On the estimation of conditional probabilities

Marco Cattaneo
Department of Statistics, LMU Munich

WPMSIIP 2012, Munich, Germany
13 September 2012

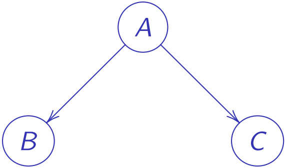
problem description

- given: **probabilistic model** $\{P_\theta : \theta \in \Theta\}$ for the random objects X and Y

problem description

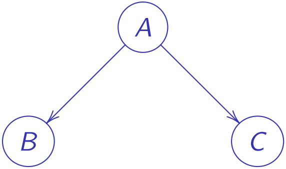
- given: **probabilistic model** $\{P_\theta : \theta \in \Theta\}$ for the random objects X and Y

- example:



with $X = A$ and $Y = (B, C)$

problem description

- given: **probabilistic model** $\{P_\theta : \theta \in \Theta\}$ for the random objects X and Y
- example: 

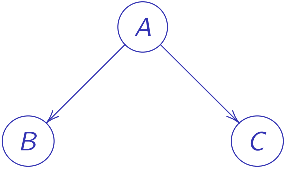
with $X = A$ and $Y = (B, C)$
- given: exchangeable/independent **observations**

$$\underbrace{(X_1, Y_1) = (x_1, y_1), \ldots (X_n, Y_n) = (x_n, y_n)}_{\mathcal{D}}, Y_0 = y_0$$

problem description

- given: **probabilistic model** $\{P_\theta : \theta \in \Theta\}$ for the random objects X and Y

- example:



with $X = A$ and $Y = (B, C)$

- given: exchangeable/independent **observations**

$$\underbrace{(X_1, Y_1) = (x_1, y_1), \ldots (X_n, Y_n) = (x_n, y_n)}_{\mathcal{D}}, Y_0 = y_0$$

- goal: **estimate** $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0)$

estimation of $P(X_0 = x \mid \mathcal{D},\ Y_0 = y_0)$

- **Bayesian** with prior π on θ:

estimation of $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0)$

- **Bayesian** with prior π on θ:
  - if π is a conjugate prior for the probabilistic model P_θ,
    then the posterior $\pi \,|\, \mathcal{D}$ can be easily calculated,
    while calculation of the posterior $\pi \,|\, (\mathcal{D},\ Y_0 = y_0)$ can be more difficult

estimation of $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0)$

- **Bayesian** with prior π on θ:
  - if π is a conjugate prior for the probabilistic model P_θ, then the posterior $\pi \,|\, \mathcal{D}$ can be easily calculated, while calculation of the posterior $\pi \,|\, (\mathcal{D},\ Y_0 = y_0)$ can be more difficult
  - $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0) =$

$$\frac{E_{\pi \,|\, \mathcal{D}}\left(P_\theta(X_0 = x,\ Y_0 = y_0)\right)}{E_{\pi \,|\, \mathcal{D}}\left(P_\theta(Y_0 = y_0)\right)} = E_{\pi \,|\, (\mathcal{D},\ Y_0 = y_0)}\left(P_\theta(X_0 = x \,|\, Y_0 = y_0)\right)$$

estimation of $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0)$

- **Bayesian** with prior π on θ:
  - if π is a conjugate prior for the probabilistic model P_θ,
    then the posterior $\pi \,|\, \mathcal{D}$ can be easily calculated,
    while calculation of the posterior $\pi \,|\, (\mathcal{D},\ Y_0 = y_0)$ can be more difficult
  - $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0) =$

$$\frac{E_{\pi \,|\, \mathcal{D}}\left(P_\theta(X_0 = x,\ Y_0 = y_0)\right)}{E_{\pi \,|\, \mathcal{D}}\left(P_\theta(Y_0 = y_0)\right)} = E_{\pi \,|\, (\mathcal{D},\ Y_0 = y_0)}\left(P_\theta(X_0 = x \,|\, Y_0 = y_0)\right)$$

- **maximum likelihood**:

estimation of $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0)$

- **Bayesian** with prior π on θ:
  - if π is a conjugate prior for the probabilistic model P_θ,
    then the posterior $\pi \,|\, \mathcal{D}$ can be easily calculated,
    while calculation of the posterior $\pi \,|\, (\mathcal{D},\ Y_0 = y_0)$ can be more difficult
  - $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0) =$

$$\frac{E_{\pi \,|\, \mathcal{D}}\left(P_\theta(X_0 = x,\ Y_0 = y_0)\right)}{E_{\pi \,|\, \mathcal{D}}\left(P_\theta(Y_0 = y_0)\right)} = E_{\pi \,|\, (\mathcal{D},\ Y_0 = y_0)}\left(P_\theta(X_0 = x \,|\, Y_0 = y_0)\right)$$

