1 Technical Details

In this appendix we provide some further details to our arguments.

1.1 Integrating over the Quadrants of a Normal Distribution

In order to derive the formula, we first have to prove an auxiliary lemma:

Lemma 1.1.1. *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:

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

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

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

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.

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

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

Now we can prove our proposition of interest:

Proposition 1.1.1. *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:

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

□

1.2 Prerequisites for Theorem ??

Lemma 1.2.1. *Let*

$$\boldsymbol{M} := \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{C} \end{bmatrix}$$

be a matrix consisting of submatrices $\boldsymbol{A} \in \mathbb{R}^{k\times k}$, $\boldsymbol{B} \in \mathbb{R}^{k\times \ell}$, and $\boldsymbol{C} \in \mathbb{R}^{\ell\times \ell}$. Let $\boldsymbol{A}$ be an irreducible aperiodic Markov transition matrix, and let $\boldsymbol{C}^n \xrightarrow{n\to\infty} \boldsymbol{0}$ with exponential decay. Then, $\boldsymbol{M}^n$ decays exponentially in n towards a matrix $\boldsymbol{M}'$.

Proof. The proof proceeds in three steps. First, we establish a closed-form expression for $\boldsymbol{M}^n$. Second, we determine the limit matrix $\boldsymbol{M}'$. Third, we prove that the convergence to this limit is exponential.

1. The Form of $\boldsymbol{M}^n$ Since $\boldsymbol{M}$ is a block upper triangular matrix, its powers take a specific form. By induction, we can show that:

$$\boldsymbol{M}^n = \begin{bmatrix} \boldsymbol{A}^n & \boldsymbol{X}_n \\ \boldsymbol{0} & \boldsymbol{C}^n \end{bmatrix} \quad \text{where} \quad \boldsymbol{X}_n = \sum_{j=0}^{n-1} \boldsymbol{A}^{n-1-j}\boldsymbol{B}\boldsymbol{C}^j \quad .$$

2. The Limit Matrix $\boldsymbol{M}'$ We analyze the limit of each block of $\boldsymbol{M}^n$ as $n \to \infty$.

- **Block $\boldsymbol{A}^n$:** Since $\boldsymbol{A}$ is an irreducible aperiodic Markov transition matrix, the Perron-Frobenius theorem for stochastic matrices guarantees that $\boldsymbol{A}^n$ converges to a rank-one matrix $\boldsymbol{A}' = \boldsymbol{\pi}\boldsymbol{1}^T$, where $\boldsymbol{\pi}$ is the unique stationary distribution. The convergence is exponential, so there exist constants $K_A > 0$ and $0 \leq \lambda < 1$ such that $\|\boldsymbol{A}^n - \boldsymbol{A}'\| \leq K_A\lambda^n$.

- **Block $\boldsymbol{C}^n$:** By hypothesis, $\boldsymbol{C}^n \to \boldsymbol{0}$ with exponential decay. This is equivalent to its spectral radius being less than one, $\rho(\boldsymbol{C}) < 1$. Thus, there exist constants $K_C > 0$ and $0 \leq \gamma < 1$ such that $\|\boldsymbol{C}^n\| \leq K_C\gamma^n$.

- **Block $\boldsymbol{X}_n$:** Let $\boldsymbol{E}_k = \boldsymbol{A}^k - \boldsymbol{A}'$. We have $\|\boldsymbol{E}_k\| \leq K_A \lambda^k$. We can rewrite $\boldsymbol{X}_n$ as:

$$\boldsymbol{X}_n = \sum_{j=0}^{n-1} (\boldsymbol{A}' + \boldsymbol{E}_{n-1-j}) \boldsymbol{B} \boldsymbol{C}^j = \boldsymbol{A}'\boldsymbol{B} \left(\sum_{j=0}^{n-1} \boldsymbol{C}^j \right) + \sum_{j=0}^{n-1} \boldsymbol{E}_{n-1-j} \boldsymbol{B} \boldsymbol{C}^j \quad .$$

