KL divergence of max-of-n

Jacob Hilton

December 3, 2023

Abstract

We analyze the Kullback–Leibler divergence from a probability distribution to the maximum of n i.i.d. samples from that distribution. When the distribution is absolutely continuous, then the KL divergence is exactly $\log n - \frac{n-1}{n}$. When the distribution is discrete with finite support and the maximum probability of any single value is $o\left(n^{-\frac{3}{2}}\right)$, then the KL divergence is at most $\log n - \frac{n-1}{n}$ and differs from this by an error term that tends to zero. We conjecture based on numerical evidence that this can be improved to requiring the maximum probability to be $o\left(n^{-1}\right)$. In practice, as long as it is rare for n i.i.d. samples to contain a repetition, it is reasonable to treat the $\log n - \frac{n-1}{n}$ formula as exact for all practical purposes.

1 Introduction

In reinforcement learning from human feedback, a *reward model* is used to assign a numerical score to outcomes taken from from some sample space S. Given some distribution P over S from which we can sample, we would like to obtain a sample with high score. The simplest way to achieve this is to take n i.i.d. samples from P and choose the one with the highest score. This is known as *best-of-n sampling*, *rejection sampling* or *reranking*.

It is natural to ask about the Kullback–Leibler divergence from P to the best-of-n distribution, which in some sense measures the "amount of optimization" that has been performed. One might naively guess that this KL divergence is $\log n$, since best-of-n sampling is similar to taking the top $\frac{1}{n}$ of the distribution. In fact, if the distribution over scores is absolutely continuous, then the KL divergence is exactly $\log n - \frac{n-1}{n}$. This somewhat surprising result appears without proof in Stiennon et al. [2020, Appendix G.3] and later in Gao et al. [2023, Section 2]. In this work we will explain this and a number of other related results.

- In Section 2, we observe that the KL divergence depends only on the distribution over scores under P up to a strictly increasing monotonic transformation, assuming for simplicity that there are no ties. This lets us reduce best-of-n to max-of-n, and explains why the KL divergence is the same for all absolutely continuous distributions.
- In Section 3, we show directly that if P is absolutely continuous, then the KL divergence from P to the max-of-n distribution is exactly $\log n - \frac{n-1}{n}$.
- In Section 4, we study the case in which P is discrete with finite support. In this case the KL divergence is no longer exactly $\log n - \frac{n-1}{n}$, but is instead slightly less than this by some amount $\varepsilon \geq 0$ that depends on P and n. We prove that

$$\varepsilon < np_{\max} + \frac{n^3 p_{\max}^2}{4},$$

> where $p_{\max}$ is the maximum probability of any single value under P. This implies that if $p_{\max} = o\left(n^{-\frac{3}{2}}\right)$ then $\varepsilon \to 0$. In fact, numerical evidence leads us to conjecture the stronger inequality
>
> $$\varepsilon < \frac{np_{\max}}{4},$$
>
> which would imply that if $p_{\max} = o\left(n^{-1}\right)$ then $\varepsilon \to 0$. This would in turn imply that if n i.i.d. samples contain a repetition with probability tending to zero, then $\varepsilon \to 0$.

In a typical language modeling setup with multi-sentence responses, P is discrete with finite support, but $n \leq 10^3$ and $p_{\max} \ll 10^{-6}$. In this case the error term $\varepsilon \ll 1.25 \times 10^{-3}$ regardless of whether our conjecture holds, and so it is reasonable to treat the $\log n - \frac{n-1}{n}$ formula as exact for all practical purposes.

2 Monotonic invariance

Recall that if P and P' are discrete probability distributions over the sample space S, then the *Kullback–Leiber divergence* or *KL divergence* from P to P' is defined as

$$D_{\mathrm{KL}}\left(P' \parallel P\right) = \sum_{x \in S} P'\left(x\right) \log\left(\frac{P'\left(x\right)}{P\left(x\right)}\right).$$

Suppose that P' is the best-of-n distribution according to some scoring function, and assume for simplicity that no two distinct samples have the same score (otherwise the KL divergence depends on how such ties are broken). It is clear from the above definition that this KL divergence does not depend on the sample space S itself, nor on the exact values of the scores, but only on the probability of the highest score, the probability of the second-highest score, and so on.

One consequence of this is that we may assume without loss of generality that the sample space is $\mathbb{R}$ and that the scoring function is the identity function. In this case, best-of-n sampling reduces to taking the maximum of n i.i.d. samples.