- **maximum likelihood**:
  - $\hat{P}(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0)$:

$$\frac{P_{\hat{\theta}_\mathcal{D}}(X_0 = x,\ Y_0 = y_0)}{P_{\hat{\theta}_\mathcal{D}}(Y_0 = y_0)} \neq P_{\hat{\theta}_{(\mathcal{D},\ Y_0 = y_0)}}(X_0 = x \,|\, Y_0 = y_0)$$

estimation of $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0)$

- **Bayesian** with prior π on θ:
  - if π is a conjugate prior for the probabilistic model P_θ, then the posterior $\pi \,|\, \mathcal{D}$ can be easily calculated, while calculation of the posterior $\pi \,|\, (\mathcal{D},\ Y_0 = y_0)$ can be more difficult
  - $P(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0) =$

$$\frac{E_{\pi \,|\, \mathcal{D}}\left(P_\theta(X_0 = x,\ Y_0 = y_0)\right)}{E_{\pi \,|\, \mathcal{D}}\left(P_\theta(Y_0 = y_0)\right)} = E_{\pi \,|\, (\mathcal{D},\ Y_0 = y_0)}\left(P_\theta(X_0 = x \,|\, Y_0 = y_0)\right)$$

- **maximum likelihood**:
  - $\hat{P}(X_0 = x \,|\, \mathcal{D},\ Y_0 = y_0)$:

$$\frac{P_{\hat{\theta}_{\mathcal{D}}}(X_0 = x,\ Y_0 = y_0)}{P_{\hat{\theta}_{\mathcal{D}}}(Y_0 = y_0)} \neq P_{\hat{\theta}_{(\mathcal{D},\ Y_0 = y_0)}}(X_0 = x \,|\, Y_0 = y_0)$$

  - likelihood functions $lik_{\mathcal{D}}(\theta) \propto P_\theta(\mathcal{D})$ and $lik_{(\mathcal{D},\ Y_0 = y_0)}(\theta) \propto P_\theta(\mathcal{D},\ Y_0 = y_0)$ on Θ have maxima at the points $\hat{\theta}_{\mathcal{D}}$ and $\hat{\theta}_{(\mathcal{D},\ Y_0 = y_0)}$, respectively

estimation of $P(X_0 = x \mid \mathcal{D},\ Y_0 = y_0)$

- **Bayesian** with prior π on θ:
  - if π is a conjugate prior for the probabilistic model P_θ,
    then the posterior $\pi \mid \mathcal{D}$ can be easily calculated,
    while calculation of the posterior $\pi \mid (\mathcal{D},\ Y_0 = y_0)$ can be more difficult
  - $P(X_0 = x \mid \mathcal{D},\ Y_0 = y_0) =$

$$\frac{E_{\pi \mid \mathcal{D}}\left(P_\theta(X_0 = x,\ Y_0 = y_0)\right)}{E_{\pi \mid \mathcal{D}}\left(P_\theta(Y_0 = y_0)\right)} = E_{\pi \mid (\mathcal{D},\ Y_0 = y_0)}\left(P_\theta(X_0 = x \mid Y_0 = y_0)\right)$$

- **maximum likelihood**:
  - $\hat{P}(X_0 = x \mid \mathcal{D},\ Y_0 = y_0)$:

$$P_{\hat{\theta}_{\mathcal{D}}}(X_0 = x \mid Y_0 = y_0) \neq P_{\hat{\theta}_{(\mathcal{D},\ Y_0 = y_0)}}(X_0 = x \mid Y_0 = y_0)$$