  As $n \to \infty$, the first term converges to $\boldsymbol{A}'\boldsymbol{B}(\boldsymbol{I} - \boldsymbol{C})^{-1}$, since the series $\sum_{j=0}^{\infty} \boldsymbol{C}^j$ converges to $(\boldsymbol{I} - \boldsymbol{C})^{-1}$. The second term converges to $\boldsymbol{0}$ because its norm is bounded by a vanishing convolution sum. Thus, the limit of $\boldsymbol{X}_n$ is $\boldsymbol{X}' = \boldsymbol{A}'\boldsymbol{B}(\boldsymbol{I} - \boldsymbol{C})^{-1}$.

Combining these limits, the limit matrix is $\boldsymbol{M}' = \left[\begin{smallmatrix} \boldsymbol{A}' & \boldsymbol{X}' \\ \boldsymbol{0} & \boldsymbol{0} \end{smallmatrix}\right]$.

3. Exponential Rate of Convergence We analyze the norm of the difference matrix $\boldsymbol{M}^n - \boldsymbol{M}'$:

$$\boldsymbol{M}^n - \boldsymbol{M}' = \begin{bmatrix} \boldsymbol{A}^n - \boldsymbol{A}' & \boldsymbol{X}_n - \boldsymbol{X}' \\ \boldsymbol{0} & \boldsymbol{C}^n \end{bmatrix} \quad .$$

The blocks $\|\boldsymbol{A}^n - \boldsymbol{A}'\|$ and $\|\boldsymbol{C}^n\|$ decay exponentially by definition. We examine the convergence of the off-diagonal block:

$$\boldsymbol{X}_n - \boldsymbol{X}' = -\boldsymbol{A}'\boldsymbol{B} \left(\sum_{j=n}^{\infty} \boldsymbol{C}^j \right) + \sum_{j=0}^{n-1} \boldsymbol{E}_{n-1-j} \boldsymbol{B} \boldsymbol{C}^j \quad .$$

The norm of each part is bounded by an exponentially decaying function:

- $\left\| -\boldsymbol{A}'\boldsymbol{B} \left(\sum_{j=n}^{\infty} \boldsymbol{C}^j \right) \right\| \leq \|\boldsymbol{A}'\| \|\boldsymbol{B}\| \sum_{j=n}^{\infty} \|\boldsymbol{C}^j\| \leq \|\boldsymbol{A}'\| \|\boldsymbol{B}\| K_C \frac{\gamma^n}{1-\gamma}$. This decays with rate γ.
- $\left\| \sum_{j=0}^{n-1} \boldsymbol{E}_{n-1-j} \boldsymbol{B} \boldsymbol{C}^j \right\| \leq \sum_{j=0}^{n-1} K_A \lambda^{n-1-j} \|\boldsymbol{B}\| K_C \gamma^j = K_A \|\boldsymbol{B}\| K_C \sum_{j=0}^{n-1} \lambda^{n-1-j} \gamma^j$. This convolution sum is bounded by $K' n \mu^n$ where $\mu = \max(\lambda, \gamma)$, which decays exponentially.

Since $\|\boldsymbol{X}_n - \boldsymbol{X}'\|$ is bounded by a sum of exponentially decaying terms, it also decays exponentially. As all blocks of $\boldsymbol{M}^n - \boldsymbol{M}'$ converge to zero exponentially, the norm $\|\boldsymbol{M}^n - \boldsymbol{M}'\|$ does as well. This completes the proof. □

1.2.1 Convergence of Mutual Information

Theorem 1.2.1 (Element-Wise Exponential Convergence Implies Exponential Convergence)**.** *Let $f : \mathcal{D} \to \mathbb{R}^m$ be a function defined on a convex domain $\mathcal{D} \subseteq \mathbb{R}^n$ that is a Cartesian product of real intervals, i.e., $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \times \cdots \times \mathcal{D}_n$ where each $\mathcal{D}_i \subseteq \mathbb{R}$ is an interval. Let $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset \mathcal{D}$ be a sequence converging exponentially fast to $\boldsymbol{x}_0 \in \mathcal{D}$.*

