Learning from soft evidence

A mathematical view on Jeffrey's rule and Pearl's rule

Wietze Koops

January 2022

Overview

- Updating beliefs
- Soft evidence
- Mathematical framework
- Jeffrey's rule and Pearl's rule
- Example
- Properties of Jeffrey's rule and Pearl's rule
- Open questions and conclusion

Updating beliefs

- We form beliefs, e.g. about the color of a car.
- Formally, a belief is a probability distribution.
- When learning new evidence, we update our beliefs.

Updating beliefs

- We form beliefs, e.g. about the color of a car.
- Formally, a belief is a probability distribution.
- When learning new evidence, we update our beliefs.

- Hard evidence: A statement that some event happened with certainty, e.g. 'The car is red.'
- Soft evidence: A statement that some event happened with some uncertainty, e.g. 'I'm 70% sure the car is red.'

Soft evidence

- Soft evidence: A statement that some event happened with some uncertainty, e.g. 'I'm 70% sure the car is red.'
- Two ways to deal with soft evidence, giving very different results:
  - Jeffrey's rule (1965)
  - Pearl's rule (1988)
- Main challenges:
  - Common mathematical framework
  - When to use which rule?
- Studied by Jacobs (2019) and Jacobs (2021).

States and channels

- State $\omega \in \mathcal{D}(X)$: probability distribution over X

$$\omega = r_1|x_1\rangle + \cdots + r_n|x_n\rangle$$

where $x_i \in X$, $r_i \in [0,1]$ and $\sum_i r_i = 1$.
We also write $\omega(x_i) = r_i$.

States and channels

- State $\omega \in \mathcal{D}(X)$: probability distribution over X

$$\omega = r_1|x_1\rangle + \cdots + r_n|x_n\rangle$$

  where $x_i \in X$, $r_i \in [0,1]$ and $\sum_i r_i = 1$.
  We also write $\omega(x_i) = r_i$.
- Channel: A function $c \colon X \to \mathcal{D}(Y)$.
  We also write $c \colon X \rightsquigarrow Y$.

States and channels

- State $\omega \in \mathcal{D}(X)$: probability distribution over X

$$\omega = r_1|x_1\rangle + \cdots + r_n|x_n\rangle$$

  where $x_i \in X$, $r_i \in [0,1]$ and $\sum_i r_i = 1$.
  We also write $\omega(x_i) = r_i$.
- Channel: A function $c\colon X \to \mathcal{D}(Y)$.
  We also write $c\colon X \rightsquigarrow Y$.
- State transformation: given $\omega \in \mathcal{D}(X)$ and $c\colon X \rightsquigarrow Y$.
  The predicted state $c \gg \omega \in \mathcal{D}(Y)$ is given by

$$(c \gg \omega)(y) = \sum_x \omega(x) \cdot c(x)(y).$$

  This is the 'law of total probability'.

Example: burglar and alarm

Each night, there is a 0.1% chance that a burglar will break into my house. If a burglar is in my house, the alarm goes off with 80% probability. If there is no burglar, the alarm goes off with a 1% probability.

Example: burglar and alarm

Each night, there is a 0.1% chance that a burglar will break into my house. If a burglar is in my house, the alarm goes off with 80% probability. If there is no burglar, the alarm goes off with a 1% probability.

- Let $X = \{b, b^{\perp}\}$: there is a burglar (b) or not ($b^{\perp}$).
- Let $Y = \{a, a^{\perp}\}$: the alarm goes off (a) or not ($a^{\perp}$).

Example: burglar and alarm

Each night, there is a 0.1% chance that a burglar will break into my house. If a burglar is in my house, the alarm goes off with 80% probability. If there is no burglar, the alarm goes off with a 1% probability.

- Let $X = \{b, b^\perp\}$: there is a burglar (b) or not ($b^\perp$).
- Let $Y = \{a, a^\perp\}$: the alarm goes off (a) or not ($a^\perp$).
- State: $\omega = 0.001|b\rangle + 0.999|b^\perp\rangle \in \mathcal{D}(X)$.

Example: burglar and alarm

Each night, there is a 0.1% chance that a burglar will break into my house. If a burglar is in my house, the alarm goes off with 80% probability. If there is no burglar, the alarm goes off with a 1% probability.

