Stick-Breaking Processes' Divergence-Based Prior Analysis

BNP Grupo de Trabajo
IIMAS - UNAM

27 de marzo de 2023

Trabajo conjunto con

Mario Diaz y Ramsés H. Mena

Motivación

Estudiar la divergencia a priori de diferentes stick-breaking priors

$$P = \sum_{n=1}^{\infty} w_n \delta_{z_n}$$

$$w_1 = v_1, \qquad w_n = v_n \prod_{j<n} (1 - v_j)$$

Ejemplos

1. Variables independientes e.g. proceso Dirichlet

$$v_i \sim \text{Be}(1,\theta)$$

2. Variables dependientes e.g. proceso geométrico

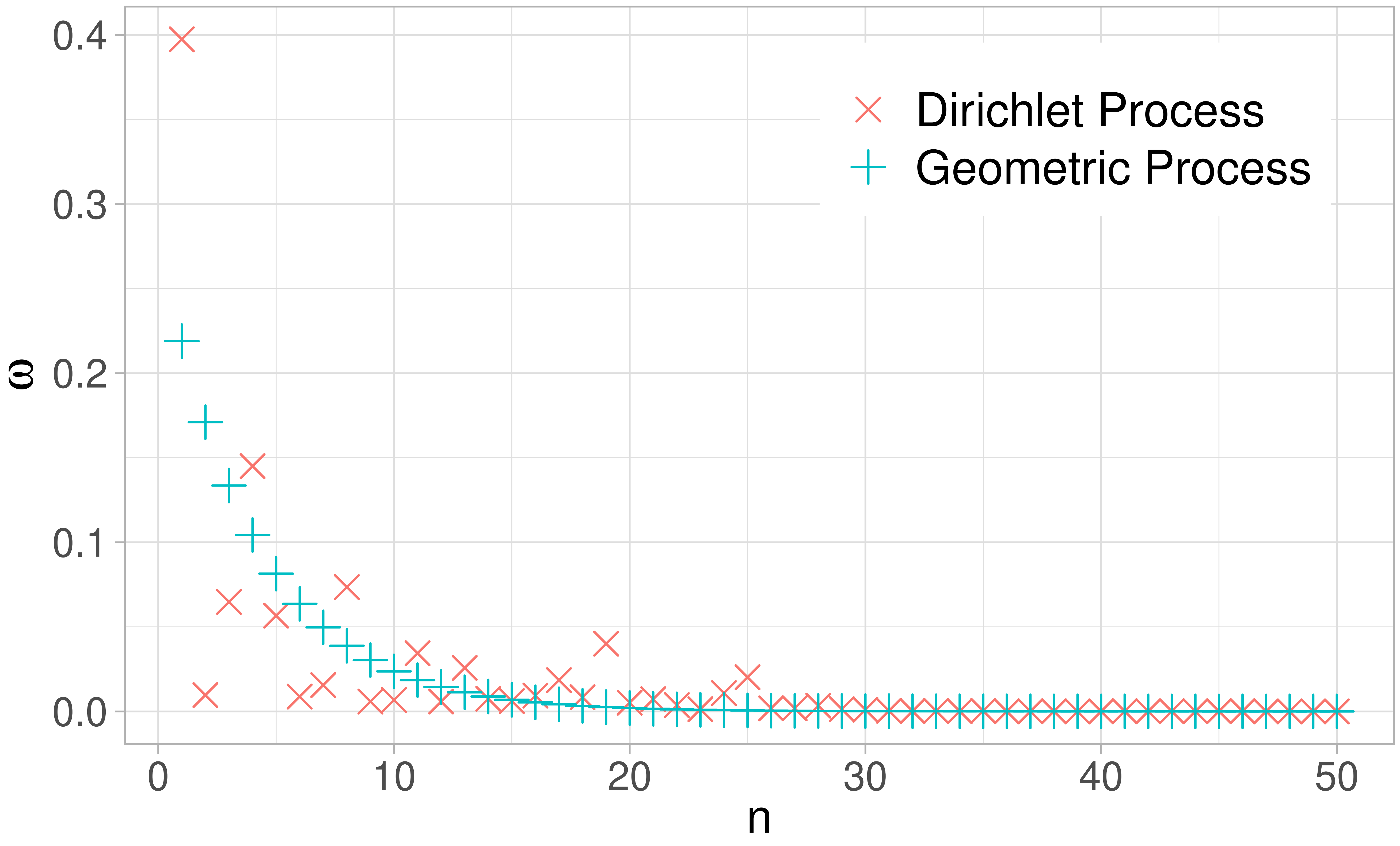
$$v \sim \text{Be}(a,b)$$

3. Variables intercambiables e.g. Dirichlet-driven length variables

$$\begin{aligned} v_i \,|\, \nu &\sim \nu \\ \nu &\sim \text{DP}(\beta,\nu_0)\,. \end{aligned}$$

