10.5 Informational Divergence

Definition 10.39 Let f and g be two pdf's defined on $\Re^n$ with supports $\mathcal{S}_f$ and $\mathcal{S}_g$, respectively. The informational divergence between f and g is defined as

$$D(f\|g) = \int_{\mathcal{S}_f} f(x) \log \frac{f(x)}{g(x)} dx = E_f \log \frac{f(X)}{g(X)},$$

where E_f denotes expectation with respect to f.

Definition 10.39 Let f and g be two pdf's defined on $\Re^n$ with supports $\mathcal{S}_f$ and $\mathcal{S}_g$, respectively. The informational divergence between f and g is defined as

$$D(f\|g) = \int_{\mathcal{S}_f} f(x) \log \frac{f(x)}{g(x)} dx = E_f \log \frac{f(X)}{g(X)},$$

where E_f denotes expectation with respect to f.

Remark If $D(f\|g) < \infty$, then

$$\mathcal{S}_f \setminus \mathcal{S}_g = \{x : f(x) > 0 \text{ and } g(x) = 0\}$$

has zero Lebesgue measure, i.e., $\mathcal{S}_f$ is essentially a subset of $\mathcal{S}_g$.

Theorem 10.40 (Divergence Inequality) Let f and g be two pdf's defined on $\Re^n$. Then

$$D(f\|g) \geq 0,$$

with equality if and only if $f = g$ a.e.

10.6 Maximum Differential Entropy Distributions

2.9 Maximum Entropy Distributions

Consider the maximization problem:

Maximize $H(p)$ over all probability distributions p defined on a countable subset $\mathcal{S}$ of the set of real numbers, subject to

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \le i \le m, \tag{1}$$

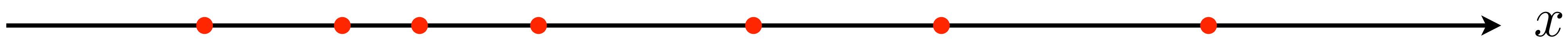
where $\mathcal{S}_p \subset \mathcal{S}$ and $r_i(x)$ is defined for all $x \in \mathcal{S}$.

Consider the maximization problem:

Maximize $H(p)$ over all probability distributions p defined on a countable subset $\mathcal{S}$ of the set of real numbers, subject to

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \le i \le m, \tag{1}$$

where $\mathcal{S}_p \subset \mathcal{S}$ and $r_i(x)$ is defined for all $x \in \mathcal{S}$.

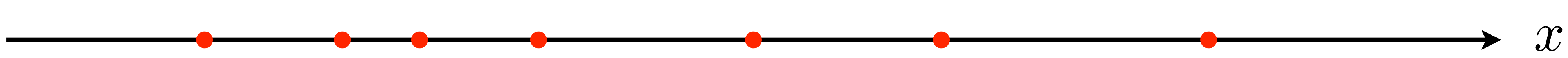


Consider the maximization problem:

Maximize $H(p)$ over all probability distributions p defined on a countable subset $\mathcal{S}$ of the set of real numbers, subject to

$$E_p\, r_i(X) = \sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \le i \le m, \tag{1}$$

where $\mathcal{S}_p \subset \mathcal{S}$ and $r_i(x)$ is defined for all $x \in \mathcal{S}$.



Consider the maximization problem:

Maximize $H(p)$ over all probability distributions p defined on a countable subset $\mathcal{S}$ of the set of real numbers, subject to

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \le i \le m, \tag{1}$$

where $\mathcal{S}_p \subset \mathcal{S}$ and $r_i(x)$ is defined for all $x \in \mathcal{S}$.

Consider the maximization problem:

Maximize $H(p)$ over all probability distributions p defined on a countable subset $\mathcal{S}$ of the set of real numbers, subject to

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \leq i \leq m, \tag{1}$$

where $\mathcal{S}_p \subset \mathcal{S}$ and $r_i(x)$ is defined for all $x \in \mathcal{S}$.

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that the constraints in (1) are satisfied. Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$, subject to the constraints in (1).

Consider the maximization problem:

> Maximize $H(p)$ over all probability distributions p defined on a countable subset $\mathcal{S}$ of the set of real numbers, subject to
>
> $$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \leq i \leq m, \tag{1}$$
>
> where $\mathcal{S}_p \subset \mathcal{S}$ and $r_i(x)$ is defined for all $x \in \mathcal{S}$.

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that the constraints in (1) are satisfied. Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$, subject to the constraints in (1).

Remark Let $q_i = e^{-\lambda_i}$. Then we can write

Consider the maximization problem:

> Maximize $H(p)$ over all probability distributions p defined on a countable subset $\mathcal{S}$ of the set of real numbers, subject to
>
> $$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \leq i \leq m, \tag{1}$$
>
> where $\mathcal{S}_p \subset \mathcal{S}$ and $r_i(x)$ is defined for all $x \in \mathcal{S}$.

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^m \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that the constraints in (1) are satisfied. Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$, subject to the constraints in (1).

