Concentration without Independence via Information Measures

Kiarash Joolaei

Contents

Introduction

McDiarmid's Inequality

Consider a sequence of independent random variables $(X_1, \cdots, X_n)$.

If f satisfies the bounded difference property with parameters c_i, then:

$$\mathbb{P}\Big(|f(X^n) - \mathbb{E}[f(X^n)]| \geqslant t\Big) \leqslant 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right)$$

Bounded difference property:

$$\forall x^n, \hat{x}, i \in [1, n] : |f(x_1, \cdots, x_i, \cdots, x_n) - f(x_1, \cdots, \hat{x}, \cdots, x_n)| \leqslant c_i$$

Introduction

McDiarmid's Inequality

Consider a sequence of independent random variables $(X_1, \cdots, X_n)$.

If f satisfies the bounded difference property with parameters c_i, then:

$$\mathbb{P}\Big(|f(X^n) - \mathbb{E}[f(X^n)]| \geqslant t\Big) \leqslant 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right)$$

- The random variables can come from different distributions and must be independent

- What if they are dependent?

Intuition

- Deriving similar bounds for “semi-independent” random variables
- Paying a multiplicative price related to how far the random variables are from being independent in the upper bound
- Calculating the distance between distributions using divergences

Notations and Preliminaries

$$\mathcal{P}_X(f) := \mathbb{E}_{x \sim \mathcal{P}_X(.)}[f(x)]$$

$$\mu \ll \nu \iff \Big(\forall A \in \mathcal{F}; \nu(A) = 0 \implies \mu(A) = 0\Big)$$

absolute continuity

$p, q \in (1, +\infty)$ $\quad \dfrac{1}{p} + \dfrac{1}{q} = 1 \quad$ Hölder Conjugates

Notations and Preliminaries

$\varphi - \text{divergence}$

$\text{Probability Spaces} \ \ (\Omega, \mathcal{F}, \mathcal{P}) \ \ \& \ \ (\Omega, \mathcal{F}, \mathcal{Q})$

$\varphi : \mathbb{R}^+ \longrightarrow \mathbb{R} \ \ \text{convex function} \ \ \varphi(1) = 0$

$\mathcal{P} \ll \mathcal{Q}$

$$\mathrm{D}_\varphi(\mathcal{P}||\mathcal{Q}) := \int \varphi\Big(\frac{d\mathcal{P}}{d\mathcal{Q}}\Big) d\mathcal{Q}$$

Notations and Preliminaries

Hellinger Integral

Probability Spaces $(\Omega, \mathcal{F}, \mathcal{P})$ & $(\Omega, \mathcal{F}, \mathcal{Q})$

$\varphi_\alpha : \mathbb{R}^+ \longrightarrow \mathbb{R}$, $\varphi_\alpha(x) = x^\alpha$, $\alpha > 1$

$\mathcal{P} \ll \mathcal{Q}$

$$\mathrm{H}_\alpha(\mathcal{P}||\mathcal{Q}) := \mathrm{D}_{\varphi_\alpha}(\mathcal{P}||\mathcal{Q}) = \int \left(\frac{d\mathcal{P}}{d\mathcal{Q}}\right)^\alpha d\mathcal{Q}$$

Notations and Preliminaries

Renyi divergence

Probability Spaces $(\Omega, \mathcal{F}, \mathcal{P})$ & $(\Omega, \mathcal{F}, \mathcal{Q})$

$$p = \frac{d\mathcal{P}}{d\mu} \ , \ q = \frac{d\mathcal{Q}}{d\mu}$$ Radon-Nikodym derivative

$\mathcal{P} \ll \mu$ & $\mathcal{Q} \ll \mu$ (μ always exists)

$$\mathrm{D}_\alpha(\mathcal{P}||\mathcal{Q}) := \frac{1}{\alpha - 1} \log \int p^\alpha q^{1-\alpha} d\mu \quad \alpha \in \mathbb{R}^+ - \{1\}$$