Divergencia KL del proceso Dirichlet con respecto al proceso geométrico

1. Asumimos que las localizaciones son las mismas y solo difieren en los pesos.



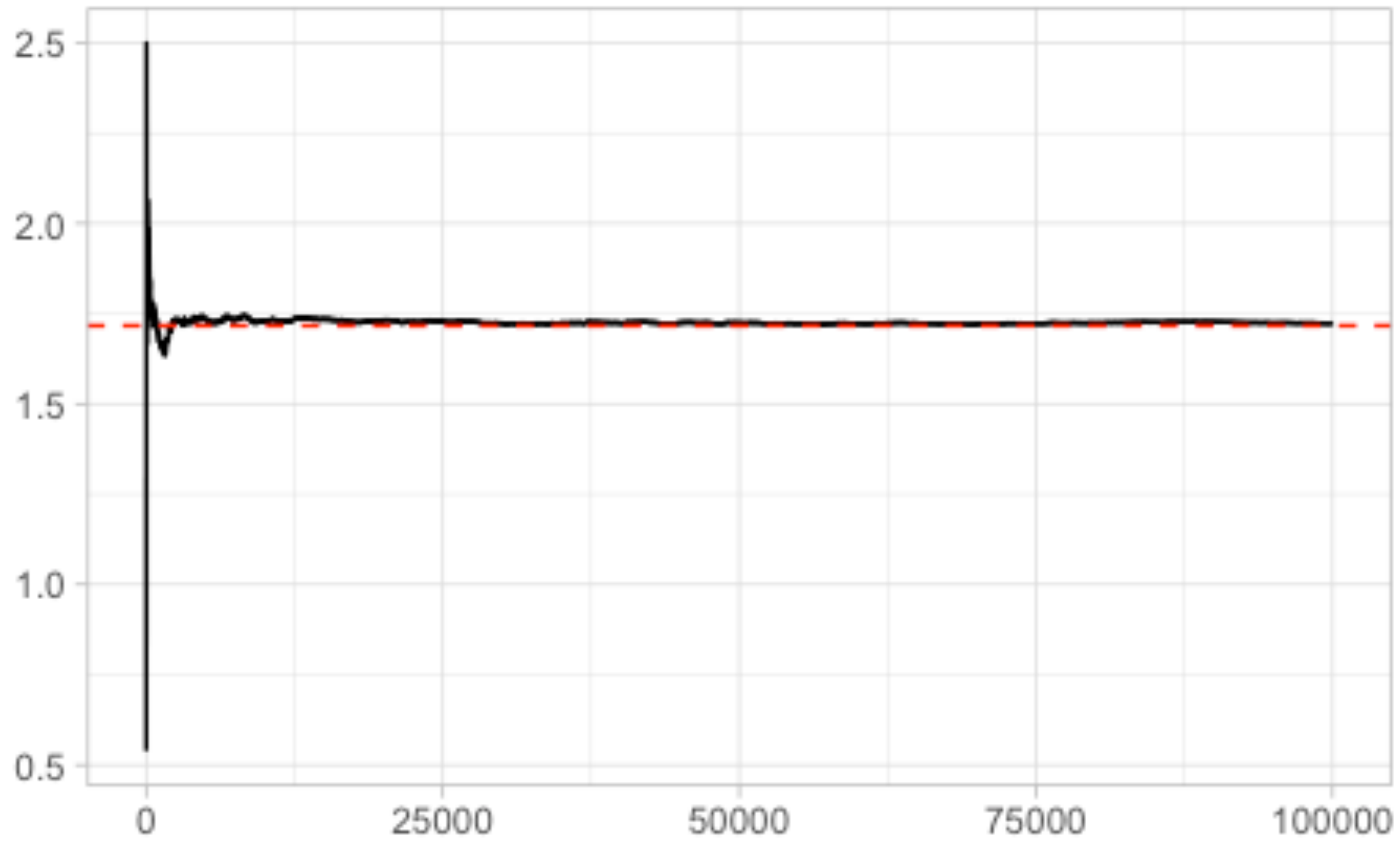
Divergencia KL del proceso Dirichlet con respecto al proceso geométrico

$$D_{\mathsf{KL}}(P\,||\,P') = \sum_{n\geq 1} w_n \log\left(\frac{w_n}{w'_n}\right) = \sum_{n\geq 1}\left[\prod_{j<n}(1-v_j)\right] d(v_n\,||\,v)$$

$$d(v_n\,||\,v) = v_n \log\left(\frac{v_n}{v}\right) + (1-v_n)\log\left(\frac{1-v_n}{1-v}\right)$$

