i \partial_j l(\boldsymbol{\theta}))$$

- Jeffrey's prior

$$f(\boldsymbol{\theta}) \propto \sqrt{\det J(\boldsymbol{\theta})}$$

Jeffrey's prior

- Für multivariate Verteilungen: Erwartete Fisher Information Matrix

$$[J(\boldsymbol{\theta})]_{ij} = E(-\partial_i \partial_j l(\boldsymbol{\theta}))$$

- Jeffrey's prior

$$f(\boldsymbol{\theta}) \propto \sqrt{\det J(\boldsymbol{\theta})}$$

Satz

Jeffrey's prior ist invariant unter Variablentransformation, d.h. wenn $\phi = g(\theta)$ *und* $f_\theta(\theta) \propto \sqrt{J_\theta(\theta)}$, *dann ist* $f_\phi(\phi) \propto \sqrt{J_\phi(\phi)}$.

Jeffrey's prior für $Bin(n, \theta)$

- $X \sim Bin(n, \theta)$, also $f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$

Jeffrey's prior für $Bin(n, \theta)$

- $X \sim Bin(n, \theta)$, also $f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$
- $l(\theta) = x\,ln(\theta) + (n-x)\,ln(1-\theta) + const$

Jeffrey's prior für $Bin(n, \theta)$

- $X \sim Bin(n, \theta)$, also $f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$
- $l(\theta) = x\,ln(\theta) + (n-x)\,ln(1-\theta) + const$
- $-\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta^2} = \frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}$

Jeffrey's prior für $Bin(n, \theta)$

- $X \sim Bin(n, \theta)$, also $f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$
- $l(\theta) = x\, ln(\theta) + (n-x)\, ln(1-\theta) + const$
- $-\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta^2} = \frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}$
- $E(-\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta^2}) = \frac{n\theta}{\theta^2} - \frac{n(1-\theta)}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)}$

Jeffrey's prior für $Bin(n, \theta)$

- $X \sim Bin(n, \theta)$, also $f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$
- $l(\theta) = x\, ln(\theta) + (n-x)\, ln(1-\theta) + const$
- $-\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta^2} = \frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}$
- $E(-\frac{\mathrm{d}^2 l(\theta)}{\mathrm{d}\theta^2}) = \frac{n\theta}{\theta^2} - \frac{n(1-\theta)}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)}$
- $\Rightarrow \sqrt{J(\theta)} \propto Be(\frac{1}{2}, \frac{1}{2})$

Jeffrey's prior für Multinomialverteilung

- $\boldsymbol{x} = (x_1, \ldots, x_p)\,,\ \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ mit $\sum_{i=1}^{p} \theta_p = 1$

Jeffrey's prior für Multinomialverteilung

- $\boldsymbol{x} = (x_1, \ldots, x_p)\,,\ \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ mit $\sum_{i=1}^{p} \theta_p = 1$
- Multinomialverteilung $M_p(n, \boldsymbol{\theta})$:

$$f(\boldsymbol{x}) = \begin{cases} \frac{n!}{\prod_{i=1}^{p} x_i!} \prod_{i=1}^{p} \theta_i^{x_i} & \text{falls } \sum_{i=1}^{p} x_i = n \\ 0 & \text{sonst} \end{cases}$$

Jeffrey's prior für Multinomialverteilung

- $\boldsymbol{x} = (x_1, \ldots, x_p)\,,\ \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ mit $\sum_{i=1}^{p} \theta_p = 1$
- Multinomialverteilung $M_p(n, \boldsymbol{\theta})$:

$$f(\boldsymbol{x}) = \begin{cases} \frac{n!}{\prod_{i=1}^{p} x_i!} \prod_{i=1}^{p} \theta_i^{x_i} & \text{falls } \sum_{i=1}^{p} x_i = n \\ 0 & \text{sonst} \end{cases}$$

