Generalized options, utilities and information measures

Reading circle on Bayesian theory, October 10th 2006

Jarno Vanhatalo

HELSINKI UNIVERSITY OF TECHNOLOGY
Laboratory of Computational Engineering

Outline

Exercise from last week

Generalized options and utilities

Generalized information measures

Intrinsic Estimation

Exercise from last week

Find the expectation of $f(x) = \frac{1}{1+x^2}$

- Answer:

Exercise from last week

Find the expectation of $f(x) = \frac{1}{1+x^2}$

- Answer:
  - $E[x] = \int_{-\infty}^{\infty} \frac{x}{1+x^2} \mathrm{d}x$.
  - Now, let $F(x) = \ln(1+x^2)$, in which case $\int_a^b \frac{x}{1+x^2} \mathrm{d}x = \frac{F(b)}{2} - \frac{F(b)}{2}$
  - Here $a = -\infty$ and $b = \infty$ and thus $E[x] = \frac{F(\infty)}{2} - \frac{F(-\infty)}{2}$, which is not defined

Generalized options and utilities

Motivation and preliminaries

- The chapter 3.2 extended the quantitative coherence theory (chapters 2.1-2.4) to the infinite domain
- The chapters 3.3 and 3.4 generalize the theory of options, utilities and information measure (chapters 2.5-2.7) to the infinite domain

Generalized options and utilities

Motivation and preliminaries

- The chapter 3.2 extended the quantitative coherence theory (chapters 2.1-2.4) to the infinite domain
- The chapters 3.3 and 3.4 generalize the theory of options, utilities and information measure (chapters 2.5-2.7) to the infinite domain

key features are

- Convergence in expected utility
- Definition of decision
- propositions that ensure the existence of sequencies of options *a* that converge in expected utility to the decision
  - needed in defining (finding) the decision

Generalized options and utilities

Generalized preferences

- Extension of the preference relation
- Maximization of expected utility for decision
  - also in the case of conditional

The value of information

- Optimal experimental design
- The value of additional information
- Expected value of perfect information
  - Additive decomposition

Generalized information measures

The utility of general probability distribution

- Score function
  - mapping u $\mathcal{Q} \times \mathbf{\Omega} \rightarrow \mathcal{R}$
- Prober score function
  - if expected score is maximized when $q_\omega(.|D) = p_\omega(.|D)$

Generalized information measures

The utility of general probability distribution

- Score function
  - mapping $u\ \mathcal{Q} \times \mathbf{\Omega} \to \mathcal{R}$
- Prober score function
  - if expected score is maximized when $q_\omega(.|D) = p_\omega(.|D)$
- quadratic score function
  - Proved to be proper
- Local score function
  - u is local if there exist functions u_ω such that $u(q_\omega(.|D)) = u_\omega(q(\omega|D))$
- Proper local score functions are logarithmic score function $A \log\left(q(\omega|D) + B(\omega)\right)$

Generalized information measures

Generalized approximation and discrepancy

- Expected loss in probability reporting
  - Based on the logarithmic score function
- Discrepancy of an approximation
  - the expected loss in approximating p with $\hat{p}$

Generalized information

- Information from data
  - Is the expected utility of data (given the prior information) in the sense of logarithmic score function
- Expected information from an experiment

Intrinsic Estimation

- The final result of Bayesian inference is the posterior distribution of the quantity of interest
- However, often it is necessary to be able to give a good point estimate
  - Good estimate should be objective and invariant under one-to-one transformations
  - How to find a good estimate?
- The problem can be formulated as a decision problem
  - let $p(\mathbf{x}\,|\theta)$ be a probability model assumed to describe data $\mathbf{x}$,
  - now action space is $A = \{\theta^e\}$, where θ^e is possible point estimate

Intrinsic Estimation

Loss function

- Let $l(\theta^e, \theta^a)$ be a loss function measuring the consecuence of estimating θ^a (the actual true parameter value) with θ^e
- conventional loss functions compare θ^e to θ^a
  - for example squared, zero-one and absolute error loss
  - Problems in invariance under one-to-one transformations and generalization into higher dimensions
- Intrinsic loss function compares $p(\mathbf{x}\,|\theta^e)$ to $p(\mathbf{x}\,|\theta^a)$
  - for example, the one based on logarithmic score function mentioned earlier (Kullback-Leibler divergence)

$$k(\theta_2|\theta_1) = \int_X p(\mathbf{x}\,|\theta_1) \log \frac{p(\mathbf{x}\,|\theta_1)}{p(\mathbf{x}\,|\theta_2)} \mathrm{d}\,\mathbf{x} \tag{1}$$

  - may diverge when the support of approximation and true distribution are different
  - is not symmetric

Intrinsic Estimation

- Intrinsic discrepancy loss

$$\delta_X(\theta_1, \theta_2) = \min\{k(\theta_2|\theta_1), k(\theta_1|\theta_2)\} \tag{2}$$

- does not diverge even if the support of approximation and true distribution are different
- Symmetric, non-negative, invariant under one-to-one transformations
- does not represent local (lograrithmic form) score function

Intrinsic Estimation

- Intrinsic discrepancy loss

$$\delta_X(\theta_1, \theta_2) = \min\{k(\theta_2|\theta_1), k(\theta_1|\theta_2)\} \tag{2}$$

- does not diverge even if the support of approximation and true distribution are different
- Symmetric, non-negative, invariant under one-to-one transformations
- does not represent local (lograrithmic form) score function
- Intrinsic estimator

$$\theta^*(\mathbf{x}) = \arg\min_{\theta^e \in \Theta} \int_\Theta \delta(\theta^e, \theta)\pi_\delta(\theta|\,\mathbf{x})d\theta \tag{3}$$

- $\pi_\delta(\theta|\,\mathbf{x})d\theta$ is the reference posterior obtained using non-informative prior $\pi(\theta)$
  - Ensures (some kind of) objectivity

Intrinsic Estimation

- For more information see:
  - Bernardo and Juárez. Intrinsic Estimation, Bayesian Statistics 7, 2003
  - http://www.uv.es/ bernardo/BernardoJuarez.pdf

Excersize

Let $\mathbf{x} = \{x_1, ..., x_n\}$, be a random sample from from the uniform distribution $\mathrm{Un}(x|0,\theta) = \theta^{-1}, 0 < x < \theta$. Let $\mathrm{Un}(x|0,\theta_1) = \theta_1^{-1}, 0 < x < \theta_1$ be the approximate distribution.

- What is the discrepancy of an approximation by definition 3.20 (Kullback-Leibler divergence)?
- What is the intrinsic discrepancy loss (2) of an approximation?
- How did the above results illustrate the behaviour of these different discrepancies in the case of different supports of an approximation and the true distribution?