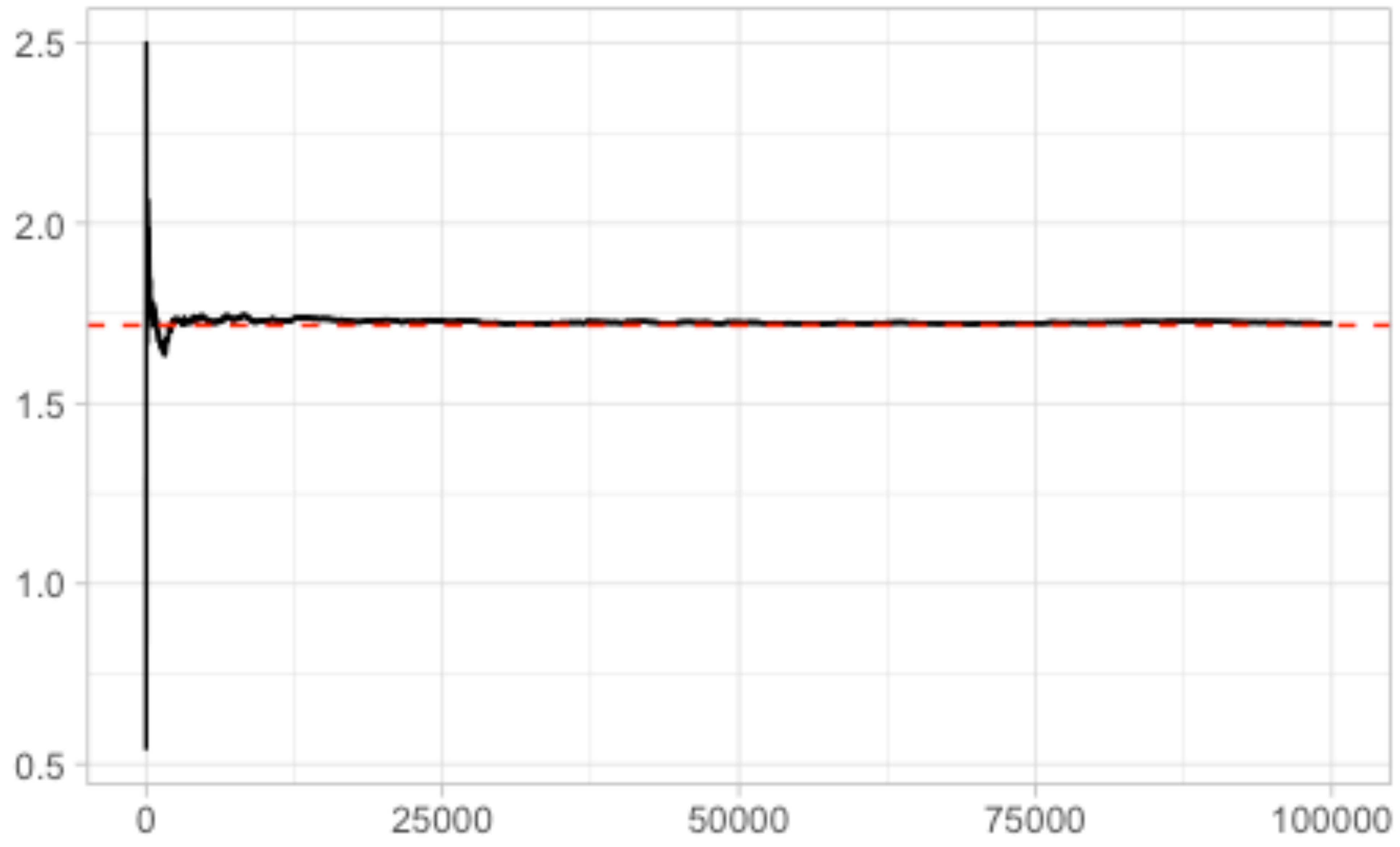
Propiedades

$$\mathbb{E}\left[D_{\mathrm{KL}}(P\,||\,P')\right] = (\theta+1)\mathbb{E}[d(v_1\,||\,v)]$$



