



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** σ_{opt}^2 for the sub-Gaussianity of Beta distribution, improves recent conjecture σ_0^2 by Elder (2016): $\sigma_{opt}^2 \leq \sigma_0^2 = \frac{1}{4(\alpha+\beta+1)}$.

- 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.

Optimal proxy variance for the Beta

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

$$\left\{ \begin{array}{l} \sigma_{opt}^2(\alpha, \beta) = \frac{\alpha}{(\alpha + \beta)x_0} \left(\frac{{}_1F_1(\alpha + 1; \alpha + \beta + 1; x_0)}{{}_1F_1(\alpha; \alpha + \beta; x_0)} - 1 \right) \\ \text{where } x_0 \text{ is the unique solution of the equation} \\ \log({}_1F_1(\alpha; \alpha + \beta; x_0)) = \frac{\alpha x_0}{2(\alpha + \beta)} \left(1 + \frac{{}_1F_1(\alpha + 1; \alpha + \beta + 1; x_0)}{{}_1F_1(\alpha; \alpha + \beta; x_0)} \right). \end{array} \right.$$

Simple explicit upper bound to $\sigma_{opt}^2(\alpha, \beta)$ is $\sigma_0^2(\alpha, \beta) = \frac{1}{4(\alpha+\beta+1)}$:

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

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:

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

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

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

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

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

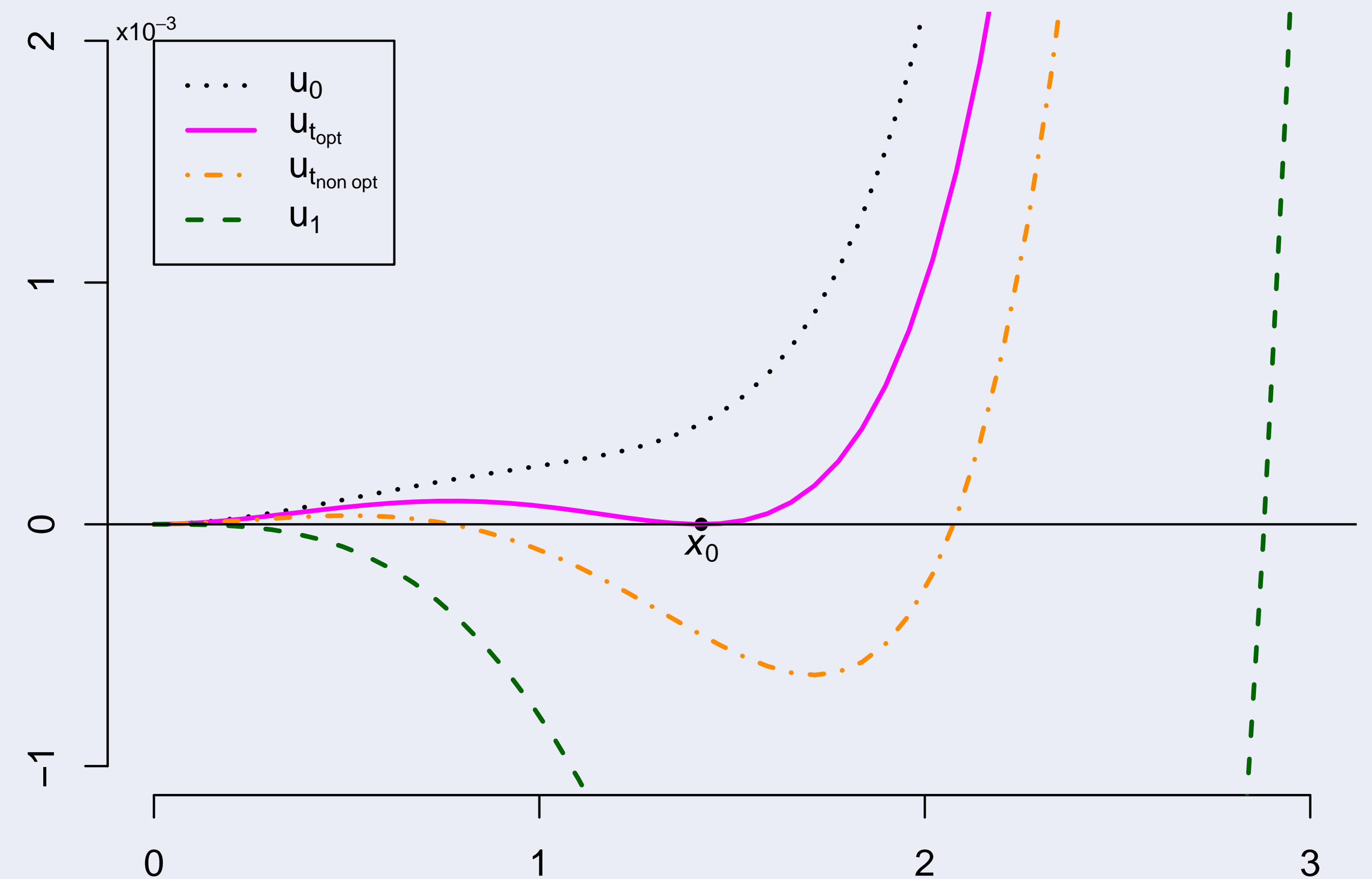
$$y(x) \stackrel{\text{def}}{=} \mathbb{E}[\exp(xX)] = {}_1F_1(\alpha; \alpha + \beta; \lambda) = \sum_{j=0}^{\infty} \frac{\Gamma(\alpha + j)\Gamma(\alpha + \beta)}{(j!) \Gamma(\alpha)\Gamma(\alpha + \beta + j)} x^j.$$

