

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 family of normal distributions $\mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ with parameters $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$.

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

Well, we can just set $\boldsymbol{\mu}_n := \mathbf{0}$, and discretize our probability space by integrating over the quadrants. Thus, we created a probability measure 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 $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_n)$ be a family of normal distributions with a positive definite parameter matrix $\boldsymbol{\Sigma}_n$, and let $p_n(\mathbf{x})$ denote the associated probability density functions. We define a model $S_{n, \boldsymbol{\Sigma}_n}$ over $\{-1, 1\}^n$ with

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

where Q_w is the quadrant

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

We immediately see that our model is indeed valid. Hence, we now analyze how the pairwise mutual information behaves.

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

In fact, our discretized model $S_{n, \mathbf{\Sigma}_n}$ 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 $\tau := |i-j|$. Using proposition 1.1 and equation (1) it follows that

$$\begin{aligned} I(Y_i; Y_j) &= c|i-j|^{-\alpha} \\ \iff -\frac{1}{2} \log(1 - \rho_\tau^2) &= c\tau^{-\alpha} \\ \iff \log(1 - \rho_\tau^2) &= -2c\tau^{-\alpha} \\ \iff 1 - \rho_\tau^2 &= e^{-2c\tau^{-\alpha}} \\ \iff \rho_\tau^2 &= 1 - e^{-2c\tau^{-\alpha}} \\ \iff \rho_\tau &= \pm \sqrt{1 - e^{-2c\tau^{-\alpha}}} \quad . \end{aligned}$$

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

1.1.3 Ensure Positive Definiteness

What is left to do is to ensure that the so defined matrix $\mathbf{\Sigma} \equiv \mathbf{\Sigma}_n$ is indeed positive definite. How should we define, c and α , and which signs should we choose?

Note that $\mathbf{\Sigma}$ is symmetric, and has positive entries along its diagonal. Thus, it is sufficient for positive definiteness to show that $\mathbf{\Sigma}$ 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 we can freely choose the sign of ρ_τ . Hence, we may choose all entries of $\mathbf{\Sigma}$ to be positive.

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

$$\sum_{j \neq i} |\rho_{|i-j|}| \leq 2 \sum_{\tau=1}^{\infty} \rho_\tau \quad . \tag{3}$$

We need an upper bound for ρ_τ . As it turns out, we will need a lower bound later as well. Hence, we will provide both at this point:

Lemma 1.1. *Let $\rho_\tau := \sqrt{1 - e^{-2c\tau^{-\alpha}}}$. Then, we have*

$$\sqrt{\frac{2c}{2c+1}}\tau^{-0.5\alpha} \leq \rho_\tau \leq \sqrt{2c}\tau^{-0.5\alpha} \quad ,$$

where the lower bound holds for all $\tau \in \mathbb{N}_{>0}$, and the upper bound for $\tau > (2c)^{\frac{1}{\alpha}}$.

Proof. Note the inequality $x + 1 \leq e^x$. For $x > -1$ it follows that

$$e^{-x} \leq \frac{1}{x+1} \quad . \quad (4)$$

Applying equation 4 to $e^{-2c\tau^{-\alpha}}$ with $x := 2c\tau^{-\alpha}$ yields

$$e^{-2c\tau^{-\alpha}} \leq \frac{1}{2c\tau^{-\alpha} + 1} \quad ,$$

where we have $x > -1 \iff 2c\tau^{-\alpha} > -1 \iff \tau \in \mathbb{N}_{>0}$.

Plugging this into the equation for ρ_τ gives

$$\begin{aligned} \rho_\tau &\geq \sqrt{1 - \frac{1}{2c\tau^{-\alpha} + 1}} \\ &= \sqrt{\frac{2c\tau^{-\alpha}}{2c\tau^{-\alpha} + 1}} \\ &= \sqrt{2c}\tau^{-0.5\alpha} \sqrt{\frac{1}{2c\tau^{-\alpha} + 1}} \\ &\geq \sqrt{2c}\tau^{-0.5\alpha} \sqrt{\frac{1}{2c1^{-\alpha} + 1}} \\ &= \sqrt{\frac{2c}{2c+1}}\tau^{-0.5\alpha} \quad . \end{aligned}$$

This proves the lower bound.

Let's now focus on the upper bound. Applying equation 4 to $e^{2c\tau^{-\alpha}}$ with $x := -2c\tau^{-\alpha}$ yields

$$e^{2c\tau^{-\alpha}} \leq \frac{1}{-2c\tau^{-\alpha} + 1} \quad ,$$

where we have $x > -1 \iff -2c\tau^{-\alpha} > -1 \iff \tau > (2c)^{\frac{1}{\alpha}}$.