- Dirichlet-Verteilung $D(\alpha_1, \ldots, \alpha_p)$:

$$f(\boldsymbol{\theta}) \propto \prod_{i=1}^{p} \theta_i^{\alpha_i - 1}$$

Jeffrey's prior für Multinomialverteilung

- $\boldsymbol{x} = (x_1, \ldots, x_p)\,,\ \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ mit $\sum_{i=1}^p \theta_p = 1$
- Multinomialverteilung $M_p(n, \boldsymbol{\theta})$:

$$f(\boldsymbol{x}) = \begin{cases} \frac{n!}{\prod_{i=1}^p x_i!} \prod_{i=1}^p \theta_i^{x_i} & \text{falls } \sum_{i=1}^p x_i = n \\ 0 & \text{sonst} \end{cases}$$

- Dirichlet-Verteilung $D(\alpha_1, \ldots, \alpha_p)$:

$$f(\boldsymbol{\theta}) \propto \prod_{i=1}^p \theta_i^{\alpha_i - 1}$$

- Jeffrey's prior für $M_p(n, \boldsymbol{\theta})$:

$$\boldsymbol{\theta} \sim D\left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$$

KOALA

KOALA

- **Koal**itions-**A**nalyse
- Ansatz für Mehr-Parteien-Systeme mit besonderem Fokus auf Koalitionswahrscheinlichkeiten
- Meinungsbild (wenn die Wahl heute wäre)
- Monte-Carlo-Ansatz basierend auf einem Bayesschen Multinomial-Dirichlet Model
- Vorstellung neuer Visualisierungstechnik

KOALA

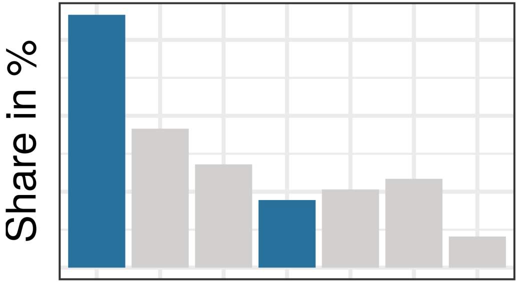


Abbildung: Stimmanteile der Parteien

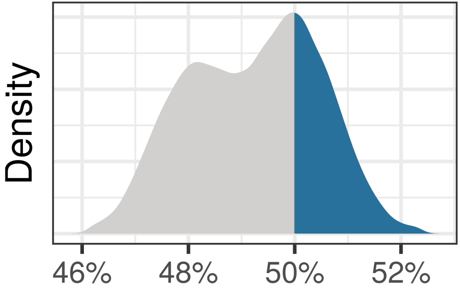
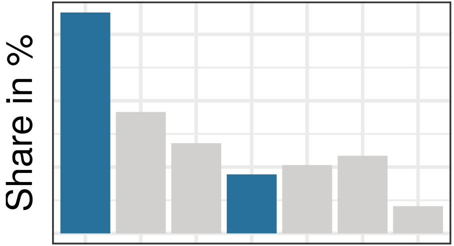


Abbildung: Sitzanteil für Koalition

KOALA

Klassische Wahlberichterstattung

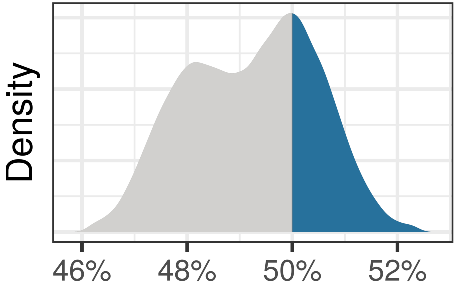
- Darstellung der Stimmanteile der Parteien
- Fokus auf Ergebnisse einzelner Parteien
- Unzureichende Vermittlung von Unsicherheit



KOALA

Vorgeschlagene Wahlberichtserstattung

- Darstellung der Wahrscheinlichkeit eines Ereignisses durch Dichteplots
- Fokus auf spezifischem Ereignis (z. B. Koalitionsbildung)
- Vermittlung der Unsicherheit und Bandbreite der Ergebnisse



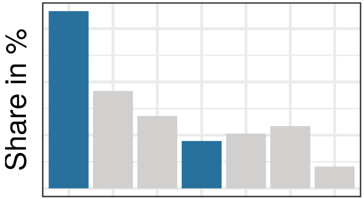
KOALA

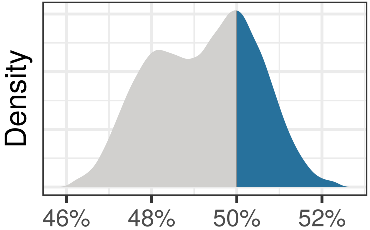
Typische Überschrift:

- „Union-FDP verliert die Mehrheit"

Vorgeschlagene Überschrift:

- „Union-FDP erreicht zu 26% Sitzmehrheit "
- „FDP zieht zu 51% ins Parlament ein "





Asymptotisches Verhalten

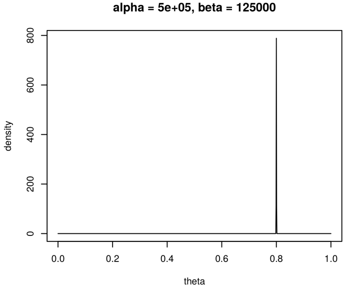
- Parametermenge $\Theta = \{\theta_0, \theta_1, \ldots\}$, θ_0 ist wahrer Parameter, $X_{1:n}$ Stichprobe von Verteilung mit Dichtefunktion $f(x|\theta)$, $\theta \in \Theta$. Dann gilt:

$$\lim_{n\to\infty} f(\theta_0|x_{1:n}) = 1 \quad \text{und} \quad \lim_{n\to\infty} f(\theta_i|x_{1:n}) = 0 \; \forall i \neq 0$$

Asymptotisches Verhalten

- Parametermenge $\Theta = \{\theta_0, \theta_1, \dots\}$, θ_0 ist wahrer Parameter, $X_{1:n}$ Stichprobe von Verteilung mit Dichtefunktion $f(x|\theta)$, $\theta \in \Theta$. Dann gilt:

$$\lim_{n\to\infty} f(\theta_0|x_{1:n}) = 1 \quad \text{und} \quad \lim_{n\to\infty} f(\theta_i|x_{1:n}) = 0 \; \forall i \neq 0$$



Empirische Bayes Methoden

- Wählen Priori Verteilung in Abhängigkeit von erhaltenen Daten

Empirische Bayes Methoden

- Wählen Priori Verteilung in Abhängigkeit von erhaltenen Daten
- Bsp.: Analyse der Trefferquoten von Baseballspielern:

Empirische Bayes Methoden

- Wählen Priori Verteilung in Abhängigkeit von erhaltenen Daten
- Bsp.: Analyse der Trefferquoten von Baseballspielern:
- Spieler A hat 4 von 10 Bällen getroffen, Spieler B hat 300 von 1000 Bällen getroffen. Welcher Spieler ist besser?

Empirische Bayes Methoden

- Wählen Priori Verteilung in Abhängigkeit von erhaltenen Daten
- Bsp.: Analyse der Trefferquoten von Baseballspielern:
- Spieler A hat 4 von 10 Bällen getroffen, Spieler B hat 300 von 1000 Bällen getroffen. Welcher Spieler ist besser?
- Trefferzahl $X \sim Bin(n, \theta)$, Treffwahrscheinlichkeit $\theta \sim Be(\alpha, \beta)$

Empirische Bayes Methoden

- Bilde Priori Verteilung durch Interpolation über Werte aller Baseballspieler

Empirische Bayes Methoden

- Bilde Priori Verteilung durch Interpolation über Werte aller Baseballspieler

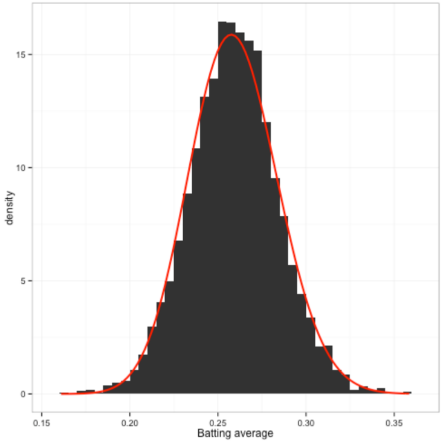


Abbildung: varianceexplained.org

Empirische Bayes Methoden

- Bilde Priori Verteilung durch Interpolation über Werte aller Baseballspieler
- $\Rightarrow \theta \sim Be(78,661\,;\,224,875)$

