

1 A Model with Power-Law Behavior

Based on our definitions it is very easy to define models without power-law behavior (see chapter ??). However, one might ask the question whether there actually exists a model that satisfies strong power-law behavior.

1.1 The Model

Which simple model could we try to analyze? Well, one intuitive option is to use a normal distribution. The model S_n is then simply a normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

There is one problem, though. We defined S_n to be a probability measure, but $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a *probability density function* over \mathbb{R}^n . So, how can we solve this?

Well, we can just set $\boldsymbol{\mu} := \mathbf{0}$, and discretize our probability space by integrating over the quadrants. Thus, we created a model S_n over $\{-1, 1\}^n$, where for example $S_2(11)$ is defined as the integral over the first quadrant, $S_2(-11)$ as the integral over the second quadrant, and so on.

Definition 1.1 (The Model). Let $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ be a normal distribution with a positive definite parameter matrix $\boldsymbol{\Sigma}$, and let $p(\mathbf{x})$ denote the associated probability density function. We define a model $S_{n, \boldsymbol{\Sigma}}$ over $\{-1, 1\}^n$ with

$$S_{n, \boldsymbol{\Sigma}}(w) = \int_{Q_w} p(\mathbf{x}) d\mathbf{x} \quad ,$$

where Q_w is the quadrant

$$Q_w = \{\mathbf{x} \in \mathbb{R}^n \mid \forall i \in [n] : w_i x_i \geq 0\} \quad .$$

Remark 1.1. It is obvious that this is a valid probability distribution.

Note that we are not really interested in the exact probabilities of our model. Instead, we focus on the pairwise mutual information.

1.1.1 A Formula for Mutual Information

Let's take a step back and consider the continuous normal distribution again. The mutual information is also well defined in this case. We will make use of some properties of the normal distribution, like the following (without proof):

Proposition 1.1 (Marginal Distributions of a Normal Distribution). *Let the n -dimensional random vector \mathbf{X} follow a multivariate normal distribution, $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We partition \mathbf{X} into two sub-vectors, $\mathbf{X}_1 \in \mathbb{R}^k$ and $\mathbf{X}_2 \in \mathbb{R}^{n-k}$, with the corresponding partitions of the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as:*

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \quad , \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad , \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \quad ,$$

where $\boldsymbol{\mu}_1 \in \mathbb{R}^k$ and $\boldsymbol{\Sigma}_{11}$ is a $k \times k$ matrix.

Then the marginal distribution of the sub-vector \mathbf{X}_1 is also a multivariate normal distribution given by:

$$\mathbf{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \quad .$$

Proposition 1.2 (Entropy of a Multivariate Normal Distribution). *Let the random vector $\mathbf{X} \in \mathbb{R}^n$ follow a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with a positive definite covariance matrix $\boldsymbol{\Sigma}$. The entropy of \mathbf{X} is given by*

$$H(\mathbf{X}) = \frac{1}{2} \log((2\pi e)^n \det(\boldsymbol{\Sigma})) \quad ,$$

where \log is the natural logarithm.

Proof. The Entropy is defined as $H(\mathbf{X}) = E[-\log p(\mathbf{X})]$, where $p(\mathbf{x})$ is the probability density function (PDF). The PDF for $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad .$$

Taking the natural logarithm of the PDF gives:

$$\log p(\mathbf{x}) = -\frac{1}{2} \log((2\pi)^n \det(\boldsymbol{\Sigma})) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \quad .$$

Now, we take the expectation of $-\log p(\mathbf{X})$:

$$\begin{aligned} H(\mathbf{X}) &= E\left[\frac{1}{2} \log((2\pi)^n \det(\boldsymbol{\Sigma})) + \frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right] \\ &= \frac{1}{2} \log((2\pi)^n \det(\boldsymbol{\Sigma})) + \frac{1}{2} E[(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})] \quad . \end{aligned}$$

We evaluate the expectation term. Since the quadratic form is a scalar, it equals its trace.

$$\begin{aligned} E[(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})] &= E[\text{tr}((\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}))] \\ &= E[\text{tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T)] \\ &= \text{tr}(\boldsymbol{\Sigma}^{-1} E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]) \\ &= \text{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}) \\ &= \text{tr}(\mathbf{I}_n) = n \quad . \end{aligned}$$

