On the estimation of imprecise probabilities

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

WPMSIIP 2010, Durham, UK
10 September 2010

the fundamental problem of statistical inference

- given the realizations of the independent random variables $X_1, \ldots, X_n$ with $X_i \sim Ber(p)$, estimate $p \in [0, 1]$

the fundamental problem of statistical inference

- given the realizations of the independent random variables $X_1, \ldots, X_n$ with $X_i \sim Ber(p)$, estimate $p \in [0, 1]$
  - the estimation is **unproblematic** when n is sufficiently large

the fundamental problem of statistical inference

- given the realizations of the independent random variables $X_1, \ldots, X_n$ with $X_i \sim Ber(p)$, estimate $p \in [0, 1]$
  - the estimation is **unproblematic** when n is sufficiently large

- given the realizations of the independent random variables $X_1, \ldots, X_n$ with $X_i \sim Ber(p_i)$ for some $p_i \in [\underline{p}, \overline{p}]$, estimate $[\underline{p}, \overline{p}] \subseteq [0, 1]$

the fundamental problem of statistical inference

- given the realizations of the independent random variables $X_1, \ldots, X_n$ with $X_i \sim Ber(p)$, estimate $p \in [0, 1]$
  - the estimation is **unproblematic** when n is sufficiently large

- given the realizations of the independent random variables $X_1, \ldots, X_n$ with $X_i \sim Ber(p_i)$ for some $p_i \in [\underline{p}, \overline{p}]$, estimate $[\underline{p}, \overline{p}] \subseteq [0, 1]$
  - for example, the estimation of the transition matrix of an **imprecise Markov chain** is a strictly related problem

the fundamental problem of statistical inference

- given the realizations of the independent random variables $X_1, \ldots, X_n$ with $X_i \sim Ber(p)$, estimate $p \in [0, 1]$
  - the estimation is **unproblematic** when n is sufficiently large

- given the realizations of the independent random variables $X_1, \ldots, X_n$ with $X_i \sim Ber(p_i)$ for some $p_i \in [\underline{p}, \overline{p}]$, estimate $[\underline{p}, \overline{p}] \subseteq [0, 1]$
  - for example, the estimation of the transition matrix of an **imprecise Markov chain** is a strictly related problem
  - note that the **IDM model** describes some imprecise knowledge about the precise probability p (for instance, in the IDM model the lower probability of $X_i = X_{i+1} = 1$ is in general not $\underline{p}^2$)

the fundamental problem of statistical inference

- given the realizations of the independent random variables $X_1, \dots, X_n$ with $X_i \sim Ber(p)$, estimate $p \in [0, 1]$
  - the estimation is **unproblematic** when n is sufficiently large

- given the realizations of the independent random variables $X_1, \dots, X_n$ with $X_i \sim Ber(p_i)$ for some $p_i \in [\underline{p}, \overline{p}]$, estimate $[\underline{p}, \overline{p}] \subseteq [0, 1]$
  - for example, the estimation of the transition matrix of an **imprecise Markov chain** is a strictly related problem
  - note that the **IDM model** describes some imprecise knowledge about the precise probability p (for instance, in the IDM model the lower probability of $X_i = X_{i+1} = 1$ is in general not $\underline{p}^2$)
  - no estimator of $[\underline{p}, \overline{p}]$ can be **consistent** under all sequences $(p_i) \in [\underline{p}, \overline{p}]^{\mathbb{N}}$, and the same holds for the estimators of $[\inf p_i, \sup p_i]$ or $[\liminf p_i, \limsup p_i]$ (since for example the deterministic model with $p_i = x_i$ can never be excluded)

maximum likelihood estimation

- basic question: given two imprecise probability models P_1, P_2 and the observed event A, which one of the two models was **best** in forecasting A?

maximum likelihood estimation

- basic question: given two imprecise probability models P_1, P_2 and the observed event A, which one of the two models was **best** in forecasting A?
- in a certain sense, the likelihood function induced by the data A on a set of imprecise probability models is **bidimensional**: $lik(P) = \big(\underline{P}(A), \overline{P}(A)\big)$

maximum likelihood estimation

- basic question: given two imprecise probability models P_1, P_2 and the observed event A, which one of the two models was **best** in forecasting A?
- in a certain sense, the likelihood function induced by the data A on a set of imprecise probability models is **bidimensional**: $lik(P) = \big(\underline{P}(A), \overline{P}(A)\big)$
- when the set of imprecise probability models is large enough, the "upper likelihood function" $\overline{lik}(P) = \overline{P}(A)$ is maximized by the **vacuous** model,

maximum likelihood estimation

- basic question: given two imprecise probability models P_1, P_2 and the observed event A, which one of the two models was **best** in forecasting A?
- in a certain sense, the likelihood function induced by the data A on a set of imprecise probability models is **bidimensional**: $lik(P) = \big(\underline{P}(A), \overline{P}(A)\big)$
- when the set of imprecise probability models is large enough, the "upper likelihood function" $\overline{lik}(P) = \overline{P}(A)$ is maximized by the **vacuous** model, while the "lower likelihood function" $\underline{lik}(P) = \underline{P}(A)$ is maximized by the **precise** ML estimate