- Let $X = \{b, b^\perp\}$: there is a burglar (b) or not ($b^\perp$).
- Let $Y = \{a, a^\perp\}$: the alarm goes off (a) or not ($a^\perp$).
- State: $\omega = 0.001|b\rangle + 0.999|b^\perp\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \rightarrow\!\!\circ\, Y$, $\begin{cases} c(b) = 0.8|a\rangle + 0.2|a^\perp\rangle \\ c(b^\perp) = 0.01|a\rangle + 0.99|a^\perp\rangle. \end{cases}$
- Gives the probability distribution over the states of the alarm based on whether or not there is a burglar.

Example: burglar and alarm

- Let $X = \{b, b^{\perp}\}$: there is a burglar (b) or not ($b^{\perp}$).
- Let $Y = \{a, a^{\perp}\}$: the alarm goes off (a) or not ($a^{\perp}$).
- State: $\omega = 0.001|b\rangle + 0.999|b^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \rightsquigarrow Y$, $\begin{cases} c(b) = 0.8|a\rangle + 0.2|a^{\perp}\rangle \\ c(b^{\perp}) = 0.01|a\rangle + 0.99|a^{\perp}\rangle. \end{cases}$

Example: burglar and alarm

- Let $X = \{b, b^\perp\}$: there is a burglar (b) or not ($b^\perp$).
- Let $Y = \{a, a^\perp\}$: the alarm goes off (a) or not ($a^\perp$).
- State: $\omega = 0.001|b\rangle + 0.999|b^\perp\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \multimap Y$, $\begin{cases} c(b) = 0.8|a\rangle + 0.2|a^\perp\rangle \\ c(b^\perp) = 0.01|a\rangle + 0.99|a^\perp\rangle. \end{cases}$
- Predicted state:

$$\begin{aligned}(c \gg \omega)(a) &= \omega(b)c(b)(a) + \omega(b^\perp)c(b^\perp)(a) \\ &= 0.001 \cdot 0.8 + 0.999 \cdot 0.01 = 0.01079.\end{aligned}$$

Similarly, $(c \gg \omega)(a^\perp) = 0.98921$.
So $c \gg \omega = 0.01079|a\rangle + 0.98921|a^\perp\rangle \in \mathcal{D}(Y)$.

Fuzzy predicates

- Fuzzy predicate $p\colon X \to [0,1]$: assigns to each $x \in X$ a measure to what extend some property p is true.

Fuzzy predicates

- Fuzzy predicate $p\colon X \to [0,1]$: assigns to each $x \in X$ a measure to what extend some property p is true.
- Validity of predicate p in state $\sigma \in \mathcal{D}(X)$:

$$\sigma \models p = \sum_x \sigma(x) \cdot p(x).$$

This can also be seen as an 'expected value'.

Fuzzy predicates

- Fuzzy predicate $p\colon X \to [0,1]$: assigns to each $x \in X$ a measure to what extend some property p is true.
- Validity of predicate p in state $\sigma \in \mathcal{D}(X)$:

$$\sigma \models p = \sum_x \sigma(x) \cdot p(x).$$

  This can also be seen as an 'expected value'.
- Updated state $\sigma|_p \in \mathcal{D}(X)$, given that predicate p holds:

$$\sigma|_p(x) = \frac{\sigma(x) \cdot p(x)}{\sigma \models p}.$$

Predicate transformation

- Predicates can also be transformed over a channel.
- In the opposite direction: given a channel $c\colon X \multimap\!\!\!\to Y$ and a predicate $q\colon Y \to [0,1]$, define a predicate $c \ll q\colon X \to [0,1]$:

$$(c \ll q)(x) = \sum_y c(x)(y) \cdot q(y).$$

Predicate transformation

- Predicates can also be transformed over a channel.
- In the opposite direction: given a channel $c\colon X \multimap\!\!\rightarrow Y$ and a predicate $q\colon Y \to [0,1]$, define a predicate $c \ll q\colon X \to [0,1]$:

$$(c \ll q)(x) = \sum_y c(x)(y) \cdot q(y).$$

Lemma

Let $c\colon X \multimap\!\!\rightarrow Y$ be a channel, let $q\colon Y \to [0,1]$ be a predicate and let $\omega \in \mathcal{D}(X)$. Then $(c \gg \omega) \models q = \omega \models (c \ll q)$.