Under this assumption, both sides of the inequality are positive, and hence it follows that

$$e^{-2c\tau^{-\alpha}} \geq 1 - 2c\tau^{-\alpha} \quad .$$

Plugging this into the equation for ρ_τ gives

$$\begin{aligned}\rho_\tau &\leq \sqrt{1 - (1 - 2c\tau^{-\alpha})} \\ &= \sqrt{2c\tau^{-\alpha}} \\ &= \sqrt{2c}\tau^{-0.5\alpha} \quad .\end{aligned}$$

□

Having this result established, we can continue on equation (3). To this end, set $\alpha := 4$, and $c < \frac{9}{2\pi^4}$, like $c := 0.045$. For $\tau > (2c)^{\frac{1}{\alpha}}$ we can now use the upper bound, which translates to all $\tau \in \mathbb{N}_{>0}$. Hence:

$$\begin{aligned}\sum_{j \neq i} |\rho_{|i-j|}| &\leq 2 \sum_{\tau=1}^{\infty} \rho_\tau \\ &\leq 2 \sum_{\tau=1}^{\infty} \frac{\sqrt{2 \cdot 0.045}}{\tau^2} \\ &= 2 \cdot 0.3 \sum_{\tau=1}^{\infty} \frac{1}{\tau^2} \\ &= 2 \cdot 0.3 \cdot \frac{\pi^2}{6} \\ &= \frac{\pi^2}{10} \\ &< 1 \quad ,\end{aligned}$$

which proves equation (2).

1.1.4 Strong Power-Law Behavior

Now that we have defined our model and shown that it is a valid probability distribution, we can finally prove that it has the desired property of strong power-law behavior according to definition ??.

We motivated our model with the continuous normal distribution like presented in section 1.1.2. However, our model S_{n,Σ_n} is a discretized version of it (defined in definition 1.1).

Another way to think about this is that the continuous normal distribution has an associated vector $(Y_1, \dots, Y_n) \in \mathbb{R}^n$ of random variables, and in order to get the random variables $(X_1, \dots, X_n) \in \{-1, 1\}^n$ of our model, we *process* each Y_i in the following manner:

$$X_i = \begin{cases} 1 & \text{if } Y_i \geq 0 \\ -1 & \text{else} \end{cases} \quad .$$

We can make use of the data processing inequality (see theorem ??), which also works in the continuous case, to derive that

$$I(X_i; X_j) \leq I(X_i; Y_j) \leq I(Y_i; Y_j) = c \cdot |i - j|^{-\alpha} \quad ,$$

i.e. the model has strong upper bound power-law behavior.

What is left to do is to prove strong lower bound power-law behavior. In order to calculate $I(X_i; X_j)$, we need the joint distribution of these two variables when marginalizing over the other ones.

Luckily, we can use proposition 1.1 again, and we can focus on two dimensional normal distributions of the form

$$\begin{pmatrix} Y_i \\ Y_j \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho_\tau \\ \rho_\tau & 1 \end{pmatrix}\right) \quad .$$

Of course, we get the probabilities for the joint distribution of (X_i, X_j) by integrating over the quadrants. But to do this, we first prove an auxiliary lemma:

Lemma 1.2. *Let X and Y have a bivariate normal distribution where X and Y are standard normal variables, $X, Y \sim \mathcal{N}(0, 1)$, with correlation ρ . The variable $Z = \frac{Y - \rho X}{\sqrt{1 - \rho^2}}$ is a standard normal variable, and X and Z are independent.*

Proof. First, we show that Z is a standard normal variable. Since Z is a linear combination of the jointly normal variables X and Y , Z is also a normal variable. We compute its mean and variance.