Remark Let $q_i = e^{-\lambda_i}$. Then we can write

$$p^*(x) \quad = \quad e^{-\lambda_0} \, e^{-\lambda_1 r_1(x)} \, \cdots \, e^{-\lambda_m r_m(x)}$$

Consider the maximization problem:

> Maximize $H(p)$ over all probability distributions p defined on a countable subset $\mathcal{S}$ of the set of real numbers, subject to
>
> $$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \le i \le m, \tag{1}$$
>
> where $\mathcal{S}_p \subset \mathcal{S}$ and $r_i(x)$ is defined for all $x \in \mathcal{S}$.

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^m \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that the constraints in (1) are satisfied. Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$, subject to the constraints in (1).

Remark Let $q_i = e^{-\lambda_i}$. Then we can write

$$\begin{aligned} p^*(x) &= e^{-\lambda_0} \, e^{-\lambda_1 r_1(x)} \, \cdots \, e^{-\lambda_m r_m(x)} \\ &= q_0 \, {q_1}^{r_1(x)} \, \cdots \, {q_m}^{r_m(x)} \end{aligned}$$

Consider the maximization problem:

> Maximize $H(p)$ over all probability distributions p defined on a countable subset $\mathcal{S}$ of the set of real numbers, subject to
>
> $$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \le i \le m, \tag{1}$$
>
> where $\mathcal{S}_p \subset \mathcal{S}$ and $r_i(x)$ is defined for all $x \in \mathcal{S}$.

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that the constraints in (1) are satisfied. Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$, subject to the constraints in (1).

Remark Let $q_i = e^{-\lambda_i}$. Then we can write

$$\begin{aligned} p^*(x) &= e^{-\lambda_0} \, e^{-\lambda_1 r_1(x)} \cdots e^{-\lambda_m r_m(x)} \\ &= q_0 \, {q_1}^{r_1(x)} \cdots {q_m}^{r_m(x)} \end{aligned}$$

where q_0 is the normalization constant.

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \leq i \leq m. \qquad (1)$$

Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$ subject to (1).

Sketch of Proof

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \leq i \leq m. \tag{1}$$

Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$ subject to (1).

Sketch of Proof

$$H(p^*) - H(p)$$

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \leq i \leq m. \qquad (1)$$

Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$ subject to (1).

Sketch of Proof

$$H(p^*) - H(p)$$
$$= \quad -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)$$

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \le i \le m. \qquad (1)$$

Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$ subject to (1).

Sketch of Proof

$$\begin{aligned} &H(p^*) - H(p) \\ &\quad = \quad -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\ &\quad = \quad -\sum_{x \in \mathcal{S}_p} p(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \end{aligned}$$

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^m \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \leq i \leq m. \qquad (1)$$

Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$ subject to (1).

Sketch of Proof

$$\begin{aligned} & H(p^*) - H(p) \\ &= -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\ &= -\sum_{x \in \mathcal{S}_p} p(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\ &= \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{p^*(x)} \end{aligned}$$

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \leq i \leq m. \qquad (1)$$

Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$ subject to (1).

Sketch of Proof

$$\begin{aligned} H(p^*) - H(p) & \\ &= -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\ &= -\sum_{x \in \mathcal{S}_p} p(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\ &= \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{p^*(x)} \\ &= D(p \| p^*) \end{aligned}$$

Theorem 2.50 Let

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \leq i \leq m. \qquad (1)$$

Then p^* maximizes $H(p)$ over all probability distribution p on $\mathcal{S}$ subject to (1).

Sketch of Proof

$$\begin{aligned} & H(p^*) - H(p) \\ &= -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\ &= -\sum_{x \in \mathcal{S}_p} p(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\ &= \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{p^*(x)} \\ &= D(p \| p^*) \\ &\geq 0. \end{aligned}$$

Proof of Theorem 2.50

Proof of Theorem 2.50

$H(p^*) - H(p)$

Proof of Theorem 2.50

$$H(p^*) - H(p)$$
$$= \; -\sum_{x\in\mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x)$$

Proof of Theorem 2.50

$$H(p^*) - H(p)$$

$$= \quad -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)$$

$$p^*(x) \quad = \quad e^{-\lambda_0 - \sum_i \lambda_i r_i(x)}$$

Proof of Theorem 2.50

$$H(p^*) - H(p)$$

$$= -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)$$

$$p^*(x) = e^{-\lambda_0 - \sum_i \lambda_i r_i(x)}$$

$$\ln p^*(x) = -\lambda_0 - \sum_i \lambda_i r_i(x)$$

Proof of Theorem 2.50

$$H(p^*) - H(p)$$
$$= -\sum_{x\in\mathcal{S}} p^*(x) \underline{\ln p^*(x)} + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)$$

$$\begin{aligned} p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\ \ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x) \end{aligned}$$