Notations and Preliminaries

Markov Kernel: Let $(\Omega, \mathcal{F})$ be a measurable space.
A Markov Kernel K is a mapping $K : \mathcal{F} \times \Omega \longrightarrow [0, 1]$ such that:

1. $\forall x \in \Omega;\ K(.|x)$ is a probability measure on $(\Omega, \mathcal{F})$
2. $\forall E \in \mathcal{F};\ K(E|.)$ is an $\mathcal{F}-$measurable real-valued function

Notations and Preliminaries

$$\mu K(E) := \mu(K(E|.)) = \int K(E|x)d\mu(x)$$ A measure

$$Kf(x) := \int f(y)dK(y|x)$$ A function

In a Markov Chain :

$$\mathbb{P}(X_i \in E|X_1, \ldots, X_{i-1}) = \mathbb{P}(X_i \in E|X_{i-1}) = K_i(E|X_{i-1})$$

In a time-homogeneous Markov Chain with discrete state space :

$$\forall i \geqslant 2; K_i = P \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|} \text{ stochastic matrix}$$

Notations and Preliminaries

$$||K||_{\alpha\to\gamma} := \sup_{f\neq 0} \frac{||Kf||_\alpha}{||f||_\gamma} \;\; , \; \gamma \leqslant \alpha$$

$$K \text{ is contractive iff } \; ||K||_{\alpha\to\alpha} \leqslant 1$$

$$K \text{ is hyper-contractive iff } \; ||K||_{\alpha\to\gamma} \leqslant 1 \text{ for some } \gamma < \alpha$$

$$\gamma^\star_K(\alpha) := \arg\min\{\gamma \mid \gamma < \alpha \; , \; ||K||_{\alpha\to\gamma} \leqslant 1\}$$

$$K^{\leftarrow}_\mu(y|x) = \frac{K(y|x)\mu(x)}{\mu K(y)} \;\; \text{adjoint of kernel } K \text{ (discrete setting)}$$

Main Result

$\mathcal{P}_{X^n}$:   joint distribution of $(X_1, \cdots, X_n)$

$\mathcal{P}_{X_i}$:   marginal distribution corresponding to X_i

$\mathcal{P}_{\bigotimes_{i=1}^{n} X_i}$:   the joint measure induced by the product of the marginals

Main Result

Theorem. If f satisfies bounded difference property with parameters c_i and $\mathcal{P}_{X^n} \ll \mathcal{P}_{\bigotimes_{i=1}^n X_i}$, for all $t > 0$ and $\alpha > 1$:

$$\mathcal{P}_{X^n}\left(|f - \mathcal{P}_{\bigotimes_{i=1}^n X_i}(f)| \geqslant t\right) \leqslant 2^{\frac{1}{\beta}} \exp\left(\frac{-2t^2}{\beta \sum_{i=1}^n c_i^2}\right) \mathrm{H}_{\alpha}^{\frac{1}{\alpha}}(\mathcal{P}_{X^n} || \mathcal{P}_{\bigotimes_{i=1}^n X_i})$$

$$\beta = \frac{\alpha}{\alpha - 1} \quad \text{(Hölder Conjugate of } \alpha\text{)}$$

Multiplicative Price

Remarks

1. For any measurable set E :

$$\mathcal{P}_{X^n}(E) \leqslant \mathcal{P}^{\frac{1}{\beta}}_{\bigotimes_{i=1}^n X_i}(E) \mathrm{H}^{\frac{1}{\alpha}}_{\alpha}(\mathcal{P}_{X^n} || \mathcal{P}_{\bigotimes_{i=1}^n X_i})$$

2. $\mathcal{P}_{X^n}\left(|f - \mathcal{P}_{\bigotimes_{i=1}^n X_i}(f)| \geqslant t\right) \leqslant 2^{\frac{1}{\beta}} \exp\left(\frac{-2t^2}{\beta \sum_{i=1}^n c_i^2}\right) \prod_{i=2}^n \max_{x^{i-1}} \mathrm{H}^{\frac{1}{\alpha}}_{\alpha}(\mathcal{P}_{X_i | X^{i-1} = x^{i-1}} || \mathcal{P}_{X_i})$

