

# 1 Technical Details

In this appendix we provide additional details to our arguments.

## 1.1 Quadrant Probabilities for the Bivariate Normal Distribution

In order to derive a formula for the quadrant probabilities of  $\mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ , 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  $\rho$ . 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  $\rho$ . 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 1.1.1.** *For bivariate standard normal variables  $X$  and  $Y$  with correlation  $\rho$ ,*

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

*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*

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

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

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

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

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

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

- **Block  $\mathbf{A}^n$ :** Since  $\mathbf{A}$  is an irreducible aperiodic Markov transition matrix, the Perron-Frobenius theorem for stochastic matrices guarantees that  $\mathbf{A}^n$  converges to a rank-one matrix  $\mathbf{A}' = \boldsymbol{\pi} \mathbf{1}^T$ , 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  $\|\mathbf{A}^n - \mathbf{A}'\| \leq K_A \lambda^n$ .
- **Block  $\mathbf{C}^n$ :** By hypothesis,  $\mathbf{C}^n \rightarrow \mathbf{0}$  with exponential decay. This is equivalent to its spectral radius being less than one,  $\rho(\mathbf{C}) < 1$ . Thus, there exist constants  $K_C > 0$  and  $0 \leq \gamma < 1$  such that  $\|\mathbf{C}^n\| \leq K_C \gamma^n$ .

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

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

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

Combining these limits, the limit matrix is  $\mathbf{M}' = \begin{bmatrix} \mathbf{A}' & \mathbf{X}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ .

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

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

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

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

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

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

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

### 1.2.1 Convergence of Mutual Information

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

*Let  $\mathbf{e}_j$  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  $\{\mathbf{u}_k\}_{k=1}^\infty \subset \mathcal{D}$  converging to  $\mathbf{u}_0$  where the difference  $\mathbf{u}_\ell - \mathbf{u}_{\ell'}$  is parallel to  $\mathbf{e}_j$  (i.e., they only differ in the  $j$ -th coordinate) that satisfies  $|\mathbf{u}_0 - \mathbf{u}_k| \leq C' \rho^k$  for all  $k$  and some  $\rho \in [0, 1)$ ,  $C' > 0$ , we have for all  $k \geq K_j(C', \rho)$ :*

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

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

$$\|f(\mathbf{x}_0) - f(\mathbf{x}_k)\| \leq C \sigma^k .$$

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

$$\|\mathbf{x}_k - \mathbf{x}_0\| \leq C_x \rho^k .$$

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

$$|x_{k,j} - x_{0,j}| \leq \|\mathbf{x}_k - \mathbf{x}_0\|_\infty \leq \|\mathbf{x}_k - \mathbf{x}_0\| \leq C_x \rho^k ,$$

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

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

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

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

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

By the triangle inequality, we have:

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

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

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

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

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

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

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  $\|f(\mathbf{u}_0) - f(\mathbf{u}_m)\| \leq C_j(C_x, \rho) \rho^m m^{P_j(C_x, \rho)}$ . Applying this for  $m = k$ :

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

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

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

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

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

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

**Lemma 1.2.2.** *Let the function  $f : [0, \infty) \rightarrow \mathbb{R}$  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  $k \geq \log_\rho \left( \frac{e^{-1}}{C} \right) =: K$ .*

*Proof.* We consider two cases for the limit  $x_\infty$ :

**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  $f : [0, \infty) \rightarrow \mathbb{R}$  be defined as  $f(x) = x \log x$  for  $x \in (0, 1)$ , and  $f(x) = 0$  everywhere else. Define the function  $I : [0, 1]^{m \times n} \rightarrow \mathbb{R}$  for a matrix  $\mathbf{M} \in [0, 1]^{m \times n}$  as:*

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

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

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

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

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

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

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

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

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

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

Hence, by lemma ??, we have:

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

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  $\rho$ .  $\square$

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

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