Definition. Let P be a probability distribution over $\mathbb{R}$ and $n \in \mathbb{N}$. The *max-of-n-from-P* distribution, denoted $P_{(n)}$, is the probability distribution of

$$\max\left(X_1, \ldots, X_n\right) \quad \text{for} \quad X_1, \ldots, X_n \sim_{\text{i.i.d.}} P.$$

This maximum is also sometimes known as the *largest order statistic*.

Another consequence of this is that if P is instead an absolutely continuous probability distribution over $\mathbb{R}$, then the KL divergence $D_{\mathrm{KL}}\left(P_{(n)} \parallel P\right)$ does not depend on the choice of distribution for P. Intuitively, this is because P may be written as a limit of discrete uniform distributions, which all lead to the same KL divergence regardless of how they are spaced. More formally, the the change-of-variables formula for probability density functions can be used to show that applying a strictly increasing monotonic function to P leaves the KL divergence unchanged.

3 The continuous case

We now explain why $D_{\mathrm{KL}}\left(P_{(n)} \parallel P\right)$ is exactly $\log n - \frac{n-1}{n}$ when P is absolutely continuous. The simplest way to see this is to use the fact that this KL divergence does not depend on the choice of distribution for P, as explained in the previous section. It is then a straightforward calculation to check the formula in the special case that P has a uniform distribution on the interval $[0,1]$.

For the sake of completeness, we give a careful, self-contained proof. This can be thought of as applying a change of variables to reduce to the case of the uniform distribution on $[0,1]$. The proof will also be useful for extending the result to the discrete case in the next section.

Proposition 1. *Let P be an absolutely continuous probability distribution and $n \in \mathbb{N}$. Then the KL divergence from P to the max-of-n-from-P distribution,*

$$D_{KL}\left(P_{(n)} \parallel P\right) = \log n - \frac{n-1}{n}.$$

Proof. Let f and F be the probability density function of P and the cumulative distribution function of P respectively, and let $f_{(n)}$ and $F_{(n)}$ be the probability density function of $P_{(n)}$ and the cumulative distribution function of $P_{(n)}$ respectively.

Given $X_1, \ldots, X_n \sim_{\text{i.i.d.}} P$ and $x \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}\left(\max\left(X_1, \ldots, X_n\right) \leq x\right) &= \mathbb{P}\left(X_1 \leq x \text{ and } \ldots \text{ and } X_n \leq x\right) \\ &= \prod_{i=1}^{n} \mathbb{P}\left(X_i \leq x\right) = \mathbb{P}\left(X_1 \leq x\right)^n,\end{aligned}$$

and so $F_{(n)}(x) = F(x)^n$. Differentiating, we obtain $f_{(n)}(x) = nF(x)^{n-1} f(x)$.

Hence

$$\begin{aligned}D_{\mathrm{KL}}\left(P_{(n)} \parallel P\right) &= \int_{-\infty}^{\infty} f_{(n)}(x) \log\left(\frac{f_{(n)}(x)}{f(x)}\right) \mathrm{d}x \\ &= \int_{-\infty}^{\infty} f_{(n)}(x) \log\left(nF(x)^{n-1}\right) \mathrm{d}x \\ &= \log n + (n-1) \int_{-\infty}^{\infty} f_{(n)}(x) \log\left(F(x)\right) \mathrm{d}x \\ &= \log n + (n-1)\left(\left[F_{(n)}(x) \log\left(F(x)\right)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F_{(n)}(x) \frac{1}{F(x)} f(x)\, \mathrm{d}x\right) \\ &= \log n + (n-1)\left(\left[F(x)^n \log\left(F(x)\right)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F(x)^{n-1} f(x)\, \mathrm{d}x\right) \\ &= \log n - (n-1) \int_{0}^{1} u^{n-1} \mathrm{d}u \\ &= \log n - \frac{n-1}{n}.\end{aligned}$$

□

4 The discrete case

If P is a discrete distribution with finite support, then the KL divergence $D_{\mathrm{KL}}\left(P_{(n)} \parallel P\right)$ is no longer exactly $\log n - \frac{n-1}{n}$. Instead, it is less than or equal to this, but the difference is small, as shown by the following result.