Remarks

$$3.\ \mathcal{P}_{X^n}\left(|f - \mathcal{P}_{\bigotimes_{i=1}^n X_i}(f)| \geqslant t\right) \leqslant 2^{\frac{1}{\beta}} \exp\left(-n\left(\frac{2t^2}{\beta} - \frac{1}{n\alpha}\log \mathrm{H}_\alpha(\mathcal{P}_{X^n}||\mathcal{P}_{\bigotimes_{i=1}^n X_i})\right)\right)$$

$$c_i = \frac{1}{n}$$

$$t > \sqrt{\frac{\beta}{2n\alpha}\ln \mathrm{H}_\alpha(\mathcal{P}_{X^n}||\mathcal{P}_{\bigotimes_{i=1}^n X_i})}$$

Exponential Decay of Upper Bound

Applications

Discrete Time Markov Chain

Non-contracting Markov Chain (SSRW)

Non-Markovian Process

Markov Chain Monte Carlo

Discrete Time Markov Chain

$$\text{Markov Chain} \quad (X_n)_{n\in\mathbb{N}}$$

$$\text{Transition Matrices} \quad (K_n)_{n\in\mathbb{N}}$$

$$X_1 \sim P_1 \qquad\qquad X_i \sim (P_1 K_1 \cdots K_{i-1} = P_i)$$

Discrete Time Markov Chain

Theorem. Consider discrete-valued Markov kernels K_i.

If function f satisfies bounded difference property with parameters $c_i = \frac{1}{n}$,

for every $\alpha > 1$ we have:

$$\mathcal{P}_{X^n}\left(\left|f - \mathcal{P}_{\bigotimes_{i=1}^{n} X_i}(f)\right| \geq t\right) \leq 2^{\frac{1}{\beta}} \exp\left(-\frac{2nt^2}{\beta} + \sum_{i=1}^{n-1}\left(\log \|K_i^{\leftarrow}\|_{\alpha \to \gamma_i^\star(\alpha)} - \frac{1}{\underbrace{\bar{\gamma}_i^\star(\alpha)}_{\text{Hölder Conjugate}}} \min_{j \in supp(\mathcal{P}_i)} \log P_i(j)\right)\right)$$

Discrete Time Markov Chain

For a time-homogeneous Markov Chain with $K_i = K$:

$$\mathcal{P}_{X^n}\left(\left|f - \mathcal{P}_{\bigotimes_{i=1}^n X_i}(f)\right| \geq t\right)$$

$$\leq 2^{\frac{1}{\beta}} \exp\left(-\frac{2nt^2}{\beta} + (n-1)\log\|K^{\leftarrow}\|_{\alpha\to\gamma_K^\star(\alpha)} - \frac{1}{\bar{\gamma}_K^\star(\alpha)}\sum_{i=1}^{n-1}\left(\min_{j\in supp(\mathcal{P}_i)}\log P_i(j)\right)\right)$$

$$\leq 2^{\frac{1}{\beta}} \exp\left(-\frac{2nt^2}{\beta} + (n-1)\log\|K^{\leftarrow}\|_{\alpha\to\gamma_K^\star(\alpha)} - \frac{n-1}{\bar{\gamma}_K^\star(\alpha)}\left(\min_{i=1,\ldots,(n-1)}\min_{j\in supp(\mathcal{P}_i)}\log P_i(j)\right)\right)$$

Non-contracting Markov Chain

Simple Symmetric Random Walk

$$(S_n)_{n \in \mathbb{N}} \qquad S_i = S_{i-1} + X_i$$

$$X_i \sim \mathrm{Rad} \quad \forall i \geqslant 1 \qquad \mathbb{P}(X_0 = 0) = 1$$