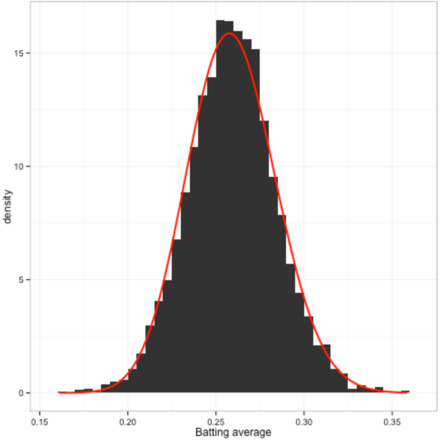


Abbildung: varianceexplained.org

Empirische Bayes Methoden

- Bilde Priori Verteilung durch Interpolation über Werte aller Baseballspieler
- $\Rightarrow \theta \sim Be(78,661\,;\,224,875)$
- Posteriori Trefferverteilung für Spieler A: $\theta|x \sim Be(82,661\,;\,230,875)$ hat Erwartungswert 0,264

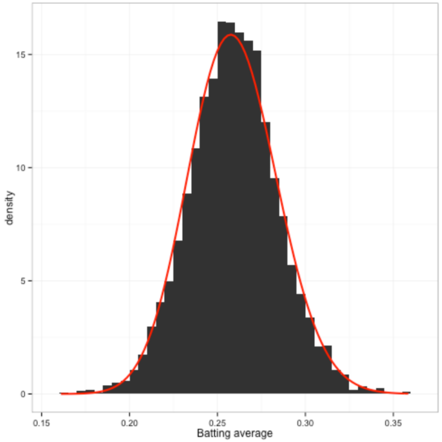


Abbildung: varianceexplained.org

Empirische Bayes Methoden

- Bilde Priori Verteilung durch Interpolation über Werte aller Baseballspieler
- $\Rightarrow \theta \sim Be(78,661\,;\,224,875)$
- Posteriori Trefferverteilung für Spieler A: $\theta|x \sim Be(82,661\,;\,230,875)$ hat Erwartungswert 0,264
- Posteriori Trefferverteilung für Spieler B: $\theta|x \sim Be(378,661\,;\,924,875)$ hat Erwartungswert 0,29

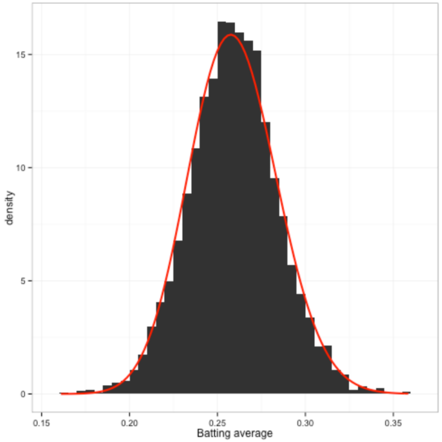


Abbildung: varianceexplained.org

Empirische Bayes Methoden

- Bilde Priori Verteilung durch Interpolation über Werte aller Baseballspieler
- $\Rightarrow \theta \sim Be(78,661\,;\,224,875)$
- Posteriori Trefferverteilung für Spieler A: $\theta|x \sim Be(82,661\,;\,230,875)$ hat Erwartungswert 0,264
- Posteriori Trefferverteilung für Spieler B: $\theta|x \sim Be(378,661\,;\,924,875)$ hat Erwartungswert 0,29
- Shrinkage

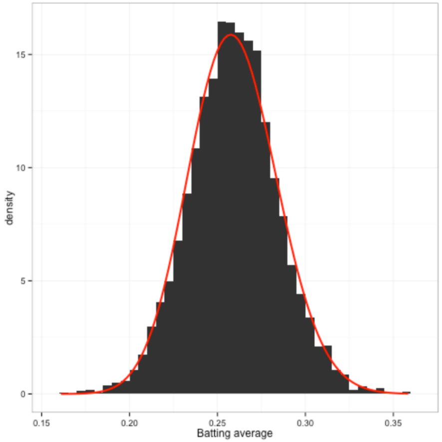


Abbildung: varianceexplained.org

Markow-Chain-Monte-Carlo-Verfahren

- Markow-Chain-Monte-Carlo-Algorithmen sind eine Klasse von Algorithmen zum Ziehen von Stichproben aus Wahrscheinlichkeitsverteilungen.
- Simulationsbasierte Bayes-Interferenz erfordert Ziehungen aus der Posteriori-Verteilung.
- Idee: Erzeugung einer Markow-Kette, deren stationäre Verteilung die Posteriori-Verteilung ist.