using the ordinary differential equation it satisfies

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

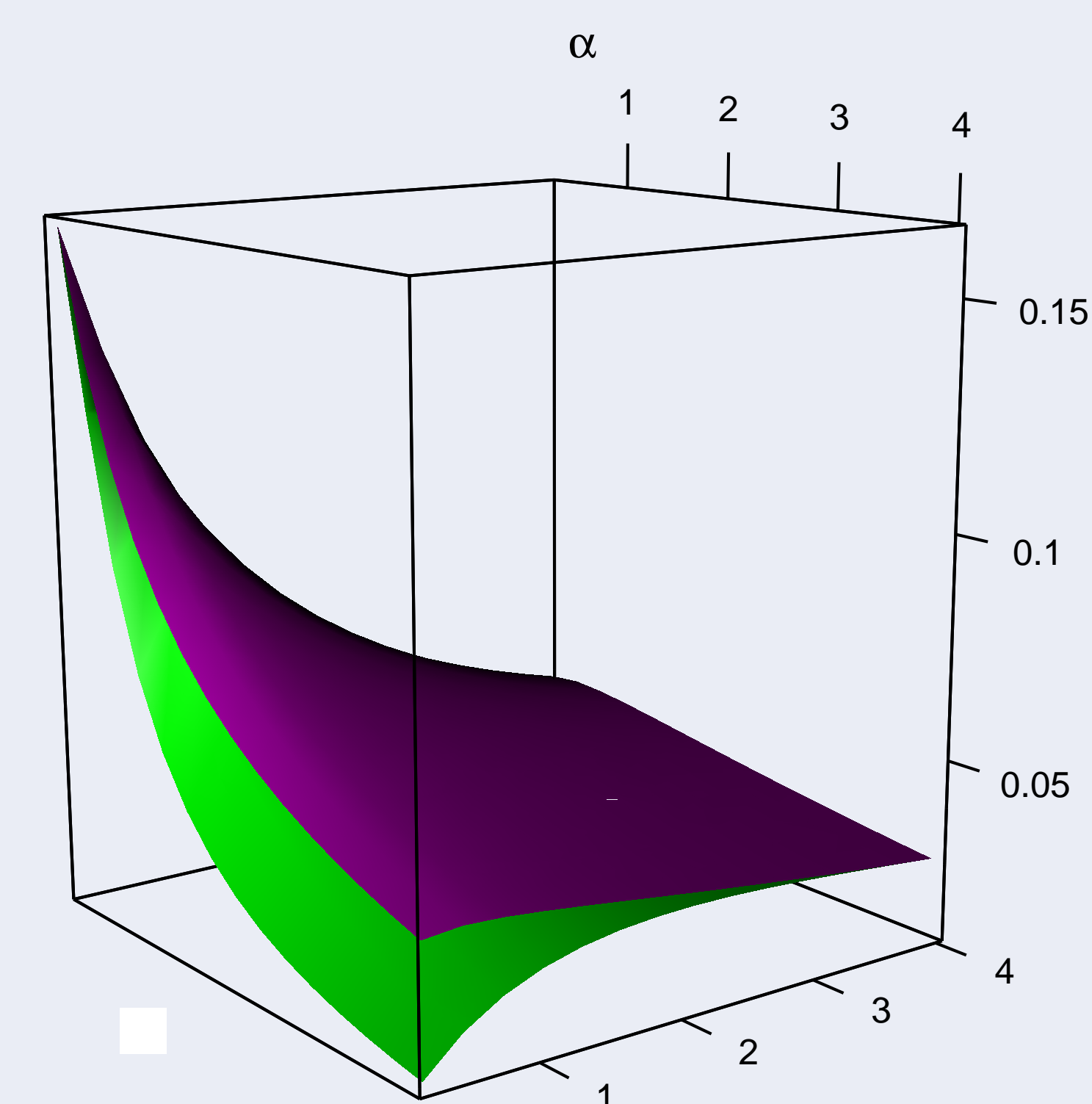
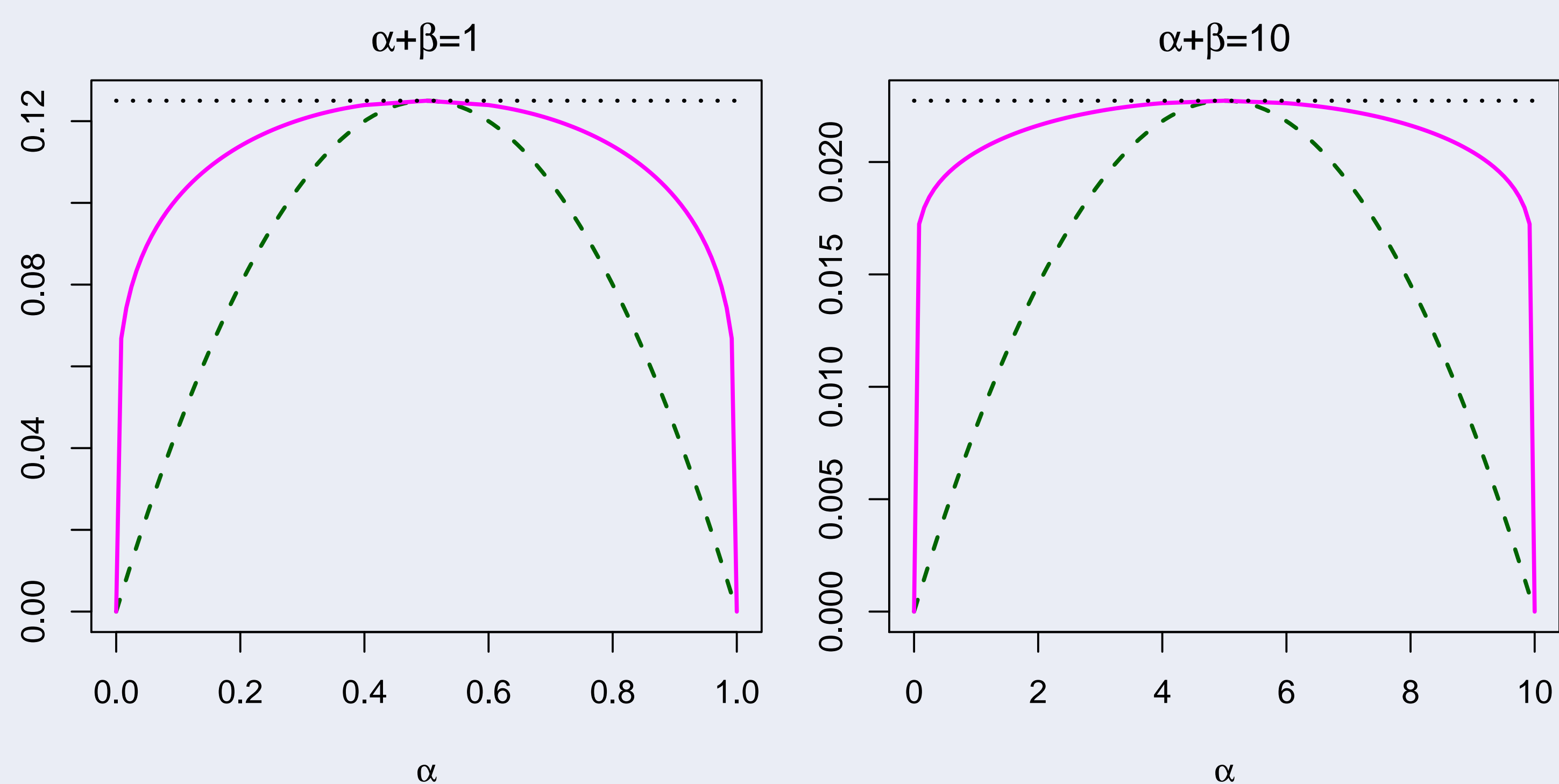
via the difference

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



- For $t = 0$ (σ_0^2), dotted black curve remains > 0
- For $t = t_{opt}$ (σ_{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 $t_{non\ opt} \in (t_{opt}, 1)$, orange, dash and dots curve is first positive, then negative, and positive again

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: $\sigma_{opt}^2(\alpha, \beta)$
 - Green: $\text{Var}[\text{Beta}(\alpha, \beta)]$
 - Black (dots): $\sigma_0^2 = \frac{1}{4(\alpha+\beta+1)}$

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*.

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**