  - likelihood functions $lik_{\mathcal{D}}(\theta) \propto P_\theta(\mathcal{D})$ and $lik_{(\mathcal{D},\ Y_0 = y_0)}(\theta) \propto P_\theta(\mathcal{D},\ Y_0 = y_0)$
    on Θ have maxima at the points $\hat{\theta}_{\mathcal{D}}$ and $\hat{\theta}_{(\mathcal{D},\ Y_0 = y_0)}$, respectively

estimation of $P(X_0 = x \mid \mathcal{D},\ Y_0 = y_0)$

- **Bayesian** with prior π on θ:
  - if π is a conjugate prior for the probabilistic model P_θ, then the posterior $\pi \mid \mathcal{D}$ can be easily calculated, while calculation of the posterior $\pi \mid (\mathcal{D},\ Y_0 = y_0)$ can be more difficult
  - $P(X_0 = x \mid \mathcal{D},\ Y_0 = y_0) =$

$$\frac{E_{\pi \mid \mathcal{D}}\left(P_\theta(X_0 = x,\ Y_0 = y_0)\right)}{E_{\pi \mid \mathcal{D}}\left(P_\theta(Y_0 = y_0)\right)} = E_{\pi \mid (\mathcal{D},\ Y_0 = y_0)}\left(P_\theta(X_0 = x \mid Y_0 = y_0)\right)$$

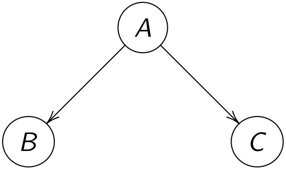
- **maximum likelihood**:
  - $\hat{P}(X_0 = x \mid \mathcal{D},\ Y_0 = y_0)$:

$$P_{\hat{\theta}_{\mathcal{D}}}(X_0 = x \mid Y_0 = y_0) \neq \boxed{P_{\hat{\theta}_{(\mathcal{D},\ Y_0 = y_0)}}(X_0 = x \mid Y_0 = y_0)}$$

  - likelihood functions $lik_{\mathcal{D}}(\theta) \propto P_\theta(\mathcal{D})$ and $lik_{(\mathcal{D},\ Y_0 = y_0)}(\theta) \propto P_\theta(\mathcal{D},\ Y_0 = y_0)$ on Θ have maxima at the points $\hat{\theta}_{\mathcal{D}}$ and $\hat{\theta}_{(\mathcal{D},\ Y_0 = y_0)}$, respectively

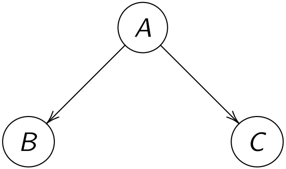
example

$A, B, C \in \{0, 1\}$



example

$A, B, C \in \{0, 1\}$

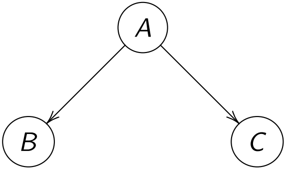


$\mathcal{D}$ ($n = 100$):

A	B	C	#
0	0	0	0
0	0	1	49
0	1	0	0
0	1	1	1
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

example

$A, B, C \in \{0, 1\}$



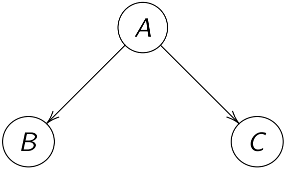
$\mathcal{D}$ $(n = 100)$:

A	B	C	#
0	0	0	0
0	0	1	49
0	1	0	0
0	1	1	1
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:

example

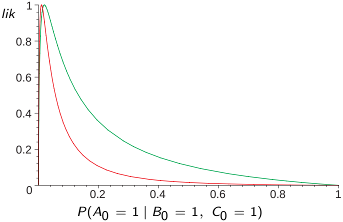
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ ($n = 100$):

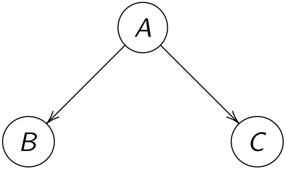
A	B	C	#
0	0	0	0
0	0	1	49
0	1	0	0
0	1	1	1
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



example

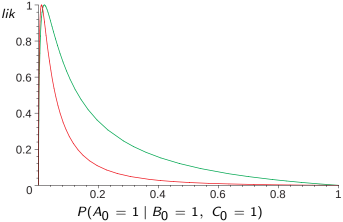
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ $(n = 100)$:

A	B	C	#
0	0	0	0
0	0	1	49
0	1	0	0
0	1	1	1
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

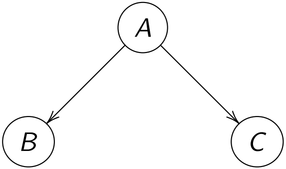
estimation of $P(A_0 = 1 \,|\, \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \,|\, \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.038$

example

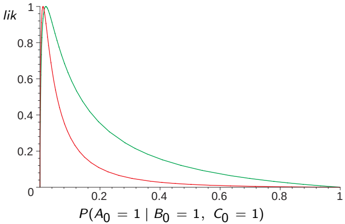
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ $(n = 100)$:

A	B	C	#
0	0	0	0
0	0	1	49
0	1	0	0
0	1	1	1
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

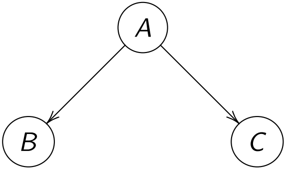
estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.038$
- **maximum likelihood** $lik_{(\mathcal{D},\ B_0=1,\ C_0=1)}$:
  $P_{\hat{\theta}_{(\mathcal{D},\ B_0=1,\ C_0=1)}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) \approx 0.010$

example

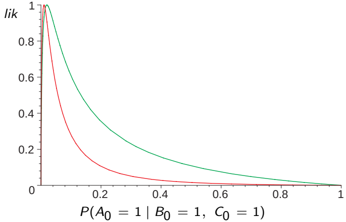
$A, B, C \in \{0,1\}$



$\mathcal{D}$ $(n = 100)$:

A	B	C	#
0	0	0	0
0	0	1	49
0	1	0	0
0	1	1	1
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

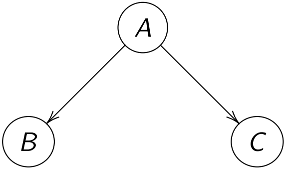
estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.038$
- **maximum likelihood** $lik_{(\mathcal{D},\ B_0=1,\ C_0=1)}$:
  $P_{\hat{\theta}_{(\mathcal{D},\ B_0=1,\ C_0=1)}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) \approx 0.010$
- **maximum likelihood** $lik_{\mathcal{D}}$:
  $P_{\hat{\theta}_{\mathcal{D}}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) \approx 0.020$

example

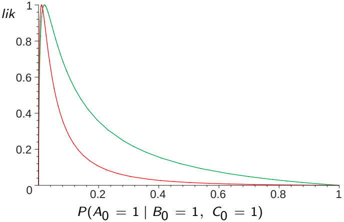
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ $(n = 100)$:

A	B	C	#
0	0	0	0
0	0	1	49
0	1	0	0
0	1	1	1
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

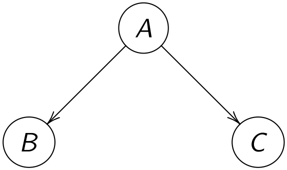
estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.038$
- **maximum likelihood** $lik_{(\mathcal{D},\ B_0=1,\ C_0=1)}$:
  $P_{\hat{\theta}_{(\mathcal{D},\ B_0=1,\ C_0=1)}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) \approx 0.010$
- **maximum likelihood** $lik_{\mathcal{D}}$:
  $P_{\hat{\theta}_{\mathcal{D}}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) \approx 0.020$
- **imprecise Bayesian** with IDM_2 priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx [0.0066, 0.15]$

example

$A, B, C \in \{0, 1\}$

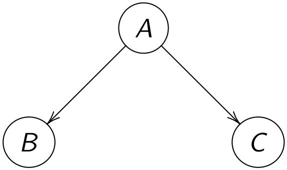


$\mathcal{D}$ ($n = 100$):

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

example

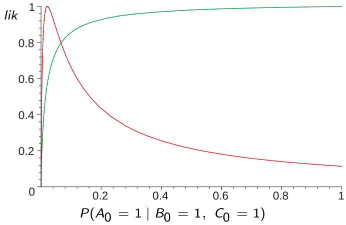
$A, B, C \in \{0,1\}$



$\mathcal{D}$ ($n = 100$):

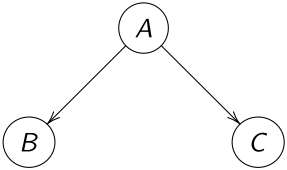
A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



example

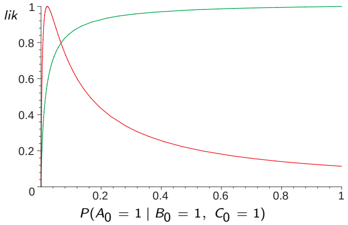
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ ($n = 100$):