Let $\boldsymbol{e}_j$ denote the j-th standard basis vector in $\mathbb{R}^n$. Suppose that for each input coordinate $j \in \{1, 2, \dots, n\}$ there exist functions $K_j(C', \rho)$, $C_j(C', \rho)$, $P_j(C', \rho)$ s.t. for every sequence $\{\boldsymbol{u}_k\}_{k=1}^{\infty} \subset \mathcal{D}$ converging to $\boldsymbol{u}_0$ where the difference $\boldsymbol{u}_\ell - \boldsymbol{u}_{\ell'}$ is parallel to $\boldsymbol{e}_j$ (i.e., they only differ in the j-th coordinate) that satisfies $|\boldsymbol{u}_0 - \boldsymbol{u}_k| \le C'\rho^k$ for all k and some $\rho \in [0, 1)$, $C' > 0$, we have for all $k \ge K_j(C', \rho)$:

$$\|f(\boldsymbol{u}_0) - f(\boldsymbol{u}_k)\| \le C_j(C', \rho)\rho^k k^{P_j(C', \rho)} \quad .$$

Then, there exist constants $C > 0$ and $\sigma \in [0, 1)$ such that for all sufficiently large k:

$$\|f(\boldsymbol{x}_0) - f(\boldsymbol{x}_k)\| \le C\sigma^k \quad .$$

Proof. Let the sequence $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset \mathcal{D}$ converge exponentially to $\boldsymbol{x}_0 \in \mathcal{D}$. By definition, there exist constants $C_x > 0$ and $\rho \in [0, 1)$ such that for all k,

$$\|\boldsymbol{x}_k - \boldsymbol{x}_0\| \le C_x \rho^k \quad .$$

Let $\boldsymbol{x}_k = (x_{k,1}, \dots, x_{k,n})^T$ and $\boldsymbol{x}_0 = (x_{0,1}, \dots, x_{0,n})^T$. An immediate consequence is that each coordinate also converges exponentially, i.e., for each $j \in \{1, \dots, n\}$:

$$|x_{k,j} - x_{0,j}| \le \|\boldsymbol{x}_k - \boldsymbol{x}_0\|_\infty \le \|\boldsymbol{x}_k - \boldsymbol{x}_0\| \le C_x \rho^k \quad ,$$

where we use the equivalence of norms in $\mathbb{R}^n$.

To bound $\|f(\boldsymbol{x}_0) - f(\boldsymbol{x}_k)\|$, we define a sequence of $n+1$ intermediate points that form a path from $\boldsymbol{x}_k$ to $\boldsymbol{x}_0$ by changing one coordinate at a time. For each k, let:

$$\begin{aligned}
\boldsymbol{z}_{k,0} &:= \boldsymbol{x}_k = (x_{k,1}, x_{k,2}, \dots, x_{k,n}) \\
\boldsymbol{z}_{k,1} &:= (x_{0,1}, x_{k,2}, \dots, x_{k,n}) \\
&\vdots \\
\boldsymbol{z}_{k,j} &:= (x_{0,1}, \dots, x_{0,j}, x_{k,j+1}, \dots, x_{k,n}) \\
&\vdots \\
\boldsymbol{z}_{k,n} &:= (x_{0,1}, \dots, x_{0,n}) = \boldsymbol{x}_0 \quad .
\end{aligned}$$