Idee: Markow-Kette

- Spezielle Klasse stochastischer Prozesse

Idee: Markow-Kette

- Spezielle Klasse stochastischer Prozesse
- Prozesse mit zyklischem Verhalten

Idee: Markow-Kette

- Spezielle Klasse stochastischer Prozesse
- Prozesse mit zyklischem Verhalten
- Es gibt eine gewisse Anzahl an diskreten Zuständen, zwischen denen der stochastische Prozess im Laufe des Prozesses wechselt

Idee: Markow-Kette

- Spezielle Klasse stochastischer Prozesse
- Prozesse mit zyklischem Verhalten
- Es gibt eine gewisse Anzahl an diskreten Zuständen, zwischen denen der stochastische Prozess im Laufe des Prozesses wechselt
- Beschreibung dessen über längeren Zeitraum ohne größeren Aufwand

Idee: Markow-Kette

- Spezielle Klasse stochastischer Prozesse
- Prozesse mit zyklischem Verhalten
- Es gibt eine gewisse Anzahl an diskreten Zuständen, zwischen denen der stochastische Prozess im Laufe des Prozesses wechselt
- Beschreibung dessen über längeren Zeitraum ohne größeren Aufwand
- Interessant für zukünftige Entwicklungen

Definition: Markow-Kette

Ein stochastischer Prozess $\{X_t, t \in X_0\}$ *mit abzählbaren Zustandsraum* E *heißt Markow-Kette genau dann, wenn*
$P(X_n + 1 = k | X_0 = e_0, X_1 = e_1, \ldots, X_n = e_n) = P(X_n + 1 = k | X_n = e_n)$ *für alle* $n \in \mathbb{N}$ *und alle* $k, e_0, \ldots, e_n \in E$.

Definition: Markow-Kette

Ein stochastischer Prozess $\{X_t, t \in X_0\}$ *mit abzählbaren Zustandsraum* E *heißt Markow-Kette genau dann, wenn* $P(X_n + 1 = k | X_0 = e_0, X_1 = e_1, \ldots, X_n = e_n) = P(X_n + 1 = k | X_n = e_n)$ *für alle* $n \in \mathbb{N}$ *und alle* $k, e_0, \ldots, e_n \in E$*.*

- Also: Wahrscheinlichkeit eines Übergangs zu Zustand $X_n + 1$ lediglich abhängig vom vorherigen Zustand X_n und nicht von früheren Zuständen.

Charakteristika: Markow-Ketten

- Zustandsraum.

Charakteristika: Markow-Ketten

- Zustandsraum.
- Anfangsverteilung. $\vec{p} = (p_1(0), p_2(0), \ldots)^T$

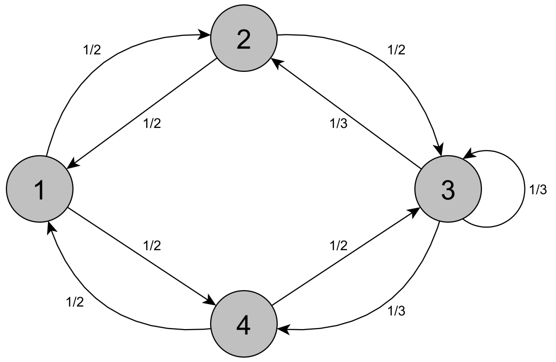
Charakteristika: Markow-Ketten

- Zustandsraum.
- Anfangsverteilung. $\vec{p} = (p_1(0), p_2(0), \dots)^T$
- Übergangswahrscheinlichkeiten. $P(e_i \to e_k) = P(e_i|e_k) = p_{ik}$

Charakteristika: Markow-Ketten

- Zustandsraum.
- Anfangsverteilung. $\vec{p} = (p_1(0), p_2(0), \dots)^T$
- Übergangswahrscheinlichkeiten. $P(e_i \to e_k) = P(e_i | e_k) = p_{ik}$
- Übergangsmatrix. $P = (p_{ik})_{i,k \in E}$