- No stationary distribution
- Not contracting

Useful Lemmas

Lemma 1. *Let* $i \geq 1$, $x \in supp(S_{i-1})$, $0 \leq j \leq i$, *and* $\alpha \geq 1$. *Then,*

$$2^{\frac{1}{\beta}\left(-1+i(1-h_2(\frac{i+1}{2i})+\frac{1}{2}\log_2\left(\frac{\pi}{2}\left(\frac{i^2-1}{i}\right)\right)\right)} \leq H_{\alpha}^{\frac{1}{\alpha}}(\mathcal{P}_{S_i|S_{i-1}=x}\|\mathcal{P}_{S_n}) \leq 2^{i\frac{1}{\beta}-1+\frac{1}{\alpha}},$$

where $h_2(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ *denotes the binary entropy. Thus one has that:*

$$\frac{n-2}{4\beta} \leq \log_2 H_{\alpha}^{\frac{1}{\alpha}}(\mathcal{P}_{S^n}\|\mathcal{P}_{\bigotimes_{j=1}^{n} S_j}) \leq \frac{n(n-1)}{2\beta}.$$

Useful Lemmas

Lemma 2. *Let* $\alpha > 1$. *Denote with* $t_\alpha = \sqrt{\frac{\beta \ln(H_\alpha(\mathcal{P}_{X^n} \| \mathcal{P}_{\otimes X_i}))}{2\alpha n}}$. *Then, the following holds true:*

$$\left|\mathcal{P}_{X^n}(f) - \mathcal{P}_{\otimes_{i=1}^n X_i}(f)\right| \leq t_\alpha + \frac{\sqrt{\beta}}{2^{\frac{1}{\alpha}} \sqrt{\frac{2n}{\alpha} \ln(H_\alpha(\mathcal{P}_{X^n} \| \mathcal{P}_{\otimes_{i=1}^n X_i}))}}.$$

$o_n(1)$

Prior work and McDiarmid-like Inequalities bound deviation with respect to **joint measure expected value**, unlike the main result of the paper.

Non-contracting Markov Chain

$$\frac{\sqrt{\beta}}{2^{\frac{1}{\alpha}}\sqrt{\frac{2n}{\alpha}\ln(H_\alpha(\mathcal{P}_{X^n}\|\mathcal{P}_{\bigotimes_{i=1}^n X_i}))}} \leq \frac{2^{\frac{1}{\beta}}\beta}{n\sqrt{\ln(2)(2-\frac{4}{n})}} = o_n(1)$$

$$\begin{aligned}|f-\mathcal{P}_{S^n}(f)| &= \left|f-\mathcal{P}_{S^n}(f)-\mathcal{P}_{\bigotimes_{i=1}^n S_i}(f)+\mathcal{P}_{\bigotimes_{i=1}^n S_i}(f)\right| \\ &\leq \left|f-\mathcal{P}_{\bigotimes_{i=1}^n S_i}(f)\right| + \left|\mathcal{P}_{X^n}(f)-\mathcal{P}_{\bigotimes_{i=1}^n S_i}(f)\right| \\ &\leq \left|f-\mathcal{P}_{\bigotimes_{i=1}^n S_i}(f)\right| + t_\alpha + o_n(1).\end{aligned}$$

Non-contracting Markov Chain

Prior work:

$$\mathbb{P}(|f - \mathcal{P}_{S^n}(f)| \geqslant t) \leqslant 2\exp\left(-\frac{t^2}{2n}\right)$$

Paper's Result:

$$\mathbb{P}\left(|f - \mathcal{P}_{S^n}(f)| \geq t + t_\alpha + o_n(1)\right) \leq$$

$$\mathbb{P}\left(\left|f - \mathcal{P}_{\bigotimes_{i=1}^n S_i}(f)\right| \geq t\right) \leq 2^{\frac{1}{\beta}} \exp\left(\frac{-2nt^2}{\beta} + \frac{n(n-1)}{2\beta}\ln 2\right)$$

Non-contracting Markov Chain

Prior work: $\tilde{t} = 2\sqrt{n}$

$$\mathbb{P}\left(|f - \mathcal{P}_{S^n}(f)| \geq \tilde{t}\right) \leq 2\exp(-2)$$