Since $\mathcal{D}$ is a cartesian product intervals and both $\boldsymbol{x}_k$ and $\boldsymbol{x}_0$ are in $\mathcal{D}$, all intermediate points $\boldsymbol{z}_{k,j}$ are also contained in $\mathcal{D}$. We can express the total difference $f(\boldsymbol{x}_0) - f(\boldsymbol{x}_k)$ as a telescoping sum:

$$f(\boldsymbol{x}_0) - f(\boldsymbol{x}_k) = f(\boldsymbol{z}_{k,n}) - f(\boldsymbol{z}_{k,0}) = \sum_{j=1}^{n} \left(f(\boldsymbol{z}_{k,j}) - f(\boldsymbol{z}_{k,j-1})\right) \quad .$$

By the triangle inequality, we have:

$$\|f(\boldsymbol{x}_0) - f(\boldsymbol{x}_k)\| \le \sum_{j=1}^{n} \|f(\boldsymbol{z}_{k,j}) - f(\boldsymbol{z}_{k,j-1})\| \quad .$$

Now, we analyze each term $\|f(\boldsymbol{z}_{k,j}) - f(\boldsymbol{z}_{k,j-1})\|$ for a fixed $j \in \{1, \ldots, n\}$. The points $\boldsymbol{z}_{k,j}$ and $\boldsymbol{z}_{k,j-1}$ differ only in their j-th coordinate.

Let us define a sequence $\{\boldsymbol{u}_m\}_{m=1}^{\infty}$ and a limit point $\boldsymbol{u}_0$ that fit the condition in the theorem's hypothesis. For the given j and k, let

$$\begin{aligned} \boldsymbol{u}_m &:= (x_{0,1}, \ldots, x_{0,j-1}, x_{m,j}, x_{k,j+1}, \ldots, x_{k,n}) \\ \boldsymbol{u}_0 &:= (x_{0,1}, \ldots, x_{0,j-1}, x_{0,j}, x_{k,j+1}, \ldots, x_{k,n}) \quad . \end{aligned}$$

Note that $\boldsymbol{u}_0 = \boldsymbol{z}_{k,j}$ and by setting $m = k$, we get $\boldsymbol{u}_k = \boldsymbol{z}_{k,j-1}$. The sequence $\{\boldsymbol{u}_m\}$ lies on a line parallel to the j-th coordinate axis. As $m \to \infty$, $\boldsymbol{u}_m \to \boldsymbol{u}_0$ because $x_{m,j} \to x_{0,j}$. The convergence is exponential:

$$\|\boldsymbol{u}_m - \boldsymbol{u}_0\| = |x_{m,j} - x_{0,j}| \leq C_x \rho^m \quad .$$

The hypothesis states that for any such sequence, there exist constants $K_j(C_x, \rho)$, $C_j(C_x, \rho)$, $P_j(C_x, \rho)$ which are independent of the specific line, such that for all $k \geq K_j(C_x, \rho)$ we have $\|f(\boldsymbol{u}_0) - f(\boldsymbol{u}_m)\| \leq C_j(C_x, \rho)\rho^m m^{P_j(C_x,\rho)}$. Applying this for $m = k$:

$$\|f(\boldsymbol{z}_{k,j}) - f(\boldsymbol{z}_{k,j-1})\| = \|f(\boldsymbol{u}_0) - f(\boldsymbol{u}_k)\| \leq C_j(C_x, \rho)\rho^k k^{P_j(C_x,\rho)} \quad .$$

This inequality holds for each $j = 1, \ldots, n$. Substituting these bounds back into the sum:

$$\|f(\boldsymbol{x}_0) - f(\boldsymbol{x}_k)\| \leq \sum_{j=1}^{n} C_j(C_x, \rho)\rho^k k^{P_j(C_x,\rho)} \quad .$$

Let $K := \max_{j\in\{1,\ldots,n\}}\{K_j(C_x, \rho)\}$, $C := \sum_{j=1}^{n} C_j(C_x, \rho)$ and $P := \max_{j\in\{1,\ldots,n\}}\{P_j(C_x, \rho)\}$. Hence, for all $k \geq K$:

$$\|f(\boldsymbol{x}_0) - f(\boldsymbol{x}_k)\| \leq \sum_{j=1}^{n} C_j(C_x, \rho)\rho^k k^P = \left(\sum_{j=1}^{n} C_j(C_x, \rho)\right)\rho^k k^P = C\rho^k k^P \quad .$$

