Imprecise probability for statistical problems: is it worth the candle?

Marco Cattaneo
Department of Statistics, LMU Munich

26 November 2012

introduction

- **comparison** of conventional and imprecise probability approaches to statistics
- theoretical perspective **and** pragmatical perspective (application to the "fundamental problem of practical statistics")
- only **statistics** (not personal decision making)
- only statistical **inference**: given a statistical model $\{P_\theta : \theta \in \Theta\}$ on $\mathcal{X}$
- a **personal** viewpoint

fundamental problem of practical statistics (Pearson, 1920): An "event" has occurred p times out of $p + q = n$ trials, where we have no a priori knowledge of the frequency of the event in the total population of occurrences. What is the probability of its occurring r times in a further $r + s = m$ trials?

- sequence of binary **random variables**: $(X_1, X_2, \ldots) \in \mathcal{X} = \{0, 1\}^{\mathbb{N}}$
- **statistical model**: $X_1, X_2, \ldots \overset{i.i.d.}{\sim} Ber(\theta)$ with $\theta \in \Theta = [0, 1]$
- **data**: $\sum_{i=1}^{n} X_i = p$
- **quantity of interest**: $P_\theta(\sum_{i=n+1}^{n+m} X_i = r) = \binom{r}{m} \theta^r (1 - \theta)^s$

comparison

Bayesian approach

classical approach

Bayesian approach

- **central idea**: uncertainty about θ described by a probability distribution π on Θ
- **model**: $\pi \times P_\theta$ on $\Theta \times \mathcal{X}$
- **necessary choice**: prior probability distribution
- **result**: posterior probability distribution (expectation / mode, credible interval/region)
- **properties**: invariances (transformation, temporal, likelihood principle, ...)

example: fundamental problem of practical statistics

- choice of prior probability distribution: e.g., **conjugate** prior $\theta \sim Beta(\alpha, \beta)$ with $\alpha, \beta \in \mathbb{R}_{>0}$
- $\beta = \alpha$ from symmetry, but choice of α is **difficult** (Bayes: 1, Jeffreys: $\frac{1}{2}$, Haldane: 0)
- **posterior** probability distribution: $\theta \sim Beta(\alpha + p, \beta + q)$
- expectation and credible interval **for** $\binom{r}{m} \theta^r (1 - \theta)^s$ analytically or numerically

classical approach

- **central idea**: comparison of inference methods on the basis of their repeated sampling performance (as a function of θ)
- **model**: $\{P_\theta : \theta \in \Theta\}$ on $\mathcal{X}$
- **necessary choice**: inference method (or comparison criterion)
- **result**: inference (point estimate, confidence interval/region)
- **properties**: repeated sampling calibration

example: fundamental problem of practical statistics

- choice of repetition: e.g., n **fixed** (binomial experiment)
- choice of comparison criterion: e.g., **maximum** mean squared error
- optimal inference method (minimax MSE estimator **of** $\binom{r}{m}\theta^r(1-\theta)^s$) analytically
- confidence interval for $\binom{r}{m}\theta^r(1-\theta)^s$ is more **difficult**

comparison

$$\begin{tikzcd} \textbf{Bayesian approach} \arrow[r, bend left, "\text{calibration, no prior}"] & \textbf{classical approach} \arrow[l, bend left, "\text{simplicity, invariances}"] \end{tikzcd}$$

IP approach

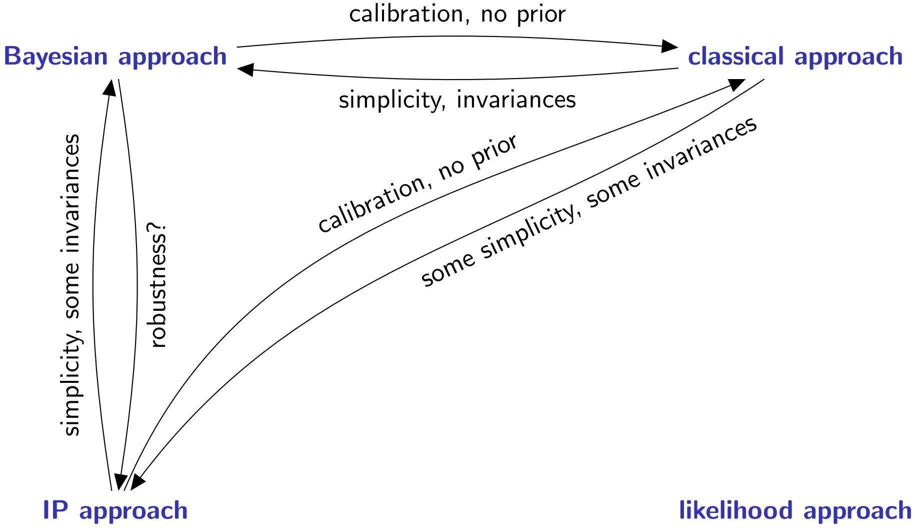
IP approach

- **central idea**: uncertainty about θ described by a lower/upper prevision (with core Γ) on Θ
- **model**: $\{\pi \times P_\theta : \pi \in \Gamma\}$ on $\Theta \times \mathcal{X}$
- **necessary choice**: prior lower/upper prevision (in particular: amount of imprecision)
- **result**: posterior lower/upper prevision (point estimate?, credible interval/region?)
- **properties**: invariances (transformation, likelihood principle, ...)