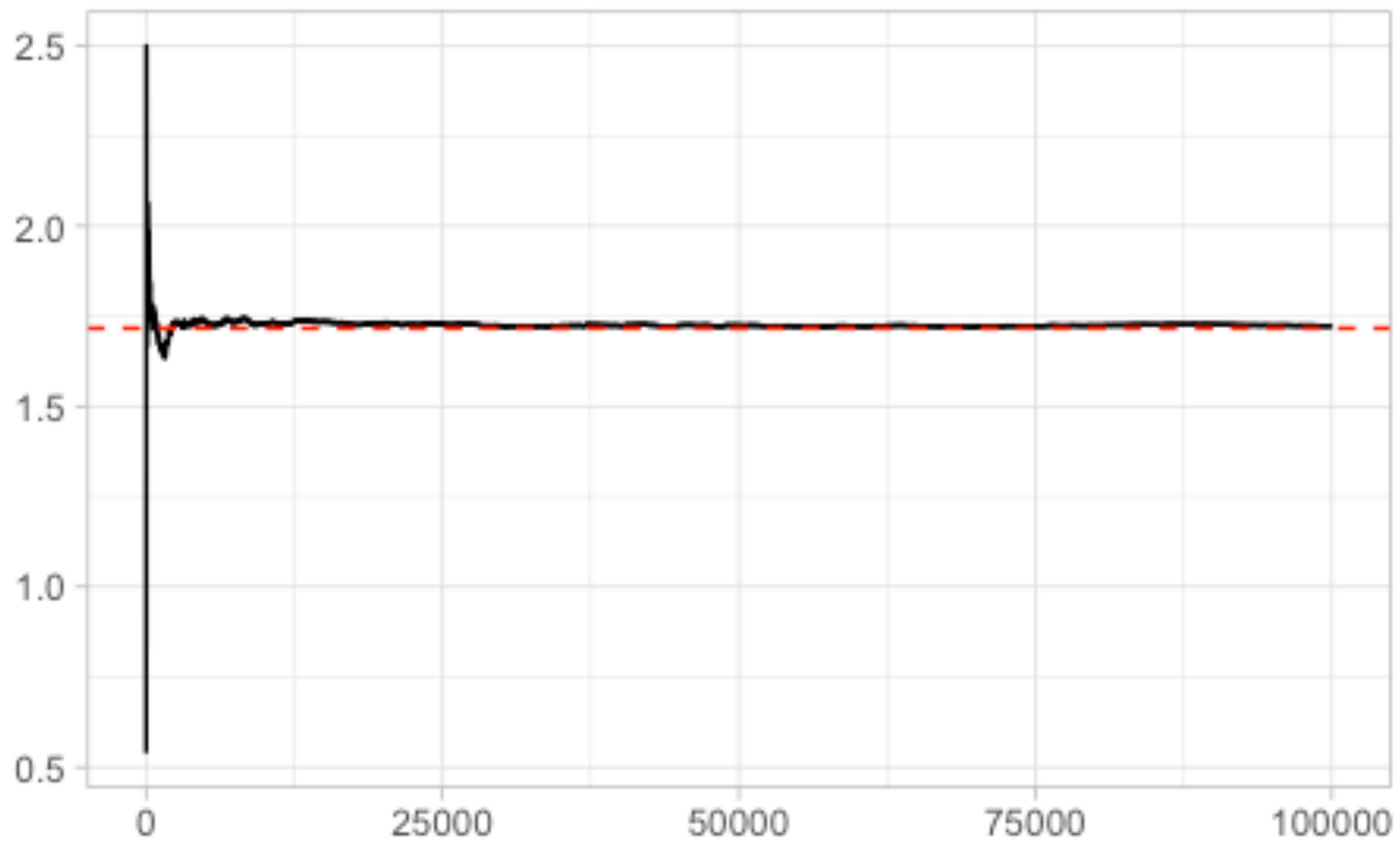
Propiedades

$$\mathbb{E}\left[D_{\mathsf{KL}}(P\,||\,P')\right] = (\theta + 1)\mathbb{E}[d(v_1\,||\,v)]$$



OBS. Si $a = 1$ y $b = \theta$ entonces $\mathbb{E}\left[D_{\mathsf{KL}}(P\,||\,P')\right] = 1$

Propiedades



$$\mathbb{E}\left[D_{\mathsf{KL}}(P\,||\,P')\right] = (\theta+1)\mathbb{E}[d(v_1\,||\,v)]$$

OBS. Si $a = 1$ y $b = \theta$ entonces $\mathbb{E}\left[D_{\mathsf{KL}}(P\,||\,P')\right] = 1$

$$\mathbb{E}\left[D_{KL}(P\,||\,P')^2\right] = \frac{(\theta+2)}{2}\mathbb{E}[d^2(v_1\,||\,v)] + (\theta+1)(\theta+2)\mathbb{E}[(1-v_1)d(v_1\,||\,v)d(v_2\,||\,v)]$$