Beispiel: Markow-Ketten



$$\begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

Monte-Carlo-Simulation

- Verfahren zur Lösung von Problemen mit der Wahrscheinlichkeitstheorie

Monte-Carlo-Simulation

- Verfahren zur Lösung von Problemen mit der Wahrscheinlichkeitstheorie
- Basis: große Zahl gleichartiger Zufallsexperimente

Beispiel: Monte-Carlo-Simulation

- Bestimmung der Zahl π mit Monte-Carlo-Simulationen

Beispiel: Monte-Carlo-Simulation

- Bestimmung der Zahl π mit Monte-Carlo-Simulationen
- Erzeugung zufälliger Punkte aus dem Intervall $(x, y \mid x, y \in [-1, 1])$

Beispiel: Monte-Carlo-Simulation

- Bestimmung der Zahl π mit Monte-Carlo-Simulationen
- Erzeugung zufälliger Punkte aus dem Intervall $(x, y \mid x, y \in [-1, 1])$
- Überprüfung, ob diese im Einheitskreis liegen

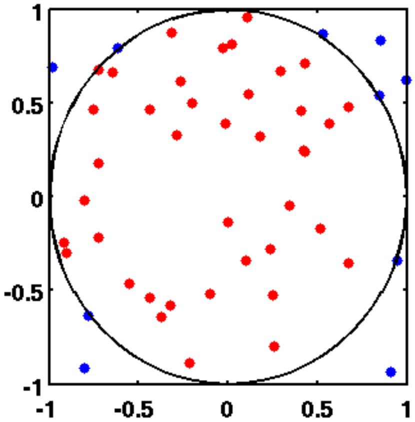
Beispiel: Monte-Carlo-Simulation

- Bestimmung der Zahl π mit Monte-Carlo-Simulationen
- Erzeugung zufälliger Punkte aus dem Intervall $(x, y \mid x, y \in [-1, 1])$
- Überprüfung, ob diese im Einheitskreis liegen
- Näherungsweise Bestimmung der Kreisfläche durch Verhältnis der Punkte innerhalb und außerhalb des Kreises

Beispiel: Monte-Carlo-Simulation

- Bestimmung der Zahl π mit Monte-Carlo-Simulationen
- Erzeugung zufälliger Punkte aus dem Intervall $(x, y \mid x, y \in [-1, 1])$
- Überprüfung, ob diese im Einheitskreis liegen
- Näherungsweise Bestimmung der Kreisfläche durch Verhältnis der Punkte innerhalb und außerhalb des Kreises
- $\frac{\pi}{4} = \frac{\text{Kreisfläche}}{\text{Quadratfläche}} = \frac{\text{Treffer in Kreisfläche}}{\text{generierte Punkte}}$

Beispiel: Monte-Carlo-Simulation



Idee: MCMC-Algorithmus

- Simulation einer Markow-Kette, die so gewählt ist, dass sie gegen die posteriori-Verteilung konvergiert.

Idee: MCMC-Algorithmus

- Simulation einer Markow-Kette, die so gewählt ist, dass sie gegen die posteriori-Verteilung konvergiert.
- Bei Konvergenz erhalten wir Stichproben der posteriori-Verteilung

Idee: MCMC-Algorithmus

- Simulation einer Markow-Kette, die so gewählt ist, dass sie gegen die posteriori-Verteilung konvergiert.
- Bei Konvergenz erhalten wir Stichproben der posteriori-Verteilung
- Allerdings: Diese Stichproben sind voneinander abhängig.

Gemeinsamkeiten: MCMC-Algorithmus

- Viele Designs des MCMC-Algorithmus.

Gemeinsamkeiten: MCMC-Algorithmus

- Viele Designs des MCMC-Algorithmus.
- In jedem Durchlauf m werden mit der sogenannten Vorschlagsverteilung potenzielle neue Zustände erzeugt.

Gemeinsamkeiten: MCMC-Algorithmus

- Viele Designs des MCMC-Algorithmus.
- In jedem Durchlauf m werden mit der sogenannten Vorschlagsverteilung potenzielle neue Zustände erzeugt.
- Diese sind abhängig vom aktuellen Zustand der Markow-Kette.

