Fuzzy probabilities based on the likelihood function

Marco Cattaneo
Department of Statistics, LMU Munich
cattaneo@stat.uni-muenchen.de

September 8, 2008

precise, imprecise, and fuzzy probabilities

- imprecise probability (IP) $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ generalizes probability P;

precise, imprecise, and fuzzy probabilities

- imprecise probability (IP) $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ generalizes probability P;

  e.g. interval $[\min_{\theta\in\Theta} P_\theta(A), \max_{\theta\in\Theta} P_\theta(A)] = [0.3, 0.9]$ instead of precise value $P(A) = 0.5$

precise, imprecise, and fuzzy probabilities

- imprecise probability (IP) $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ generalizes probability P;

  e.g. interval $[\min_{\theta \in \Theta} P_\theta(A), \max_{\theta \in \Theta} P_\theta(A)] = [0.3, 0.9]$ instead of precise value $P(A) = 0.5$

- usually not all values in the probability interval are considered equally plausible: to accommodate this fact, the interval can be generalized to a fuzzy number;

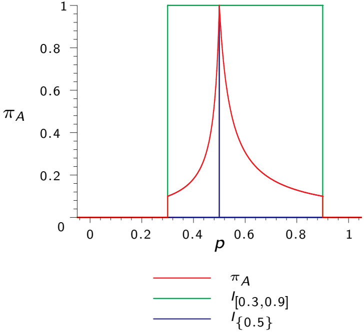
precise, imprecise, and fuzzy probabilities

- imprecise probability (IP) $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ generalizes probability P;

  e.g. interval $[\min_{\theta\in\Theta} P_\theta(A), \max_{\theta\in\Theta} P_\theta(A)] = [0.3, 0.9]$ instead of precise value $P(A) = 0.5$

- usually not all values in the probability interval are considered equally plausible: to accommodate this fact, the interval can be generalized to a fuzzy number;

  e.g. the membership function $\pi_A : [0,1] \to [0,1]$ describes the fuzzy probability of A (the above examples correspond to the crisp membership functions $\pi_A = I_{\{0.5\}}$ and $\pi_A = I_{[0.3,0.9]}$)



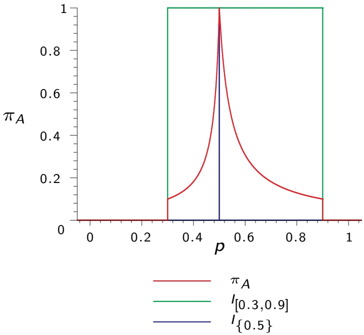
precise, imprecise, and fuzzy probabilities

- imprecise probability (IP) $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ generalizes probability P;

  e.g. interval $[\min_{\theta\in\Theta} P_\theta(A), \max_{\theta\in\Theta} P_\theta(A)] = [0.3, 0.9]$ instead of precise value $P(A) = 0.5$

- usually not all values in the probability interval are considered equally plausible: to accommodate this fact, the interval can be generalized to a fuzzy number;

  e.g. the membership function $\pi_A : [0,1] \to [0,1]$ describes the fuzzy probability of A (the above examples correspond to the crisp membership functions $\pi_A = I_{\{0.5\}}$ and $\pi_A = I_{[0.3,0.9]}$)



- models of "fuzzy probability" have been proposed in particular by
  - Walley (1997) and De Cooman (2005)
  - Viertl and Hareter (2006) and Buckley (2006)

updating and likelihood

- when an event A is observed, the IP model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is usually updated by conditioning each P_θ on A;

updating and likelihood

- when an event A is observed, the IP model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is usually updated by conditioning each P_θ on A;
  the normalization of the conditional probability $P_\theta(B \mid A) = \frac{P_\theta(B \cap A)}{P_\theta(A)}$ deletes the information about the value $P_\theta(A)$

updating and likelihood

- when an event A is observed, the IP model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is usually updated by conditioning each P_θ on A;
  the normalization of the conditional probability $P_\theta(B \mid A) = \frac{P_\theta(B \cap A)}{P_\theta(A)}$ deletes the information about the value $P_\theta(A)$
- the **likelihood function** $lik(\theta) \propto P_\theta(A)$ describes the *relative* ability of the probabilistic models P_θ to forecast the observed data;