Predicate and state transformation

Lemma

Let $c\colon X \rightsquigarrow Y$ *be a channel, let* $q\colon Y \to [0,1]$ *be a predicate and let* $\omega \in \mathcal{D}(X)$. *Then* $(c \gg \omega) \models q = \omega \models (c \ll q)$.

Predicate and state transformation

Lemma

Let $c\colon X \multimap\!\!\rightarrow Y$ *be a channel, let* $q\colon Y \to [0,1]$ *be a predicate and let* $\omega \in \mathcal{D}(X)$. *Then* $(c \gg \omega) \vDash q = \omega \vDash (c \ll q)$.

Proof.

$$(c \gg \omega) \vDash q \;=\; \sum_y (c \gg \omega)(y) q(y)$$

Predicate and state transformation

Lemma

Let $c\colon X \multimap\!\!\rightarrow Y$ *be a channel, let* $q\colon Y \to [0,1]$ *be a predicate and let* $\omega \in \mathcal{D}(X)$*. Then* $(c \gg \omega) \models q = \omega \models (c \ll q)$*.*

Proof.

$$\begin{aligned}(c \gg \omega) \models q &= \sum_y (c \gg \omega)(y) q(y) \\ &= \sum_y \left(\sum_x \omega(x) \cdot c(x)(y) \right) q(y)\end{aligned}$$

Predicate and state transformation

Lemma

Let $c\colon X \to\!\!\circ\!\!\to Y$ *be a channel, let* $q\colon Y \to [0,1]$ *be a predicate and let* $\omega \in \mathcal{D}(X)$*. Then* $(c \gg \omega) \models q = \omega \models (c \ll q)$*.*

Proof.

$$\begin{aligned}(c \gg \omega) \models q &= \sum_y (c \gg \omega)(y) q(y) \\ &= \sum_y \left(\sum_x \omega(x) \cdot c(x)(y) \right) q(y) \\ &= \sum_x \left(\sum_y c(x)(y) \cdot q(y) \right) \omega(x)\end{aligned}$$

Predicate and state transformation

Lemma

Let $c\colon X \to\!\!\circ\!\!\to Y$ be a channel, let $q\colon Y \to [0,1]$ be a predicate and let $\omega \in \mathcal{D}(X)$. Then $(c \gg \omega) \models q = \omega \models (c \ll q)$.

Proof.

$$\begin{aligned}
(c \gg \omega) \models q &= \sum_y (c \gg \omega)(y) q(y) \\
&= \sum_y \left(\sum_x \omega(x) \cdot c(x)(y) \right) q(y) \\
&= \sum_x \left(\sum_y c(x)(y) \cdot q(y) \right) \omega(x) \\
&= \sum_x (c \ll q)(x) \omega(x) = \omega \models (c \ll q).
\end{aligned}$$

Bayesian inversion

- Given a (prior) state $\sigma \in \mathcal{D}(X)$ and a channel $c \colon X \rightsquigarrow Y$, it is possible to define an inverted channel $c^\dagger_\sigma \colon Y \rightsquigarrow X$:

$$c^\dagger_\sigma(y) = \sum_x \frac{\sigma(x) \cdot c(x)(y)}{(c \gg \sigma)(y)} |x\rangle.$$

- Let $\mathbf{1}_y \colon Y \to [0,1]$ be a predicate on satisfying

$$\mathbf{1}_y(y') = \begin{cases} 1 & \text{if } y' = y \\ 0 & \text{otherwise.} \end{cases}$$

Then $c^\dagger_\sigma(y) = \sigma|_{c \ll \mathbf{1}_y}$.

Example: burglar and alarm

- Let $X = \{b, b^{\perp}\}$ and $Y = \{a, a^{\perp}\}$.
- State: $\omega = 0.001|b\rangle + 0.999|b^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \multimap Y$, $\begin{cases} c(b) = 0.8|a\rangle + 0.2|a^{\perp}\rangle \\ c(b^{\perp}) = 0.01|a\rangle + 0.99|a^{\perp}\rangle. \end{cases}$
- Predicted state: $c \gg \omega = 0.01079|a\rangle + 0.98921|a^{\perp}\rangle \in \mathcal{D}(Y)$.