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

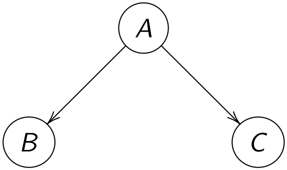
estimation of $P(A_0 = 1 \,|\, \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \,|\, \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.073$

example

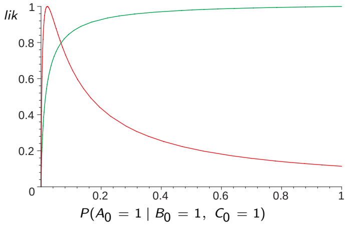
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ ($n = 100$):

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

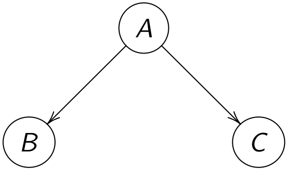
estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.073$
- **maximum likelihood** $lik_{(\mathcal{D},\ B_0=1,\ C_0=1)}$:
  $P_{\hat{\theta}_{(\mathcal{D},\ B_0=1,\ C_0=1)}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) \approx 0.021$

example

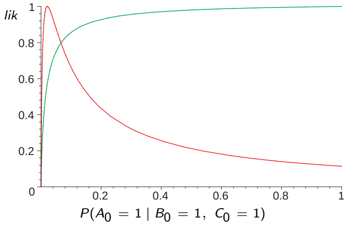
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ $(n = 100)$:

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

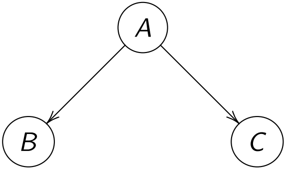
estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.073$
- **maximum likelihood** $lik_{(\mathcal{D},\ B_0=1,\ C_0=1)}$:
  $P_{\hat{\theta}_{(\mathcal{D},\ B_0=1,\ C_0=1)}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) \approx 0.021$
- **maximum likelihood** $lik_{\mathcal{D}}$:
  $P_{\hat{\theta}_{\mathcal{D}}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) = 1$

example

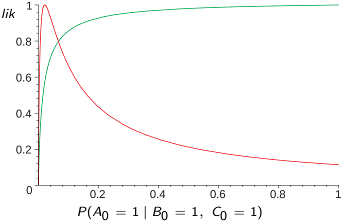
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ ($n = 100$):

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	48
1	0	1	1
1	1	0	1
1	1	1	0
			100

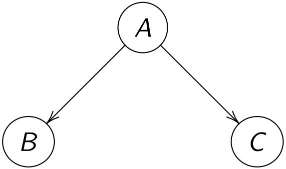
estimation of $P(A_0 = 1 \,|\, \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \,|\, \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.073$
- **maximum likelihood** $lik_{(\mathcal{D},\ B_0=1,\ C_0=1)}$:
  $P_{\hat{\theta}_{(\mathcal{D},\ B_0=1,\ C_0=1)}}(A_0 = 1 \,|\, B_0 = 1,\ C_0 = 1) \approx 0.021$
- **maximum likelihood** $lik_{\mathcal{D}}$:
  $P_{\hat{\theta}_{\mathcal{D}}}(A_0 = 1 \,|\, B_0 = 1,\ C_0 = 1) = 1$
- **imprecise Bayesian** with IDM_2 priors:
  $P(A_0 = 1 \,|\, \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx [0.0099, 1]$

example

$A, B, C \in \{0, 1\}$

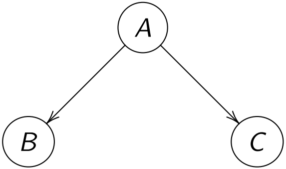


$\mathcal{D}$ ($n = 100$):

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	49
1	0	1	1
1	1	0	0
1	1	1	0
			100

example

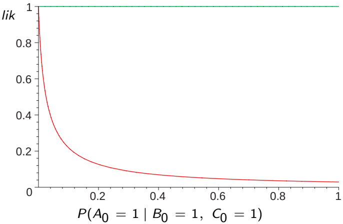
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ $(n = 100)$:

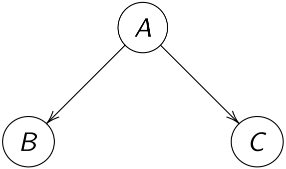
A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	49
1	0	1	1
1	1	0	0
1	1	1	0
			100

estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



example

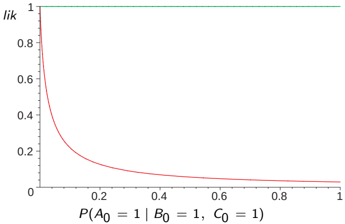
$A, B, C \in \{0,1\}$



$\mathcal{D}$ ($n = 100$):