Theorem 1. *Let P be a discrete distribution over $\mathbb{R}$ with finite support and $n \in \mathbb{N}$. Then the KL divergence from P to the max-of-n-from-P distribution,*

$$D_{KL}\left(P_{(n)} \parallel P\right) = \log n - \frac{n-1}{n} - \varepsilon \qquad \textit{for some} \qquad 0 \leq \varepsilon < np_{max} + \frac{n^3 p_{max}^2}{4},$$

where p_{max} is the maximum probability of any single value under P.

The proof of Theorem 1 follows the same outline as the proof of Proposition 1, but makes discretized approximations to the integrals. All of the technical difficulty involves tracking and bounding the discretization error. The full proof can be found in Appendix A.

In order to better understand how the error term ε in this result behaves empirically, we calculated it numerically for each value of n from 2 to 64, for different discrete uniform distributions P. The results of these calculations are shown in Figure 1. Based on this numerical evidence, we make the following conjecture.

Conjecture. *The inequality in Theorem 1 can be improved to $0 \leq \varepsilon < \frac{np_{max}}{4}$.*

It is straightforward to show that the probability that n i.i.d. samples from P contain a repetition is at least $1 - e^{-np_{\max}}\left(1 + np_{\max}e^{p_{\max}}\right)$. Hence if this conjecture holds and $\varepsilon \geq 1$, say, then the probability of a repetition is greater than $1 - e^{-4}\left(1 + 4e\right) \approx 0.78$. This provides a practical way to verify that ε is small when performing best-of-n sampling, by checking that it is rare for the n i.i.d. samples to contain a repetition.

Acknowledgements

Thanks to John Schulman for helpful discussions.

References

L. Gao, J. Schulman, and J. Hilton. Scaling laws for reward model overoptimization. In *International Conference on Machine Learning*, pages 10835–10866. PMLR, 2023.

N. Stiennon, L. Ouyang, J. Wu, D. Ziegler, R. Lowe, C. Voss, A. Radford, D. Amodei, and P. F. Christiano. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021, 2020.