The mean of Z is:

$$E[Z] = E\left[\frac{Y - \rho X}{\sqrt{1 - \rho^2}}\right] = \frac{E[Y] - \rho E[X]}{\sqrt{1 - \rho^2}} = \frac{0 - \rho \cdot 0}{\sqrt{1 - \rho^2}} = 0 \quad .$$

The variance of Z is:

$$\begin{aligned} \text{Var}(Z) &= \text{Var}\left(\frac{Y - \rho X}{\sqrt{1 - \rho^2}}\right) = \frac{1}{1 - \rho^2} \text{Var}(Y - \rho X) \\ &= \frac{1}{1 - \rho^2} (\text{Var}(Y) + \rho^2 \text{Var}(X) - 2\rho \text{Cov}(X, Y)) \quad . \end{aligned}$$

Since X and Y are standard normal variables, $\text{Var}(X) = 1$, $\text{Var}(Y) = 1$, and their covariance $\text{Cov}(X, Y)$ is equal to their correlation ρ . Hence:

$$\text{Var}(Z) = \frac{1}{1 - \rho^2} (1 + \rho^2(1) - 2\rho(\rho)) = \frac{1 - \rho^2}{1 - \rho^2} = 1 \quad .$$

Thus, Z is a standard normal variable, $Z \sim \mathcal{N}(0, 1)$.

To show that X and Z are independent, we compute their covariance. Since they are jointly normal, zero covariance implies independence.

$$\begin{aligned} \text{Cov}(X, Z) &= \text{Cov}\left(X, \frac{Y - \rho X}{\sqrt{1 - \rho^2}}\right) = \frac{1}{\sqrt{1 - \rho^2}} \text{Cov}(X, Y - \rho X) \\ &= \frac{1}{\sqrt{1 - \rho^2}} (\text{Cov}(X, Y) - \rho \text{Cov}(X, X)) \\ &= \frac{1}{\sqrt{1 - \rho^2}} (\rho - \rho \text{Var}(X)) = \frac{1}{\sqrt{1 - \rho^2}} (\rho - \rho \cdot 1) = 0 \quad . \end{aligned}$$

Since $\text{Cov}(X, Z) = 0$ and they are jointly normal, X and Z are independent. \square

Now we can prove our proposition of interest:

Proposition 1.3. *For bivariate standard normal variables X and Y with correlation ρ ,*

$$P(X > 0, Y > 0) = \frac{1}{4} + \frac{1}{2\pi} \arcsin(\rho) \quad .$$

Proof. Define the random variable Z like in the previous lemma. Then, the event $\{X > 0, Y > 0\}$ is the same as the event $\{X > 0, Z > \frac{-\rho}{\sqrt{1 - \rho^2}} X\}$, where X and Z are independent standard normal variables as shown above. Writing $a := \frac{-\rho}{\sqrt{1 - \rho^2}}$ for brevity, the desired probability is expressible as a double integral involving the joint density of (X, Z) :

$$\begin{aligned} P(X > 0, Y > 0) &= P(X > 0, Z > aX) \\ &= \int_{x=0}^{\infty} \int_{z=ax}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz dx \quad . \end{aligned}$$

Switching to polar coordinates ($x = r \cos \theta, z = r \sin \theta$), the integral becomes:

$$\int_{\theta=\arctan(a)}^{\pi/2} \int_{r=0}^{\infty} \frac{1}{2\pi} e^{-r^2/2} r dr d\theta = \int_{\theta=\arctan(a)}^{\pi/2} \frac{1}{2\pi} \left[-e^{-r^2/2} \right]_0^{\infty} d\theta \quad .$$

This equals:

$$\int_{\theta=\arctan(a)}^{\pi/2} \frac{1}{2\pi} d\theta = \frac{1}{2\pi} \left(\frac{\pi}{2} - \arctan(a) \right) = \frac{1}{4} - \frac{1}{2\pi} \arctan\left(\frac{-\rho}{\sqrt{1 - \rho^2}}\right) \quad .$$

Using the fact that the arctan function is odd, i.e. $\arctan(-u) = -\arctan(u)$, we get:

$$\frac{1}{4} + \frac{1}{2\pi} \arctan\left(\frac{\rho}{\sqrt{1 - \rho^2}}\right) \quad .$$

To finish, we use the identity $\arcsin(\rho) = \arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)$. To see this, let $\phi = \arcsin(\rho)$ for $\phi \in [-\pi/2, \pi/2]$. Then $\sin(\phi) = \rho$ and $\cos(\phi) = \sqrt{1-\rho^2}$. Thus, $\tan(\phi) = \frac{\sin(\phi)}{\cos(\phi)} = \frac{\rho}{\sqrt{1-\rho^2}}$, which implies $\phi = \arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)$. Substituting this into our expression gives the final result:

$$P(X > 0, Y > 0) = \frac{1}{4} + \frac{1}{2\pi} \arcsin(\rho) \quad .$$

□

Now, set $\delta_\tau := \frac{\arcsin(\rho_\tau)}{2\pi}$. Note that since $\rho_\tau \in (0, 1)$, we have $\delta_\tau \in (0, \frac{1}{4})$. Based on the previous result, and using the symmetry of the normal distribution, we derive the joint probability distribution of (X_i, X_j) :