Proof of Theorem 2.50

$$
\begin{aligned}
&H(p^*) - H(p) \\
&\quad = -\sum_{x\in\mathcal{S}} p^*(x)\,\underline{\ln p^*(x)} + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&\quad = -\sum_{x\in\mathcal{S}} p^*(x)\left(\underline{-\lambda_0 - \sum_i \lambda_i r_i(x)}\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) \\
&= -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= -\sum_{x \in \mathcal{S}} p^*(x) \left(-\underline{\lambda_0} - \sum_i \lambda_i r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$\begin{aligned}
H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\underline{\lambda_0} - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \underline{\lambda_0}\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i \left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}$$

$$\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}$$

Proof of Theorem 2.50

$$
\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i \left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x) \left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i \left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) \\
&= -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= -\underline{\sum_{x \in \mathcal{S}} p^*(x)} \left(-\lambda_0 - \sum_i \lambda_i r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \left(\sum_{x \in \mathcal{S}} p^*(x) \right) + \sum_i \lambda_i \left(\underline{\sum_{x \in \mathcal{S}} p^*(x)} r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x) \left(-\lambda_0 - \sum_i \lambda_i \underline{r_i(x)}\right) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i \left(\sum_{x\in\mathcal{S}} p^*(x) \underline{r_i(x)}\right) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) & \\
= & -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
= & -\sum_{x \in \mathcal{S}} p^*(x) \left(-\lambda_0 - \sum_i \lambda_i r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
= & \ \lambda_0 \left(\underline{\sum_{x \in \mathcal{S}} p^*(x)} \right) + \sum_i \lambda_i \left(\sum_{x \in \mathcal{S}} p^*(x) r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
&H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0 \left(\underline{\sum_{x\in\mathcal{S}} p^*(x)}\right) + \sum_i \lambda_i \left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0 \cdot \underline{1} + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
&H(p^*) - H(p)\\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)\\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)\\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\underline{\sum_{x\in\mathcal{S}} p^*(x) r_i(x)}\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)\\
&= \lambda_0\cdot 1 + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)}\\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) \\
&= -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= -\sum_{x \in \mathcal{S}} p^*(x) \left(-\lambda_0 - \sum_i \lambda_i r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \left(\sum_{x \in \mathcal{S}} p^*(x) \right) + \sum_i \lambda_i \left(\underline{\sum_{x \in \mathcal{S}} p^*(x) r_i(x)} \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \cdot 1 + \sum_i \lambda_i \underline{a_i} + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= -\sum_{x \in \mathcal{S}} p^*(x) \left(-\lambda_0 - \sum_i \lambda_i r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \left(\sum_{x \in \mathcal{S}} p^*(x) \right) + \sum_i \lambda_i \left(\sum_{x \in \mathcal{S}} p^*(x) r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \cdot 1 + \sum_i \lambda_i a_i + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x) \left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i \left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \cdot \underline{1} + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) &- H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\cdot\underline{1} + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\underline{\sum_{x\in\mathcal{S}_p} p(x)}\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}_p} p(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) & \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\cdot 1 + \sum_i \lambda_i \underline{a_i} + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}_p} p(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}_p} p(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) & \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\cdot 1 + \sum_i \lambda_i \underline{a_i} + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}_p} p(x)\right) + \sum_i \lambda_i\left(\underline{\sum_{x\in\mathcal{S}_p} p(x) r_i(x)}\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0 \cdot 1 + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{\underline{x\in\mathcal{S}_p}} p(x)\right) + \sum_i \lambda_i\left(\sum_{\underline{x\in\mathcal{S}_p}} p(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{\underline{x\in\mathcal{S}_p}} p(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\cdot 1 + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}_p} \underline{p(x)}\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}_p} \underline{p(x)} r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}_p} \underline{p(x)}\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= -\sum_{x \in \mathcal{S}} p^*(x) \left(-\lambda_0 - \sum_i \lambda_i r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \left(\sum_{x \in \mathcal{S}} p^*(x) \right) + \sum_i \lambda_i \left(\sum_{x \in \mathcal{S}} p^*(x) r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \cdot 1 + \sum_i \lambda_i a_i + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= \lambda_0 \left(\sum_{x \in \mathcal{S}_p} p(x) \right) + \sum_i \lambda_i \left(\sum_{x \in \mathcal{S}_p} p(x) r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
&= -\sum_{x \in \mathcal{S}_p} p(x) \left(-\lambda_0 - \sum_i \lambda_i r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\cdot 1 + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \underline{\lambda_0}\left(\sum_{x\in\mathcal{S}_p} p(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}_p} p(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \underline{-}\sum_{x\in\mathcal{S}_p} p(x)\left(-\underline{\lambda_0} - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i \left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0 \cdot 1 + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}_p} p(x)\right) + \underline{\sum_i \lambda_i} \left(\sum_{x\in\mathcal{S}_p} p(x) \underline{r_i(x)}\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \underline{-}\sum_{x\in\mathcal{S}_p} p(x)\left(-\lambda_0 \underline{- \sum_i \lambda_i r_i(x)}\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}$$

