Beta and Dirichlet sub-Gaussianity
Olivier Marchal, Julyan Arbel

To cite this version:
Olivier Marchal, Julyan Arbel. Beta and Dirichlet sub-Gaussianity. Bayesian learning theory for complex data modelling Workshop, Sep 2018, Grenoble, France. hal-01950665

HAL Id: hal-01950665
https://hal.archives-ouvertes.fr/hal-01950665
Submitted on 11 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract

**Optimal proxy variance** \( \sigma_{\text{opt}}^2 \) for the sub-Gaussianity of Beta distribution, improves recent conjecture \( \sigma_0^2 \) by Elder (2016): \( \sigma_{\text{opt}}^2 \leq \sigma_0^2 = \frac{1}{\log(1+\alpha)} \).

- Provide different proof techniques for (i) symmetrical case \( \alpha = \beta \): direct coef. comparison of entire series (ii) non-symmetrical case \( \alpha \neq \beta \): ordinary differential equation satisfied by moment-generating function, aka confluent hypergeometric function
- Derive optimal proxy variance for Dirichlet.

Sub-Gaussianity & optimal proxy variance

- A random variable \( X \) with finite mean \( \mu = \mathbb{E}[X] \) is sub-Gaussian if there is a positive number \( \sigma^2 \) (called a proxy variance) such that:
  \[
  \mathbb{E}[\exp(x(X - \mu))] \leq \exp \left( \frac{x^2\sigma^2}{2} \right) \text{ for all } x \in \mathbb{R}.
  \]  
  (1)

- If \( X \) is sub-Gaussian, the **optimal proxy variance** is:
  \[
  \sigma_{\text{opt}}^2(X) = \min\{\sigma^2 \geq 0 \text{ such that } X \text{ is } \sigma^2\text{-sub-Gaussian}\}.
  \]

Optimal proxy variance illustrated & compared

- **Top**: varying \( \alpha \) on x-axis, \( \alpha + \beta = 1 \) (left) and \( \alpha + \beta = 10 \) (right).
- **Left**: \( \alpha, \beta \in [0,2,4,\ldots] \)
  - Magenta: \( \sigma_{\text{opt}}^2(\alpha, \beta) \)
  - Green: \( \text{Var}[	ext{Beta}(\alpha, \beta)] \)
  - Black (dots): \( \sigma_0^2 = \frac{1}{\log(1+\alpha)} \)

Optimal proxy variance for the Beta

**Theorem** Beta(\( \alpha, \beta \)) is \( \sigma_{\text{opt}}^2(\alpha, \beta) \)-sub-Gaussian with:

\[
\sigma_{\text{opt}}^2(\alpha, \beta) = \frac{\alpha}{\alpha + \beta} \left( \frac{F_1(\alpha + 1; \alpha + \beta + 1; x_0)}{F_1(\alpha; \alpha + \beta, x_0)} - 1 \right)
\]

where \( x_0 \) is the unique solution of the equation

\[
\log(F_1(\alpha; \alpha + \beta; x_0)) = \frac{\alpha x_0}{2(\alpha + \beta)} \left( 1 + \frac{F_1(\alpha + 1; \alpha + \beta + 1; x_0)}{F_1(\alpha; \alpha + \beta, x_0)} \right).
\]

Simple explicit upper bound to \( \sigma_{\text{opt}}^2(\alpha, \beta) \) is \( \sigma_0^2(\alpha, \beta) = \frac{1}{\alpha + \beta} \).

Var[Beta(\( \alpha, \beta \))] \leq \sigma_{\text{opt}}^2(\alpha, \beta) \leq \sigma_0^2 \text{ (strict when } \alpha \neq \beta \text{)}

Sketch of proof

- \( \alpha = \beta \): direct coef. comparison of entire series representations of (1)
- \( \alpha \neq \beta \): study the MGF aka confluent hypergeometric function or Kummer’s function

\[
y(x) \overset{\text{def}}{=} \mathbb{E}[\exp(xX)] = \sum_{\alpha} \Gamma(\alpha + j) / \Gamma(\alpha + \beta + j) x^\alpha.
\]

using the ordinary differential equation it satisfies

\[
x y'(x) + (\alpha + \beta - x) y'(x) - \alpha y(x) = 0
\]

via the difference

\[
u_t(x) \overset{\text{def}}{=} \mathbb{E}[\mu x + \sigma^2 x^2 / 2] - \mathbb{E}[\exp(xX)], \sigma_t^2 = t \text{Var}[X] + (1 - t)\sigma_0^2
\]

- For \( t = 0 (\sigma_0^2) \), dotted black curve remains > 0
- For \( t = t_{\text{opt}} (\sigma_{\text{opt}}^2) \), magenta curve has minimum = 0 at \( x_0 \)
- For \( t = 1 (\text{Var}[X]) \), dashed green curve has negative second derivative at \( x = 0 \), directly negative around 0
- For \( t_{\text{union opt}} \in (t_{\text{opt}}, 1) \), orange, dash and dots curve is first positive, then negative, and positive again

Looking for connections with literature

- Compare with Bernoulli setting (Berend and Kontorovich, 2013; Buldygin and Moskovtsova, 2013)
- Possible links of our non-uniform sub-Gaussian result with
  - transportation inequalities
  - logarithmic Sobolev inequalities

References