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	49
1	0	1	1
1	1	0	0
1	1	1	0
			100

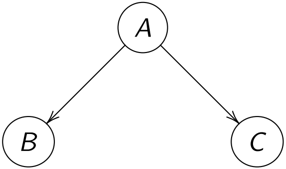
estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.038$

example

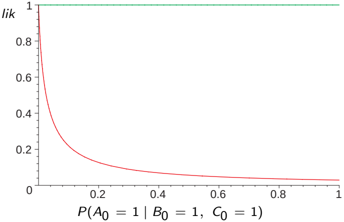
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ $(n = 100)$:

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	49
1	0	1	1
1	1	0	0
1	1	1	0
			100

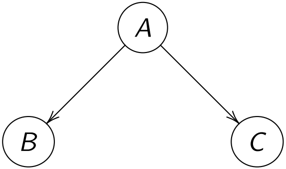
estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.038$
- **maximum likelihood** $lik_{(\mathcal{D},\ B_0=1,\ C_0=1)}$:
  $P_{\hat{\theta}_{(\mathcal{D},\ B_0=1,\ C_0=1)}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) = 0$

example

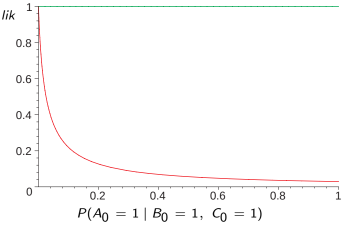
$A, B, C \in \{0, 1\}$



$\mathcal{D}$ $(n = 100)$:

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	49
1	0	1	1
1	1	0	0
1	1	1	0
			100

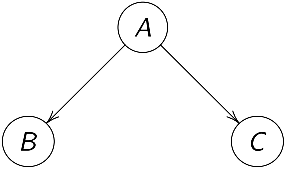
estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.038$
- **maximum likelihood** $lik_{(\mathcal{D},\ B_0=1,\ C_0=1)}$:
  $P_{\hat{\theta}_{(\mathcal{D},\ B_0=1,\ C_0=1)}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) = 0$
- **maximum likelihood** $lik_{\mathcal{D}}$:
  $P_{\hat{\theta}_{\mathcal{D}}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) = [0, 1]$

example

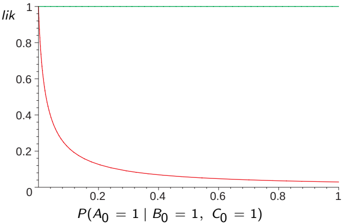
$A, B, C \in \{0,1\}$



$\mathcal{D}$ ($n = 100$):

A	B	C	#
0	0	0	0
0	0	1	50
0	1	0	0
0	1	1	0
1	0	0	49
1	0	1	1
1	1	0	0
1	1	1	0
			100

estimation of $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1)$:



- **Bayesian** with uniform priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) \approx 0.038$
- **maximum likelihood** $lik_{(\mathcal{D},\ B_0=1,\ C_0=1)}$:
  $P_{\hat{\theta}_{(\mathcal{D},\ B_0=1,\ C_0=1)}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) = 0$
- **maximum likelihood** $lik_{\mathcal{D}}$:
  $P_{\hat{\theta}_{\mathcal{D}}}(A_0 = 1 \mid B_0 = 1,\ C_0 = 1) = [0, 1]$
- **imprecise Bayesian** with IDM_2 priors:
  $P(A_0 = 1 \mid \mathcal{D},\ B_0 = 1,\ C_0 = 1) = [0, 1]$

references

- Cattaneo (2010). **Likelihood-based inference for probabilistic graphical models: Some preliminary results**. In: *PGM 2010, Proceedings of the Fifth European Workshop on Probabilistic Graphical Models*, HIIT Publications, pp. 57–64.
- Antonucci, Cattaneo, and Corani (2011). **Likelihood-based naive credal classifier**. In: *ISIPTA '11, Proceedings of the Seventh International Symposium on Imprecise Probability: Theories and Applications*, SIPTA, pp. 21–30.
- Antonucci, Cattaneo, and Corani (2012). **Likelihood-based robust classification with Bayesian networks**. In: *Advances in Computational Intelligence*, Part 3, Springer, pp. 491–500.