香港中文大學
The Chinese University of Hong Kong

10.3 Joint Differential Entropy, Conditional (Differential) Entropy, and Mutual Information

Definition 10.17 The joint differential entropy $h(\mathbf{X})$ of a random vector $\mathbf{X}$ with joint pdf $f(\mathbf{x})$ is defined as

$$h(\mathbf{X}) = -\int_{\mathcal{S}} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} = -E \log f(\mathbf{X}).$$

Definition 10.17 The joint differential entropy $h(\mathbf{X})$ of a random vector $\mathbf{X}$ with joint pdf $f(\mathbf{x})$ is defined as

$$h(\mathbf{X}) = -\int_{\mathcal{S}} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} = -E \log f(\mathbf{X}).$$

Corollary If $X_1, X_2, \cdots, X_n$ are mutually independent, then

$$h(\mathbf{X}) = \sum_{i=1}^{n} h(X_i).$$

Definition 10.17 The joint differential entropy $h(\mathbf{X})$ of a random vector $\mathbf{X}$ with joint pdf $f(\mathbf{x})$ is defined as

$$h(\mathbf{X}) = -\int_{\mathcal{S}} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} = -E \log f(\mathbf{X}).$$

Corollary If $X_1, X_2, \cdots, X_n$ are mutually independent, then

$$h(\mathbf{X}) = \sum_{i=1}^{n} h(X_i).$$

Theorem 10.18 (Translation) $h(\mathbf{X} + \mathbf{c}) = h(\mathbf{X})$.

Definition 10.17 The joint differential entropy $h(\mathbf{X})$ of a random vector $\mathbf{X}$ with joint pdf $f(\mathbf{x})$ is defined as

$$h(\mathbf{X}) = -\int_{\mathcal{S}} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} = -E \log f(\mathbf{X}).$$

Corollary If $X_1, X_2, \cdots, X_n$ are mutually independent, then

$$h(\mathbf{X}) = \sum_{i=1}^{n} h(X_i).$$

Theorem 10.18 (Translation) $h(\mathbf{X} + \mathbf{c}) = h(\mathbf{X})$.

Theorem 10.19 (Scaling) $h(A\mathbf{X}) = h(\mathbf{X}) + \log |\det(A)|$.

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q \Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q \Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q \Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q \Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.
2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).
3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.
4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.
5. Now consider

$$h(\mathbf{X}) = h(Q\mathbf{Y})$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned} h(\mathbf{X}) &= h(Q\mathbf{Y}) \\ &= h(\mathbf{Y}) + \log |\det(Q)| \quad \boxed{\text{Theorem 10.19}} \end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q \Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned} h(\mathbf{X}) &= h(Q\mathbf{Y}) \\ &= h(\mathbf{Y}) + \log |\det(Q)| \\ &= h(\mathbf{Y}) + 0 \end{aligned}$$

$\det(Q) = \pm 1$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q \Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned} h(\mathbf{X}) &= h(Q\mathbf{Y}) \\ &= h(\mathbf{Y}) + \log |\det(Q)| \\ &= h(\mathbf{Y}) + 0 \\ &= \sum_{i=1}^{n} h(Y_i) \end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned} h(\mathbf{X}) &= h(Q\mathbf{Y}) \\ &= h(\mathbf{Y}) + \log |\det(Q)| \\ &= h(\mathbf{Y}) + 0 \\ &= \sum_{i=1}^{n} h(Y_i) \\ &= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned} h(\mathbf{X}) &= h(Q\mathbf{Y}) \\ &= h(\mathbf{Y}) + \log|\det(Q)| \\ &= h(\mathbf{Y}) + 0 \\ &= \sum_{i=1}^{n} h(Y_i) \\ &= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned}
h(\mathbf{X}) &= h(Q\mathbf{Y}) \\
&= h(\mathbf{Y}) + \log |\det(Q)| \\
&= h(\mathbf{Y}) + 0 \\
&= \sum_{i=1}^{n} h(Y_i) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \\
&= \frac{1}{2} \log \left[(2\pi e)^n \prod_{i=1}^{n} \lambda_i \right]
\end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q \Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned} h(\mathbf{X}) &= h(Q\mathbf{Y}) \\ &= h(\mathbf{Y}) + \log |\det(Q)| \\ &= h(\mathbf{Y}) + 0 \\ &= \sum_{i=1}^{n} h(Y_i) \\ &= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \\ &= \frac{1}{2} \log \left[(2\pi e)^n \prod_{i=1}^{n} \lambda_i \right] \end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned}
h(\mathbf{X}) &= h(Q\mathbf{Y}) \\
&= h(\mathbf{Y}) + \log |\det(Q)| \\
&= h(\mathbf{Y}) + 0 \\
&= \sum_{i=1}^{n} h(Y_i) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \underline{\lambda_i}) \\
&= \frac{1}{2} \log \left[(2\pi e)^n \underline{\prod_{i=1}^{n} \lambda_i} \right]
\end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q \Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned} h(\mathbf{X}) &= h(Q\mathbf{Y}) \\ &= h(\mathbf{Y}) + \log |\det(Q)| \\ &= h(\mathbf{Y}) + 0 \\ &= \sum_{i=1}^n h(Y_i) \\ &= \sum_{i=1}^n \frac{1}{2} \log(2\pi e \lambda_i) \\ &= \frac{1}{2} \log \left[(2\pi e)^n \underline{\prod_{i=1}^n \lambda_i} \right] \\ &= \frac{1}{2} \log[(2\pi e)^n \underline{|\Lambda|}] \end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned}
h(\mathbf{X}) &= h(Q\mathbf{Y}) \\
&= h(\mathbf{Y}) + \log |\det(Q)| \\
&= h(\mathbf{Y}) + 0 \\
&= \sum_{i=1}^{n} h(Y_i) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \\
&= \frac{1}{2} \log \left[(2\pi e)^n \prod_{i=1}^{n} \lambda_i \right] \\
&= \frac{1}{2} \log[(2\pi e)^n \underline{|\Lambda|}] \\
&= \frac{1}{2} \log[(2\pi e)^n \underline{|K|}].
\end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.
2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).
3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.
4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.
5. Now consider

$$\begin{aligned}
h(\mathbf{X}) &= h(Q\mathbf{Y}) \\
&= h(\mathbf{Y}) + \log |\det(Q)| \\
&= h(\mathbf{Y}) + 0 \\
&= \sum_{i=1}^{n} h(Y_i) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \\
&= \frac{1}{2} \log \left[(2\pi e)^n \prod_{i=1}^{n} \lambda_i \right] \\
&= \frac{1}{2} \log[(2\pi e)^n |\Lambda|] \\
&= \frac{1}{2} \log[(2\pi e)^n |K|].
\end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned}
h(\mathbf{X}) &= h(Q\mathbf{Y}) \\
&= h(\mathbf{Y}) + \log|\det(Q)| \\
&= h(\mathbf{Y}) + 0 \\
&= \sum_{i=1}^{n} h(Y_i) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \\
&= \frac{1}{2} \log \left[(2\pi e)^n \prod_{i=1}^{n} \lambda_i \right] \\
&= \frac{1}{2} \log[(2\pi e)^n \underline{|\Lambda|}] \\
&= \frac{1}{2} \log[(2\pi e)^n \underline{|K|}].
\end{aligned}$$

$$K = Q\Lambda Q^\top$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q \Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned}
h(\mathbf{X}) &= h(Q\mathbf{Y}) \\
&= h(\mathbf{Y}) + \log |\det(Q)| \\
&= h(\mathbf{Y}) + 0 \\
&= \sum_{i=1}^{n} h(Y_i) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \\
&= \frac{1}{2} \log \left[(2\pi e)^n \prod_{i=1}^{n} \lambda_i \right] \\
&= \frac{1}{2} \log[(2\pi e)^n \underline{|\Lambda|}] \\
&= \frac{1}{2} \log[(2\pi e)^n \underline{|K|}].
\end{aligned}$$

$$\begin{aligned}
K &= Q \Lambda Q^\top \\
|K| &= |Q||\Lambda||Q^\top|
\end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.
2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).
3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.
4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.
5. Now consider

$$\begin{aligned} h(\mathbf{X}) &= h(Q\mathbf{Y}) \\ &= h(\mathbf{Y}) + \log |\det(Q)| \\ &= h(\mathbf{Y}) + 0 \\ &= \sum_{i=1}^{n} h(Y_i) \\ &= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \\ &= \frac{1}{2} \log \left[(2\pi e)^n \prod_{i=1}^{n} \lambda_i \right] \\ &= \frac{1}{2} \log[(2\pi e)^n \underline{|\Lambda|}] \\ &= \frac{1}{2} \log[(2\pi e)^n \underline{|K|}]. \end{aligned}$$

$$\begin{aligned} K &= Q\Lambda Q^\top \\ |K| &= |Q||\Lambda||Q^\top| \\ &= |Q|^2 |\Lambda| \end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\mathrm{var}\, Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

$$\begin{aligned}
h(\mathbf{X}) &= h(Q\mathbf{Y}) \\
&= h(\mathbf{Y}) + \log |\det(Q)| \\
&= h(\mathbf{Y}) + 0 \\
&= \sum_{i=1}^{n} h(Y_i) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \\
&= \frac{1}{2} \log \left[(2\pi e)^n \prod_{i=1}^{n} \lambda_i \right] \\
&= \frac{1}{2} \log[(2\pi e)^n \underline{|\Lambda|}] \\
&= \frac{1}{2} \log[(2\pi e)^n \underline{|K|}].
\end{aligned}$$

$$\begin{aligned}
K &= Q\Lambda Q^\top \\
|K| &= |Q||\Lambda||Q^\top| \\
&= |Q|^2 |\Lambda| \\
&= |\Lambda|
\end{aligned}$$

Theorem 10.20 (Multivariate Gaussian Distribution) Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$. Then

$$h(\mathbf{X}) = \frac{1}{2} \log \left[(2\pi e)^n |K| \right].$$

Proof

1. Let K be diagonalizable as $Q\Lambda Q^\top$.

2. Write $\mathbf{X} = Q\mathbf{Y}$, where the random variables in $\mathbf{Y}$ are uncorrelated with $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ (cf. Corollary 10.7).

