Note on Flow Matching and Diffusion Models

Anderson Lee

September 5, 2025

Abstract

Information Theory

Contents

Chapter 1

Shannon's Entropy

1.1 Entropy Function

Definition 1.1.1. Entropy function is introduced to determine the *uncertainty of a random variable* X. Providing two **intuitions**

1. The number of bits of information that we don't know about X
2. The number of bits needed to describe X

$$\begin{aligned} H(X) &= -\sum_x P(X=x)\log P(X=x) && (1.1)\\ &= -\mathbb{E}_X\left[\log P(X)\right] && (1.2)\\ &= \mathbb{E}_X\left[\log \frac{1}{P(X)}\right] && (1.3)\end{aligned}$$

Remark. Entropy function H is

1. **Symmetric**
2. **Continuous**
3. **Concave**

1.2 Deriving Entropy Function

We start with thinking about the desired properties of H.

1. H is a continuous symmetric function that only depends on the distribution
2. $H(\frac{1}{2}, \frac{1}{2}) = 1$ (**Scaling**)
3. $H(\frac{1}{2}, \frac{1}{2}) \le H(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \le H(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) \dots$
4. **The Grouping Axiom**

   Example (Simple Case). Let $X \in \{0, 1, 2\}$ and $P_X = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ We need one bit to tell whether $X = 0$ or $X \in \{1, 2\}$. If $X = 0$, then we only need one bit. If $X \in \{1, 2\}$, we need more bits to determine X. Therefore,

   $$H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot H(\frac{1}{2}, \frac{1}{2}) = 1.5$$

   The first term is the cost of determining whether $X = 0$ or $X \in \{1, 2\}$. The second term is the cost of $X = 0$, which is simply 0. The third term is the cost of identifying X when $X \in \{1, 2\}$

Example (Special Case). Let $X \in \{1, \ldots, 9\}$, $P_X = (p_1, \ldots, p_9)$, and $q_1 = p_1 + p_2 + p_3, q_2 = p_4 + p_5 + p_6, q_3 = p_7 + p_8 + p_9$. Then,

$$H(p_1, \ldots, p_9) = H(q_1, q_2, q_3) + q_1 \cdot H(\frac{p_1}{q_1}, \frac{p_2}{q_1}, \frac{p_3}{q_1}) + q_2 \cdot H(\frac{p_4}{q_2}, \frac{p_5}{q_2}, \frac{p_6}{q_2}) + H(\frac{p_7}{q_3}, \frac{p_8}{q_3}, \frac{p_9}{q_3})$$

Example (Another Special Case).

$$H(p_1, p_2, p_3, \ldots, p_m) = H(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2) \cdot H(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2})$$

This form implies the general case

1.2.1 Compute $H(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

Firstly, using grouping axiom, we know that $H(U_{3^n}) = n \cdot H(U_3)$. Let k be such that $2^k < 3^n < 2^{k+1}$, that is $k = \lfloor n \cdot \log_2 3 \rfloor$ and $k + 1 = \lceil n \cdot \log_2 3 \rceil$. By the third axiom (increasing entropy with more uniformly distributed elements), we have

$$\begin{aligned} k = H(U_{2^k}) \leq& H(U_{3^n}) \leq H(U_{2^{k+1}}) = k + 1 \\ \frac{\lfloor n \cdot \log_2 3 \rfloor}{n} \leq& H(U_3) \leq \frac{\lceil n \cdot \log_2 3 \rceil}{n} \end{aligned}$$

for all n. Therefore,

$$H(U_3) = \log_2 3$$

1.2.2 Compute $H(p, q)$

Assume $p = \frac{k}{m}, q = \frac{m-k}{m}$, where k, m are integers. Take $U_m = (\frac{1}{m}, \ldots, \frac{1}{m})$. By partitioning $m = k + (m - k)$, we have

$$\begin{aligned} H(U_m) &= H(p, q) + p \cdot H(U_k) + q \cdot H(U_{m-k}) \\ H(p, q) &= H(U_m) - p \cdot H(U_k) - q \cdot H(U_{m-k}) \\ &= \log m - p \cdot \log k - q \cdot (\log(m - k)) \\ &= p \cdot (\log m - \log k) + q \cdot (\log m - \log(m - k)) \\ &= -p \cdot \log p - q \cdot \log q \end{aligned}$$

1.3 Fundamental Inequalities

Theorem 1.3.1. If X is a random variable over n elements,

$$0 \leq H(X) \leq \log n$$

The first inequality comes from log being a concave function trivially. The second inequality comes from Jenson's inequality

Note. The intuition of this inequality is that any random variable of n elements should require at most the number of bits required to describe a uniform distribution over n elements.

