Beta and Dirichlet sub-Gaussianity

Olivier Marchal, Julyan Arbel

Beta and Dirichlet sub-Gaussianity

Olivier Marchal, Univ. Lyon, Univ. Monnet, Institut Camille Jordan, France

Julyan Arbel, Inria, Mistis, Grenoble, France, www.julyanarbel.com

Abstract

- **Optimal proxy variance** σ^2_{opt} for the sub-Gaussianity of Beta distribution, improves recent conjecture σ^2_0 by Elder (2016): $\sigma^2_{opt} \leq \sigma^2_0 = \frac{1}{4(\alpha+\beta+1)}$.
- 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 σ^2 (called a proxy variance) such that:

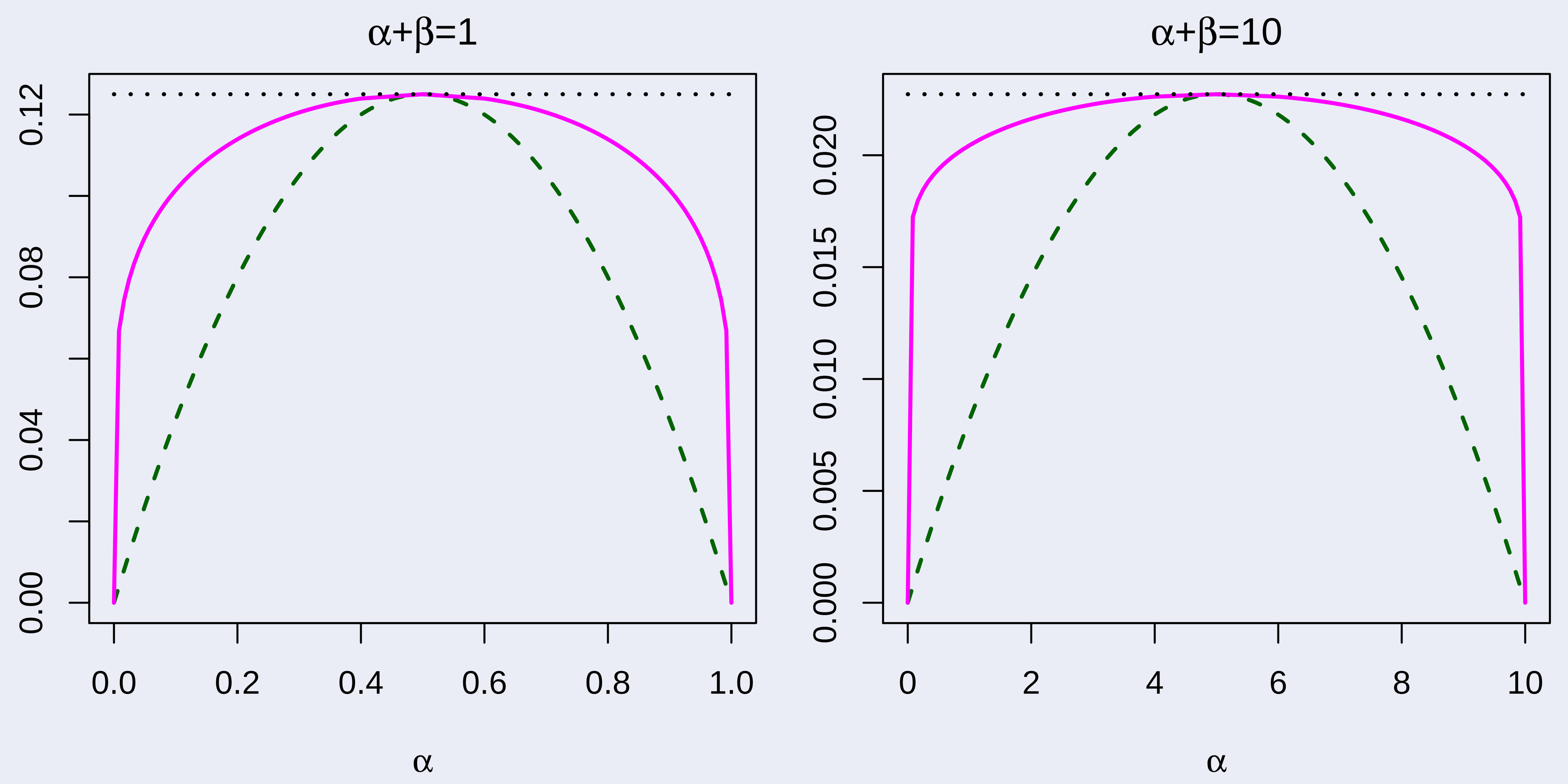
$$\mathbb{E}[\exp(x(X-\mu))] \leq \exp\left(\frac{x^2\sigma^2}{2}\right) \text{ for all } x \in \mathbb{R}. \qquad (1)$$