3. Since $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$.

4. Then the random variables in $\mathbf{Y}$ are mutually independent because they are uncorrelated.

5. Now consider

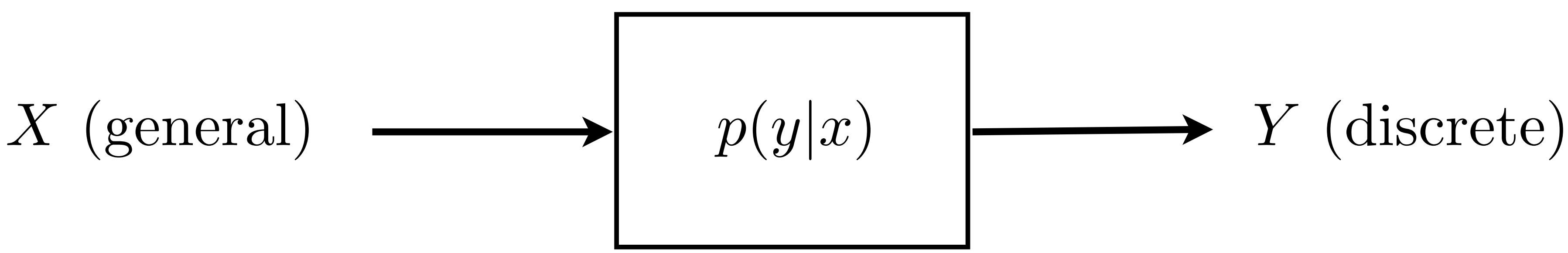
$$\begin{aligned}
h(\mathbf{X}) &= h(Q\mathbf{Y}) \\
&= h(\mathbf{Y}) + \log|\det(Q)| \\
&= h(\mathbf{Y}) + 0 \\
&= \sum_{i=1}^{n} h(Y_i) \\
&= \sum_{i=1}^{n} \frac{1}{2} \log(2\pi e \lambda_i) \\
&= \frac{1}{2} \log \left[(2\pi e)^n \prod_{i=1}^{n} \lambda_i \right] \\
&= \frac{1}{2} \log[(2\pi e)^n |\Lambda|] \\
&= \frac{1}{2} \log[(2\pi e)^n |K|].
\end{aligned}$$

The Model of a "Channel" with Discrete Output

Definition 10.21 The random variable Y is related to the random variable X through a conditional distribution $p(y|x)$ defined for all x means

The Model of a "Channel" with Discrete Output

Definition 10.21 The random variable Y is related to the random variable X through a conditional distribution $p(y|x)$ defined for all x means

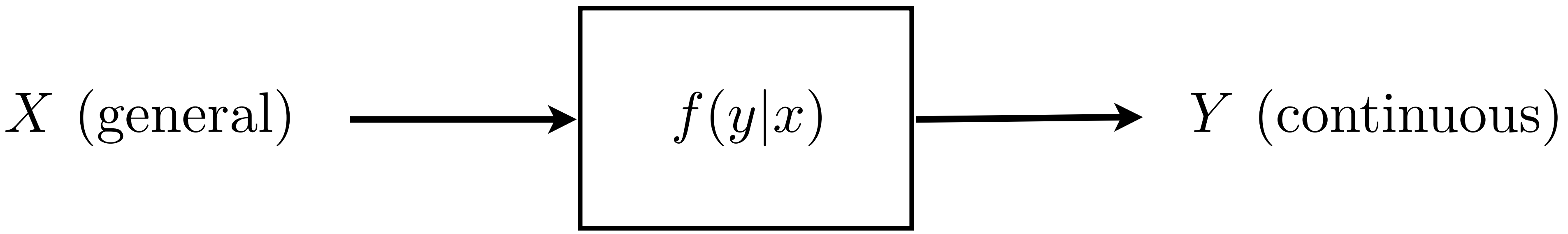


The Model of a "Channel" with Continuous Output

Definition 10.22 The random variable Y is related to the random variable X through a conditional pdf $f(y|x)$ defined for all x means

The Model of a "Channel" with Continuous Output

Definition 10.22 The random variable Y is related to the random variable X through a conditional pdf $f(y|x)$ defined for all x means

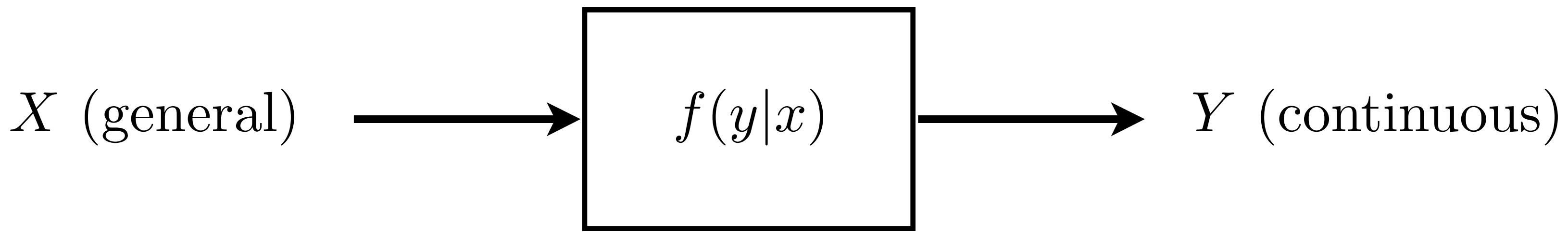


Conditional Differential Entropy

Definition 10.23 Let X and Y be jointly distributed random variables where Y is continuous and is related to X through a conditional pdf $f(y|x)$ defined for all x. The conditional differential entropy of Y given $\{X = x\}$ is defined as

Conditional Differential Entropy

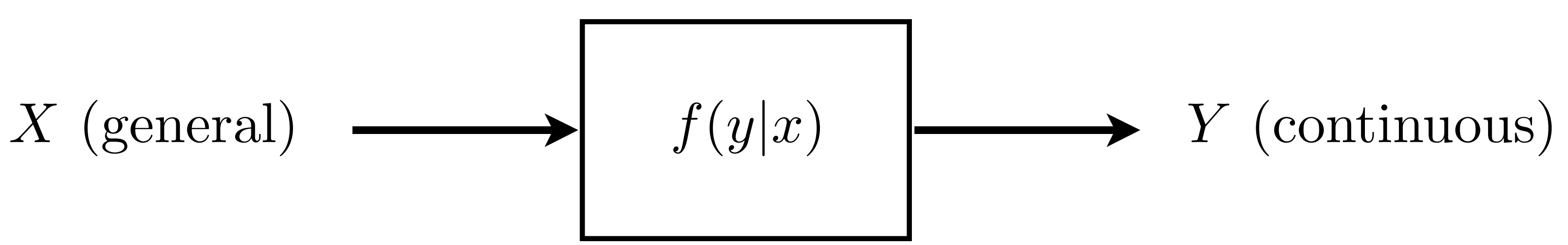
Definition 10.23 Let X and Y be jointly distributed random variables where Y is continuous and is related to X through a conditional pdf $f(y|x)$ defined for all x. The conditional differential entropy of Y given $\{X = x\}$ is defined as



Conditional Differential Entropy

Definition 10.23 Let X and Y be jointly distributed random variables where Y is continuous and is related to X through a conditional pdf $f(y|x)$ defined for all x. The conditional differential entropy of Y given $\{X = x\}$ is defined as

$$h(Y|X = x) = -\int_{\mathcal{S}_Y(x)} f(y|x) \log f(y|x)\, dy$$



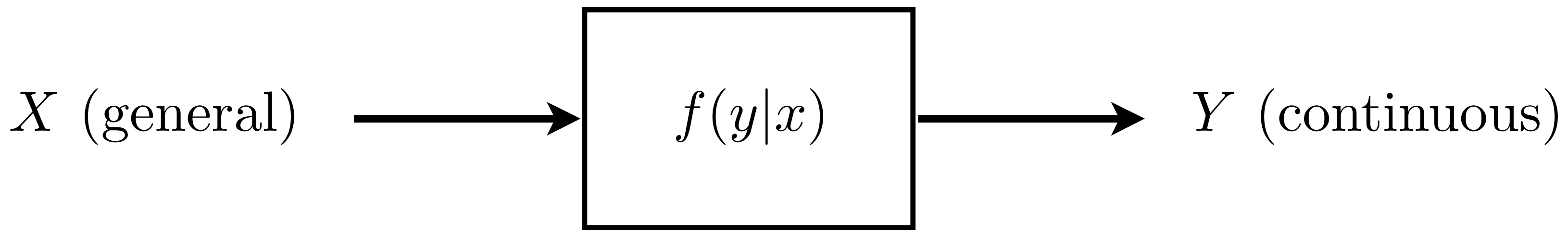
Conditional Differential Entropy

Definition 10.23 Let X and Y be jointly distributed random variables where Y is continuous and is related to X through a conditional pdf $f(y|x)$ defined for all x. The conditional differential entropy of Y given $\{X = x\}$ is defined as