maximum likelihood estimation

- basic question: given two imprecise probability models P_1, P_2 and the observed event A, which one of the two models was **best** in forecasting A?
- in a certain sense, the likelihood function induced by the data A on a set of imprecise probability models is **bidimensional**: $lik(P) = \big(\underline{P}(A), \overline{P}(A)\big)$
- when the set of imprecise probability models is large enough, the "upper likelihood function" $\overline{lik}(P) = \overline{P}(A)$ is maximized by the **vacuous** model, while the "lower likelihood function" $\underline{lik}(P) = \underline{P}(A)$ is maximized by the **precise** ML estimate
- the weighted geometric mean $lik_\alpha(P) = \overline{lik}(P)^\alpha \, \underline{lik}(P)^{1-\alpha}$ of the upper and lower likelihood functions is an interesting **compromise**, where $\alpha \in [0, 1]$ can perhaps be interpreted as a degree of optimism;

maximum likelihood estimation

- basic question: given two imprecise probability models P_1, P_2 and the observed event A, which one of the two models was **best** in forecasting A?
- in a certain sense, the likelihood function induced by the data A on a set of imprecise probability models is **bidimensional**: $lik(P) = \big(\underline{P}(A), \overline{P}(A)\big)$
- when the set of imprecise probability models is large enough, the "upper likelihood function" $\overline{lik}(P) = \overline{P}(A)$ is maximized by the **vacuous** model, while the "lower likelihood function" $\underline{lik}(P) = \underline{P}(A)$ is maximized by the **precise** ML estimate

- the weighted geometric mean $lik_\alpha(P) = \overline{lik}(P)^\alpha \, \underline{lik}(P)^{1-\alpha}$ of the upper and lower likelihood functions is an interesting **compromise**, where $\alpha \in [0, 1]$ can perhaps be interpreted as a degree of optimism; of particular interest is the case with $\alpha = \frac{1}{2}$, since

$$\frac{lik_{\frac{1}{2}}(P_1)}{lik_{\frac{1}{2}}(P_2)} = \sqrt{\frac{\overline{P}_1(A)}{\underline{P}_2(A)} \, \frac{\underline{P}_1(A)}{\overline{P}_2(A)}}$$

  is the geometric mean of the likelihood ratio most favorable to P_1 and the one most favorable to P_2

imprecise Bernoulli problem

- maximum *lik*$_\alpha$ estimates:

$$\alpha \leq \tfrac{1}{2} \quad \Rightarrow \quad \underline{\widehat{p}}_\alpha = \overline{\widehat{p}}_\alpha = \widehat{p} = \frac{n_1}{n_0 + n_1}$$

$$\alpha > \tfrac{1}{2} \quad \Rightarrow \quad \underline{\widehat{p}}_\alpha = \frac{(1-\alpha)\, n_1}{\alpha\, n_0 + (1-\alpha)\, n_1}, \quad \overline{\widehat{p}}_\alpha = \frac{\alpha\, n_1}{(1-\alpha)\, n_0 + \alpha\, n_1}$$

imprecise Bernoulli problem

- maximum *lik_α* estimates:

$$\alpha \leq \tfrac{1}{2} \quad \Rightarrow \quad \underline{\widehat{p}}_\alpha = \overline{\widehat{p}}_\alpha = \widehat{p} = \frac{n_1}{n_0 + n_1}$$

$$\alpha > \tfrac{1}{2} \quad \Rightarrow \quad \underline{\widehat{p}}_\alpha = \frac{(1-\alpha)\, n_1}{\alpha\, n_0 + (1-\alpha)\, n_1}, \quad \overline{\widehat{p}}_\alpha = \frac{\alpha\, n_1}{(1-\alpha)\, n_0 + \alpha\, n_1}$$

- $\alpha \geq \tfrac{1}{2} \quad \Rightarrow \quad [\operatorname{logit} \underline{\widehat{p}}_\alpha, \operatorname{logit} \overline{\widehat{p}}_\alpha] = [\operatorname{logit} \widehat{p} \pm \operatorname{logit} \alpha]$

imprecise Bernoulli problem

- maximum lik_α estimates:

$$\alpha \leq \tfrac{1}{2} \quad \Rightarrow \quad \underline{\widehat{p}}_\alpha = \overline{\widehat{p}}_\alpha = \widehat{p} = \frac{n_1}{n_0 + n_1}$$

$$\alpha > \tfrac{1}{2} \quad \Rightarrow \quad \underline{\widehat{p}}_\alpha = \frac{(1-\alpha)\, n_1}{\alpha\, n_0 + (1-\alpha)\, n_1}, \quad \overline{\widehat{p}}_\alpha = \frac{\alpha\, n_1}{(1-\alpha)\, n_0 + \alpha\, n_1}$$

- $\alpha \geq \tfrac{1}{2} \quad \Rightarrow \quad [\operatorname{logit} \underline{\widehat{p}}_\alpha, \operatorname{logit} \overline{\widehat{p}}_\alpha] = [\operatorname{logit} \widehat{p} \pm \operatorname{logit} \alpha]$

- in the case with k categories, the maximum lik_α estimates are precise if and only if $\alpha \leq \frac{1}{k}$