Substituting this result back into the entropy equation:

$$\begin{aligned}
H(\mathbf{X}) &= \frac{1}{2} \log((2\pi)^n \det(\mathbf{\Sigma})) + \frac{1}{2}n \\
&= \frac{1}{2} (n \log(2\pi) + \log(\det(\mathbf{\Sigma})) + n \log(e)) \\
&= \frac{1}{2} (n \log(2\pi e) + \log(\det(\mathbf{\Sigma}))) \\
&= \frac{1}{2} \log((2\pi e)^n \det(\mathbf{\Sigma})) \quad .
\end{aligned}$$

□

Let's now consider

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \quad .$$

With our previous results in mind, we calculate the mutual information:

$$\begin{aligned}
I(X; Y) &= H(X) + H(Y) - H(X, Y) \\
&= \left[\frac{1}{2} \log(2\pi e \cdot 1) \right] + \left[\frac{1}{2} \log(2\pi e \cdot 1) \right] - \left[\frac{1}{2} \log\left((2\pi e)^2 \det\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \right] \\
&= \log(2\pi e) - \frac{1}{2} \log((2\pi e)^2 (1 - \rho^2)) \\
&= \log(2\pi e) - \frac{1}{2} (\log((2\pi e)^2) + \log(1 - \rho^2)) \\
&= \log(2\pi e) - \frac{1}{2} (2 \log(2\pi e) + \log(1 - \rho^2)) \\
&= \log(2\pi e) - \log(2\pi e) - \frac{1}{2} \log(1 - \rho^2) \\
&= -\frac{1}{2} \log(1 - \rho^2) \quad .
\end{aligned} \tag{1}$$

1.1.2 Initializing Parameters

Our discretized model $S_{n, \mathbf{\Sigma}}$ behaves differently than the continuous normal distribution, and so do the pairwise mutual information measures. However, let's act like they behave similar.

Let (Y_1, \dots, Y_n) denote the random variables of the n-dimensional normal distribution

$$\mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{n-2} \\ \vdots & & \ddots & & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & 1 \end{pmatrix}\right) \quad .$$

We want the mutual information $I(Y_i; Y_j)$ to follow a power-law, i.e.

$$I(Y_i; Y_j) \stackrel{!}{=} c|i-j|^{-\alpha} \quad ,$$

for some $c, \alpha \in \mathbb{R}_{>0}$.

Let $\Delta := |i-j|$. Using equation (1) it follows that

$$\begin{aligned} -\frac{1}{2} \log(1 - \rho_\Delta^2) &= c\Delta^{-\alpha} \\ \log(1 - \rho_\Delta^2) &= -2c\Delta^{-\alpha} \\ 1 - \rho_\Delta^2 &= e^{-2c\Delta^{-\alpha}} \\ \rho_\Delta^2 &= 1 - e^{-2c\Delta^{-\alpha}} \\ \rho_\Delta &= \pm \sqrt{1 - e^{-2c\Delta^{-\alpha}}} \quad . \end{aligned}$$

Thus, upon defining the constants c and α , we can directly calculate the parameters ρ_Δ , where we have the freedom to chose the sign.

1.1.3 Ensure Positive Definiteness

What is left to do is to ensure that the so defined matrix Σ is indeed positive definite. How should we define, c and α , and which signs should we choose?

Note that Σ is symmetric, and has positive entries along its diagonal. Thus, it is sufficient for positive definiteness to show that Σ is strictly diagonally dominant, i.e.

$$\begin{aligned} \forall i \in [n] : |\Sigma_{ii}| &> \sum_{j \neq i} |\Sigma_{ij}| \\ \iff \forall i \in [n] : 1 &> \sum_{j \neq i} |\rho_{|i-j|}| \quad . \end{aligned} \tag{2}$$

When using this method, we immediately see that the sign of ρ_Δ is of no meaning. Hence, we may chose all entries of Σ to be positive.

Note that a specific entry ρ_Δ can occur two times in the same row. This happens especially in the middle rows of the matrix. However, for a fixed Δ , ρ_Δ can occur at most two times in the same row. Hence, we derive the following bound:

$$\sum_{j \neq i} |\rho_{|i-j|}| \leq 2 \sum_{\Delta=1}^{\infty} \rho_\Delta \quad .$$