Divergencia KL del proceso Dirichlet con respecto al proceso geométrico 2.0

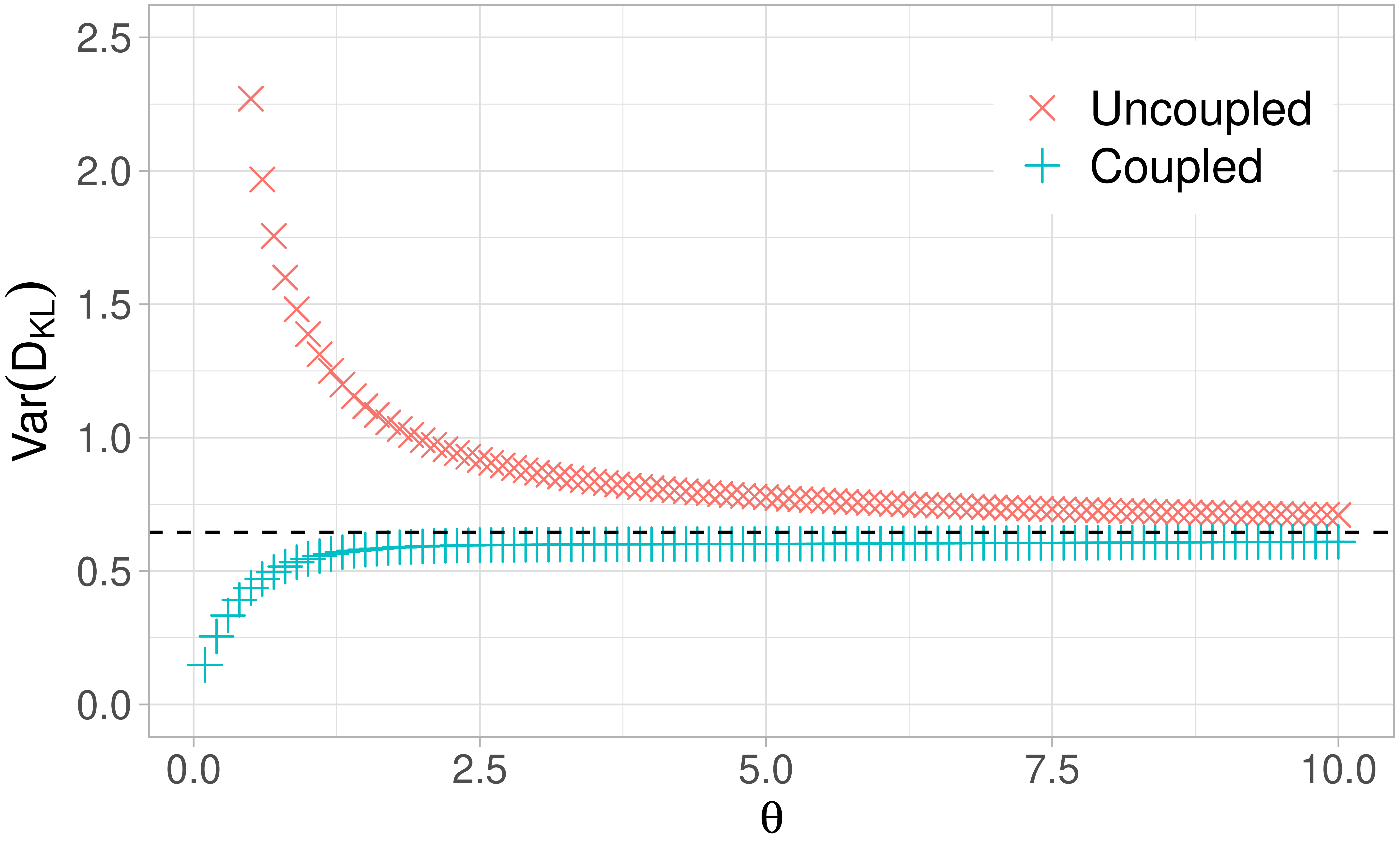
1. Asumimos los dos procesos empiezan en el mismo punto, i.e., $v_1 = v$

$$\mathbb{E}\left[D_{\mathsf{KL}}(P\,||\,P')\right] = \frac{\theta}{\theta+1}$$

OBS.

$$\lim_{\theta\to\infty}\mathbb{E}\left[D_{\mathsf{KL}}(P\,||\,P')\right] = 1 \qquad \lim_{\theta\to 0}\mathbb{E}\left[D_{\mathsf{KL}}(P\,||\,P')\right] = 0$$