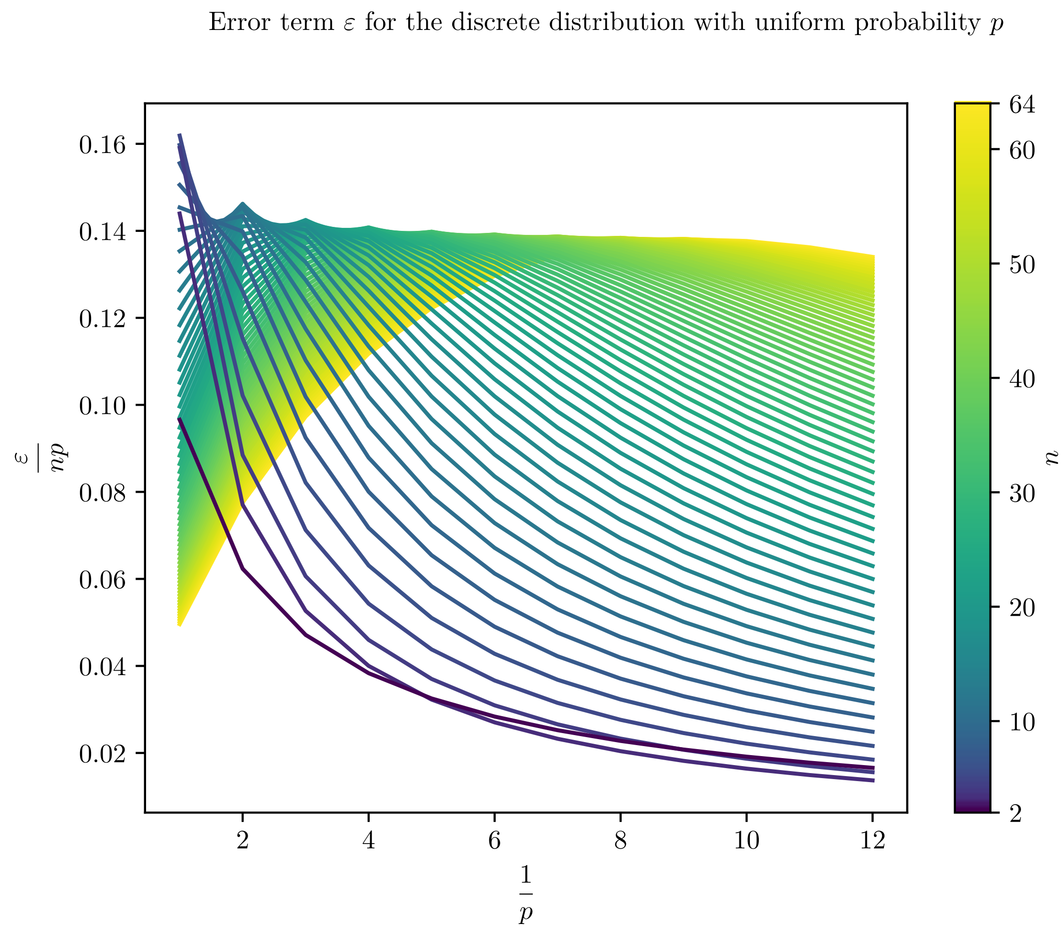

Figure 1: The error term $\varepsilon = \left(\log n - \frac{n-1}{n}\right) - D_{\mathrm{KL}}\left(P_{(n)} \parallel P\right)$ as a function of n and p when P is the discrete uniform distribution that takes on $\frac{1}{p}$ values with probability p each. Each line shows how $\frac{\varepsilon}{np}$ varies with $\frac{1}{p}$ for fixed n. The trend of the upper frontier suggests that $\frac{\varepsilon}{np} < 0.17$ for all n and p, motivating our conjecture that $\varepsilon < \frac{np_{\max}}{4}$ for all P.

A Proof for the discrete case

Proof of Theorem 1. The proof follows the same outline as the proof of Proposition 1, but makes discretized approximations to the integrals. In order to bound the discretization error, we use the following inequalities for $x, \delta > 0$:

$$(x+\delta)^n - x^n \leq n\delta\,(x+\delta)^{n-1} \tag{1}$$

$$(x+\delta)^n - x^n \geq n\delta\left(x+\frac{\delta}{2}\right)^{n-1} \geq n\delta x^{n-1} \tag{2}$$

$$(x+\delta)^n - x^n \geq n\delta x^{n-1} + \frac{n\,(n-1)}{2}\delta^2 x^{n-2} \tag{3}$$

$$(x+\delta)^n - x^n \geq n\delta\,(x+\delta)^{n-1} - \frac{n\,(n-1)}{2}\delta^2\,(x+\delta)^{n-2} \tag{4}$$

$$\log(x+\delta) - \log(x) \leq \frac{\delta}{x} \tag{5}$$

$$\log(x+\delta) - \log(x) \leq \frac{\delta}{x+\delta} + \frac{\delta^2}{2x^2} \tag{6}$$

These all follow from Taylor's theorem with Lagrange's form of the remainder, or by Jensen's inequality for (2).

Let k be the number of points in the support of P. For $i = 1, \ldots, k$, let p_i be the probability under P of the ith point, ordered from smallest to largest, and let $p_{(n)i}$ be the probability under $P_{(n)}$ of the ith point, ordered from smallest to largest.

As in Proposition 1, given $X_1, \ldots, X_n \sim_{\text{i.i.d.}} P$ and $x \in \mathbb{R}$, we have $\mathbb{P}(\max(X_1, \ldots, X_n) \leq x) = \mathbb{P}(X_1 \leq x)^n$, and so

$$p_{(n)i} = (p_1 + \cdots + p_i)^n - (p_1 + \cdots + p_{i-1})^n$$

for $i = 1, \ldots, k$.

We first prove the lower bound for ε. We have

$$\begin{aligned}
D_{\mathrm{KL}}\left(P_{(n)} \parallel P\right) &= \sum_{i=1}^{k} p_{(n)i} \log\left(\frac{(p_1+\cdots+p_i)^n - (p_1+\cdots+p_{i-1})^n}{p_i}\right) \\
&\leq \sum_{i=1}^{k} p_{(n)i} \log\left(\frac{np_i\,(p_1+\cdots+p_i)^{n-1}}{p_i}\right) \qquad \text{by (1)} \\
&= \log n + (n-1)\sum_{i=1}^{k} p_{(n)i} \log(p_1+\cdots+p_i) \\
&= \log n + (n-1)\sum_{i=1}^{k} \left((p_1+\cdots+p_i)^n - (p_1+\cdots+p_{i-1})^n\right)\log(p_1+\cdots+p_i) \\
&= \log n + (n-1)\Bigg((p_1+\cdots+p_k)^n \log(p_1+\cdots+p_k) \\
&\qquad - \sum_{i=1}^{k-1} (p_1+\cdots+p_i)^n \left(\log(p_1+\cdots+p_{i+1}) - \log(p_1+\cdots+p_i)\right)\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&\leq \log n-(n-1)\sum_{i=1}^{k-1}\left(p_1+\cdots+p_i\right)^n \frac{p_{i+1}}{p_1+\cdots+p_i} && \text{by (5)}\\
&=\log n-(n-1)\sum_{i=1}^{k-1} p_{i+1}\left(p_1+\cdots+p_i\right)^{n-1}\\
&\leq \log n-(n-1)\sum_{i=1}^{k-1}\frac{1}{n}\left(\left(p_1+\cdots+p_{i+1}\right)^n-\left(p_1+\cdots+p_i\right)^n\right) && \text{by (2)}\\
&=\log n-\frac{n-1}{n}\left(\left(p_1+\cdots+p_k\right)^n-p_1^n\right)\\
&\leq \log n-\frac{n-1}{n},
\end{aligned}
$$