Konzept: Metropolis-Hastings-Algorithmus

- Akzeptanz des neuen Vorschlags mit der sogenannten Vorschlagswahrscheinlichkeit.

Konzept: Metropolis-Hastings-Algorithmus

- Akzeptanz des neuen Vorschlags mit der sogenannten Vorschlagswahrscheinlichkeit.
- Man kann zeigen: Unter bestimmten Voraussetzungen konvergiert dieser Algorithmus gegen die Zielverteilung.

Konzept: Metropolis-Hastings-Algorithmus

- Akzeptanz des neuen Vorschlags mit der sogenannten Vorschlagswahrscheinlichkeit.
- Man kann zeigen: Unter bestimmten Voraussetzungen konvergiert dieser Algorithmus gegen die Zielverteilung.
- Unabhängig von der Wahl der Vorschlagsverteilung.

Effizienz: Metropolis-Hastings-Algorithmus

- Konvergenzgeschwindigkeit und Abhängigkeit zw. Stichproben variiert stark mit der Vorschlagsverteilung.

Effizienz: Metropolis-Hastings-Algorithmus

- Konvergenzgeschwindigkeit und Abhängigkeit zw. Stichproben variiert stark mit der Vorschlagsverteilung.
- Effizienz entscheidend abhängig von der Akzeptanzrate.

Effizienz: Metropolis-Hastings-Algorithmus

- Konvergenzgeschwindigkeit und Abhängigkeit zw. Stichproben variiert stark mit der Vorschlagsverteilung.
- Effizienz entscheidend abhängig von der Akzeptanzrate.
- Oftmals empfohlen: Akzeptraten zw. $0,3$ und $0,5$

Praxis: MCMC-Algorithmus

- Wissen: MCMC-Algorithmus konvergiert gegen Zielverteilung

Praxis: MCMC-Algorithmus

- Wissen: MCMC-Algorithmus konvergiert gegen Zielverteilung
- Entscheidung: Wann ist die Zielverteilung erreicht?

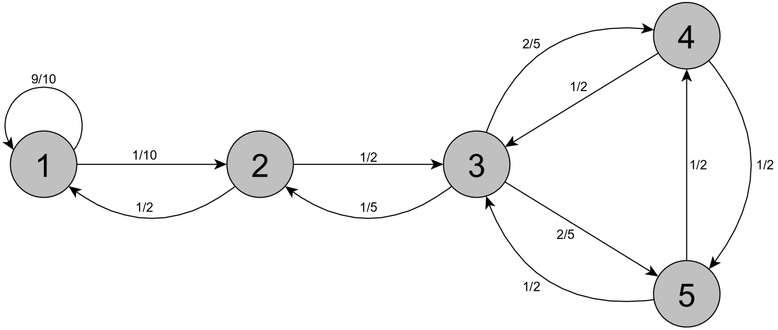
Praxis: MCMC-Algorithmus

- Wissen: MCMC-Algorithmus konvergiert gegen Zielverteilung
- Entscheidung: Wann ist die Zielverteilung erreicht?
- Burn-in-Phase. Nicht-Berücksichtigung der ersten Zustände.

Praxis: MCMC-Algorithmus

- Wissen: MCMC-Algorithmus konvergiert gegen Zielverteilung
- Entscheidung: Wann ist die Zielverteilung erreicht?
- Burn-in-Phase. Nicht-Berücksichtigung der ersten Zustände.
- Dann: Betrachtung der folgenden Werte als Stichprobe der Zielverteilung.