Divergencia KL del proceso Dirichlet con respecto al proceso geométrico 2.0



Divergencia KL del proceso geométrico con respecto al proceso Dirichlet

$$D_{KL}(P' || P) = \sum_{n \geq 1} (1 - v)^{n-1} d(v || v_n)$$

OBS. La esperanza solo existe si $a > 1$

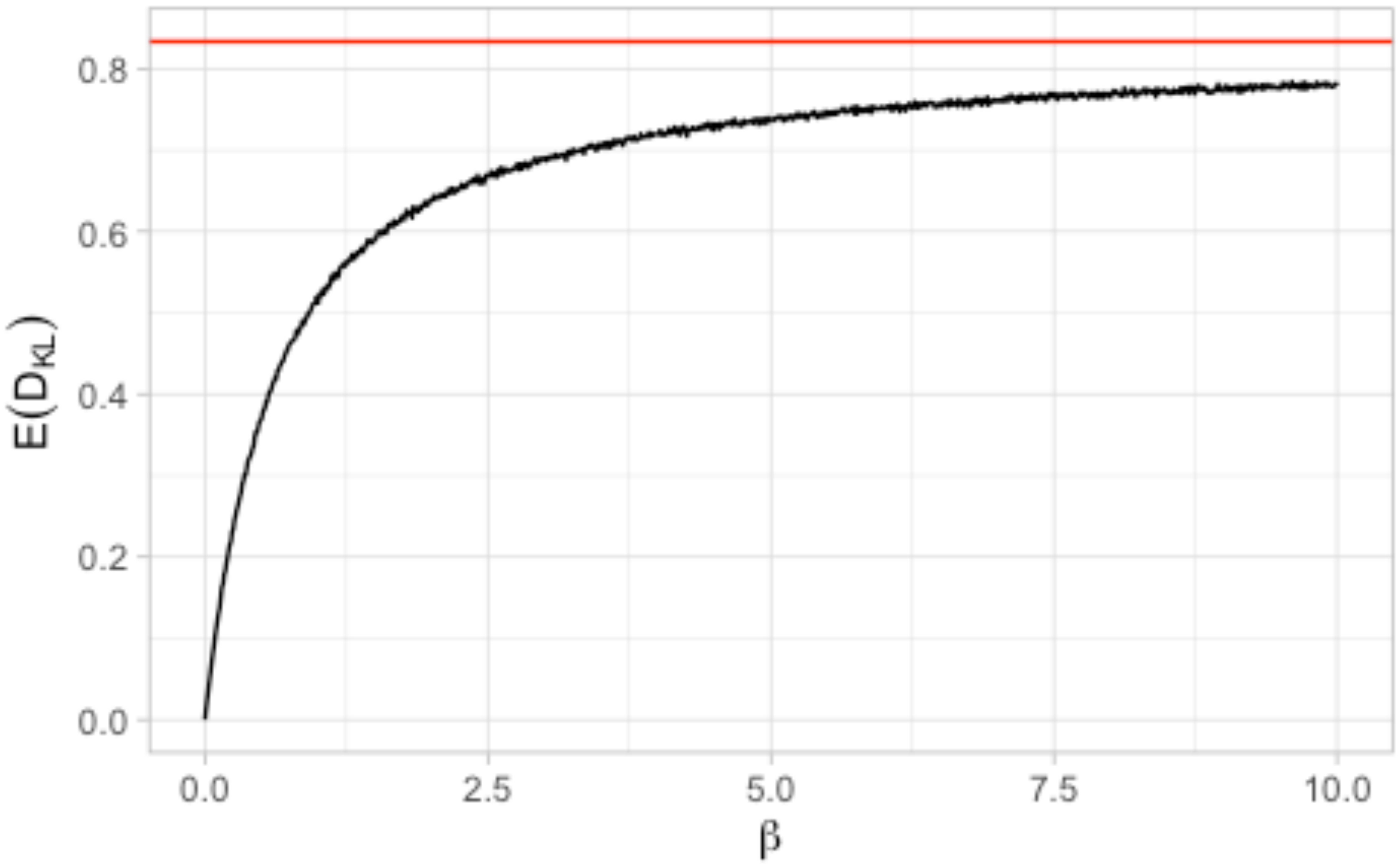
Divergencia KL para procesos intercambiables con respecto al proceso geométrico