This shows that $\{f(\boldsymbol{x}_k)\}$ converges exponentially to $f(\boldsymbol{x}_0)$ with a rate of σ s.t. $\rho < \sigma < 1$ (like $\sigma := \sqrt{\rho}$). This completes the proof. $\square$

Lemma 1.2.2. *Let the function $f : [0, \infty) \to \mathbb{R}$ be defined as $f(x) = x \log x$ for $x \in (0, 1)$, and $f(x) = 0$ everywhere else. If a sequence $\{x_k\}_{k=1}^{\infty} \subset [0, 1]$ converging to a limit $x_\infty \in [0, 1]$ satisfies $|x_k - x_\infty| \leq C\rho^k$ for some $C \in \mathbb{R}_{>0}$, $\rho \in [0, 1)$, then the sequence $\{f(x_k)\}$ converges to $f(x_\infty)$ with $|f(x_k) - f(x_\infty)| \leq C\rho^k \; |\log C + k \log \rho|$ for $k \geq \log_\rho \left(\dfrac{e^{-1}}{C}\right) =: K$.*

Proof. We consider two cases for the limit x_∞:

Case 1: $x_\infty = 0$
In this case, $|x_k - 0| = x_k \leq C\rho^k$. We want to bound the difference $|f(x_k) - f(0)| =$

$|f(x_k)|$. Note that $|f(x)|$ is monotonically increasing on $[0, e^{-1}]$. For $k \geq K$ we have $x_k \leq C\rho^k \leq e^{-1}$. Thus, for all $k \geq K$:

$$|f(x_k)| \leq |f(C\rho^k)| = |C\rho^k \log(C\rho^k)| = C\rho^k \, |\log C + k \log \rho| \quad .$$

Case 2: $x_\infty > 0$
Similarly, for $k \geq K$ we have $|x_k - x_\infty| \leq C\rho^k \leq e^{-1}$. Based on the function graph, it follows that for $k \geq K$ we have:

$$|f(x_k) - f(x_\infty)| \leq |f(C\rho^k) - f(0)| = |f(C\rho^k)| = C\rho^k \, |\log C + k \log \rho| \quad .$$

□

Theorem 1.2.2 (Element-Wise Exponential Convergence Property of Mutual Information)**.** *Let the function $f : [0, \infty) \to \mathbb{R}$ be defined as $f(x) = x \log x$ for $x \in (0, 1)$, and $f(x) = 0$ everywhere else. Define the function $I : [0, 1]^{m \times n} \to \mathbb{R}$ for a matrix $\boldsymbol{M} \in [0, 1]^{m \times n}$ as:*

$$I(\boldsymbol{M}) = \sum_{i=1}^{m} \sum_{j=1}^{n} f(M_{ij}) - \sum_{i=1}^{m} f\left(\sum_{j=1}^{n} M_{ij}\right) - \sum_{j=1}^{n} f\left(\sum_{i=1}^{m} M_{ij}\right) \quad .$$

This function exhibits element-wise exponential convergence. That is, for any single component (i_0, j_0), if a sequence of matrices $\{\boldsymbol{U}_k\}_{k=1}^{\infty} \subset [0, 1]^{m \times n}$ converges to a limit $\boldsymbol{U}_\infty$, varies only in the (i_0, j_0)-th component and satisfies $\|\boldsymbol{U}_k - \boldsymbol{U}_\infty\| \leq C\rho^k$ for some $C > 0$, $\rho \in [0, 1)$ and all k, then the sequence of values $\{I(\boldsymbol{U}_k)\}$ converges to $I(\boldsymbol{U}_\infty)$ with $|I(\boldsymbol{U}_k) - I(\boldsymbol{U}_\infty)| \leq C'(C, \rho)\rho^k k^{P(C,\rho)}$ for all $k \geq K(C, \rho)$.