updating and likelihood

- when an event A is observed, the IP model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is usually updated by conditioning each P_θ on A;
  the normalization of the conditional probability $P_\theta(B \mid A) = \frac{P_\theta(B \cap A)}{P_\theta(A)}$ deletes the information about the value $P_\theta(A)$
- the **likelihood function** $lik(\theta) \propto P_\theta(A)$ describes the *relative* ability of the probabilistic models P_θ to forecast the observed data;
  it is a (the?) central concept in statistical inference

updating and likelihood

- when an event A is observed, the IP model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is usually updated by conditioning each P_θ on A;
  the normalization of the conditional probability $P_\theta(B \mid A) = \frac{P_\theta(B \cap A)}{P_\theta(A)}$ deletes the information about the value $P_\theta(A)$
- the **likelihood function** $lik(\theta) \propto P_\theta(A)$ describes the *relative* ability of the probabilistic models P_θ to forecast the observed data;
  it is a (the?) central concept in statistical inference
- *lik* is usually interpreted as a measure of the **relative plausibility** of the probabilistic models P_θ in the light of the observed data alone;

updating and likelihood

- when an event A is observed, the IP model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is usually updated by conditioning each P_θ on A;
  the normalization of the conditional probability $P_\theta(B \mid A) = \frac{P_\theta(B \cap A)}{P_\theta(A)}$ deletes the information about the value $P_\theta(A)$
- the **likelihood function** $lik(\theta) \propto P_\theta(A)$ describes the *relative* ability of the probabilistic models P_θ to forecast the observed data;
  it is a (the?) central concept in statistical inference
- *lik* is usually interpreted as a measure of the **relative plausibility** of the probabilistic models P_θ in the light of the observed data alone;
  the usual updating rule for IP disregards the information contained in *lik* (the elements of $\mathcal{P}$ are always considered equally plausible), and this can lead to problems such as statistical inconsistency or discontinuities

updating and likelihood

- when an event A is observed, the IP model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is usually updated by conditioning each P_θ on A;
  the normalization of the conditional probability $P_\theta(B \mid A) = \frac{P_\theta(B \cap A)}{P_\theta(A)}$ deletes the information about the value $P_\theta(A)$
- the **likelihood function** $lik(\theta) \propto P_\theta(A)$ describes the *relative* ability of the probabilistic models P_θ to forecast the observed data;
  it is a (the?) central concept in statistical inference
- *lik* is usually interpreted as a measure of the **relative plausibility** of the probabilistic models P_θ in the light of the observed data alone;
  the usual updating rule for IP disregards the information contained in *lik* (the elements of $\mathcal{P}$ are always considered equally plausible), and this can lead to problems such as statistical inconsistency or discontinuities
- alternative updating rules for IP making use of some information contained in *lik* have been proposed in particular by Moral (1992), Wilson (2001), Held, Kriegler, and Augustin (2008);

updating and likelihood

- when an event A is observed, the IP model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is usually updated by conditioning each P_θ on A;
  the normalization of the conditional probability $P_\theta(B \mid A) = \frac{P_\theta(B \cap A)}{P_\theta(A)}$ deletes the information about the value $P_\theta(A)$
- the **likelihood function** $lik(\theta) \propto P_\theta(A)$ describes the *relative* ability of the probabilistic models P_θ to forecast the observed data;
  it is a (the?) central concept in statistical inference
- *lik* is usually interpreted as a measure of the **relative plausibility** of the probabilistic models P_θ in the light of the observed data alone;
  the usual updating rule for IP disregards the information contained in *lik* (the elements of $\mathcal{P}$ are always considered equally plausible), and this can lead to problems such as statistical inconsistency or discontinuities
- alternative updating rules for IP making use of some information contained in *lik* have been proposed in particular by Moral (1992), Wilson (2001), Held, Kriegler, and Augustin (2008);
  however, to completely solve the problems of IP updating, it seems necessary to store more information than it is possible in the framework of IP

probabilistic-possibilistic hierarchical model

- let us interpret *normalized* possibility distributions as likelihood functions;

probabilistic-possibilistic hierarchical model

- let us interpret *normalized* possibility distributions as likelihood functions;
  this is an established interpretation of possibility theory, studied in particular by Hisdal (1988)

probabilistic-possibilistic hierarchical model

- let us interpret *normalized* possibility distributions as likelihood functions;
  this is an established interpretation of possibility theory, studied in particular by Hisdal (1988)
- the set $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of probability measures and the normalized possibility measure Π on Θ with possibility distribution π are the two levels of a probabilistic-possibilistic hierarchical description of uncertain knowledge;