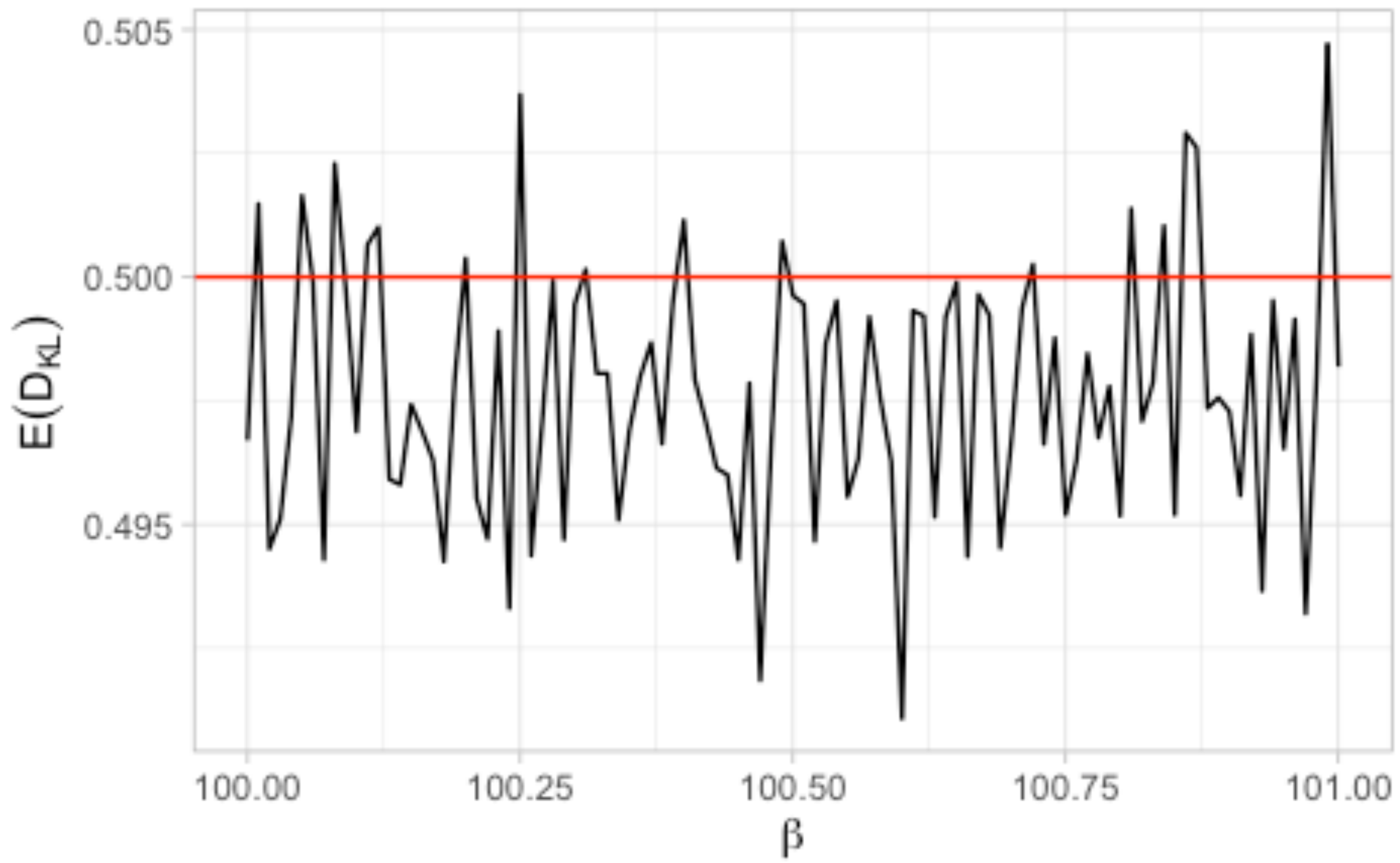
$$D_\theta(\beta) = \mathbb{E}[D_{KL}(P_\beta \,||\, P')] = \sum_{n=1}^{\infty} \sum_{\pi \in P([n])} F_\theta(\pi) p_\beta(\pi)$$

$$F_\theta(\pi) p_\beta(\pi) = \mathbb{E}\left[\left(\prod_{j=1}^{k-1} (1 - v_j)^{|A_j|} \right) (1 - v_{k_n})^{|A_k|-1} d(v_k \,||\, v) \right] \frac{\beta^k}{(\beta)^{(n)}} \prod_{j=1}^{k} (|A_j| - 1)!$$

Muy complicado de analizar !