Example: burglar and alarm

- Let $X = \{b, b^\perp\}$ and $Y = \{a, a^\perp\}$.
- State: $\omega = 0.001|b\rangle + 0.999|b^\perp\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \rightsquigarrow Y$, $\begin{cases} c(b) = 0.8|a\rangle + 0.2|a^\perp\rangle \\ c(b^\perp) = 0.01|a\rangle + 0.99|a^\perp\rangle. \end{cases}$
- Predicted state: $c \gg \omega = 0.01079|a\rangle + 0.98921|a^\perp\rangle \in \mathcal{D}(Y)$.
- Inverted channel $c^\dagger_\omega\colon Y \rightsquigarrow X$:

$$\begin{aligned} c^\dagger_\omega(a) &= \frac{\omega(b)\cdot c(b)(a)}{(c \gg \omega)(a)}|b\rangle + \frac{\omega(b^\perp)\cdot c(b^\perp)(a)}{(c \gg \omega)(a)}|b^\perp\rangle \\ &= \tfrac{0.001\cdot 0.8}{0.01079}|b\rangle + \tfrac{0.999\cdot 0.01}{0.01079}|b^\perp\rangle \\ &\approx 0.0741|b\rangle + 0.9259|b^\perp\rangle. \end{aligned}$$

Similarly, $c^\dagger_\omega(a^\perp) \approx 0.0002|b\rangle + 0.9998|b^\perp\rangle$.

Jeffrey's rule and Pearl's rule

- Let $c\colon X \multimap Y$ be a channel with prior state $\sigma \in \mathcal{D}(X)$.
- For Jeffrey's rule, consider evidence in the form of a state $\rho \in \mathcal{D}(Y)$. Then Jeffrey's update is given by

$$c_\sigma^\dagger \gg \rho \in \mathcal{D}(X).$$

Jeffrey's rule and Pearl's rule

- Let $c \colon X \rightsquigarrow Y$ be a channel with prior state $\sigma \in \mathcal{D}(X)$.
- For Jeffrey's rule, consider evidence in the form of a state $\rho \in \mathcal{D}(Y)$. Then Jeffrey's update is given by

$$c_\sigma^\dagger \gg \rho \in \mathcal{D}(X).$$

- For Pearl's rule, consider evidence in the form of a predicate $q \colon Y \to [0, 1]$. Then Pearl's update is given by

$$\sigma|_{c \ll q} \in \mathcal{D}(X).$$

Interpretation of Jeffrey's and Pearl's rule

- Jeffrey's rule assumes that evidence is of the type 'all things considered', i.e. taking into account other evidence.
- Jeffrey's rule can be interpreted as a correction.

- Pearl's rule assumes that evidence is of the type 'nothing else considered', i.e. not taking into account other evidence.
- Pearl's rule can be interpreted as an improvement.

Example: soft evidence

Around 3% of the cars in a city is red. One day, a car drives way too fast through the city center. The police asks for the car color. A witness: I'm 70% sure the car is red.

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \multimap X$ defined by $c(x) = 1|x\rangle$.

Example: soft evidence

Around 3% of the cars in a city is red. One day, a car drives way too fast through the city center. The police asks for the car color. A witness: I'm 70% sure the car is red.

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \to X$ defined by $c(x) = 1|x\rangle$.
- Pearl's rule: predicate $q\colon X \to [0,1]$ defined by $q(r) = \frac{7}{10}$ and $q(r^{\perp}) = \frac{3}{10}$. To compute: $\omega|_{c \ll q}$.

Example: soft evidence

Around 3% of the cars in a city is red. One day, a car drives way too fast through the city center. The police asks for the car color. A witness: I'm 70% sure the car is red.

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \multimap X$ defined by $c(x) = 1|x\rangle$.
- Pearl's rule: predicate $q\colon X \to [0,1]$ defined by $q(r) = \frac{7}{10}$ and $q(r^{\perp}) = \frac{3}{10}$. To compute: $\omega|_{c \ll q}$.
- We have

$$(c \ll q)(x) = \sum_y c(x)(y) \cdot q(y) = \sum_y \delta_{xy} \cdot q(y) = q(x),$$

so $c \ll q = q$.