$$\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) & \\
= & -\sum_{x \in \mathcal{S}} p^*(x) \ln p^*(x) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
= & -\sum_{x \in \mathcal{S}} p^*(x) \left(-\lambda_0 - \sum_i \lambda_i r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
= & \lambda_0 \left(\sum_{x \in \mathcal{S}} p^*(x) \right) + \sum_i \lambda_i \left(\sum_{x \in \mathcal{S}} p^*(x) r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
= & \lambda_0 \cdot 1 + \sum_i \lambda_i a_i + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
= & \lambda_0 \left(\sum_{x \in \mathcal{S}_p} p(x) \right) + \sum_i \lambda_i \left(\sum_{x \in \mathcal{S}_p} p(x) r_i(x) \right) + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x) \\
= & -\sum_{x \in \mathcal{S}_p} p(x) \underline{\left(-\lambda_0 - \sum_i \lambda_i r_i(x) \right)} + \sum_{x \in \mathcal{S}_p} p(x) \ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0 \cdot 1 + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}_p} p(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}_p} p(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}_p} p(x)\underline{\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right)} + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}_p} p(x)\,\underline{\ln p^*(x)} + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
& H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0 \cdot 1 + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}_p} p(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}_p} p(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}_p} p(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}_p} p(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \sum_{x\in\mathcal{S}_p} p(x)\ln\frac{p(x)}{p^*(x)}
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
&H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0 \cdot 1 + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}_p} p(x)\right) + \sum_i \lambda_i\left(\sum_{x\in\mathcal{S}_p} p(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}_p} p(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}_p} p(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \sum_{x\in\mathcal{S}_p} p(x)\ln\frac{p(x)}{p^*(x)} \\
&= D(p\|p^*)
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Proof of Theorem 2.50

$$
\begin{aligned}
H(p^*) - H(p) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}} p^*(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}} p^*(x)\right) + \sum_i \lambda_i \left(\sum_{x\in\mathcal{S}} p^*(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0 \cdot 1 + \sum_i \lambda_i a_i + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \lambda_0\left(\sum_{x\in\mathcal{S}_p} p(x)\right) + \sum_i \lambda_i \left(\sum_{x\in\mathcal{S}_p} p(x) r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}_p} p(x)\left(-\lambda_0 - \sum_i \lambda_i r_i(x)\right) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= -\sum_{x\in\mathcal{S}_p} p(x)\ln p^*(x) + \sum_{x\in\mathcal{S}_p} p(x)\ln p(x) \\
&= \sum_{x\in\mathcal{S}_p} p(x)\ln \frac{p(x)}{p^*(x)} \\
&= D(p\|p^*) \\
&\geq 0.
\end{aligned}
$$

$$
\begin{aligned}
p^*(x) &= e^{-\lambda_0 - \sum_i \lambda_i r_i(x)} \\
\ln p^*(x) &= -\lambda_0 - \sum_i \lambda_i r_i(x)
\end{aligned}
$$

Corollary 2.51 Let p^* be a probability distribution defined on $\mathcal{S}$ with

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$. Then p^* maximizes $H(p)$ over all probability distribution p defined on $\mathcal{S}$, subject to the constraints

Corollary 2.51 Let p^* be a probability distribution defined on $\mathcal{S}$ with

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^m \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$. Then p^* maximizes $H(p)$ over all probability distribution p defined on $\mathcal{S}$, subject to the constraints

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = \sum_{x \in \mathcal{S}} p^*(x) r_i(x) \quad \text{for } 1 \leq i \leq m.$$

Corollary 2.51 Let p^* be a probability distribution defined on $\mathcal{S}$ with

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$. Then p^* maximizes $H(p)$ over all probability distribution p defined on $\mathcal{S}$, subject to the constraints

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = \sum_{x \in \mathcal{S}} p^*(x) r_i(x) \quad \text{for } 1 \le i \le m.$$

Proof

Corollary 2.51 Let p^* be a probability distribution defined on $\mathcal{S}$ with

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$. Then p^* maximizes $H(p)$ over all probability distribution p defined on $\mathcal{S}$, subject to the constraints

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = \sum_{x \in \mathcal{S}} p^*(x) r_i(x) \quad \text{for } 1 \leq i \leq m.$$

Proof

1. Let $\sum_{x \in \mathcal{S}} p^*(x) r_i(x) = a_i$ for $1 \leq i \leq m$.

Corollary 2.51 Let p^* be a probability distribution defined on $\mathcal{S}$ with

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$. Then p^* maximizes $H(p)$ over all probability distribution p defined on $\mathcal{S}$, subject to the constraints

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = \underbrace{\sum_{x \in \mathcal{S}} p^*(x) r_i(x)}_{a_i} \quad \text{for } 1 \le i \le m.$$

Proof

1. Let $\sum_{x \in \mathcal{S}} p^*(x) r_i(x) = a_i$ for $1 \le i \le m$.

Corollary 2.51 Let p^* be a probability distribution defined on $\mathcal{S}$ with

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$. Then p^* maximizes $H(p)$ over all probability distribution p defined on $\mathcal{S}$, subject to the constraints

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = \underbrace{\sum_{x \in \mathcal{S}} p^*(x) r_i(x)}_{a_i} \quad \text{for } 1 \le i \le m.$$