	$X_i = 1$	$X_i = -1$
$X_j = 1$	$\frac{1}{4} + \delta_\tau$	$\frac{1}{4} - \delta_\tau$
$X_j = -1$	$\frac{1}{4} - \delta_\tau$	$\frac{1}{4} + \delta_\tau$

Let's finally compute the mutual information $I(X_i; X_j)$. First, note that the marginalized distribution of a single X_i is just the uniform distribution on $\{-1, 1\}$. Furthermore, let's substitute $f : (0, \frac{1}{2}) \rightarrow \mathbb{R}$, $x \mapsto x \log x$, where \log is the natural logarithm. Thus:

$$\begin{aligned} I(X_i; X_j) &= H(X_i) + H(X_j) - H(X_i, X_j) \\ &= \log 2 + \log 2 - \left[- \sum_{(x_i, x_j) \in \{-1, 1\}^2} p(x_i, x_j) \log p(x_i, x_j) \right] \\ &= \log 4 + 2 \cdot f\left(\frac{1}{4} - \delta_\tau\right) + 2 \cdot f\left(\frac{1}{4} + \delta_\tau\right) \quad . \end{aligned}$$

Using that $\log 4 = -\log \frac{1}{4} = -4 \cdot f(\frac{1}{4})$, we arrive at

$$\begin{aligned} I(X_i; X_j) &= 2 \cdot f\left(\frac{1}{4} - \delta_\tau\right) + 2 \cdot f\left(\frac{1}{4} + \delta_\tau\right) - 4 \cdot f\left(\frac{1}{4}\right) \\ &= 4 \cdot \left[\frac{1}{2} \cdot f\left(\frac{1}{4} - \delta_\tau\right) + \frac{1}{2} \cdot f\left(\frac{1}{4} + \delta_\tau\right) - f\left(\frac{1}{4}\right) \right] \quad . \end{aligned} \quad (5)$$

This expression looks like it measures how *convex* the function f is at the point $\frac{1}{4}$. Hence, we might guess that we can lower bound the expression by substituting a less convex function g for f . This idea is formalized in the following lemma:

Lemma 1.3. *Let $I \subseteq \mathbb{R}$ be an interval, and let $f, g \in \mathcal{C}^2(I, \mathbb{R})$ be two functions that are twice differentiable s.t. $f''(x) \geq g''(x)$ for all $x \in I$. Furthermore, let $x_0 \in I$ and $\epsilon \in \mathbb{R}_{>0}$ s.t. $U_\epsilon(x_0) \subseteq I$. Then, it follows that for all $\delta \in (-\epsilon, \epsilon)$ we have:*

$$\frac{1}{2} \cdot f(x_0 - \delta) + \frac{1}{2} \cdot f(x_0 + \delta) - f(x_0) \geq \frac{1}{2} \cdot g(x_0 - \delta) + \frac{1}{2} \cdot g(x_0 + \delta) - g(x_0) \quad .$$

Proof. Let's define a new function $h : I \rightarrow \mathbb{R}$ as the difference between f and g :

$$h(x) := f(x) - g(x) \quad .$$

Since both $f, g \in \mathcal{C}^2(I, \mathbb{R})$, their difference h is also twice continuously differentiable on I . We compute the second derivative of $h(x)$:

$$h''(x) = \frac{d^2}{dx^2}(f(x) - g(x)) = f''(x) - g''(x) \quad .$$

From the lemma's premise, we know that $f''(x) \geq g''(x)$ for all $x \in I$. Therefore, it follows that

$$h''(x) \geq 0 \quad \forall x \in I \quad ,$$

which means that $h(x)$ is a convex function on I .

By the definition of convexity, for any points $a, b \in I$, we have:

$$\frac{h(a) + h(b)}{2} \geq h\left(\frac{a + b}{2}\right) \quad .$$

Let us choose our points $a = x_0 - \delta$ and $b = x_0 + \delta$. Since $\delta \in (-\epsilon, \epsilon)$ and the neighborhood $U_\epsilon(x_0) \subseteq I$, both a and b are in the interval I . Their midpoint is $\frac{(x_0 - \delta) + (x_0 + \delta)}{2} = x_0$.