probabilistic-possibilistic hierarchical model

- let us interpret *normalized* possibility distributions as likelihood functions;
  this is an established interpretation of possibility theory, studied in particular by Hisdal (1988)
- the set $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of probability measures and the normalized possibility measure Π on Θ with possibility distribution π are the two levels of a probabilistic-possibilistic hierarchical description of uncertain knowledge;
  this hierarchical model generalizes the IP model, which corresponds to the case with $\pi \equiv 1$ (complete ignorance at the possibilistic level)

probabilistic-possibilistic hierarchical model

- let us interpret *normalized* possibility distributions as likelihood functions;
  this is an established interpretation of possibility theory, studied in particular by Hisdal (1988)
- the set $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of probability measures and the normalized possibility measure Π on Θ with possibility distribution π are the two levels of a probabilistic-possibilistic hierarchical description of uncertain knowledge;
  this hierarchical model generalizes the IP model, which corresponds to the case with $\pi \equiv 1$ (complete ignorance at the possibilistic level)
- when A is observed, the hierarchical model is updated by conditioning each P_θ on A, and replacing π with the pointwise product of π and *lik*;

probabilistic-possibilistic hierarchical model

- let us interpret *normalized* possibility distributions as likelihood functions;
  this is an established interpretation of possibility theory, studied in particular by Hisdal (1988)
- the set $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of probability measures and the normalized possibility measure Π on Θ with possibility distribution π are the two levels of a probabilistic-possibilistic hierarchical description of uncertain knowledge;
  this hierarchical model generalizes the IP model, which corresponds to the case with $\pi \equiv 1$ (complete ignorance at the possibilistic level)
- when A is observed, the hierarchical model is updated by conditioning each P_θ on A, and replacing π with the pointwise product of π and *lik*;
  unlike for IP updating, the information contained in *lik* is used

probabilistic-possibilistic hierarchical model

- let us interpret *normalized* possibility distributions as likelihood functions;
  this is an established interpretation of possibility theory, studied in particular by Hisdal (1988)
- the set $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of probability measures and the normalized possibility measure Π on Θ with possibility distribution π are the two levels of a probabilistic-possibilistic hierarchical description of uncertain knowledge;
  this hierarchical model generalizes the IP model, which corresponds to the case with $\pi \equiv 1$ (complete ignorance at the possibilistic level)
- when A is observed, the hierarchical model is updated by conditioning each P_θ on A, and replacing π with the pointwise product of π and *lik*;
  unlike for IP updating, the information contained in *lik* is used
- the uncertain knowledge about the value $g(\theta)$ of a function $g : \Theta \to \mathcal{G}$ is described by the induced possibility measure $\Pi \circ g^{-1}$ on Θ (where $g^{-1} : 2^{\mathcal{G}} \to 2^{\Theta}$);

probabilistic-possibilistic hierarchical model

- let us interpret *normalized* possibility distributions as likelihood functions;
  this is an established interpretation of possibility theory, studied in particular by Hisdal (1988)
- the set $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ of probability measures and the normalized possibility measure Π on Θ with possibility distribution π are the two levels of a probabilistic-possibilistic hierarchical description of uncertain knowledge;
  this hierarchical model generalizes the IP model, which corresponds to the case with $\pi \equiv 1$ (complete ignorance at the possibilistic level)
- when A is observed, the hierarchical model is updated by conditioning each P_θ on A, and replacing π with the pointwise product of π and *lik*;
  unlike for IP updating, the information contained in *lik* is used
- the uncertain knowledge about the value $g(\theta)$ of a function $g : \Theta \to \mathcal{G}$ is described by the induced possibility measure $\Pi \circ g^{-1}$ on Θ (where $g^{-1} : 2^{\mathcal{G}} \to 2^{\Theta}$);
  in particular, if $g(\theta) = P_\theta(B)$, then the induced possibility measure $\Pi \circ g^{-1}$ on $[0, 1]$ describes the **fuzzy probability** of B

example

draws (with replacement) from an urn containing 10 balls, each either black or white

example

draws (with replacement) from an urn containing 10 balls, each either black or white

let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be the IP model with complete ignorance about the number of black balls in the urn: Θ corresponds to the set of all prior probability measures on $\{0, 1, \ldots, 10\}$

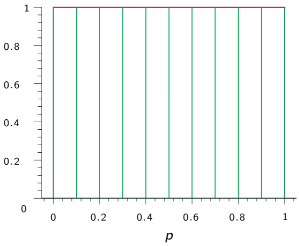
example

draws (with replacement) from an urn containing 10 balls, each either black or white

let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be the IP model with complete ignorance about the number of black balls in the urn: Θ corresponds to the set of all prior probability measures on $\{0, 1, \ldots, 10\}$

