Bayesian nonparametric tests based on the Imprecise Dirichlet process

Francesca Mangili, Alessio Benavoli, Giorgio Corani, and Marco Zaffalon

Istituto "Dalle Molle" di Studi sull'Intelligenza Artificiale
Lugano (Switzerland)

http://www.ipg.idsia.ch/

ISIPTA 2015, Pescara

Bayesian nonparametric tests based on the Imprecise Dirichlet process

Francesca Mangili, Alessio Benavoli, Giorgio Corani, and Marco Zaffalon

IDSIA - Istituto Dalle Molle di studi sull'Intelligenza Artificiale, USI - SUPSI, Lugano - Switzerland

Introduction

Frequentist procedures are usually adopted in nonparametric hypothesis tests. However,

- p-value are easily misinterpreted;
- no principled way for making decisions (α=0.05 or α=0.01).
- Bayesian tests compute the posterior probability of the hypotheses and allow taking optimal decisions.

Need of Bayesian alternatives to frequentist nonparametric tests.

Contradiction of Bayesian nonparametrics:

- wish to minimize assumptions;
- require infinitesimally minute details for prior specification.

We propose a default (near)-ignorance mechanism for prior elicitation.

Dirichlet Process (DP)

- Assigns a distribution over probability distributions.
- It provides a Bayesian justification of some traditional nonparametric estimators.

Definition:

- Let $P \sim \mathrm{Dp}(s, g_0)$, with:
  - s = prior strength;
  - g_0 = prior base probability measure.
- Given a finite partition $B_1, B_2, \dots, B_m$ of $\mathbb{X}$, it holds $(P(B_1), \dots, P(B_m)) \sim \mathrm{Dir}(s\, g_0(B_1), \dots, s\, g_0(B_m))$.
- $\mathcal{E}[P(B_i)] = g_0(B_i)$
- $Var[P(B_i)] = \dfrac{g_0(B_i)[1 - g_0(B_i)]}{s+1}$

Conjugacy:

Let $X_1, X_2, \dots$ be n samples from P

$$P|X_1, X_2, \dots \sim \mathrm{Dp}(s+n, \underbrace{\frac{s}{s+n} g_0 + \frac{1}{n+s} \sum_{i=1}^{n} \delta_{X_i}}_{g_n})$$

Imprecise DP (IDP)

Definition

An IDP is the set of all DP with s fixed and g_0 free to vary in the set of all probability measures $\mathbb{P}$:

$$IDP: \quad \{Dp(s, g_0), \ \ g_0 \in \mathbb{P}\}$$

$$0 = \underline{g}_0(B_i) \leq \mathcal{E}[P(B_i)] \leq \overline{g}_0(B_i) = 1$$

No prior information about P

$$f(x_l) \leq f(x) \leq f(x_u)$$

$$f(x_l) = E[f|P = \delta_{x_l}] \leq \mathcal{E}[E[f]] \leq E[f|P = \delta_{x_u}] f(x_u)$$

No prior information about $f(x)$

Choice of the prior strength s

- Fix the fraction of prior imprecision after one observation OR
- Fix the minimum number of observation necessary to make a decision.

Hypothesis testing

A simple example - the sign test

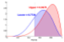
$H_0: \ P(X<0) \leq 0.5$, $H_1: \ P(X<0) > 0.5$

$n_l := \#(X_i < 0)$, $\quad n_g := \#(X_i \geq 0)$

Lower: $\underline{\mathcal{P}}(H_1) = \int_{0.5}^{1} \mathrm{Beta}(n_l, s+n_g)$

Upper: $\overline{\mathcal{P}}(H_1) = \int_{0.5}^{1} \mathrm{Beta}(s+n_l, n_g)$

Example:
$n_l = 5$,
$n_g = 2$,
$s = 1$



Decision making

L_0: loss accepting H_1 (action a_1) if H_0 true

L_1: loss accepting H_0 (action a_0) if H_1 true

Minimum risk: accept H_1 if

$$\text{Expected loss}|a_1 < \text{Expected loss}|a_0$$
$$\Rightarrow \mathrm{Prob}(H_1) > \frac{L_0}{L_0+L_1} = 1 - \alpha.$$

What if $\overline{\mathcal{P}}(H_1) > 1-\alpha$ but $\underline{\mathcal{P}}(H_1) < 1-\alpha$?

The decision is prior-dependent. No robust decision can be made.

"$Y > X$" at 95% — "$Y \not> X$" at 95% — "Indeterminate" at 95% — "$Y \not> X$" at 95%

Advantages

Tractability

- Simple elicitation.
- Sampling is easier than standard sampling from the DP.
  Samples $P^{(k)}$ from $Dp(s+n, g_n)$
  $$P^{(k)} = w_0 \delta_{X_0} + \sum_{i=1}^{n} w_i \delta_{X_i}$$
  with $(w_0, w_1, \dots, w_n) \sim \mathrm{Dir}(s, \overbrace{1, \dots, 1}^{n})$

Asymptotic consistency

The IDP tests are asymptotically consistent. Frequentist ones can have inconsistencies.

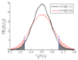
Example: Wilcoxon rank-sum test

Samples: $X_1, X_2, \dots \sim F_X$; $Y_1, Y_2, \dots \sim F_Y$.

Hypothesis: $H_0: \ F_X = F_Y$; $H_1: F_X \neq F_Y$.

Statistic: $U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I_{\{X_i < Y_j\}}$.

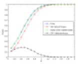
- 9% rejections with $\alpha = 0.05$,$n = 25$, $X \sim N(0,1)$.
- The power $\lll 1$ even for large n_1, n_2.
- To avoid this situation, it is assumed that F_X and F_Y have the same shape.



Indecision

We empirically observed that some frequentist test can virtually behaves as a random guesser when the IDP test is indeterminate.

- This shows that prior-dependent instances are critical.
- It makes sense to suspend the decisions in those instances.



Wilcoxon sum-rank test. $n = 20$, $\alpha = 0.1$

The IDP statistical package

The IDP project (in progress) has developed IDP version of:

1. Wilcoxon rank-sum test;
2. Wilcoxon signed test;
3. Sign test;
4. Analysis of survival data.

IDP

http://ipg.idsia.ch/software/IDP.php