$$h(Y|X = x) = -\int_{\mathcal{S}_Y(x)} f(y|x) \log f(y|x)\, dy$$

and the conditional differential entropy of Y given X is defined as

$$h(Y|X) = -\int_{\mathcal{S}_X} h(Y|X = x)\, dF(x) = -E \log f(Y|X)$$



Conditional Differential Entropy

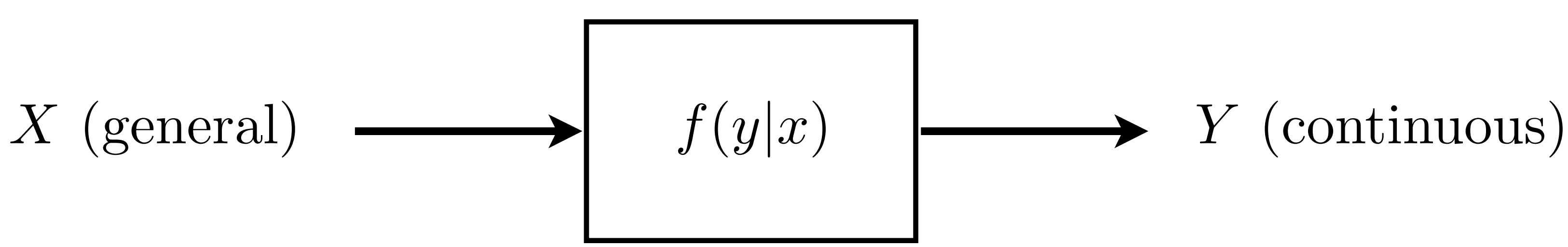
Definition 10.23 Let X and Y be jointly distributed random variables where Y is continuous and is related to X through a conditional pdf $f(y|x)$ defined for all x. The conditional differential entropy of Y given $\{X = x\}$ is defined as

$$h(Y|X = x) = -\int_{\mathcal{S}_Y(x)} f(y|x) \log f(y|x) \, dy$$

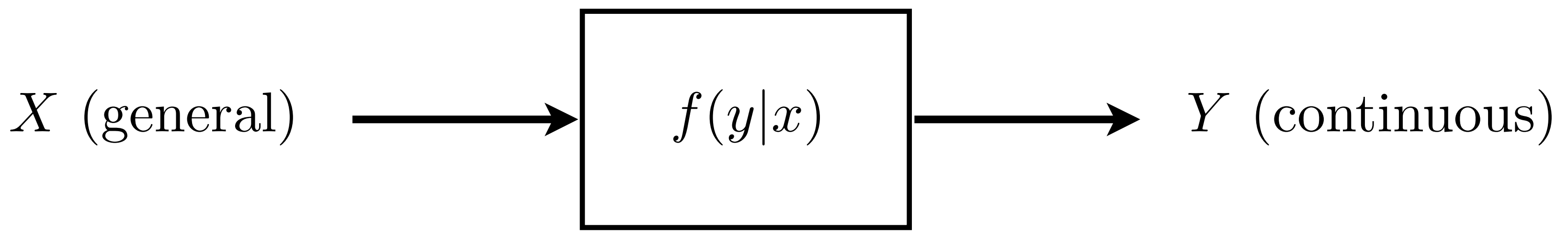
and the conditional differential entropy of Y given X is defined as

$$h(Y|X) = -\int_{\mathcal{S}_X} h(Y|X = x) \, dF(x) = -E \log f(Y|X)$$

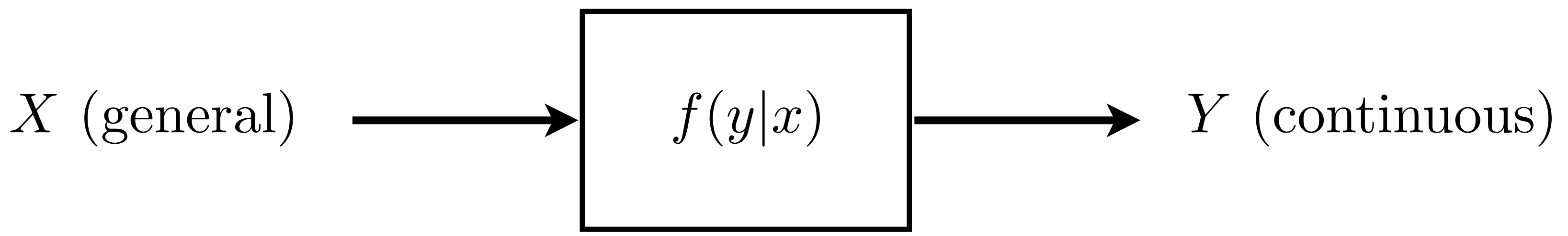
$f(x)\,dx$



Proposition 10.24 If



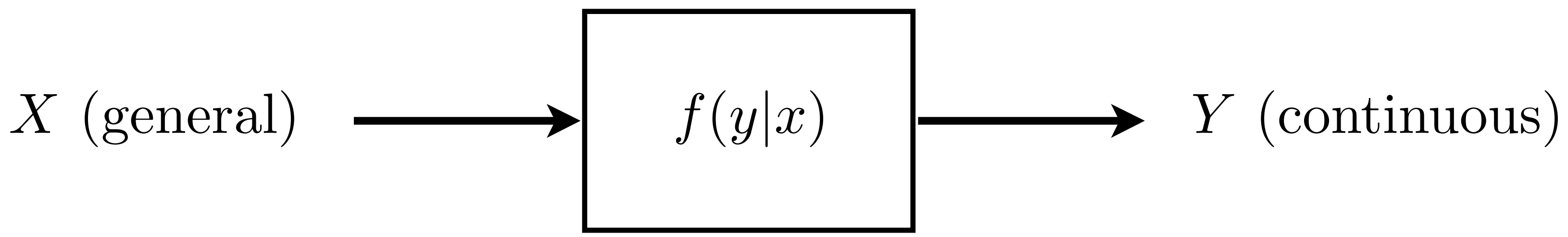
Proposition 10.24 If



then $f(y)$ exists and is given by

$$f(y) = \int f(y|x)\, dF(x).$$

Proposition 10.24 If

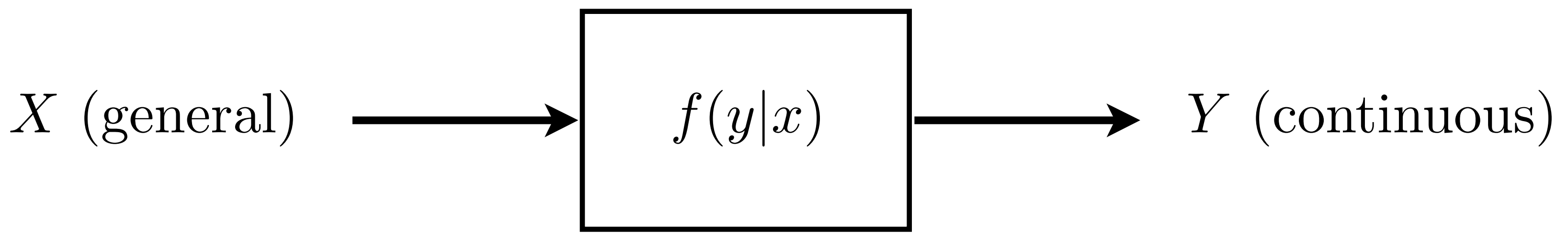


then $f(y)$ exists and is given by

$$f(y) = \int f(y|x)\, dF(x).$$

Remark Proposition 10.24 says that the pdf of Y exists regardless of the distribution of X. The next proposition is its vector generalization.

Proposition 10.24 If



then $f(y)$ exists and is given by

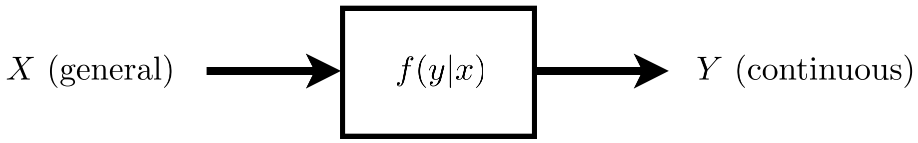
$$f(y) = \int f(y|x)\, dF(x).$$

Remark Proposition 10.24 says that the pdf of Y exists regardless of the distribution of X. The next proposition is its vector generalization.

Proposition 10.25 Let $\mathbf{X}$ and $\mathbf{Y}$ be jointly distributed random vectors where $\mathbf{Y}$ is continuous and is related to $\mathbf{X}$ through a conditional pdf $f(\mathbf{y}|\mathbf{x})$ defined for all $\mathbf{x}$. Then $f(\mathbf{y})$ exists and is given by

$$f(\mathbf{y}) = \int f(\mathbf{y}|\mathbf{x}) dF(\mathbf{x}).$$

Proposition 10.24 If

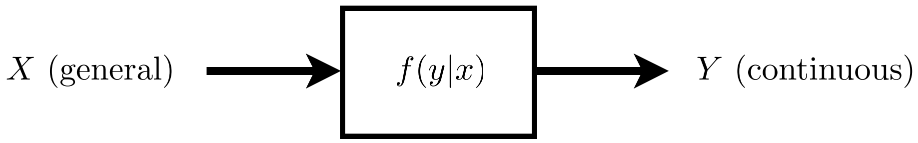


then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

Proposition 10.24 If



then $f(y)$ exists and is given by

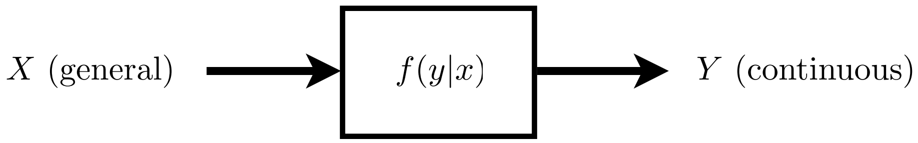
$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

$$F_Y(y) \quad = \quad F_{XY}(\infty, y)$$

Proposition 10.24 If



then $f(y)$ exists and is given by

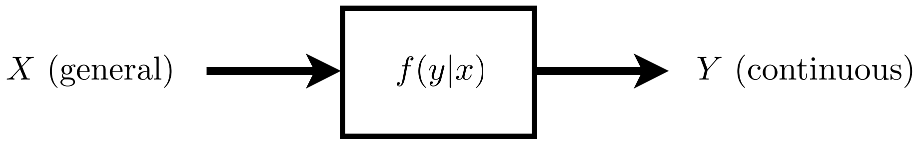
$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