Example: soft evidence

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \rightarrow\!\!\circ\!\!\rightarrow X$ defined by $c(x) = 1|x\rangle$.
- Pearl's rule: predicate $q\colon X \to [0,1]$ defined by $q(r) = \frac{7}{10}$ and $q(r^{\perp}) = \frac{3}{10}$. To compute: $\omega|_{c \ll q}$.
- We have $c \ll q = q$, so $\omega|_{c \ll q}(x) = \omega|_{q}(x) = \frac{\omega(x) q(x)}{\omega \models q}$.

Example: soft evidence

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \multimap X$ defined by $c(x) = 1|x\rangle$.
- Pearl's rule: predicate $q\colon X \to [0,1]$ defined by $q(r) = \frac{7}{10}$ and $q(r^{\perp}) = \frac{3}{10}$. To compute: $\omega|_{c \ll q}$.
- We have $c \ll q = q$, so $\omega|_{c \ll q}(x) = \omega|_{q}(x) = \frac{\omega(x) q(x)}{\omega \models q}$.
- $\omega \models q = \omega(r)q(r) + \omega(r^{\perp})q(r^{\perp}) = \frac{3}{100}\frac{7}{10} + \frac{97}{100}\frac{3}{10} = \frac{312}{1000}$.

Example: soft evidence

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \multimap X$ defined by $c(x) = 1|x\rangle$.
- Pearl's rule: predicate $q\colon X \to [0,1]$ defined by $q(r) = \frac{7}{10}$ and $q(r^{\perp}) = \frac{3}{10}$. To compute: $\omega|_{c \ll q}$.
- We have $c \ll q = q$, so $\omega|_{c \ll q}(x) = \omega|_{q}(x) = \frac{\omega(x)q(x)}{\omega \models q}$.
- $\omega \models q = \omega(r)q(r) + \omega(r^{\perp})q(r^{\perp}) = \frac{3}{100}\frac{7}{10} + \frac{97}{100}\frac{3}{10} = \frac{312}{1000}$.
- $\omega|_{c \ll q}(r) = \frac{\omega(r)q(r)}{\omega \models q} = \frac{21/1000}{312/1000} = \frac{21}{312} \approx 0.067$.

Example: soft evidence

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \multimap X$ defined by $c(x) = 1|x\rangle$.
- Pearl's rule: predicate $q\colon X \to [0,1]$ defined by $q(r) = \frac{7}{10}$ and $q(r^{\perp}) = \frac{3}{10}$. To compute: $\omega|_{c \ll q}$.
- We have $c \ll q = q$, so $\omega|_{c \ll q}(x) = \omega|_{q}(x) = \frac{\omega(x)q(x)}{\omega \models q}$.
- $\omega \models q = \omega(r)q(r) + \omega(r^{\perp})q(r^{\perp}) = \frac{3}{100}\frac{7}{10} + \frac{97}{100}\frac{3}{10} = \frac{312}{1000}$.
- $\omega|_{c \ll q}(r) = \frac{\omega(r)q(r)}{\omega \models q} = \frac{21/1000}{312/1000} = \frac{21}{312} \approx 0.067$.
- $\omega|_{c \ll q}(r^{\perp}) = \frac{\omega(r^{\perp})q(r^{\perp})}{\omega \models q} = \frac{291/1000}{312/1000} = \frac{291}{312} \approx 0.933$.

Example: soft evidence

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \rightsquigarrow X$ defined by $c(x) = 1|x\rangle$.
- Pearl's rule: predicate $q\colon X \to [0,1]$ defined by $q(r) = \frac{7}{10}$ and $q(r^{\perp}) = \frac{3}{10}$. To compute: $\omega|_{c \ll q}$.
- We have $c \ll q = q$, so $\omega|_{c \ll q}(x) = \omega|_{q}(x) = \frac{\omega(x)q(x)}{\omega \models q}$.
- $\omega \models q = \omega(r)q(r) + \omega(r^{\perp})q(r^{\perp}) = \frac{3}{100}\frac{7}{10} + \frac{97}{100}\frac{3}{10} = \frac{312}{1000}$.
- $\omega|_{c \ll q}(r) = \frac{\omega(r)q(r)}{\omega \models q} = \frac{21/1000}{312/1000} = \frac{21}{312} \approx 0.067$.
- $\omega|_{c \ll q}(r^{\perp}) = \frac{\omega(r^{\perp})q(r^{\perp})}{\omega \models q} = \frac{291/1000}{312/1000} = \frac{291}{312} \approx 0.933$.
- So $\omega|_{c \ll q} = \frac{21}{312}|r\rangle + \frac{291}{312}|r^{\perp}\rangle \in \mathcal{D}(X)$.