Ilustraciones





Teorema

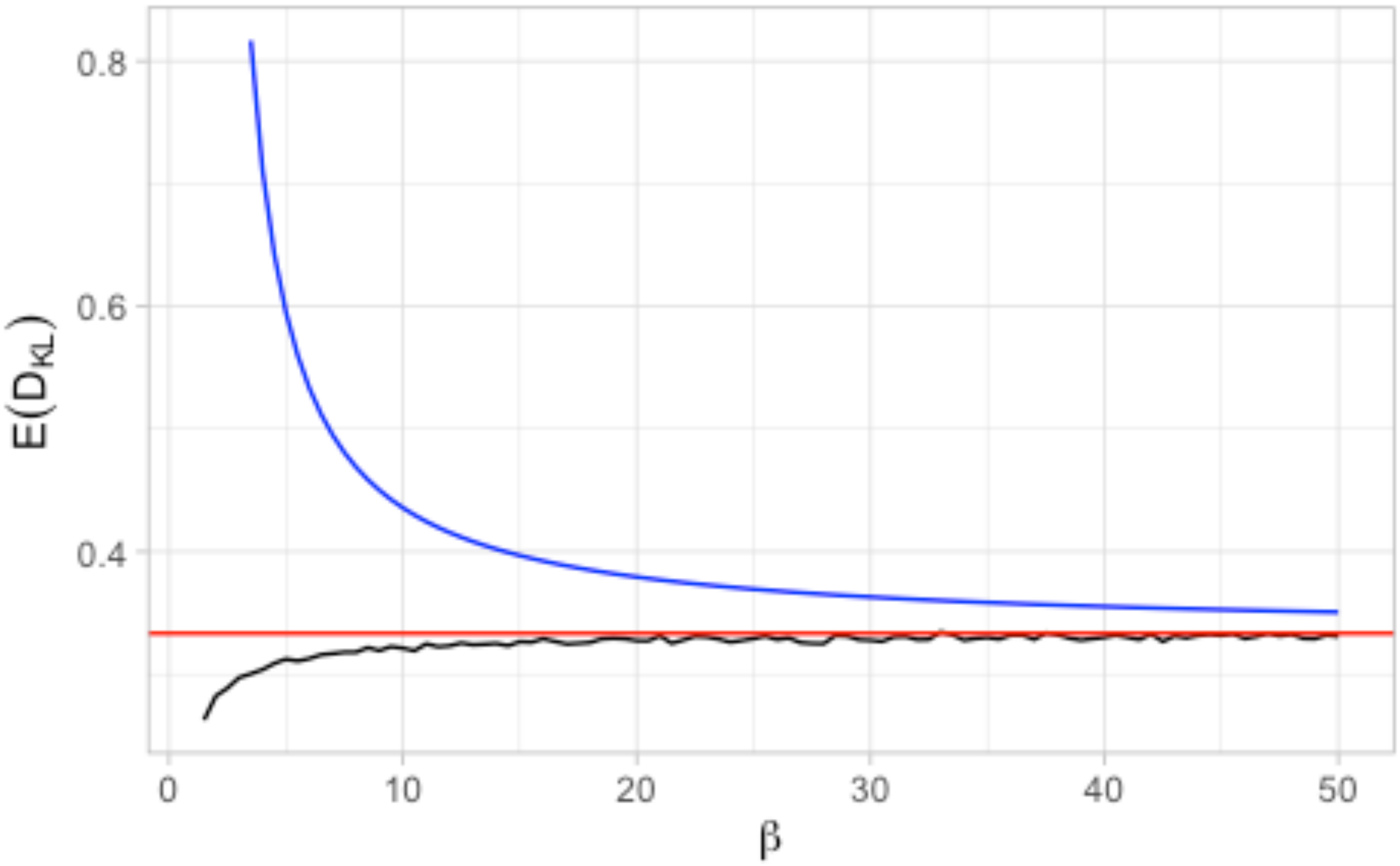
Sea P_β un proceso stick-breaking con length-variables intercambiables guiadas por un proceso Dirichlet (β, P_0) donde $\nu_0 = \mathrm{Be}(1,\theta)$, P un proceso Dirichlet (θ, P_0) y P' un proceso geométrico $(1,\theta, P_0)$ entonces para $\beta > \theta + 1$ se tiene que:

Teorema

Sea P_β un proceso stick-breaking con length-variables intercambiables guiadas por un proceso Dirichlet (β, P_0) donde $\nu_0 =$ Be$(1,\theta)$, P un proceso Dirichlet (θ, P_0) y P' un proceso geométrico $(1,\theta, P_0)$ entonces para $\beta > \theta + 1$ se tiene que:

1. $D_\theta(\beta)$ es continua
2. $D_\theta(\beta) \leq \dfrac{\theta}{\theta+1} \dfrac{\beta^2}{(\beta-1)(\beta-(\theta+1)}$
3. $\lim_{\beta \to \infty} \mathbb{E}(D_{KL}(P_\beta \,||\, P')) = \mathbb{E}(D_{KL}(P \,||\, P'))$

Ilustraciones



Observaciones

1. Por la desigualdad de Pinsker

$$\mathbb{E}[\mathsf{TV}(P_\beta, P')] \leq \sqrt{\frac{\mathbb{E}[D_{KL}(P_\beta || P')]}{2}}$$

Para β "grande" y θ pequeña las inferencias deberían ser casi las mismas.

Observaciones

1. Por la desigualdad de Pinsker

$$\mathbb{E}[\mathrm{TV}(P_\beta, P')] \leq \sqrt{\frac{\mathbb{E}[D_{KL}(P_\beta || P')]}{2}}$$

Para β "grande" y θ pequeña las inferencias deberían ser casi las mismas.

2. Resultado similar debería ser cierto para todo $\theta > 0$ o al menos para $\theta << 1$

Gracias!