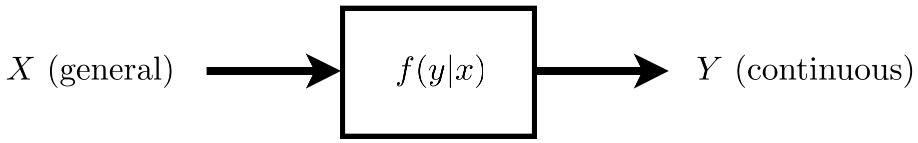
1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| dv\, dF_X(x)$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

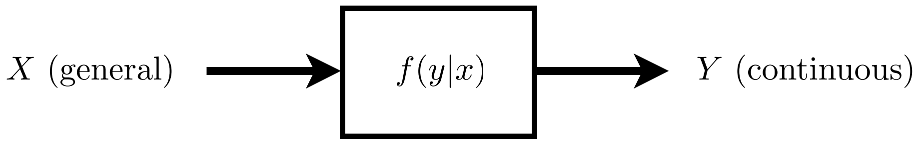
1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \end{aligned}$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

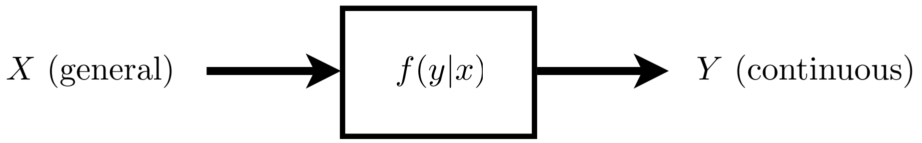
1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \end{aligned}$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

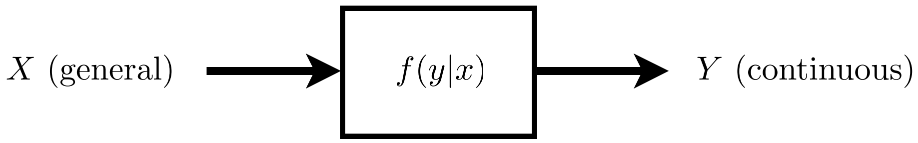
1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty}\int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| dv\, dF_X(x) \\ &\quad = \int_{-\infty}^{\underline{\infty}}\int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &\quad = F_{XY}(\underline{\infty}, y) \end{aligned}$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

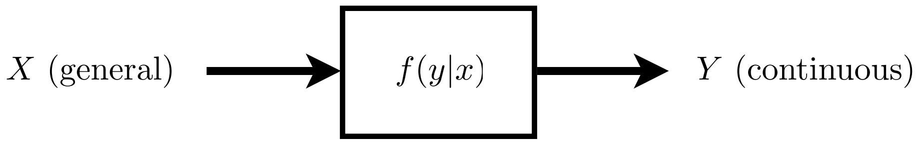
1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\underline{y}} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, \underline{y}) \end{aligned}$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

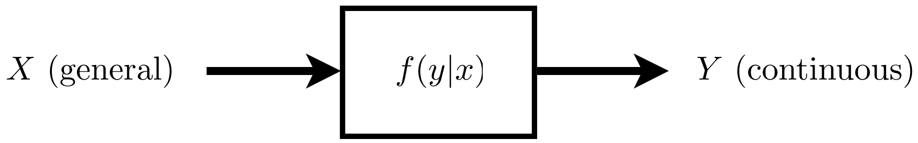
1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &\quad = \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &\quad = F_{XY}(\infty, y) \\ &\quad = F_Y(y) \end{aligned}$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

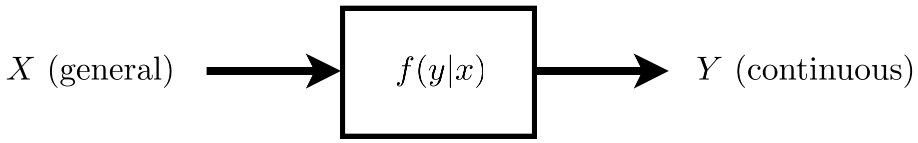
1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

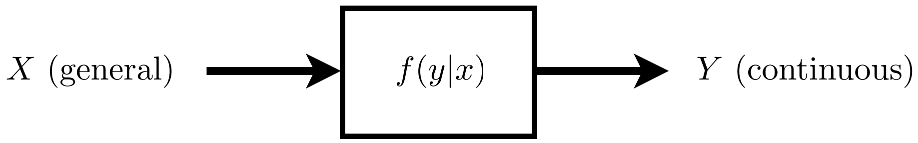
$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\le 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

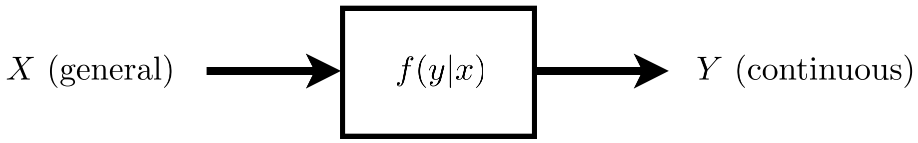
$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

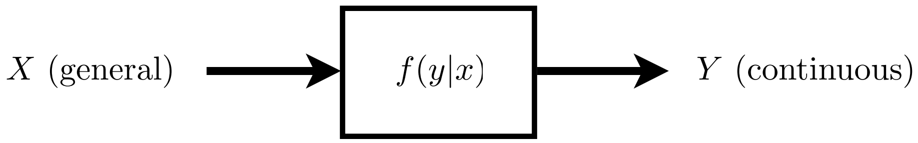
$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

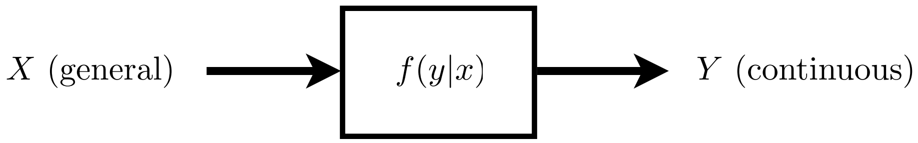
$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

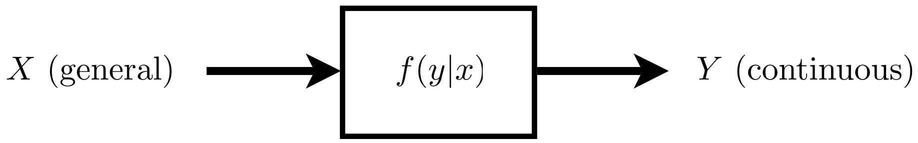
2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

3. By Fubini's theorem, the order of integration in $F_Y(y)$ can be exchanged, and so

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

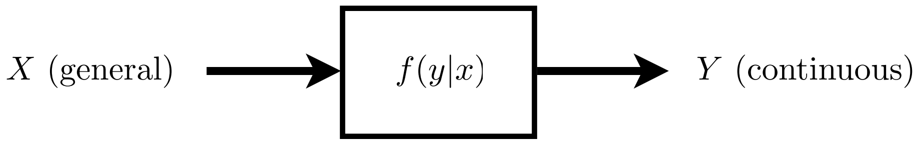
$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

3. By Fubini's theorem, the order of integration in $F_Y(y)$ can be exchanged, and so

$$F_Y(y) = \int_{-\infty}^{y} \left[\int_{-\infty}^{\infty} f_{Y|X}(v|x)\, dF_X(x) \right] dv.$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty}\int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

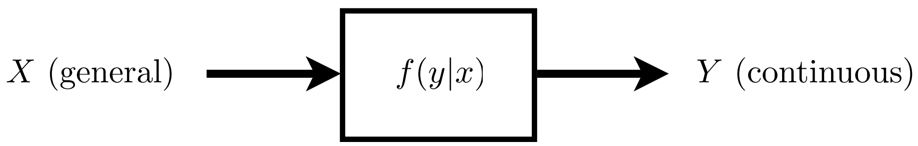
3. By Fubini's theorem, the order of integration in $F_Y(y)$ can be exchanged, and so

$$F_Y(y) = \int_{-\infty}^{y} \left[\int_{-\infty}^{\infty} f_{Y|X}(v|x)\, dF_X(x) \right] dv.$$

4. Then

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| dv\, dF_X(x) \\ &\quad = \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &\quad = F_{XY}(\infty, y) \\ &\quad = F_Y(y) \\ &\quad \leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

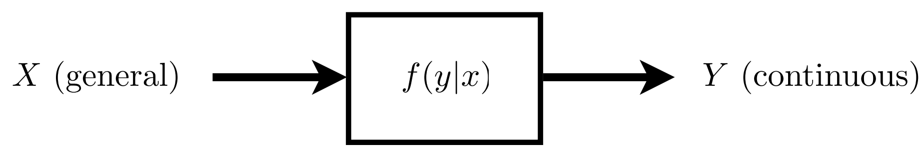
3. By Fubini's theorem, the order of integration in $F_Y(y)$ can be exchanged, and so

$$F_Y(y) = \int_{-\infty}^{y} \underbrace{\left[\int_{-\infty}^{\infty} f_{Y|X}(v|x)\, dF_X(x) \right]}_{g(v)} dv.$$

4. Then

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

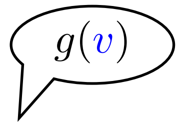
$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

3. By Fubini's theorem, the order of integration in $F_Y(y)$ can be exchanged, and so

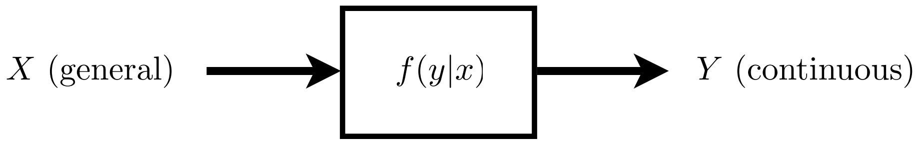


$$F_Y(y) = \int_{-\infty}^{y} \left[\int_{-\infty}^{\infty} f_{Y|X}(v|x)\, dF_X(x) \right] dv.$$

4. Then

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} \int_{-\infty}^{y} g(v)\, dv \end{aligned}$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