Proof

1. Let $\sum_{x \in \mathcal{S}} p^*(x) r_i(x) = a_i$ for $1 \le i \le m$.

2. Obviously, $\lambda_0, \lambda_1, \cdots, \lambda_m$ are such that p^* satisfies the constraints

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \le i \le m.$$

Corollary 2.51 Let p^* be a probability distribution defined on $\mathcal{S}$ with

$$p^*(x) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(x)}$$

for all $x \in \mathcal{S}$. Then p^* maximizes $H(p)$ over all probability distribution p defined on $\mathcal{S}$, subject to the constraints

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = \underbrace{\sum_{x \in \mathcal{S}} p^*(x) r_i(x)}_{a_i} \quad \text{for } 1 \le i \le m.$$

Proof

1. Let $\sum_{x \in \mathcal{S}} p^*(x) r_i(x) = a_i$ for $1 \le i \le m$.
2. Obviously, $\lambda_0, \lambda_1, \cdots, \lambda_m$ are such that p^* satisfies the constraints

$$\sum_{x \in \mathcal{S}_p} p(x) r_i(x) = a_i \quad \text{for } 1 \le i \le m.$$

3. Then the corollary is implied by Theorem 2.50.

Example 2.52 Let $\mathcal{S}$ be finite and let the set of constraints be empty. Then

$$p^*(x) = e^{-\lambda_0},$$

a constant that does not depend on x. Therefore, p^* is simply the uniform distribution over $\mathcal{S}$, i.e., $p^*(x) = |\mathcal{S}|^{-1}$ for all $x \in \mathcal{S}$.

Example 2.53

Example 2.53

1. Let $\mathcal{S} = \{0, 1, 2, \cdots\}$, and let the set of constraints be

$$\sum_x p(x)x = a \tag{1}$$

where $a \geq 0$, i.e., the mean of the distribution p is fixed.

Example 2.53

1. Let $\mathcal{S} = \{0, 1, 2, \cdots\}$, and let the set of constraints be

$$\sum_x p(x)x = a \tag{1}$$

where $a \geq 0$, i.e., the mean of the distribution p is fixed.

2. Let $q_i = e^{-\lambda_i}$ for $i = 0, 1$. Then

$$p^*(x) = e^{-\lambda_0} e^{-\lambda_1 x} = q_0 q_1^x .$$

Example 2.53

1. Let $\mathcal{S} = \{0, 1, 2, \cdots\}$, and let the set of constraints be

$$\sum_x p(x)x = a \tag{1}$$

where $a \geq 0$, i.e., the mean of the distribution p is fixed.

2. Let $q_i = e^{-\lambda_i}$ for $i = 0, 1$. Then

$$p^*(x) = e^{-\lambda_0} e^{-\lambda_1 x} = q_0 q_1^x.$$

3. Evidently, p^* is a geometric distribution, so that

$$q_1 = 1 - q_0.$$

Example 2.53

1. Let $\mathcal{S} = \{0, 1, 2, \cdots\}$, and let the set of constraints be

$$\sum_x p(x)x = a \tag{1}$$

where $a \geq 0$, i.e., the mean of the distribution p is fixed.

2. Let $q_i = e^{-\lambda_i}$ for $i = 0, 1$. Then

$$p^*(x) = e^{-\lambda_0} e^{-\lambda_1 x} = q_0 q_1^x.$$

3. Evidently, p^* is a geometric distribution, so that

$$q_1 = 1 - q_0.$$

4. Finally, we invoke the constraint (1) on p to obtain

$$q_0 = (a + 1)^{-1}.$$

10.6 Maximum Differential Entropy Distributions

Consider the maximization problem:

Maximize $h(f)$ over all pdf f defined on a subset $\mathcal{S}$ of $\Re^n$, subject to

$$\int_{\mathcal{S}_f} r_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = a_i \quad \text{for } 1 \le i \le m \tag{1}$$

where $\mathcal{S}_f \subset \mathcal{S}$ and $r_i(\mathbf{x})$ is defined for all $\mathbf{x} \in \mathcal{S}$.

Consider the maximization problem:

Maximize $h(f)$ over all pdf f defined on a subset $\mathcal{S}$ of $\Re^n$, subject to

$$E_f\, r_i(X) = \int_{\mathcal{S}_f} r_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = a_i \quad \text{for } 1 \le i \le m \tag{1}$$

where $\mathcal{S}_f \subset \mathcal{S}$ and $r_i(\mathbf{x})$ is defined for all $\mathbf{x} \in \mathcal{S}$.

Consider the maximization problem:

Maximize $h(f)$ over all pdf f defined on a subset $\mathcal{S}$ of $\Re^n$, subject to

$$\int_{\mathcal{S}_f} r_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = a_i \quad \text{for } 1 \leq i \leq m \tag{1}$$

where $\mathcal{S}_f \subset \mathcal{S}$ and $r_i(\mathbf{x})$ is defined for all $\mathbf{x} \in \mathcal{S}$.

Theorem 10.41 Let

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i=1}^{m} \lambda_i r_i(\mathbf{x})}$$

for all $\mathbf{x} \in \mathcal{S}$, where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are chosen such that the constraints in (1) are satisfied. Then f^* maximizes $h(f)$ over all pdf f defined on $\mathcal{S}$, subject to the constraints in (1).