Lemma 1.3.1 (Jensen's Inequality). A concave function f: for every t_1, t_2 (in the interval of the concavity) and every $0 \leq \lambda \leq 1$,

$$\lambda \cdot f(t_1) + (1 - \lambda) \cdot f(t_2) \leq f(\lambda \cdot t_1 + (1 - \lambda) \cdot t_2) \tag{1.4}$$

A more general form: for $t_1, \ldots, t_n$ and $\lambda_1, \ldots \lambda_n$

$$\sum_i \lambda_i f(t_i) \leq f(\sum_i \lambda_i \cdot t_i) \tag{1.5}$$

In expectation form, for every random variable X,

$$\mathbb{E}(f(X)) \leq f(\mathbb{E}(X)) \tag{1.6}$$

Proof of Second Inequality.

$$H(X) = \sum_i^n P(X = x_i) \log \frac{1}{P(X = x_i)} \leq \log\left(\sum_i^n P(X = x_i) \cdot \frac{1}{P(X = x_i)}\right) = \log n \tag{1.7}$$

■

1.3.1 Conditional Entropy

Intuitively, for random variable X, Y, the conditional entropy $H(X|Y)$ will be the uncertainty remained in X after we observe Y.

$$H(X|Y) = \mathbb{E}_{Y \to b} H(X|Y = b) = \sum_b P(Y = b) \cdot H(X|Y = b) \tag{1.8}$$

$$= -\sum_b P(Y = b) \cdot (\sum_a P(X = a|Y = b) \cdot \log P(X = a|Y = b)) \tag{1.9}$$

$$= -\sum_{a,b} P(X = a, Y = b) \cdot \log P(X = a|Y = b) \tag{1.10}$$

$$= -\mathbb{E}_{X,Y} [\log P(X|Y)] \tag{1.11}$$

Chapter 2

Known Bugs

Lecture 2: Second Lecture

2.1 Introduction

9 Sep. 08:00

Nothing is bugs-free. There are some known bugs which I don't have incentive to solve, or it is hard to solve whatsoever. Let me list some of them.

2.1.1 Footnote Environment

It's easy to let you fall into a situation that you want to keep using `footnote` to add a bunch of unrelated stuffs. However, with our environment there is a known strange behavior, which is following.

Example. Footnote![a]

Remark. Oops! footnote somehow shows up earlier than expect![a]

[a]This is a footnote!
[a]This is another footnote!

Bugs caught![b]

[b]The final footnote which is ok!

As we saw, the footnote in the `Example` environment should show at the bottom of its own box, but it's caught by `Remark` which causes the unwanted behavior. Unfortunately, I haven't found a nice way to solve this. A potential way to solve this is by using `footnotemark` with `footnotetext` placing at the bottom of the environment, but this is tedious and needs lots of manual tweaking.

Furthermore, not sure whether you notice it or not, but the color box of `Remark` is not quite right! It extends to the right, another trick bug...

2.1.2 Mdframe Environment

Though `mdframe` package is nice and is the key theme throughout this template, but it has some kind of weird behavior. Let's see the demo.

Proof of ??. We need to prove the followings.

Claim. $E = mc^2$.

Proof. Nonsense.
Nonsense,
Nonsense,
Nonsense,
Nonsense,
Nonsense. ⊛

■

I expect it should break much earlier, and this seems to be an algorithmic issue of `mdframe`. One potential solution is to use `tcolorbox` instead, but I haven't completely figure it out, hence I can't really say anything right now.

Chapter 3

Test

Lecture 2: Second Lecture

3.1 Introduction

9 Sep. 08:00

Nothing is bugs-free. There are some known bugs which I don't have incentive to solve, or it is hard to solve whatsoever. Let me list some of them.

3.1.1 Footnote Environment

It's easy to let you fall into a situation that you want to keep using `footnote` to add a bunch of unrelated stuffs. However, with our environment there is a known strange behavior, which is following.

Example. Footnote![a]

> **Remark.** Oops! footnote somehow shows up earlier than expect![a]
>
> [a]This is a footnote!
> [a]This is another footnote!

Bugs caught![b]

[b]The final footnote which is ok!

As we saw, the footnote in the `Example` environment should show at the bottom of its own box, but it's caught by `Remark` which causes the unwanted behavior. Unfortunately, I haven't found a nice way to solve this. A potential way to solve this is by using `footnotemark` with `footnotetext` placing at the bottom of the environment, but this is tedious and needs lots of manual tweaking.

Furthermore, not sure whether you notice it or not, but the color box of `Remark` is not quite right! It extends to the right, another trick bug...

3.1.2 Mdframe Environment

Though `mdframe` package is nice and is the key theme throughout this template, but it has some kind of weird behavior. Let's see the demo.

Proof of ??. We need to prove the followings.

Claim. $E = mc^2$.

Proof. Nonsense.
Nonsense,
Nonsense,
Nonsense,
Nonsense,
Nonsense. ⊛

■

I expect it should break much earlier, and this seems to be an algorithmic issue of `mdframe`. One potential solution is to use `tcolorbox` instead, but I haven't completely figure it out, hence I can't really say anything right now.

Appendix

Appendix A

Additional Proofs

A.1 Proof of ??

We can now prove **??**.

Proof of ??. See here. ■