Proof. The function $I(\boldsymbol{M})$ is a sum of terms involving f applied to the matrix entries and their row and column sums. Let $m_i(\boldsymbol{M}) = \sum_j M_{ij}$ and $m'_j(\boldsymbol{M}) = \sum_i M_{ij}$. We have:

$$I(\boldsymbol{M}) = \sum_{i,j} f(M_{ij}) - \sum_i f(m_i(\boldsymbol{M})) - \sum_j f(m'_j(\boldsymbol{M})) \quad .$$

We are given a sequence $\{\boldsymbol{U}_k\}$ that varies only in the (i_0, j_0)-th component, $u_k = \boldsymbol{U}_k(i_0, j_0)$. All other components are constant. The exponential convergence of $\{\boldsymbol{U}_k\}$ means $|u_k - u_\infty| \leq C\rho^k$.

The difference $I(\boldsymbol{U}_k) - I(\boldsymbol{U}_\infty)$ consists only of terms whose arguments change with k. These are:

1. The entry term: $f(u_k)$.
2. The row-sum term: $f(m_{i_0}(\boldsymbol{U}_k))$, where $m_{i_0}(\boldsymbol{U}_k) = u_k + \text{const}$.
3. The column-sum term: $f(m'_{j_0}(\boldsymbol{U}_k))$, where $m'_{j_0}(\boldsymbol{U}_k) = u_k + \text{const}$.

By the triangle inequality, the total error is bounded by the sum of the absolute errors of these three terms:

$$\begin{aligned} |I(\boldsymbol{U}_k) - I(\boldsymbol{U}_\infty)| \leq\ & |f(u_k) - f(u_\infty)| \\ & + |f(m_{i_0}(\boldsymbol{U}_k)) - f(m_{i_0}(\boldsymbol{U}_\infty))| \\ & + |f(m'_{j_0}(\boldsymbol{U}_k)) - f(m'_{j_0}(\boldsymbol{U}_\infty))| \quad . \end{aligned}$$

The arguments to the function f in each of these three terms converge exponentially to their limits with rate ρ and constant C, since $|m_{i_0}(\boldsymbol{U}_k) - m_{i_0}(\boldsymbol{U}_\infty)| = |u_k - u_\infty|$ and $|m'_{j_0}(\boldsymbol{U}_k) - m'_{j_0}(\boldsymbol{U}_\infty)| = |u_k - u_\infty|$.

Hence, by lemma **??**, we have:

$$\begin{aligned} |I(\boldsymbol{U}_k) - I(\boldsymbol{U}_\infty)| &\leq 3C\rho^k \ |\log C + k \log \rho| \\ &\leq 3C\rho^k k \ (|\log C| + |\log \rho|) \\ &= C'\rho^k k \quad , \end{aligned}$$

with $C' := 3C(|\log C| + |\log \rho|)$ and for all $k \geq K(C, \rho)$. Note that K, C' and P only depend on C and ρ. □

Corollary 1.2.1. *Using theorem* **??**, *we see that if a sequence $\{\boldsymbol{P}_k\}$ of joint probability distributions converges exponentially fast, then $\{I(\boldsymbol{P}_k)\}$ converges exponentially fast as well.*

Corollary 1.2.2. *A joint probability matrix $\boldsymbol{P}_{X,Y}$ can be calculated from the conditional probability matrix $\boldsymbol{P}_{Y|X}$ and the diagonal matrix $\boldsymbol{P}_X$ with the probabilities for X on its diagonal using $\boldsymbol{P}_{X,Y} = \boldsymbol{P}_{Y|X}\boldsymbol{P}_X$. Hence, if $\boldsymbol{P}_{Y|X}$ converges exponentially fast while $\boldsymbol{P}_X$ stays constant, the mutual information $I(X;Y)$ will converge exponentially fast as well.*