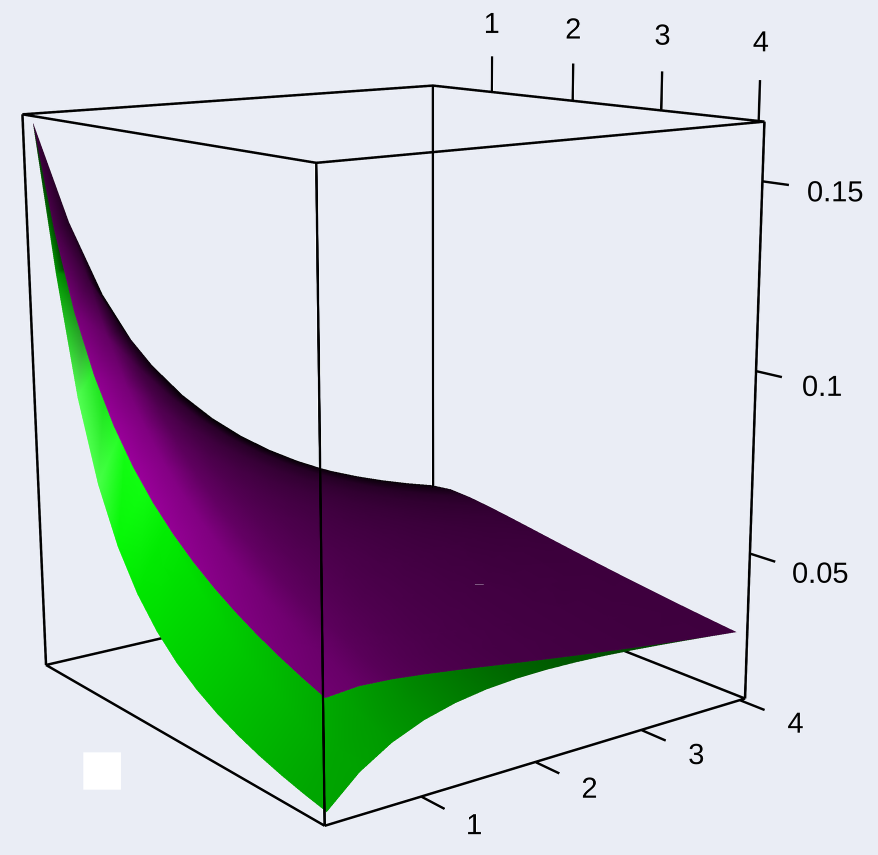
- If X is sub-Gaussian, the **optimal proxy variance** is:

$$\sigma^2_{\text{opt}}(X) = \min\{\sigma^2 \geq 0 \text{ such that } X \text{ is } \sigma^2\text{-sub-Gaussian}\}.$$

Variance lower bounds optimal proxy variance: $\text{Var}[X] \leq \sigma^2_{\text{opt}}(X)$. When $\sigma^2_{\text{opt}}(X) = \text{Var}[X]$, X is said to be strictly sub-Gaussian.

Optimal proxy variance illustrated & compared





- **Top**: varying α on x-axis, $\alpha+\beta = 1$ (left) and $\alpha+\beta = 10$ (right).
- **Left**: $\alpha, \beta \in [0.2, 4]$.
  - Magenta: $\sigma^2_{\text{opt}}(\alpha,\beta)$
  - Green: Var[Beta(α,β)]
  - Black (dots): $\sigma^2_0 = \frac{1}{4(\alpha+\beta+1)}$

Looking for connections with literature

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

Optimal proxy variance for the Beta

Theorem Beta(α,β) is $\sigma^2_{\text{opt}}(\alpha,\beta)$-sub-Gaussian with:

$$\begin{cases} \sigma^2_{\text{opt}}(\alpha,\beta) = \dfrac{\alpha}{(\alpha+\beta)x_0}\left(\dfrac{{}_1F_1(\alpha+1;\alpha+\beta+1;x_0)}{{}_1F_1(\alpha;\alpha+\beta;x_0)} - 1\right) \\ \quad\text{where } x_0 \text{ is the unique solution of the equation} \\ \log({}_1F_1(\alpha;\alpha+\beta;x_0)) = \dfrac{\alpha x_0}{2(\alpha+\beta)}\left(1 + \dfrac{{}_1F_1(\alpha+1;\alpha+\beta+1;x_0)}{{}_1F_1(\alpha;\alpha+\beta;x_0)}\right). \end{cases}$$

Simple explicit upper bound to $\sigma^2_{\text{opt}}(\alpha,\beta)$ is $\sigma^2_0(\alpha,\beta) = \frac{1}{4(\alpha+\beta+1)}$:

$$\text{Var}[\text{Beta}(\alpha,\beta)] \leq \sigma^2_{\text{opt}}(\alpha,\beta) \leq \sigma^2_0 \text{ (strict when } \alpha \neq \beta)$$