3. By Fubini's theorem, the order of integration in $F_Y(y)$ can be exchanged, and so

$g(v)$

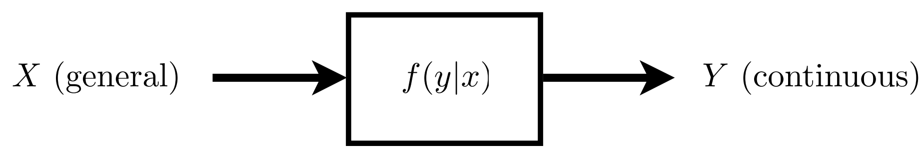
$$F_Y(y) = \int_{-\infty}^{y} \left[\int_{-\infty}^{\infty} f_{Y|X}(v|x)\, dF_X(x) \right] dv.$$

4. Then

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} \int_{-\infty}^{y} g(v)\, dv \\ &= g(y) \end{aligned}$$

Fundamental theorem of calculus

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

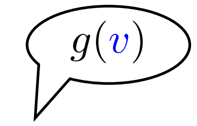
$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| \, dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

3. By Fubini's theorem, the order of integration in $F_Y(y)$ can be exchanged, and so

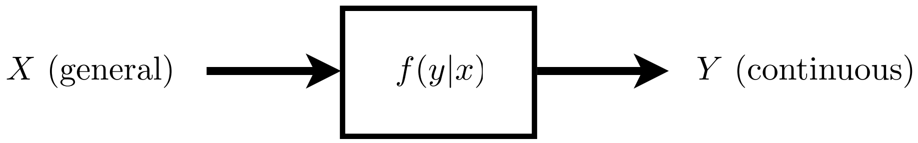


$$F_Y(y) = \int_{-\infty}^{y} \left[\int_{-\infty}^{\infty} f_{Y|X}(v|x)\, dF_X(x) \right] dv.$$

4. Then

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} \int_{-\infty}^{y} g(v)\, dv \\ &= g(y) \\ &= \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x). \end{aligned}$$

Proposition 10.24 If



then $f(y)$ exists and is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x).$$

Proof

1.

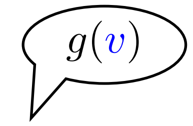
$$\begin{aligned} F_Y(y) &= F_{XY}(\infty, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x). \end{aligned}$$

2. Since

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{y} \left| f_{Y|X}(v|x) \right| dv\, dF_X(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv\, dF_X(x) \\ &= F_{XY}(\infty, y) \\ &= F_Y(y) \\ &\leq 1, \end{aligned}$$

$f_{Y|X}(v|x)$ is absolutely integrable.

3. By Fubini's theorem, the order of integration in $F_Y(y)$ can be exchanged, and so



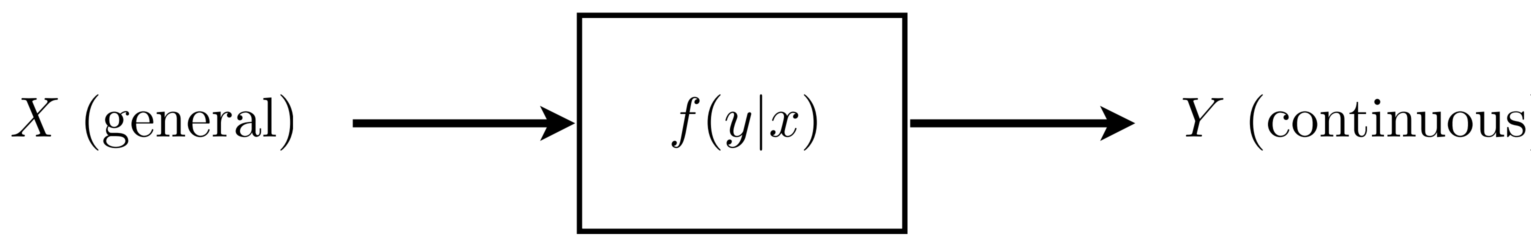
$$F_Y(y) = \int_{-\infty}^{y} \left[\int_{-\infty}^{\infty} f_{Y|X}(v|x)\, dF_X(x) \right] dv.$$

4. Then

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} \int_{-\infty}^{y} g(v)\, dv \\ &= g(y) \\ &= \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dF_X(x). \end{aligned}$$

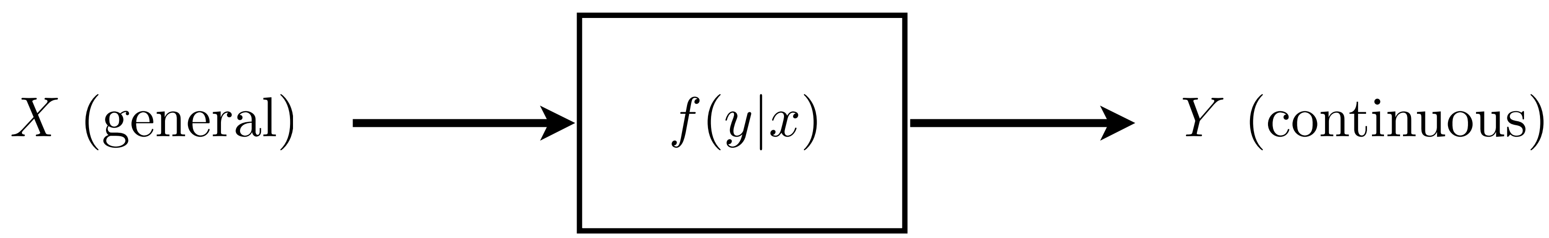
Mutual Information

Definition 10.26 Let



Mutual Information

Definition 10.26 Let

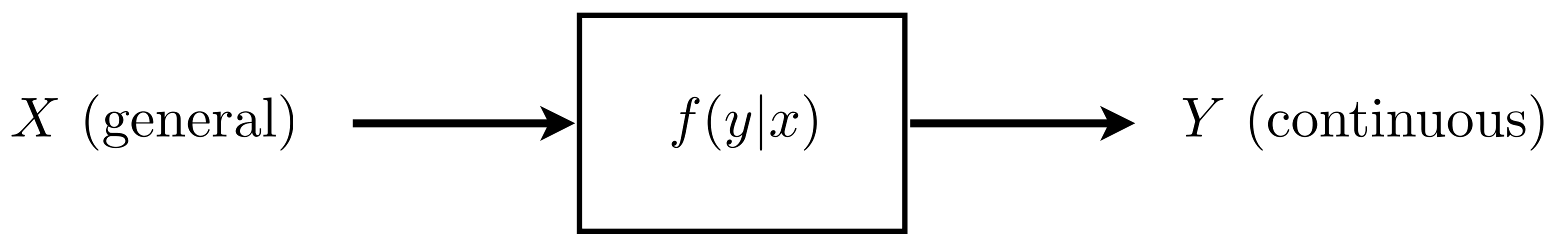


1. The mutual information between X and Y is defined as

$$I(X;Y) \quad = \quad E \log \frac{f(Y|X)}{f(Y)}$$

Mutual Information

Definition 10.26 Let

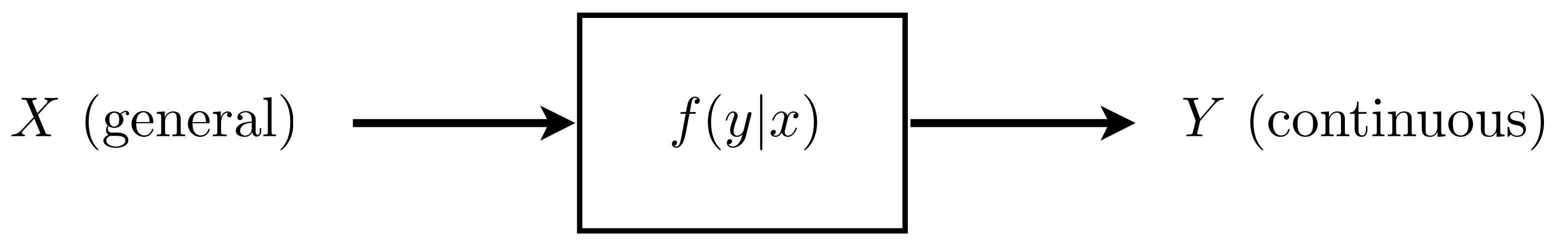


1. The mutual information between X and Y is defined as

$$
\begin{aligned}
I(X;Y) &= E \log \frac{f(Y|X)}{f(Y)} \\
&= \int_{\mathcal{S}_X} \int_{\mathcal{S}_Y(x)} f(y|x) \log \frac{f(y|x)}{f(y)} dy \, dF(x).
\end{aligned}
$$

Mutual Information

Definition 10.26 Let



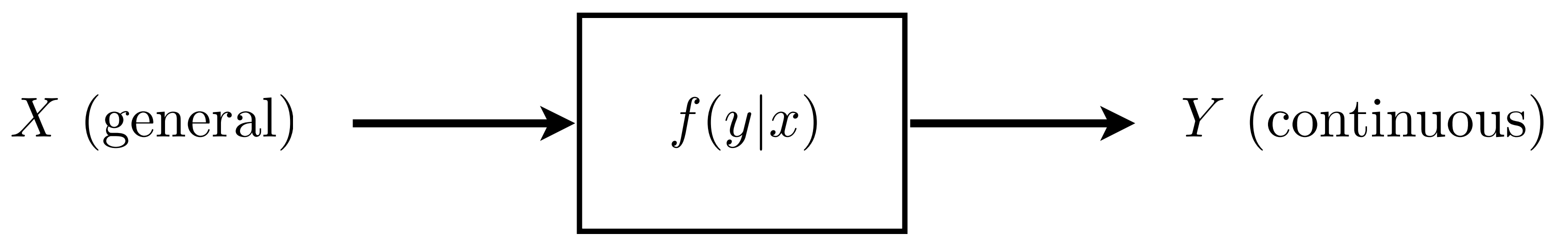
1. The mutual information between X and Y is defined as

$$
\begin{aligned}
I(X;Y) &= E \log \frac{f(Y|X)}{f(Y)} \\
&= \int_{\mathcal{S}_X} \int_{\mathcal{S}_Y(x)} f(y|x) \log \frac{f(y|x)}{f(y)} dy \, dF(x).
\end{aligned}
$$