Paper's Result:

$$\mathbb{P}\left(|f - \mathcal{P}_{S^n}(f)| \geq \tilde{t}\right) \leq 2^{\frac{1}{\beta}} \exp\left(-\frac{n^2}{\beta}\left(2 - \frac{\ln 2}{2} + \frac{\ln 2}{2n}\right)\right)$$

Non-Markovian Process

$$X_n = \begin{cases} +1, & \text{with probability } \sum_{i=0}^{n-1} p_i X_i, \\ -1, & \text{with probability } 1 - \sum_{i=0}^{n-1} p_i X_i, \end{cases}$$

$$p_i = 2^{-i-1} \qquad \mathbb{P}_{X_1}(1|x_0) = \frac{1}{2} = \mathbb{P}_{X_1}(-1|x_0)$$

Fully dependent on the past

Non-Markovian Process

$$X_n = \begin{cases} +1, & \text{with probability } \sum_{i=0}^{n-1} p_i X_i, \\ -1, & \text{with probability } 1 - \sum_{i=0}^{n-1} p_i X_i, \end{cases}$$

$$p_i = 2^{-i-1} \qquad \mathbb{P}_{X_1}(1|x_0) = \tfrac{1}{2} = \mathbb{P}_{X_1}(-1|x_0)$$

$$\mathbb{P}_{X_n}(1|x_0^{n-1}) = \frac{1}{2} + \sum_{i=1}^{n-1} p_i x_i = \sum_{i=0}^{n-1} x_i 2^{-i-1} = 1 - \mathbb{P}_{X_n}(-1|x_0^{n-1})$$

Non-Markovian Process

Claim.

$$\mathbb{P}\left(\left\{\left|f - \mathcal{P}_{\bigotimes_{i=1}^{n} X_i}(f)\right| \geq t\right\}\right) \leq \inf_{\beta > 1} 2^{\frac{1}{\beta}} \exp\left(-\frac{2n}{\beta}\left(t^2 - \frac{n-1}{n}\frac{\beta \ln 2}{2}\right)\right)$$

$$t^2 > (1 + o_n(1))\frac{\beta \ln 2}{2}$$

Exponential decay given the values for t

Non-Markovian Process

Claim.

$$\mathbb{P}\left(\left\{\left|f - \mathcal{P}_{\bigotimes_{i=1}^{n} X_i}(f)\right| \geq t\right\}\right) \leq \inf_{\beta > 1} 2^{\frac{1}{\beta}} \exp\left(-\frac{2n}{\beta}\left(t^2 - \frac{n-1}{n}\frac{\beta \ln 2}{2}\right)\right)$$

Proof Ideas:

$$H_\alpha(\mathcal{P}_{X_n}(\cdot|x_0^{n-1}) \| (1/2, 1/2)) < 2 \sum_{j=0}^{\lfloor \frac{\alpha}{2} \rfloor} \binom{\lfloor \alpha \rfloor}{2j} \left(2 \sum_{i=1}^{n-1} p_i x_i\right)^{2j}$$

$$\implies \max_{x_1^{n-1}} H_\alpha(\mathcal{P}_{X_n}(\cdot|x_0^{n-1}) \| (1/2, 1/2)) < 2^\alpha \implies H_\alpha^{\frac{1}{\alpha}}(\mathcal{P}_{X^n} \| (1/2, 1/2)^{\otimes n}) < 2^{n-1}$$

Markov Chain Monte Carlo

- A method for sampling from a target distribution
- Estimating the mean of a function using correlated samples
- Starting from an arbitrary distribution
- Forming a Markov Chain whose stationary distribution is our target distribution
- After the “burn-in phase”, the generated samples are coming from the stationary distribution

Markov Chain Monte Carlo

Prior work: If $f : \mathcal{X} \longrightarrow [a, b]$ is uniformly bounded, then:

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_{n_0+i}) - \pi(f) > t\right) \leq C(\nu, n_0, \alpha)\exp\left(-\frac{1}{\beta}\cdot\frac{1-\max\{\lambda_r, 0\}}{1+\max\{\lambda_r, 0\}}\cdot\frac{2nt^2}{(b-a)^2}\right)$$

$$\lambda_r = 1 - \sup\{|\lambda| : \lambda \in \sigma(P), \lambda \neq 1\}$$ Right spectral gap

Markov Chain Monte Carlo

Claim. If $f : \mathcal{X} \longrightarrow [a, b]$ is uniformly bounded, then:

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_{n_0+i}) - \pi(f) > t\right) \leq \exp\left(-\frac{2nt^2}{\beta(b-a)^2}\right) H_{\alpha}^{\frac{1}{\alpha}}(\nu K^{n_0} \| \pi) \prod_{i=2}^{n} \max_{x_{n_0+i-1}} H_{\alpha}^{\frac{1}{\alpha}}(K(\cdot | x_{n_0+i-1}) \| \pi)$$

$$\leq C(\nu, n_0, \alpha) \exp\left(-\frac{2nt^2}{\beta(b-a)^2}\right) \max_{x} \pi(\{x\})^{-\frac{n-1}{\beta}},$$

ν starting distribution

π stationary distribution of MC

Markov Chain Monte Carlo

$$t^2 \geq \frac{n-1}{n} \frac{(b-a)^2}{2} \frac{1+\lambda_r}{2\lambda_r} \log\left(\frac{1}{\min_x \pi(\{x\})}\right)$$

For these values of t, the new bound is tighter than the prior

If $n_0 = \Omega(\log n)$, the exponential decay is guaranteed

Tensorization

Tensorization of D_{KL} :

$$D(\mathcal{Q}\|\mathcal{P}) \leq \sum_{i=1}^{n} \int \mathrm{d}\mathcal{Q}_{\bar{X}^i} D(Q_{X_i|\bar{X}^i} \| P_{X_i})$$

$\mathcal{Q}$ is a product measure

With some changes, tensorization might hold for other divergences

Tensorization

Theorem 3. *Let $\mathcal{Q}$ and $\mathcal{P}$ be two probability measures on the space $\mathcal{X}^n$ such that $\mathcal{Q} \ll \mathcal{P}$, and assume that $\mathcal{P}$ is a product measure (i.e., $\mathcal{P} = \bigotimes_{i=1}^n P_i$). Assume also that $\mathcal{Q}$ is a Markov measure induced by Q_1 and the kernels $Q_i(\cdot|\cdot)$ with $1 \leq i \leq n$, i.e., $\mathcal{Q}(x^n) = Q_1(x_1) \prod_{i=2}^n Q_i(x_i|x_{i-1})$. Moreover, given a constant c, let $X_0 = c$ (almost surely) be an auxiliary random variable, then,*

$$D_\alpha(\mathcal{Q}\|\mathcal{P}) \leq \frac{1}{\alpha - 1} \sum_{i=1}^n \frac{1}{\beta_{i-1}} \log \mathcal{P}_{X_{i-1}} \left(\exp \left(\frac{(\alpha\alpha_i - 1)\beta_{i-1}}{\alpha_i} \left(D_{\alpha\alpha_i}(Q_i(\cdot|X_{i-1})\|P_{X_i} \right) \right) \right)$$

where $\alpha_i \geq 1$ *for* $i \geq 0$, $\beta_0 = 1$, $\alpha_n = 1$ *and* $\beta_i = \alpha_i/(\alpha_i - 1)$

Tensorization

Theorem 3. *Let $\mathcal{Q}$ and $\mathcal{P}$ be two probability measures on the space $\mathcal{X}^n$ such that $\mathcal{Q} \ll \mathcal{P}$, and assume that $\mathcal{P}$ is a product measure (i.e., $\mathcal{P} = \bigotimes_{i=1}^n P_i$). Assume also that $\mathcal{Q}$ is a Markov measure induced by Q_1 and the kernels $Q_i(\cdot|\cdot)$ with $1 \leq i \leq n$, i.e., $\mathcal{Q}(x^n) = Q_1(x_1) \prod_{i=2}^n Q_i(x_i|x_{i-1})$. Moreover, given a constant c, let $X_0 = c$ (almost surely) be an auxiliary random variable, then,*