Burn-In-Phase: MCMC-Algorithmus



Vorhersagemodell

- Dynamisches Bayessches Vorhersagemodell für Mehr-Parteien-Wahlen
- Kombination aus Regressionsmodell und Bayesschen Modell
- Vorhersage von Stimmanteilen und weiterer Kennzahlen

Regressionsmodell

- Vorhersage der Stimmenanteile lange vor der Wahl
- Nutzung von Daten vorheriger Wahlen
- Vorhersager X sagt Ergebnisse V mit Parametern θ vorher

Bayessches Modell

- Vorhersage der aktuellen Stimmenanteile
- Nutzung von Daten aus aktuellen Umfragen

Bayessches Modell

- Vorhersage der aktuellen Stimmenanteile
- Nutzung von Daten aus aktuellen Umfragen
- Backward Random Walk: $\alpha_t = \alpha_{t+1} + \omega$

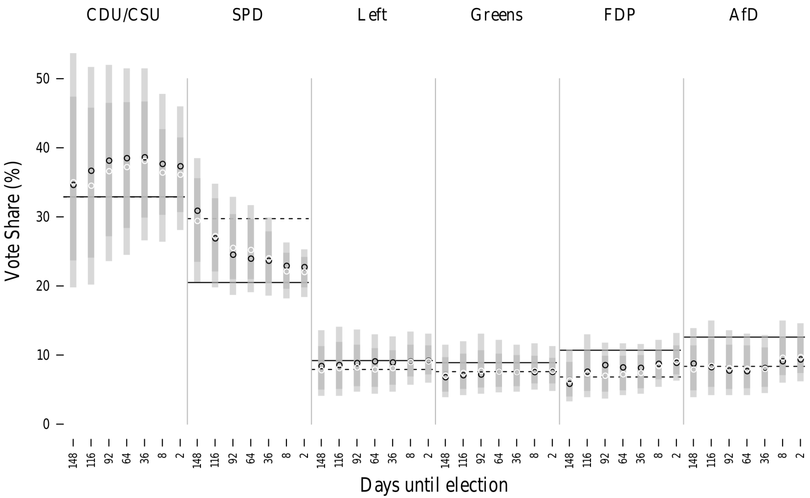
Bayessches Modell

- Vorhersage der aktuellen Stimmenanteile
- Nutzung von Daten aus aktuellen Umfragen
- Backward Random Walk: $\alpha_t = \alpha_{t+1} + \omega$
- Stimmanteile am Wahltag = Posterior predictive distribution des Regressionsmodells

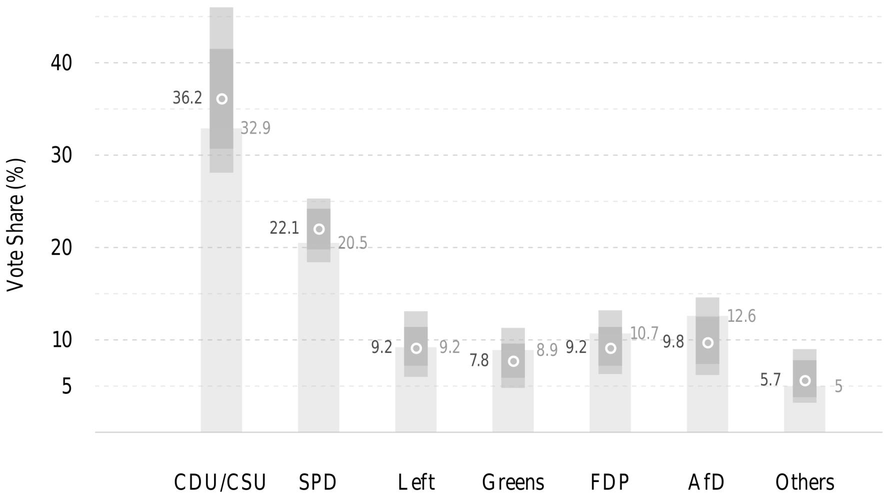
Bayessches Modell

- Vorhersage der aktuellen Stimmenanteile
- Nutzung von Daten aus aktuellen Umfragen
- Backward Random Walk: $\alpha_t = \alpha_{t+1} + \omega$
- Stimmanteile am Wahltag = Posterior predictive distribution des Regressionsmodells
- Update bei neuen Ereignissen

Vorhersagemodell



Vorhersagemodell



Bibliographie

- Bauer, Bender, Klima, Küchenhoff: KOALA: a new paradigm for election coverage. An opinion poll-based now-cast of probabilities of events in multi-party electoral systems.
- Held, Bové: Applied Statistical Inference
- Stoetzer, Neunhoeffer, Gschwend, Munzert, Sternberg: Forecasting elections in multiparty systems: A bayesian approach combining polls and fundamentals.