$f(x)\,dx$

Mutual Information

Definition 10.26 Let



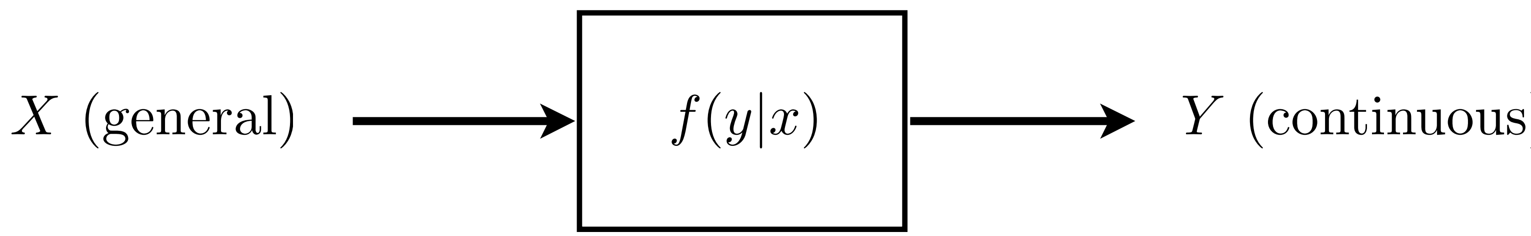
1. The mutual information between X and Y is defined as

$$
\begin{aligned}
I(X;Y) &= E \log \frac{f(Y|X)}{f(Y)} \\
&= \int_{\mathcal{S}_X} \int_{\mathcal{S}_Y(x)} f(y|x) \log \frac{f(y|x)}{f(y)} dy\, dF(x).
\end{aligned}
$$

2. When both X and Y are continuous and $f(x,y)$ exists,

Mutual Information

Definition 10.26 Let



1. The mutual information between X and Y is defined as

$$\begin{aligned} I(X;Y) &= E \log \frac{f(Y|X)}{f(Y)} \\ &= \int_{\mathcal{S}_X} \int_{\mathcal{S}_Y(x)} f(y|x) \log \frac{f(y|x)}{f(y)} dy \, dF(x). \end{aligned}$$

2. When both X and Y are continuous and $f(x,y)$ exists,

$$I(X;Y) = E \log \frac{f(Y|X)}{f(Y)} = E \log \frac{f(X,Y)}{f(X)f(Y)}.$$

Remarks

Remarks

- With Definition 10.26, the mutual information is defined when one r.v. is general and the other is continuous.

Remarks

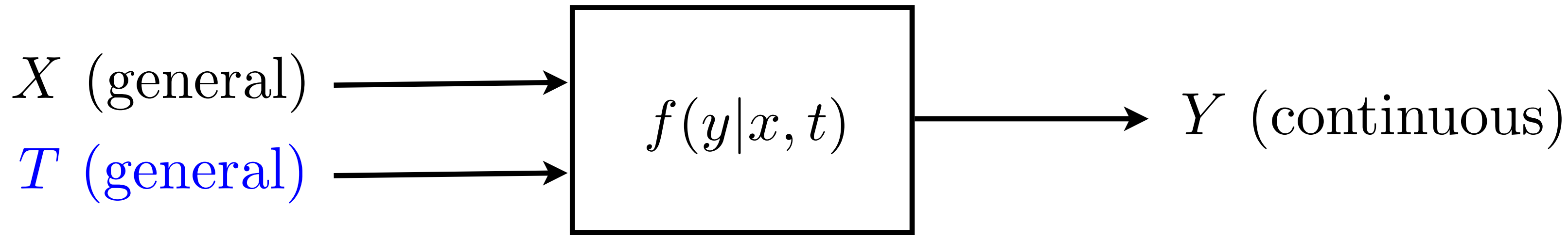
- With Definition 10.26, the mutual information is defined when one r.v. is general and the other is continuous.
- In Ch. 2, the mutual information is defined when both r.v.'s are discrete.

Remarks

- With Definition 10.26, the mutual information is defined when one r.v. is general and the other is continuous.
- In Ch. 2, the mutual information is defined when both r.v.'s are discrete.
- Thus the mutual information is defined when each of the r.v.'s is either discrete or continuous.

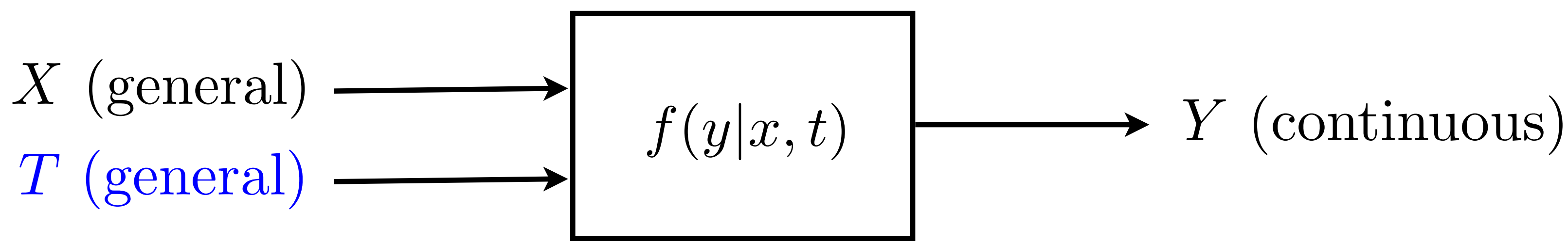
Conditional Mutual Information

Definition 10.27 Let



Conditional Mutual Information

Definition 10.27 Let

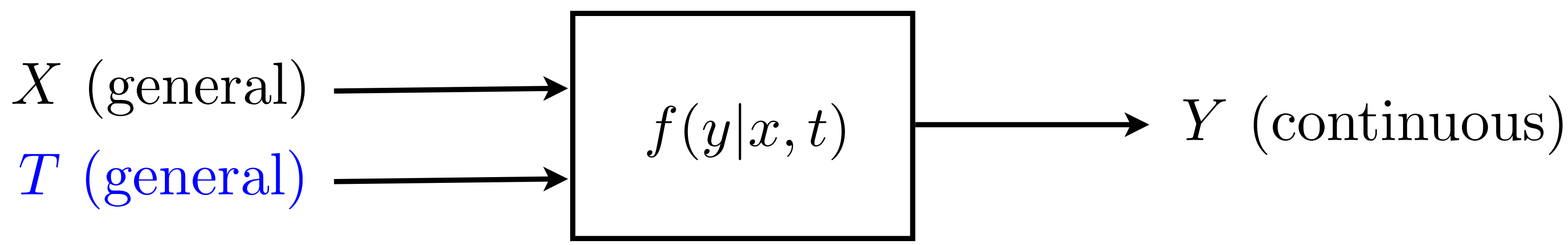


The mutual information between X and Y given T is defined as

$$I(X;Y|T) = \int_{\mathcal{S}_T} I(X;Y|T=t)dF(t) = E\log\frac{f(Y|X,T)}{f(Y|T)}$$

Conditional Mutual Information

Definition 10.27 Let



The mutual information between X and Y given T is defined as

$$I(X;Y|T) = \int_{\mathcal{S}_T} I(X;Y|T=t)\,dF(t) = E \log \frac{f(Y|X,T)}{f(Y|T)}$$

where

$$I(X;Y|T=t) = \int_{\mathcal{S}_X(t)} \int_{\mathcal{S}_Y(x,t)} f(y|x,t) \log \frac{f(y|x,t)}{f(y|t)}\,dy\,dF(x|t).$$

Interpretation of I(X;Y)

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

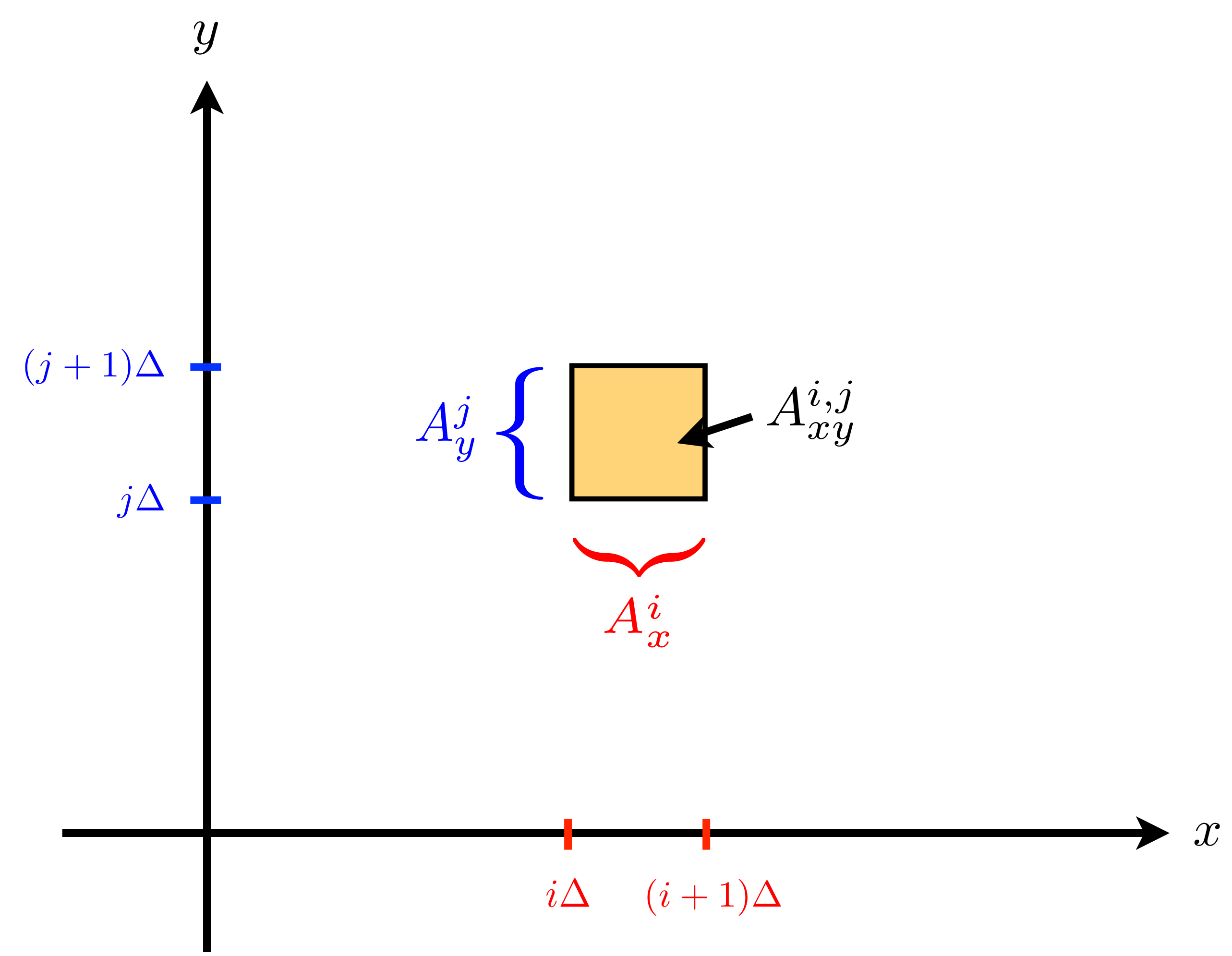
and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.