Corollary 10.42 Let f^* be a pdf defined on $\mathcal{S}$ with

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i=1}^m \lambda_i r_i(\mathbf{x})}$$

for all $\mathbf{x} \in \mathcal{S}$. Then f^* maximizes $h(f)$ over all pdf f defined on $\mathcal{S}$, subject to the constraints

$$\int_{\mathcal{S}_f} r_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{S}} r_i(\mathbf{x}) f^*(\mathbf{x}) d\mathbf{x} \quad \text{for } 1 \leq i \leq m.$$

Theorem 10.43 Let X be a continuous random variable with $EX^2 = \kappa$. Then

$$h(X) \leq \frac{1}{2} \log(2\pi e \kappa),$$

with equality if and only if $X \sim \mathcal{N}(0, \kappa)$.

Proof

Theorem 10.43 Let X be a continuous random variable with $EX^2 = \kappa$. Then

$$h(X) \le \frac{1}{2}\log(2\pi e\kappa),$$

with equality if and only if $X \sim \mathcal{N}(0, \kappa)$.

Proof

1. Maximize $h(f)$ subject to the constraint

$$\int x^2 f(x)dx = EX^2 = \kappa.$$

Theorem 10.43 Let X be a continuous random variable with $EX^2 = \kappa$. Then

$$h(X) \leq \frac{1}{2}\log(2\pi e\kappa),$$

with equality if and only if $X \sim \mathcal{N}(0, \kappa)$.

Proof

1. Maximize $h(f)$ subject to the constraint

$$\int x^2 f(x)dx = EX^2 = \kappa.$$

2. Then by Theorem 10.41, $f^*(x) = ae^{-bx^2}$, which is the Gaussian distribution with zero mean.

Theorem 10.43 Let X be a continuous random variable with $EX^2 = \kappa$. Then

$$h(X) \leq \frac{1}{2} \log(2\pi e\kappa),$$

with equality if and only if $X \sim \mathcal{N}(0, \kappa)$.

Proof

1. Maximize $h(f)$ subject to the constraint

$$\int x^2 f(x)dx = EX^2 = \kappa.$$

2. Then by Theorem 10.41, $f^*(x) = ae^{-bx^2}$, which is the Gaussian distribution with zero mean.

3. In order to satisfy the second moment constraint, the only choices are

$$a = \frac{1}{\sqrt{2\pi\kappa}} \quad \text{and} \quad b = \frac{1}{2\kappa}.$$

An Application of Corollary 10.42

An Application of Corollary 10.42

1. Consider the pdf of $\mathcal{N}(0, \sigma^2)$:

$$f^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

An Application of Corollary 10.42

1. Consider the pdf of $\mathcal{N}(0, \sigma^2)$:

$$f^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

2. Write

$$f^*(x) = e^{-\lambda_0} e^{-\lambda_1 x^2},$$

An Application of Corollary 10.42

1. Consider the pdf of $\mathcal{N}(0, \sigma^2)$:

$$f^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

2. Write

$$f^*(x) = e^{-\lambda_0} e^{-\lambda_1 x^2},$$

An Application of Corollary 10.42

1. Consider the pdf of $\mathcal{N}(0, \sigma^2)$:

$$f^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

2. Write

$$f^*(x) = e^{-\lambda_0} e^{-\lambda_1 x^2},$$

An Application of Corollary 10.42

1. Consider the pdf of $\mathcal{N}(0, \sigma^2)$:

$$f^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

2. Write

$$f^*(x) = e^{-\lambda_0} e^{-\lambda_1 x^2},$$

where

$$\lambda_0 = \frac{1}{2}\ln(2\pi\sigma^2) \quad \text{and} \quad \lambda_1 = \frac{1}{2\sigma^2}.$$

An Application of Corollary 10.42

1. Consider the pdf of $\mathcal{N}(0, \sigma^2)$:

$$f^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

2. Write

$$f^*(x) = e^{-\lambda_0} e^{-\lambda_1 x^2},$$

where

$$\lambda_0 = \frac{1}{2} \ln(2\pi\sigma^2) \quad \text{and} \quad \lambda_1 = \frac{1}{2\sigma^2}.$$

An Application of Corollary 10.42

1. Consider the pdf of $\mathcal{N}(0, \sigma^2)$:

$$f^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

2. Write

$$f^*(x) = e^{-\lambda_0} e^{-\lambda_1 x^2},$$

where

$$\lambda_0 = \frac{1}{2}\ln(2\pi\sigma^2) \quad \text{and} \quad \lambda_1 = \frac{1}{2\sigma^2}.$$

3. Then f^* maximizes $h(f)$ over all f subject to

$$\int x^2 f(x)dx = \int x^2 f^*(x)dx = EX^2 = \sigma^2.$$

Theorem 10.44 Let X be a continuous random variable with mean μ and variance σ^2. Then

$$h(X) \leq \frac{1}{2}\log(2\pi e\sigma^2)$$

with equality if and only if $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proof