$$H_\alpha(Q\|\mathcal{P}) \leq \prod_{i=1}^n \mathcal{P}_{X_{i-1}}^{\frac{1}{\beta_{i-1}}} \left(H_{\alpha\alpha_i}^{\frac{\beta_{i-1}}{\alpha_i}} (Q_i(\cdot|X_{i-1})\|P_{X_i}) \right)$$

where $\alpha_i \geq 1$ *for* $i \geq 0$, $\beta_0 = 1$, $\alpha_n = 1$ *and* $\beta_i = \alpha_i/(\alpha_i - 1)$

Tensorized Version of the Theorem

if $(X_1, \ldots, X_n)$ *are Markovian under* $\mathcal{P}_{X^n}$, *i.e.,* $\mathcal{P}_{X_i|X^{i-1}} = \mathcal{P}_{X_i|X_{i-1}}$ *almost surely, then the following holds:*

$$\mathcal{P}_{X^n}\left(\left|f - \mathcal{P}_{\bigotimes_{i=1}^n X_i}(f)\right| \geq t\right) \leq 2^{\frac{1}{\beta}} \exp\left(\frac{-2t^2}{\beta \sum_{i=1}^n c_i^2}\right)\left(\prod_{i=2}^n H_i^\alpha\right)^{\frac{1}{\alpha}},$$

with $\beta = \alpha/(\alpha-1)$, $H_i^\alpha = \mathcal{P}_{X_{i-1}}^{\frac{1}{\beta_{i-1}}}\left(H_{\alpha\alpha_i}^{\frac{\beta_{i-1}}{\alpha_i}}(\mathcal{P}_{X_i|X_{i-1}} \| \mathcal{P}_{X_i})\right)$, $\alpha_i > 1$ *for* $i \geq 0$, $\beta_0 = 1$, $\alpha_n = 1$, *and* $\beta_i = \alpha_i/(\alpha_i - 1)$.

Useful Lemma

Hölder's Inequality

Let $(X, \mathcal{F}, \mu)$ be a measure space. For all measurable functions f, g we have:

$$\int_X |fg| d\mu \leqslant \left(\int_X |f|^p \right)^{\frac{1}{p}} \left(\int_X |g|^q \right)^{\frac{1}{q}}$$

$$p, q \in [1, +\infty] \,, \ \frac{1}{p} + \frac{1}{q} = 1$$

Main Theorem Proof

Proof. Assume that $E = \{|f - \mathcal{P}_{\bigotimes_{i=1}^n X_i}(f)| \geq t\}$. Then, one has that

$$\begin{aligned}
\mathcal{P}_{X^n}(E) &= \int \mathbb{1}_E \, \mathrm{d}\mathcal{P}_{X^n} \\
&= \int \mathbb{1}_E \frac{d\mathcal{P}_{X^n}}{d\mathcal{P}_{\bigotimes_{i=1}^n X_i}} \, \mathrm{d}\mathcal{P}_{\bigotimes_{i=1}^n X_i} \\
\text{Hölder's Inequality} \qquad &\leq \left(\int \mathbb{1}_E \, \mathrm{d}\mathcal{P}_{\bigotimes_{i=1}^n X_i} \right)^{\frac{\alpha-1}{\alpha}} \left(\int \left(\frac{d\mathcal{P}_{X^n}}{d\mathcal{P}_{\bigotimes_{i=1}^n X_i}} \right)^{\alpha} \mathrm{d}\mathcal{P}_{\bigotimes_{i=1}^n X_i} \right)^{\frac{1}{\alpha}} \\
&= \mathcal{P}_{\bigotimes_{i=1}^n X_i}^{\frac{1}{\beta}}(E) H_{\alpha}^{\frac{1}{\alpha}}(\mathcal{P}_{X^n} \| \mathcal{P}_{\bigotimes_{i=1}^n X_i}),
\end{aligned}$$