Example: soft evidence

Around 3% of the cars in a city is red. One day, a car drives way too fast through the city center. The police asks for the car color. A witness: I'm 70% sure the car is red.

Example: soft evidence

Around 3% of the cars in a city is red. One day, a car drives way too fast through the city center. The police asks for the car color. A witness: I'm 70% sure the car is red.

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c \colon X \multimapinv X$ defined by $c(x) = 1|x\rangle$.

Example: soft evidence

Around 3% of the cars in a city is red. One day, a car drives way too fast through the city center. The police asks for the car color. A witness: I'm 70% sure the car is red.

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c \colon X \multimap X$ defined by $c(x) = 1|x\rangle$.
- Jeffrey's rule: state $\rho \in \mathcal{D}(X)$ defined by $\rho = \frac{7}{10}|r\rangle + \frac{3}{10}|r^{\perp}\rangle$.
  To compute: $c_{\omega}^{\dagger} \gg \rho$.

Example: soft evidence

Around 3% of the cars in a city is red. One day, a car drives way too fast through the city center. The police asks for the car color. A witness: I'm 70% sure the car is red.

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \multimap X$ defined by $c(x) = 1|x\rangle$.
- Jeffrey's rule: state $\rho \in \mathcal{D}(X)$ defined by $\rho = \frac{7}{10}|r\rangle + \frac{3}{10}|r^{\perp}\rangle$. To compute: $c_{\omega}^{\dagger} \gg \rho$.
- We have $c \gg \omega = \omega$ and hence

$$c_{\omega}^{\dagger}(y) = \sum_{x} \frac{\omega(x)c(x)(y)}{(c \gg \omega)(y)}|x\rangle = \frac{\omega(y)}{(c \gg \omega)(y)}|y\rangle = 1|y\rangle = c(y),$$

so $c_{\omega}^{\dagger} = c$.

Example: soft evidence

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \to X$ defined by $c(x) = 1|x\rangle$.
- Jeffrey's rule: state $\rho \in \mathcal{D}(X)$ defined by $\rho = \frac{7}{10}|r\rangle + \frac{3}{10}|r^{\perp}\rangle$.
  To compute: $c^{\dagger}_{\omega} \gg \rho$.
- We have $c^{\dagger}_{\omega} = c$.

Example: soft evidence

- Prior probability: $\Pr(r) = \frac{3}{100}$ and $\Pr\left(r^{\perp}\right) = \frac{97}{100}$.
- Let $X = \{r, r^{\perp}\}$ and let $\omega = \frac{3}{100}|r\rangle + \frac{97}{100}|r^{\perp}\rangle \in \mathcal{D}(X)$.
- Channel: $c\colon X \multimap X$ defined by $c(x) = 1|x\rangle$.
- Jeffrey's rule: state $\rho \in \mathcal{D}(X)$ defined by $\rho = \frac{7}{10}|r\rangle + \frac{3}{10}|r^{\perp}\rangle$. To compute: $c_{\omega}^{\dagger} \gg \rho$.
- We have $c_{\omega}^{\dagger} = c$.
- We have

$$(c_{\omega}^{\dagger} \gg \rho)(y) = \sum_{x} \rho(x) c_{\omega}^{\dagger}(x)(y) = \sum_{x} \rho(x)\delta_{xy} = \rho(y),$$

so $c_{\omega}^{\dagger} \gg \rho = \rho = \frac{7}{10}|r\rangle + \frac{3}{10}|r^{\perp}\rangle \in \mathcal{D}(X)$.

Example: soft evidence

Around 3% of the cars in a city is red. One day, a car drives way too fast through the city center. The police asks for the car color. A witness: I'm 70% sure the car is red.