Applying the convexity inequality to h with these specific points gives:

$$\frac{1}{2}h(x_0 - \delta) + \frac{1}{2}h(x_0 + \delta) \geq h(x_0) \quad .$$

Now, we substitute the definition of $h(x) = f(x) - g(x)$ back into this inequality:

$$\frac{1}{2}[f(x_0 - \delta) - g(x_0 - \delta)] + \frac{1}{2}[f(x_0 + \delta) - g(x_0 + \delta)] \geq f(x_0) - g(x_0) \quad .$$

Rearranging the terms to separate the functions f and g , we arrive at the desired result:

$$\frac{1}{2}f(x_0 - \delta) + \frac{1}{2}f(x_0 + \delta) - f(x_0) \geq \frac{1}{2}g(x_0 - \delta) + \frac{1}{2}g(x_0 + \delta) - g(x_0) \quad .$$

□

Remember, we defined $f : (0, \frac{1}{2}) \rightarrow \mathbb{R}$, $x \mapsto x \log x$. Hence:

$$\begin{aligned} f'(x) &= \log x + 1 \\ f''(x) &= \frac{1}{x} \quad . \end{aligned}$$

Note that $f''(x) \geq 2$ for all $x \in (0, \frac{1}{2})$. Thus, when we define $g : (0, \frac{1}{2}) \rightarrow \mathbb{R}$, $x \mapsto x^2$, where $g''(x) \equiv 2$, we can use lemma 1.3 in equation (5):

$$\begin{aligned} I(X_i; X_j) &= 4 \cdot \left[\frac{1}{2} \cdot f\left(\frac{1}{4} - \delta_\tau\right) + \frac{1}{2} \cdot f\left(\frac{1}{4} + \delta_\tau\right) - f\left(\frac{1}{4}\right) \right] \\ &\geq 4 \cdot \left[\frac{1}{2} \cdot g\left(\frac{1}{4} - \delta_\tau\right) + \frac{1}{2} \cdot g\left(\frac{1}{4} + \delta_\tau\right) - g\left(\frac{1}{4}\right) \right] \\ &= 4 \cdot \left[\frac{1}{2} \left(\frac{1}{4} - \delta_\tau\right)^2 + \frac{1}{2} \left(\frac{1}{4} + \delta_\tau\right)^2 - \left(\frac{1}{4}\right)^2 \right] \\ &= 2 \left[\left(\frac{1}{16} - \frac{1}{2}\delta_\tau + \delta_\tau^2\right) + \left(\frac{1}{16} + \frac{1}{2}\delta_\tau + \delta_\tau^2\right) \right] - 4 \left(\frac{1}{16}\right) \\ &= 2 \left[\frac{2}{16} + 2\delta_\tau^2 \right] - \frac{4}{16} \\ &= 2 \left[\frac{1}{8} + 2\delta_\tau^2 \right] - \frac{1}{4} \\ &= \frac{1}{4} + 4\delta_\tau^2 - \frac{1}{4} \\ &= 4\delta_\tau^2 \quad . \end{aligned}$$

Since $\delta_\tau = \frac{\arcsin(\rho_\tau)}{2\pi}$, we conclude

$$\begin{aligned} I(X_i; X_j) &\geq 4 \frac{\arcsin(\rho_\tau)^2}{4\pi^2} \\ &\geq \frac{\rho_\tau^2}{\pi^2} \quad , \end{aligned}$$

since $\arcsin(x) \geq x$ for $x \in (0, 1)$.

Finally, we use the lower bound of ρ_τ provided by lemma 1.1, to arrive at

$$I(X_i; X_j) \geq \frac{2c}{(2c+1) \cdot \pi^2} \tau^{-\alpha} \quad .$$

1.2 Summary

Let's summarize our findings in a concise theorem:

Theorem 1.1 (A Model with Strong Power-Law Behavior). *Define $\alpha := 4$, and $c := 0.045$. Furthermore, let $\rho_\tau := \sqrt{1 - e^{-2c\tau^{-\alpha}}}$. We define the matrix*

$$\Sigma_n = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{n-2} \\ \vdots & & \ddots & & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & 1 \end{pmatrix}.$$

We use Σ_n as the parameter covariance matrix of the model defined in definition 1.1. It follows that S_{n, Σ_n} is a valid model over $\{-1, 1\}$, since Σ_n is positive definite especially. Furthermore, S has strong power-law behavior according to definition ???. Specifically, for any random variables X_i, X_j sampled from S , we have:

$$\frac{2c}{(2c+1) \cdot \pi^2} |i-j|^{-\alpha} \leq I(X_i; X_j) \leq |i-j|^{-\alpha}.$$

Proof. The proof directly follows from our preliminary considerations. □