as required.

The proof of the upper bound for ε is similar but requires more care. We have

$$
\begin{aligned}
&D_{\mathrm{KL}}\left(P_{(n)} \parallel P\right)\\
&=\sum_{i=1}^{k} p_{(n)i}\log\left(\frac{\left(p_1+\cdots+p_i\right)^n-\left(p_1+\cdots+p_{i-1}\right)^n}{p_i}\right)\\
&\geq \sum_{i=1}^{k} p_{(n)i}\log\left(\frac{np_i\left(p_1+\cdots+p_{i-1}+\frac{p_i}{2}\right)^{n-1}}{p_i}\right) \qquad \text{by (2)}\\
&=\log n+(n-1)\sum_{i=1}^{k} p_{(n)i}\log\left(p_1+\cdots+p_{i-1}+\frac{p_i}{2}\right)\\
&=\log n+(n-1)\sum_{i=1}^{k}\left(\left(p_1+\cdots+p_i\right)^n-\left(p_1+\cdots+p_{i-1}\right)^n\right)\log\left(p_1+\cdots+p_{i-1}+\frac{p_i}{2}\right)\\
&=\log n+(n-1)\Bigg(\left(p_1+\cdots+p_k\right)^n\log\left(p_1+\cdots+p_{k-1}+\frac{p_k}{2}\right)\\
&\qquad\qquad -\sum_{i=1}^{k-1}\left(p_1+\cdots+p_i\right)^n\left(\log\left(p_1+\cdots+p_i+\frac{p_{i+1}}{2}\right)-\log\left(p_1+\cdots+p_{i-1}+\frac{p_i}{2}\right)\right)\Bigg)\\
&=\log n-(n-1)\left(\sum_{i=1}^{k-1}\left(p_1+\cdots+p_i\right)^n\left(\log\left(p_1+\cdots+p_i+\frac{p_{i+1}}{2}\right)-\log\left(p_1+\cdots+p_{i-1}+\frac{p_i}{2}\right)\right)+\varepsilon_1\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\varepsilon_1 &= -\left(p_1+\cdots+p_k\right)^n\log\left(p_1+\cdots+p_{k-1}+\frac{p_k}{2}\right)\\
&=-\log\left(1-\frac{p_k}{2}\right)\\
&\leq \frac{p_k}{2}+\frac{p_k^2}{4}\\
&\leq \frac{p_{\max}}{2}+\frac{p_{\max}^2}{4}.
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\sum_{i=1}^{k-1} (p_1 + \cdots + p_i)^n \left(\log\left(p_1 + \cdots + p_i + \frac{p_{i+1}}{2}\right) - \log\left(p_1 + \cdots + p_{i-1} + \frac{p_i}{2}\right)\right) \\
&= \sum_{i=1}^{k-1} (p_1 + \cdots + p_i)^n \left(\log\left(p_1 + \cdots + p_i + \frac{p_{i+1}}{2}\right) - \log\left(p_1 + \cdots + p_i\right)\right. \\
&\qquad\qquad\qquad\qquad \left. + \log\left(p_1 + \cdots + p_i\right) - \log\left(p_1 + \cdots + p_{i-1} + \frac{p_i}{2}\right)\right) \\
&\le \sum_{i=1}^{k-1} (p_1 + \cdots + p_i)^n \left(\frac{\frac{p_{i+1}}{2}}{p_1 + \cdots + p_i} + \frac{\frac{p_i}{2}}{p_1 + \cdots + p_i} + \frac{\left(\frac{p_i}{2}\right)^2}{2\left(p_1 + \cdots + p_{i-1} + \frac{p_i}{2}\right)^2}\right) \qquad \text{by (5) and (6)} \\
&= \left(\frac{1}{2} \sum_{i=1}^{k-1} (p_{i+1} + p_i)(p_1 + \cdots + p_i)^{n-1}\right) + \varepsilon_2
\end{aligned}
$$