- Pearl's rule: $\frac{21}{312}|r\rangle + \frac{291}{312}|r^\perp\rangle \in \mathcal{D}(X)$.
- Jeffrey's rule: $\frac{7}{10}|r\rangle + \frac{3}{10}|r^\perp\rangle \in \mathcal{D}(X)$.
- Pearl's rule: although the witness is 70% sure, it is much more likely that the car was not red.
- Jeffrey's rule: the witness is 70% sure taking into account there are very few red cars.
- In general the inverted channel $c_\omega^\dagger$ and hence the result of Jeffrey's rule actually depends on the prior.

Applications of Jeffrey's rule do not commute

- In the red car example, we saw that $c_\omega^\dagger \gg \rho = \rho$.
- Now consider two witnesses with states ρ_1 and ρ_2.
- First witness 1, then witness 2: $c_\omega^\dagger \gg \rho_1 = \rho_1$, then $c_{\rho_1}^\dagger \gg \rho_2 = \rho_2$. So final updated state is ρ_2.
- First witness 2, then witness 1: Final updated state is ρ_1.
- If witness 1's update is done first, witness 2 needs to take into account witness 1's testimony.

Applications of Jeffrey's rule do not commute

- In the red car example, we saw that $c_\omega^\dagger \gg \rho = \rho$.
- Now consider two witnesses with states ρ_1 and ρ_2.
- First witness 1, then witness 2: $c_\omega^\dagger \gg \rho_1 = \rho_1$, then $c_{\rho_1}^\dagger \gg \rho_2 = \rho_2$. So final updated state is ρ_2.
- First witness 2, then witness 1: Final updated state is ρ_1.
- If witness 1's update is done first, witness 2 needs to take into account witness 1's testimony.
- So applications of Jeffrey's rule do not commute.
- Applications of Pearl's rule do commute.

Improvement and correction without effect

- When using Pearl's rule with a uniform likelihood, the beliefs do not change,
  i.e. improving with no information has no effect.

- When using Jeffrey's rule with the current belief, the beliefs do not change,
  i.e. correcting beliefs to the current belief has no effect.

Pearl's rule increases validity

Theorem

Let $c\colon X \rightsquigarrow Y$ *be a channel with prior state* $\sigma \in \mathcal{D}(X)$ *and let* $q\colon Y \to [0,1]$ *be a predicate. Let* $\sigma_P = \sigma|_{c \ll q}$ *denotes Pearl's update. Pearl's rule increases validity:*

$$(c \gg \sigma_P) \models q \geq (c \gg \sigma) \models q.$$

Pearl's rule increases validity

Lemma

Let $\omega \in \mathcal{D}(X)$ *be a state and let* $p\colon X \to [0,1]$ *be a predicate. Assume that* $\omega \models p$ *is nonzero. Then*

$$\omega|_p \models p \geq \omega \models p.$$

Pearl's rule increases validity

Lemma

Let $\omega \in \mathcal{D}(X)$ *be a state and let* $p\colon X \to [0,1]$ *be a predicate. Assume that* $\omega \models p$ *is nonzero. Then*

$$\omega|_p \models p \geq \omega \models p.$$

Proof. Recall that $\omega \models p = \sum_x \omega(x)p(x)$ and $\omega|_p(x) = \frac{\omega(x)p(x)}{\omega \models p}$.
Hence, $\omega|_p \models p = \sum_x \frac{\omega(x)p(x)}{\omega \models p} p(x) = \frac{1}{\omega \models p} \sum_x \omega(x)p(x)^2$.

Pearl's rule increases validity

Lemma

Let $\omega \in \mathcal{D}(X)$ *be a state and let* $p\colon X \to [0,1]$ *be a predicate. Assume that* $\omega \models p$ *is nonzero. Then*

$$\omega|_p \models p \geq \omega \models p.$$

Proof. Recall that $\omega \models p = \sum_x \omega(x)p(x)$ and $\omega|_p(x) = \frac{\omega(x)p(x)}{\omega \models p}$. Hence, $\omega|_p \models p = \sum_x \frac{\omega(x)p(x)}{\omega \models p} p(x) = \frac{1}{\omega \models p} \sum_x \omega(x)p(x)^2$. So it is sufficient to show that

$$\left(\sum_x \omega(x)(p(x))^2\right) \geq (\omega \models p)^2 = \left(\sum_x \omega(x)p(x)\right)^2.$$