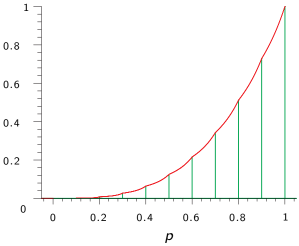
membership function of the fuzzy probability of drawing ●:

prior (vacuous)



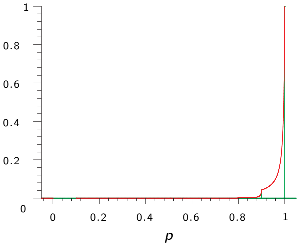
IP (0-cut): $[0, 1]$
0.036-cut: $[0, 1]$

after observing $3\times$●



IP (0-cut): $[0.1, 1]$
0.036-cut: $[0.344, 1]$

after observing $30\times$●

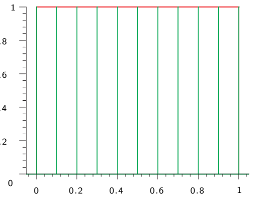


IP (0-cut): $[0.1, 1]$
0.036-cut: $[0.899, 1]$

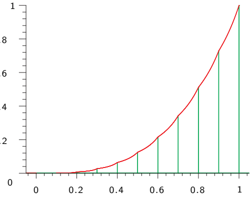
example

membership function of the fuzzy probability of drawing ●:

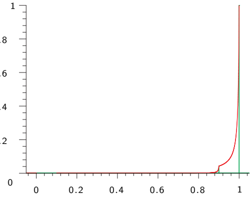
prior (vacuous)



after observing 3×●



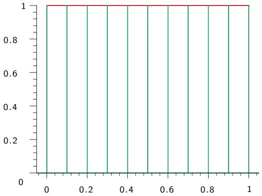
after observing 30×●



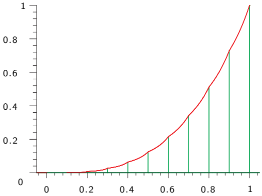
example

membership function of the fuzzy probability of drawing ●:

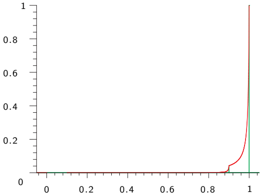
prior (vacuous)



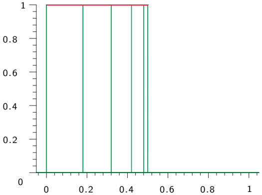
after observing 3×●

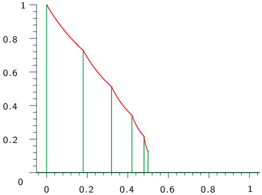


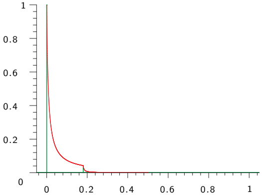
after observing 30×●



membership function of the fuzzy probability of drawing ●○ or ○●:







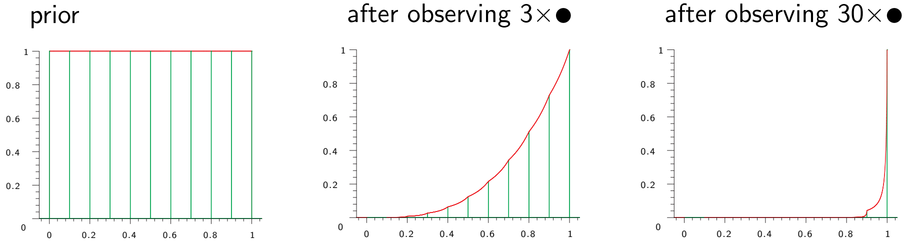
IP (0-cut): $[0, 0.5]$
0.036-cut: $[0, 0.5]$

IP (0-cut): $[0, 0.5]$
0.036-cut: $[0, 0.5]$

IP (0-cut): $[0, 0.5]$
0.036-cut: $[0, 0.181]$

example

membership function of the fuzzy probability of drawing ● without and with prior information about the number of black balls in the urn:



0.036-cut: $[0, 1]$

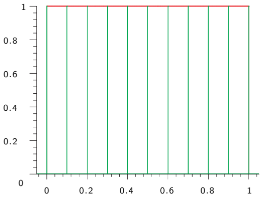
0.036-cut: $[0.344, 1]$

0.036-cut: $[0.899, 1]$

example

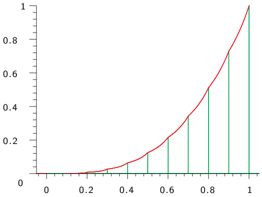
membership function of the fuzzy probability of drawing ● without and with prior information about the number of black balls in the urn:

prior



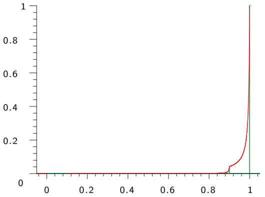
0.036-cut: $[0, 1]$

after observing 3×●

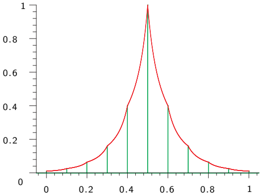


0.036-cut: $[0.344, 1]$

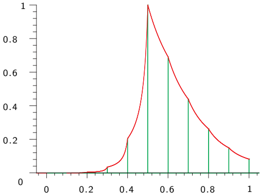
after observing 30×●



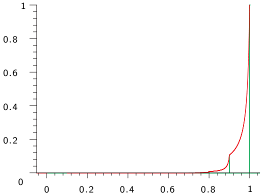
0.036-cut: $[0.899, 1]$



0.036-cut: $[0.149, 0.851]$



0.036-cut: $[0.306, 1]$

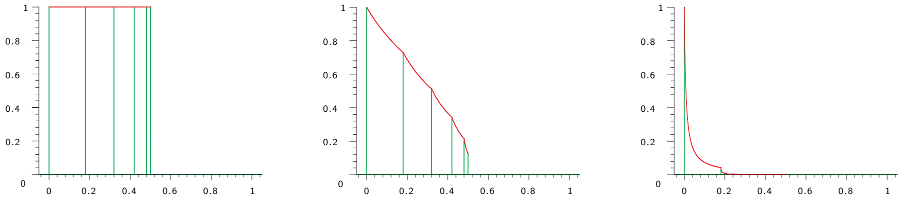


0.036-cut: $[0.885, 1]$

example

membership function of the fuzzy probability of drawing ●○ or ○● without and with prior information about the number of black balls in the urn:

prior | after observing 3×● | after observing 30×●

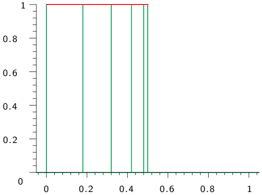


0.036-cut: $[0, 0.5]$ | 0.036-cut: $[0, 0.5]$ | 0.036-cut: $[0, 0.181]$

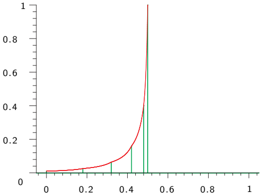
example

membership function of the fuzzy probability of drawing ●○ or ○● without and with prior information about the number of black balls in the urn:

prior

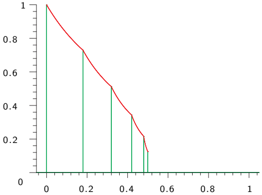


0.036-cut: $[0, 0.5]$

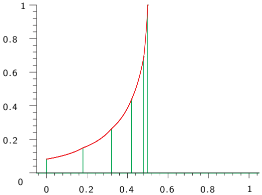


0.036-cut: $[0.249, 0.5]$

after observing 3×●

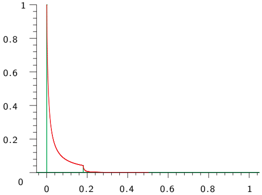


0.036-cut: $[0, 0.5]$

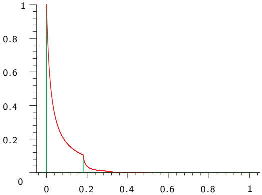


0.036-cut: $[0, 0.5]$

after observing 30×●



0.036-cut: $[0, 0.181]$



0.036-cut: $[0, 0.201]$

additional references

- Cattaneo, M. (2008). Probabilistic-possibilistic belief networks. Technical Report, LMU Munich. `http://epub.ub.uni-muenchen.de`
- Held, H., Augustin, T., and Kriegler, E. (2008). Bayesian learning for a class of priors with prescribed marginals. Int. J. Approx. Reasoning 49, 212–233.
- Walley, P. (1997). Statistical inferences based on a second-order possibility distribution. Int. J. Gen. Syst. 26, 337-383.