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

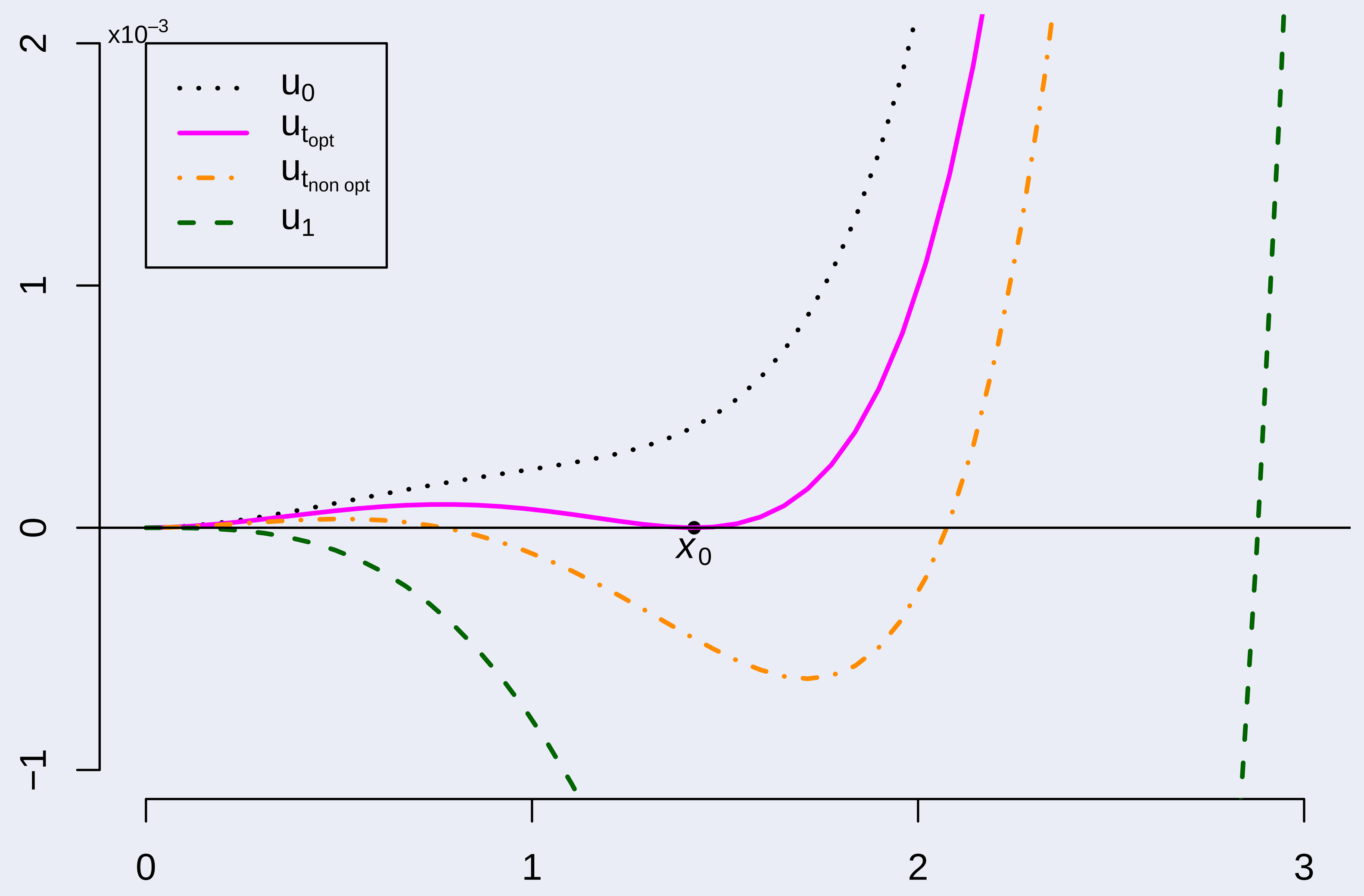
$$y(x) \overset{\text{def}}{=} \mathbb{E}[\exp(xX)] = {}_1F_1(\alpha;\alpha+\beta;\lambda) = \sum_{j=0}^{\infty} \frac{\Gamma(\alpha+j)\Gamma(\alpha+\beta)}{(j!)\Gamma(\alpha)\Gamma(\alpha+\beta+j)} x^j.$$

using the ordinary differential equation it satisfies

$$xy''(x) + (\alpha+\beta-x)y'(x) - \alpha y(x) = 0$$

via the difference

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



- For $t = 0$ (σ_0^2), dotted black curve remains > 0
- For $t = t_{\text{opt}}$ (σ^2_{opt}), magenta curve has minimum $= 0$ at x_0
- For $t = 1$ (Var[X]), dashed green curve has negative second derivative at $x = 0$, directly negative around 0
- For $t_{\text{non opt}} \in (t_{\text{opt}}, 1)$, orange, dash and dots curve is first positive, then negative, and positive again

References

Berend, D. and Kontorovich, A. (2013). On the concentration of the missing mass. *Electronic Communications in Probability*, 18(3):1–7.

Buldygin, V. V. and Moskvichova, K. (2013). The sub-Gaussian norm of a binary random variable. *Theory of probability and mathematical statistics*, 86:33–49.

Elder, S. (2016). Bayesian adaptive data analysis guarantees from subgaussianity. *arXiv preprint: arXiv:1611.00065*.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30.

Marchal, O. and Arbel, J. (2017). On the sub-Gaussianity of the Beta and Dirichlet distributions. *Under revision.*