Pearl's rule increases validity

Lemma

Let $\omega \in \mathcal{D}(X)$ be a state and let $p \colon X \to [0,1]$ be a predicate. Assume that $\omega \models p$ is nonzero. Then

$$\omega\big|_p \models p \geq \omega \models p.$$

Proof. The Cauchy-Schwarz inequality states that $\left(\sum_i y_i^2\right)\left(\sum_i z_i^2\right) \geq \left(\sum_i y_i z_i\right)^2$. Let $y_i = \sqrt{\omega(x)}p(x)$ and $z_i = \sqrt{\omega(x)}$. Then

$$\left(\sum_x \omega(x)(p(x))^2\right) \underbrace{\left(\sum_x \omega(x)\right)}_{=1} \geq \left(\sum_x \omega(x)p(x)\right)^2.$$

Pearl's rule increases validity

Theorem

Let $c \colon X \rightsquigarrow Y$ *be a channel with prior state* $\sigma \in \mathcal{D}(X)$ *and let* $q \colon Y \to [0,1]$ *be a predicate. Let* $\sigma_P = \sigma|_{c \ll q}$ *denotes Pearl's update. Pearl's rule increases validity:*

$$(c \gg \sigma_P) \models q \geq (c \gg \sigma) \models q.$$

Proof. For any state ω we have $(c \gg \omega) \models q = \omega \models (c \ll q)$. By the lemma, it follows that

$$\begin{aligned}(c \gg (\sigma|_{c \ll q})) \models q &= \sigma|_{c \ll q} \models (c \ll q) \\ &\geq \sigma \models (c \ll q) = (c \gg \sigma) \models q.\end{aligned}$$

Jeffrey's rule decreases KL-divergence

Definition (Kullback-Leibler divergence)

Let $\omega, \rho \in \mathcal{D}(X)$. *The KL divergence* $D_{KL}(\omega, \rho)$ *is defined as*

$$D_{KL}(\omega, \rho) = \sum_x \omega(x) \ln \left(\frac{\omega(x)}{\rho(x)} \right).$$

Jeffrey's rule decreases KL-divergence

Definition (Kullback-Leibler divergence)

Let $\omega, \rho \in \mathcal{D}(X)$*. The KL divergence* $D_{KL}(\omega, \rho)$ *is defined as*

$$D_{KL}(\omega, \rho) = \sum_x \omega(x) \ln\left(\frac{\omega(x)}{\rho(x)}\right).$$

- $D_{KL}(\omega, \rho) \geq 0$ and $D_{KL}(\omega, \rho) = 0$ if and only if $\omega = \rho$.
- The KL divergence is *not* symmetric, i.e. in general $D_{KL}(\omega, \rho) \neq D_{KL}(\rho, \omega)$.

Jeffrey's rule decreases KL-divergence

Theorem

Let $c\colon X \multimap Y$ be a channel with prior state $\sigma \in \mathcal{D}(X)$. Assume that Y is finite and that $c \gg \sigma$ has full support. Let $\tau \in \mathcal{D}(Y)$ be an evidence state, and let $\sigma_J = c^{\dagger}_{\sigma} \gg \tau$ denote Jeffrey's update. Jeffrey's rule decreases divergence:

$$D_{KL}(\tau, c \gg \sigma_J) \leq D_{KL}(\tau, c \gg \sigma).$$

Open questions

- Are there other measures which always increase or decrease when doing an update according to Pearl's rule or Jeffrey's rule?
- What to do when evidence takes some, but not all other information into account?
- What does soft evidence actually mean?
- Is there a cognitive difference between 'improvement' and 'correction'?

Conclusion

- Pearl's rule: improvement
  Evidence type: 'nothing else considered'.
- Jeffrey's rule: correction
  Evidence type: 'all things considered'.
- Pearl's rule increases validity of the updating predicate.
- Jeffrey's rule decreases divergence with the updating state.

References

[1] Bart Jacobs. "The Mathematics of Changing one's Mind, via Jeffrey's or via Pearl's update rule." In: *Journal of Artificial Intelligence Research* 65 (2019), pp. 783–806.

[2] Bart Jacobs. "Learning from What's Right and Learning from What's Wrong." In: *arXiv preprint arXiv:2112.14045* (2021).

[3] Richard C. Jeffrey. *The Logic of Decision*. New York, NY, USA: University of Chicago Press, 1965.

[4] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.