example: fundamental problem of practical statistics

- choice of prior lower/upper prevision: e.g., IDM (set of **conjugate** priors) $\theta \sim \{Beta(\alpha, \beta) : \alpha, \beta \in \mathbb{R}_{>0},\ \alpha + \beta = s\}$ with $s \in \mathbb{R}_{>0}$
- choice of s is **difficult** (Walley: 2 or 1)
- **posterior** lower/upper prevision: $\theta \sim \{Beta(\alpha + p, \beta + q) : \alpha, \beta \in \mathbb{R}_{>0},\ \alpha + \beta = s\}$
- (imprecise) expectation **of** $\binom{r}{m}\, \theta^r\, (1-\theta)^s$ analytically or numerically, but is neither a point estimate nor a confidence/credible interval

comparison


Bayesian approach
classical approach
IP approach
likelihood approach
calibration, no prior
simplicity, invariances
simplicity, some invariances
robustness?
calibration, no prior
some simplicity, some invariances

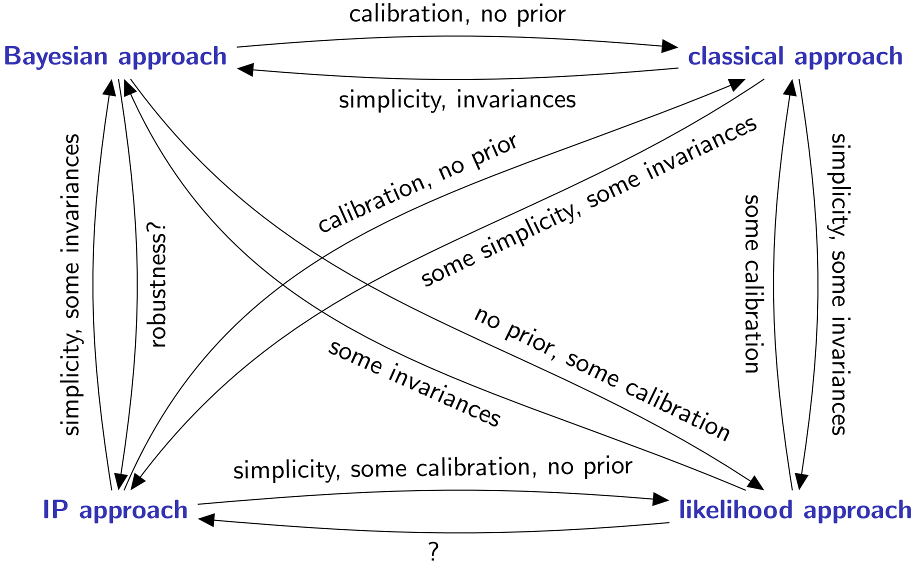
likelihood approach

- **central idea**: uncertainty about θ described by a likelihood function (possibility measure) *lik* on Θ
- **model**: $\{P_\theta : \theta \in \Theta\}$ on $\mathcal{X}$, and *lik* on Θ
- **necessary choice**: (prior likelihood function)
- **result**: (posterior) likelihood function (maximum likelihood estimate, likelihood interval/region)
- **properties**: invariances (transformation, likelihood principle, ...), sometimes repeated sampling calibration

example: fundamental problem of practical statistics

- **no** choice necessary
- (posterior) **likelihood function**: $lik(\theta) \propto \theta^p \, (1-\theta)^q$
- maximum likelihood estimate and likelihood interval **for** $\binom{r}{m} \, \theta^r \, (1-\theta)^s$ analytically or numerically
- repeated sampling calibration is easy (**regular** problem)

comparison



calibration, no prior
Bayesian approach
classical approach
simplicity, invariances
simplicity, some invariances
robustness?
calibration, no prior
some simplicity, some invariances
some calibration
simplicity, some invariances
no prior, some calibration
some invariances
simplicity, some calibration, no prior
IP approach
likelihood approach
?

conclusion

- imprecise probabilities (**as** sets of probabilities) appear naturally in many statistical problems
- conventional approaches to statistics have advantages **and** disadvantages compared to each other
- is there some good **reason** for preferring the IP approach (to statistics) to the Bayesian one?
  - imprecise expectations are often **misinterpreted** as confidence/credible intervals
  - choosing the amount of **imprecision** in prior lower/upper previsions is particularly difficult
- the likelihood approach to statistics seems to be a better **compromise** between the Bayesian and classical ones