Main Theorem Proof

$$\text{McDiarmid's Inequality} \implies \mathcal{P}^{\frac{1}{\beta}}_{\bigotimes_{i=1}^n X_i}(E) \leqslant 2^{\frac{1}{\beta}} \exp\left(-\frac{2t^2}{\beta \sum_{i=1}^n c_i^2}\right)$$

$$\implies \quad \mathcal{P}_{X^n}\Big(|f - \mathcal{P}_{\bigotimes_{i=1}^n X_i}(f)| \geqslant t\Big) \leqslant 2^{\frac{1}{\beta}} \exp\left(\frac{-2t^2}{\beta \sum_{i=1}^n c_i^2}\right) \mathrm{H}_{\alpha}^{\frac{1}{\alpha}}(\mathcal{P}_{X^n} || \mathcal{P}_{\bigotimes_{i=1}^n X_i})$$

Proof of Lemma 2

$$
\begin{aligned}
\left|\mathcal{P}_{X^n}(f) - \mathcal{P}_{\otimes_{i=1}^n X_i}(f)\right| &= \left|\mathcal{P}_{X^n}(f-c)\right| \\
&\leq \mathcal{P}_{X^n}\left(|f-c|\right) \\
&= \int_0^\infty \mathcal{P}_{X^n}(|f-c| \geq t)\,\mathrm{d}t \\
&\leq \int_0^{t_\alpha} 1\,\mathrm{d}t + \int_{t_\alpha}^\infty \mathcal{P}_{X^n}(|f-c| \geq t)\,\mathrm{d}t \\
&\leq t_\alpha + 2^{\frac{1}{\beta}} H_\alpha^{\frac{1}{\alpha}}(\mathcal{P}_{X^n} \| \mathcal{P}_{\otimes_{i=1}^n X_i}) \int_{t_\alpha}^\infty \exp\left(-\frac{2nt^2}{\beta}\right)\mathrm{d}t \\
&\leq t_\alpha + 2^{\frac{1}{\beta}} H_\alpha^{\frac{1}{\alpha}}(\mathcal{P}_{X^n} \| \mathcal{P}_{\otimes_{i=1}^n X_i}) \frac{\beta}{2nt_\alpha} \int_{t_\alpha}^\infty \frac{2nt}{\beta} \exp\left(-\frac{2nt^2}{\beta}\right)\mathrm{d}t
\end{aligned}
$$

Proof of Lemma 2

$$= t_\alpha + 2^{\frac{1}{\beta}} H_\alpha^{\frac{1}{\alpha}}(\mathcal{P}_{X^n} \| \mathcal{P}_{\otimes_{i=1}^n X_i}) \frac{\beta}{2nt_\alpha} \left(-\frac{1}{2} \exp\left(-\frac{2nt^2}{\beta} \right) \right) \Bigg|_{t_\alpha}^{\infty}$$

$$= t_\alpha + 2^{\frac{1}{\beta}} H_\alpha^{\frac{1}{\alpha}}(\mathcal{P}_{X^n} \| \mathcal{P}_{\otimes_{i=1}^n X_i}) \frac{\beta}{4nt_\alpha} \exp\left(-\frac{2nt_\alpha^2}{\beta} \right)$$

$$= t_\alpha + 2^{\frac{1}{\beta}} \frac{\beta}{4nt_\alpha}$$

$$= t_\alpha + 2^{\frac{1}{\beta}} \frac{\beta}{4n} \frac{\sqrt{2n}}{\sqrt{\frac{\beta}{\alpha} \ln(H_\alpha(\mathcal{P}_{X^n} \| \mathcal{P}_{\otimes_{i=1}^n X_i}))}}$$

$$= t_\alpha + \frac{\sqrt{\beta}}{2^{\frac{1}{\alpha}} \sqrt{\frac{2n}{\alpha} \ln(H_\alpha(\mathcal{P}_{X^n} \| \mathcal{P}_{\otimes_{i=1}^n X_i}))}},$$

Thank You !