Interpretation of I(X;Y)

1. Assume $f(x,y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j .$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$I(\hat{X}_\Delta; \hat{Y}_\Delta)$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$I(\hat{X}_\Delta; \hat{Y}_\Delta) = \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$I(\hat{X}_\Delta; \hat{Y}_\Delta) = \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned} & I(\hat{X}_\Delta; \hat{Y}_\Delta) \\ = \quad & \sum_i \sum_j \underline{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\}} \log \frac{\underline{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\}}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\ \approx \quad & \sum_i \sum_j \underline{f(x_i, y_j)\Delta^2} \log \frac{\underline{f(x_i, y_j)\Delta^2}}{(f(x_i)\Delta)(f(y_j)\Delta)} \end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned} & I(\hat{X}_\Delta; \hat{Y}_\Delta) \\ = & \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\}}{\underline{\Pr\{\hat{X}_\Delta = i\}}\Pr\{\hat{Y}_\Delta = j\}} \\ \approx & \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\Delta^2}{\underline{(f(x_i)\Delta)}(f(y_j)\Delta)} \end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned} I(\hat{X}_\Delta; \hat{Y}_\Delta) & \\ &= \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\}}{\Pr\{\hat{X}_\Delta = i\}\underline{\Pr\{\hat{Y}_\Delta = j\}}} \\ &\approx \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\Delta^2}{(f(x_i)\Delta)\underline{(f(y_j)\Delta)}} \end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x,y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned} & I(\hat{X}_\Delta; \hat{Y}_\Delta) \\ = \quad & \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\ \approx \quad & \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\cancel{\Delta^2}}{(f(x_i)\Delta)(f(y_j)\Delta)} \end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned} & I(\hat{X}_\Delta; \hat{Y}_\Delta) \\ = & \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\ \approx & \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\cancel{\Delta^2}}{(f(x_i)\cancel{\Delta})(f(y_j)\Delta)} \end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned} I(\hat{X}_\Delta; \hat{Y}_\Delta) & \\ &= \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\ &\approx \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\cancel{\Delta^2}}{(f(x_i)\cancel{\Delta})(f(y_j)\cancel{\Delta})} \end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x,y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned} & I(\hat{X}_\Delta; \hat{Y}_\Delta) \\ = & \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\ \approx & \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\cancel{\Delta^2}}{(f(x_i)\cancel{\Delta})(f(y_j)\cancel{\Delta})} \\ = & \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)}{f(x_i)f(y_j)} \end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned}
& I(\hat{X}_\Delta; \hat{Y}_\Delta) \\
&= \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i, j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\
&\approx \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\cancel{\Delta^2}}{(f(x_i)\cancel{\Delta})(f(y_j)\cancel{\Delta})} \\
&= \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)}{f(x_i)f(y_j)} \\
&\approx \int\int f(x, y) \log \frac{f(x, y)}{f(x)f(y)}\, dx\, dy
\end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x,y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\,i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\,j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned}
& I(\hat{X}_\Delta; \hat{Y}_\Delta) \\
&= \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\
&\approx \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\Delta^2}{(f(x_i)\Delta)(f(y_j)\Delta)} \\
&= \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)}{f(x_i)f(y_j)} \\
&\approx \int\int f(x,y) \log \frac{f(x,y)}{f(x)f(y)}\, dx\, dy \\
&= I(X;Y)
\end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.
2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.
6. Then

$$\begin{aligned}
&I(\hat{X}_\Delta; \hat{Y}_\Delta) \\
&= \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\
&\approx \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\cancel{\Delta^2}}{(f(x_i)\cancel{\Delta})(f(y_j)\cancel{\Delta})} \\
&= \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)}{f(x_i)f(y_j)} \\
&\approx \int \int f(x, y) \log \frac{f(x, y)}{f(x)f(y)}\, dx\, dy \\
&= I(X;Y)
\end{aligned}$$

Interpretation of I(X;Y)

1. Assume $f(x, y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

6. Then

$$\begin{aligned} I(\hat{X}_\Delta; \hat{Y}_\Delta) & \\ &= \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\ &\approx \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\cancel{\Delta^2}}{(f(x_i)\cancel{\Delta})(f(y_j)\cancel{\Delta})} \\ &= \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)}{f(x_i)f(y_j)} \\ &\approx \int\!\!\int f(x,y) \log \frac{f(x,y)}{f(x)f(y)}\, dx\, dy \\ &= I(X;Y) \end{aligned}$$

7. Therefore, $I(X;Y)$ can be interpreted as the limit of $I(\hat{X}_\Delta; \hat{Y}_\Delta)$ as $\Delta \to 0$.

Interpretation of I(X;Y)

1. Assume $f(x,y)$ exists and is continuous.

2. For a fixed Δ, for all integer i and j, define the intervals

$$A_x^i = [\, i\Delta, (i+1)\Delta) \quad \text{and} \quad A_y^j = [\, j\Delta, (j+1)\Delta)$$

and the rectangle

$$A_{xy}^{i,j} = A_x^i \times A_y^j.$$

3. Define discrete r.v.'s

$$\begin{cases} \hat{X}_\Delta = i & \text{if } X \in A_x^i \\ \hat{Y}_\Delta = j & \text{if } Y \in A_y^j \end{cases}$$

4. $\hat{X}_\Delta$ and $\hat{Y}_\Delta$ are quantizations of X and Y, respectively.

5. For all i and j, let $(x_i, y_j) \in A_x^i \times A_y^j$.

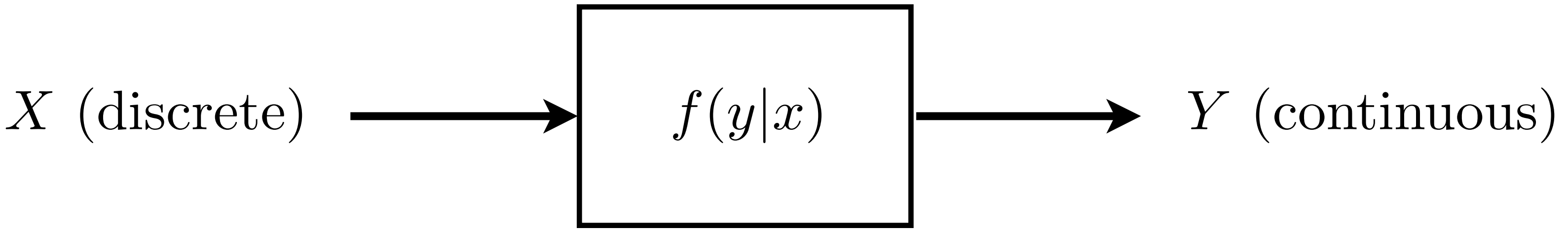
6. Then

$$\begin{aligned}
&I(\hat{X}_\Delta; \hat{Y}_\Delta) \\
&= \sum_i \sum_j \Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\} \log \frac{\Pr\{(\hat{X}_\Delta, \hat{Y}_\Delta) = (i,j)\}}{\Pr\{\hat{X}_\Delta = i\}\Pr\{\hat{Y}_\Delta = j\}} \\
&\approx \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)\cancel{\Delta^2}}{(f(x_i)\cancel{\Delta})(f(y_j)\cancel{\Delta})} \\
&= \sum_i \sum_j f(x_i, y_j)\Delta^2 \log \frac{f(x_i, y_j)}{f(x_i)f(y_j)} \\
&\approx \int\int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} \, dx\, dy \\
&= I(X;Y)
\end{aligned}$$

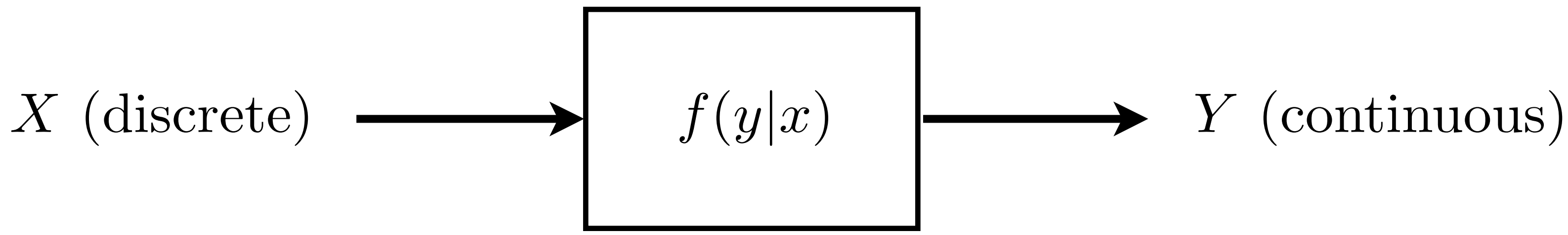
7. Therefore, $I(X;Y)$ can be interpreted as the limit of $I(\hat{X}_\Delta; \hat{Y}_\Delta)$ as $\Delta \to 0$.