Theorem 10.44 Let X be a continuous random variable with mean μ and variance σ^2. Then

$$h(X) \leq \frac{1}{2} \log(2\pi e \sigma^2)$$

with equality if and only if $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proof

1. Let $X' = X - \mu$.

Theorem 10.44 Let X be a continuous random variable with mean μ and variance σ^2. Then

$$h(X) \leq \frac{1}{2} \log(2\pi e \sigma^2)$$

with equality if and only if $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proof

1. Let $X' = X - \mu$.

2. Then $EX' = 0$ and $E(X')^2 = E(X - \mu)^2 = \text{var}X = \sigma^2$.

Theorem 10.44 Let X be a continuous random variable with mean μ and variance σ^2. Then

$$h(X) \leq \frac{1}{2}\log(2\pi e\sigma^2)$$

with equality if and only if $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proof

1. Let $X' = X - \mu$.
2. Then $EX' = 0$ and $E(X')^2 = E(X-\mu)^2 = \text{var}X = \sigma^2$.
3. By Theorem 10.14 (Translation) and then Theorem 10.43,

$$h(X) = h(X') \leq \frac{1}{2}\log(2\pi e\sigma^2).$$

Theorem 10.44 Let X be a continuous random variable with mean μ and variance σ^2. Then

$$h(X) \leq \frac{1}{2} \log(2\pi e \sigma^2)$$

with equality if and only if $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proof

1. Let $X' = X - \mu$.
2. Then $EX' = 0$ and $E(X')^2 = E(X-\mu)^2 = \mathrm{var} X = \sigma^2$.
3. By Theorem 10.14 (Translation) and then Theorem 10.43,

$$h(X) = h(X') \leq \frac{1}{2} \log(2\pi e \sigma^2).$$

Theorem 10.44 Let X be a continuous random variable with mean μ and variance σ^2. Then

$$h(X) \leq \frac{1}{2}\log(2\pi e\sigma^2)$$

with equality if and only if $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proof

1. Let $X' = X - \mu$.
2. Then $EX' = 0$ and $E(X')^2 = E(X-\mu)^2 = \text{var}X = \sigma^2$.
3. By Theorem 10.14 (Translation) and then Theorem 10.43,

$$h(X) = h(X') \leq \frac{1}{2}\log(2\pi e\sigma^2).$$

Theorem 10.44 Let X be a continuous random variable with mean μ and variance σ^2. Then

$$h(X) \leq \frac{1}{2}\log(2\pi e\sigma^2)$$

with equality if and only if $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proof

1. Let $X' = X - \mu$.
2. Then $EX' = 0$ and $E(X')^2 = E(X - \mu)^2 = \text{var}X = \sigma^2$.
3. By Theorem 10.14 (Translation) and then Theorem 10.43,

$$h(X) = h(X') \leq \frac{1}{2}\log(2\pi e\sigma^2).$$

Theorem 10.44 Let X be a continuous random variable with mean μ and variance σ^2. Then

$$h(X) \leq \frac{1}{2} \log(2\pi e \sigma^2)$$

with equality if and only if $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proof

1. Let $X' = X - \mu$.
2. Then $EX' = 0$ and $E(X')^2 = E(X - \mu)^2 = \mathrm{var}X = \sigma^2$.
3. By Theorem 10.14 (Translation) and then Theorem 10.43,

$$h(X) = h(X') \leq \frac{1}{2} \log(2\pi e \sigma^2).$$

4. Equality holds if and only if $X' \sim \mathcal{N}(0, \sigma^2)$, or $X \sim \mathcal{N}(\mu, \sigma^2)$.

A Remark

A Remark

- Theorem 10.43 says that with the constraint $EX^2 = \kappa$, the differential entropy is maximized by the distribution $\mathcal{N}(0, \kappa)$.

A Remark

- Theorem 10.43 says that with the constraint $EX^2 = \kappa$, the differential entropy is maximized by the distribution $\mathcal{N}(0, \kappa)$.

- If we impose the additional constraint that $EX = 0$, then

A Remark

- Theorem 10.43 says that with the constraint $EX^2 = \kappa$, the differential entropy is maximized by the distribution $\mathcal{N}(0, \kappa)$.

- If we impose the additional constraint that $EX = 0$, then

$$\text{var}X = EX^2 = \kappa.$$

A Remark

- Theorem 10.43 says that with the constraint $EX^2 = \kappa$, the differential entropy is maximized by the distribution $\mathcal{N}(0, \kappa)$.

- If we impose the additional constraint that $EX = 0$, then

$$\text{var}X = EX^2 = \kappa.$$

- By Theorem 10.44, the differential entropy is still maximized by $\mathcal{N}(0, \kappa)$.

Differential Entropy and Spread

Differential Entropy and Spread

1. From Theorem 10.44, we have

$$h(X) \leq \frac{1}{2} \log(2\pi e \sigma^2) = \log \sigma + \frac{1}{2} \log(2\pi e)$$

where $\sigma^2 = \mathrm{var} X$.

Differential Entropy and Spread

1. From Theorem 10.44, we have

$$h(X) \leq \frac{1}{2} \log(2\pi e \sigma^2) = \log \sigma + \frac{1}{2} \log(2\pi e)$$

   where $\sigma^2 = \text{var}X$.