where

$$
\begin{aligned}
\varepsilon_2 &= \sum_{i=1}^{k-1} (p_1 + \cdots + p_i)^n \frac{\left(\frac{p_i}{2}\right)^2}{2\left(p_1 + \cdots + p_{i-1} + \frac{p_i}{2}\right)^2} \\
&\le \sum_{i=1}^{k-1} (p_1 + \cdots + p_i)^n \frac{\left(\frac{p_i}{2}\right)^2}{2\left(\frac{p_1}{2} + \cdots + \frac{p_{i-1}}{2} + \frac{p_i}{2}\right)^2} \\
&= \sum_{i=1}^{k-1} (p_1 + \cdots + p_i)^{n-2} \frac{p_i^2}{2} \\
&\le \sum_{i=1}^{k-1} \frac{p_i p_{\max}}{2} \\
&\le \frac{p_{\max}}{2}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\frac{1}{2} \sum_{i=1}^{k-1} (p_{i+1} + p_i)(p_1 + \cdots + p_i)^{n-1} \\
&\le \frac{1}{2} \sum_{i=1}^{k-1} \left(\frac{1}{n}(p_1 + \cdots + p_{i+1})^n - \frac{1}{n}(p_1 + \cdots + p_i)^n - \frac{n-1}{2} p_{i+1}^2 (p_1 + \cdots + p_i)^{n-2}\right. \\
&\qquad\qquad \left. + \frac{1}{n}(p_1 + \cdots + p_i)^n - \frac{1}{n}(p_1 + \cdots + p_{i-1})^n + \frac{n-1}{2} p_i^2 (p_1 + \cdots + p_i)^{n-2}\right) \qquad \text{by (3) and (4)} \\
&= \frac{1}{2n}\left((p_1 + \cdots + p_k)^n - p_1^n + (p_1 + \cdots + p_{k-1})^n\right) + \frac{n-1}{4} \sum_{i=1}^{k-1} \left(p_i^2 - p_{i+1}^2\right)(p_1 + \cdots + p_i)^{n-2} \\
&\le \frac{1}{n} + \varepsilon_3
\end{aligned}
$$

where

$$
\begin{aligned}
\varepsilon_3 &= \frac{n-1}{4} \sum_{i=1}^{k-1} \left(p_i^2 - p_{i+1}^2\right) \left(p_1 + \cdots + p_i\right)^{n-2} \\
&= \frac{n-1}{4} \left(p_1^2 p_1^{n-2} - p_k^2 \left(p_1 + \cdots + p_{k-1}\right)^{n-2} + \sum_{i=2}^{k-1} p_i^2 \left(\left(p_1 + \cdots + p_i\right)^{n-2} - \left(p_1 + \cdots + p_{i-1}\right)^{n-2} \right) \right) \\
&\leq \frac{n-1}{4} \left(p_{\max}^2 + \sum_{i=2}^{k-1} (n-2)\, p_i^3 \left(p_1 + \cdots + p_i\right)^{n-3} \right) \qquad \text{by (1)} \\
&\leq \frac{n-1}{4} \left(p_{\max}^2 + (n-2) \sum_{i=2}^{k-1} p_i p_{\max}^2 \right) \\
&\leq \frac{(n-1)^2 p_{\max}^2}{4}.
\end{aligned}
$$

Piecing everything together, we obtain $D_{\mathrm{KL}}\left(P_{(n)} \parallel P\right) = \log n - \dfrac{n-1}{n} - \varepsilon$ with

$$
\begin{aligned}
\varepsilon &\leq (n-1)\left(\varepsilon_1 + \varepsilon_2 + \varepsilon_3\right) \\
&\leq (n-1) \left(\left(\frac{p_{\max}}{2} + \frac{p_{\max}^2}{4} \right) + \frac{p_{\max}}{2} + \frac{(n-1)^2 p_{\max}^2}{4} \right) \\
&= (n-1) \left(p_{\max} + \frac{\left(n^2 - 2n + 2\right) p_{\max}^2}{4} \right) \\
&< n p_{\max} + \frac{n^3 p_{\max}^2}{4}. \qquad \square
\end{aligned}
$$