8. This interpretation continues to be valid for general distribution for X and Y.

Definition 10.28 Let



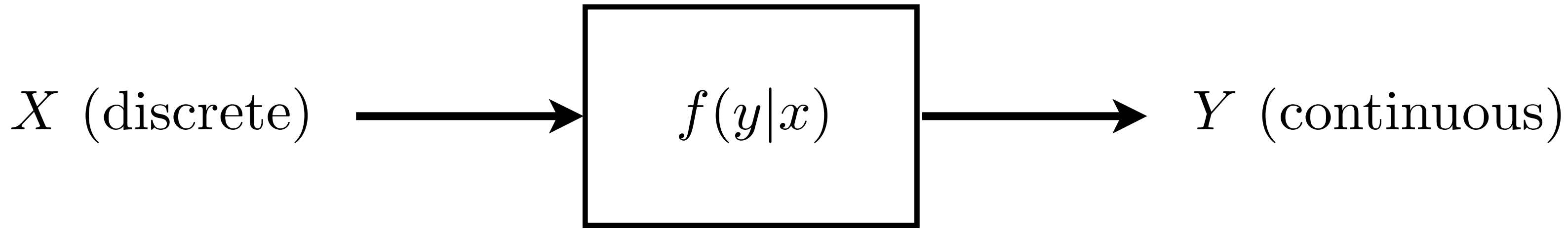
Definition 10.28 Let



The conditional entropy of X given Y is defined as

$$H(X|Y) = H(X) - I(X;Y).$$

Definition 10.28 Let

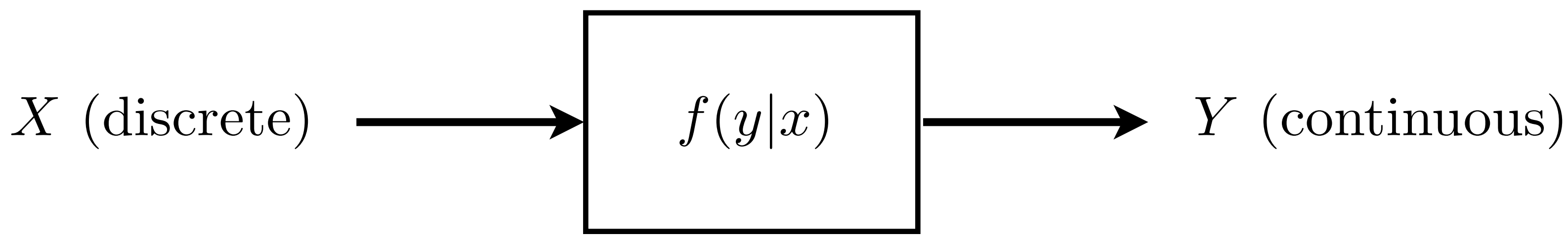


The conditional entropy of X given Y is defined as

$$H(X|Y) = H(X) - I(X;Y).$$

Proposition 10.29 For two random variables X and Y,

Definition 10.28 Let



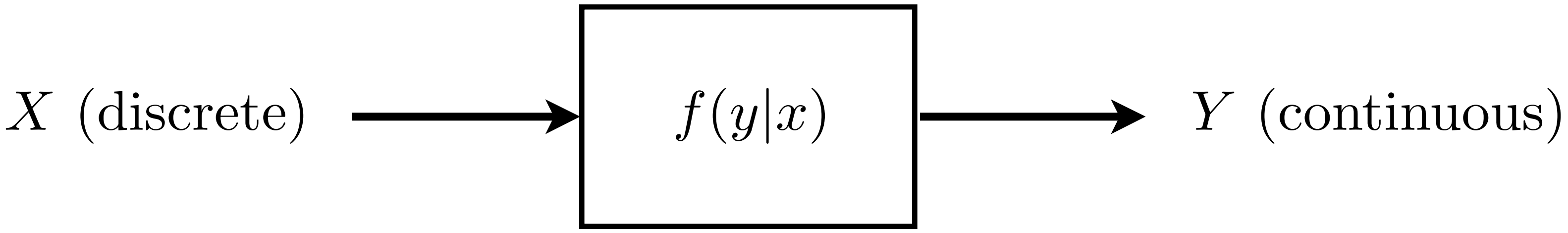
The conditional entropy of X given Y is defined as

$$H(X|Y) = H(X) - I(X;Y).$$

Proposition 10.29 For two random variables X and Y,

1. $h(Y) = h(Y|X) + I(X;Y)$ if Y is continuous;

Definition 10.28 Let



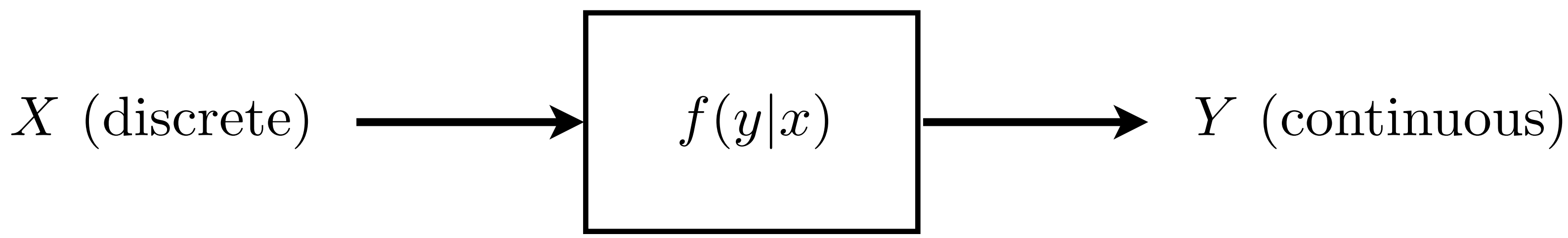
The conditional entropy of X given Y is defined as

$$H(X|Y) = H(X) - I(X;Y).$$

Proposition 10.29 For two random variables X and Y,

1. $h(Y) = h(Y|X) + I(X;Y)$ if Y is continuous;
2. $H(Y) = H(Y|X) + I(X;Y)$ if Y is discrete.

Definition 10.28 Let



The conditional entropy of X given Y is defined as

$$H(X|Y) = H(X) - I(X;Y).$$

Proposition 10.29 For two random variables X and Y,

1. $h(Y) = h(Y|X) + I(X;Y)$ if Y is continuous;
2. $H(Y) = H(Y|X) + I(X;Y)$ if Y is discrete.

Proposition 10.30 (Chain Rule)

$$h(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} h(X_i|X_1, \cdots, X_{i-1})$$

Theorem 10.31

$$I(X;Y) \geq 0,$$

with equality if and only if X is independent of Y.

Theorem 10.31

$$I(X;Y) \geq 0,$$

with equality if and only if X is independent of Y.

Corollary 10.32

$$I(X;Y|T) \geq 0,$$

with equality if and only if X is independent of Y conditioning on T.

Theorem 10.31

$$I(X;Y) \geq 0,$$

with equality if and only if X is independent of Y.

Corollary 10.32

$$I(X;Y|T) \geq 0,$$

with equality if and only if X is independent of Y conditioning on T.

Corollary 10.33 (Conditioning Does Not Increase Differential Entropy)

$$h(X|Y) \leq h(X)$$

with equality if and only if X and Y are independent.

Theorem 10.31

$$I(X;Y) \geq 0,$$

with equality if and only if X is independent of Y.

Corollary 10.32

$$I(X;Y|T) \geq 0,$$

with equality if and only if X is independent of Y conditioning on T.

Corollary 10.33 (Conditioning Does Not Increase Differential Entropy)

$$h(X|Y) \leq h(X)$$

with equality if and only if X and Y are independent.

Remarks

Theorem 10.31

$$I(X;Y) \geq 0,$$

with equality if and only if X is independent of Y.

Corollary 10.32

$$I(X;Y|T) \geq 0,$$

with equality if and only if X is independent of Y conditioning on T.

Corollary 10.33 (Conditioning Does Not Increase Differential Entropy)

$$h(X|Y) \leq h(X)$$

with equality if and only if X and Y are independent.

Remarks For continuous r.v.'s,

Theorem 10.31

$$I(X;Y) \geq 0,$$

with equality if and only if X is independent of Y.

Corollary 10.32

$$I(X;Y|T) \geq 0,$$

with equality if and only if X is independent of Y conditioning on T.

Corollary 10.33 (Conditioning Does Not Increase Differential Entropy)

$$h(X|Y) \leq h(X)$$

with equality if and only if X and Y are independent.

Remarks For continuous r.v.'s,

1. $h(X), h(X|Y) \geq 0$ DO NOT generally hold;

Theorem 10.31

$$I(X;Y) \geq 0,$$

with equality if and only if X is independent of Y.

Corollary 10.32

$$I(X;Y|T) \geq 0,$$

with equality if and only if X is independent of Y conditioning on T.

Corollary 10.33 (Conditioning Does Not Increase Differential Entropy)

$$h(X|Y) \leq h(X)$$

with equality if and only if X and Y are independent.

Remarks For continuous r.v.'s,

1. $h(X), h(X|Y) \geq 0$ DO NOT generally hold;

2. $I(X;Y), I(X;Y|Z) \geq 0$ always hold.

Corollary 10.34 (Independence Bound for Differential Entropy)

$$h(X_1, X_2, \cdots, X_n) \leq \sum_{i=1}^{n} h(X_i)$$

with equality if and only if $i = 1, 2, \cdots, n$ are mutually independent.