2. $h(X)$ is at most equal to the logarithm of the standard deviation ("spread") plus a constant.

Differential Entropy and Spread

1. From Theorem 10.44, we have

$$h(X) \leq \frac{1}{2}\log(2\pi e\sigma^2) = \log\sigma + \frac{1}{2}\log(2\pi e)$$

where $\sigma^2 = \text{var}X$.

2. $h(X)$ is at most equal to the logarithm of the standard deviation ("spread") plus a constant.

Differential Entropy and Spread

1. From Theorem 10.44, we have

$$h(X) \leq \frac{1}{2}\log(2\pi e\sigma^2) = \log\sigma + \frac{1}{2}\log(2\pi e)$$

   where $\sigma^2 = \mathrm{var}X$.

2. $h(X)$ is at most equal to the logarithm of the standard deviation ("spread") plus a constant.

3. $h(X) \to -\infty$ as $\sigma \to 0$.

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

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

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

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

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Theorem 10.46 Let $\mathbf{X}$ be a vector of n continuous random variables with mean $\boldsymbol{\mu}$ and covariance matrix K. Then

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

with equality if and only if $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$.

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.
2. Then the constraints on $f(\mathbf{x})$ are equivalent to

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.
2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\tilde{k}_{ij} = \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.
2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\tilde{k}_{ij} = \int_{\mathcal{S}_f} \underline{r_{ij}(\mathbf{x})} f(\mathbf{x}) d\mathbf{x}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} \underline{r_{ij}(\mathbf{x})} f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} \underline{x_i x_j} f(\mathbf{x}) d\mathbf{x} \end{aligned}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.
2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.
2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.
2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned}
\tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\
&= E X_i X_j
\end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \underline{\sum_{i,j} \lambda_{ij} x_i x_j}} = e^{-\lambda_0 - \underline{\mathbf{x}^\top L \mathbf{x}}},$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \underline{\sum_{i,j} \lambda_{ij} x_i x_j}} = e^{-\lambda_0 - \underline{\mathbf{x}^\top L \mathbf{x}}},$$

where $L = [\lambda_{ij}]$.

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \quad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top K^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\mathrm{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top K^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\text{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\text{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\text{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top K^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\text{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top K^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\mathrm{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top K^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \quad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned}
\tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\
&= E X_i X_j
\end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\operatorname{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2}\log\left[(2\pi e)^n |\tilde{K}|\right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\text{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}|\right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\operatorname{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\mathrm{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

6. Accordingly, λ_0 and L have the unique solution given by

$$e^{-\lambda_0} = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}}$$

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \le \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \le i, j \le n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0} {}^{-\mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \le i, j \le n$,

$$\mathrm{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

6. Accordingly, λ_0 and L have the unique solution given by

$$e^{-\lambda_0} = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}}$$

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \tag{1}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\operatorname{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

6. Accordingly, λ_0 and L have the unique solution given by

$$e^{-\lambda_0} = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}}$$

and

$$L = \frac{1}{2} \tilde{K}^{-1},$$

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \le \frac{1}{2}\log\left[(2\pi e)^n|\tilde{K}|\right], \quad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned}\tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= EX_iX_j\end{aligned}$$

for $1 \le i, j \le n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j}\lambda_{ij}x_ix_j} = e^{-\lambda_0 - \mathbf{x}^\top L\mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \le i, j \le n$,

$$\text{cov}(X_i, X_j) = EX_iX_j - (EX_i)(EX_j) = EX_iX_j.$$

Hence, $K_{f^*} = \tilde{K}$.

6. Accordingly, λ_0 and L have the unique solution given by

$$e^{-\lambda_0} = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}}$$

and

$$L = \frac{1}{2}\tilde{K}^{-1},$$

so that

$$f^*(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}} e^{-\frac{1}{2}\mathbf{x}^\top \tilde{K}^{-1}\mathbf{x}},$$

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \le \frac{1}{2}\log\left[(2\pi e)^n |\tilde{K}|\right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned}\tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j\end{aligned}$$

for $1 \le i, j \le n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \le i, j \le n$,

$$\mathrm{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

6. Accordingly, λ_0 and L have the unique solution given by

$$e^{-\lambda_0} = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}}$$

and

$$L = \frac{1}{2}\tilde{K}^{-1},$$

so that

$$f^*(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}} e^{-\frac{1}{2}\mathbf{x}^\top \tilde{K}^{-1}\mathbf{x}},$$

the joint pdf of $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2}\log\left[(2\pi e)^n |\tilde{K}|\right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= EX_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\mathrm{cov}(X_i, X_j) = EX_i X_j - (EX_i)(EX_j) = EX_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

6. Accordingly, λ_0 and L have the unique solution given by

$$e^{-\lambda_0} = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}}$$

and

$$L = \frac{1}{2}\tilde{K}^{-1},$$

so that

$$f^*(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}} e^{-\frac{1}{2}\mathbf{x}^\top \tilde{K}^{-1}\mathbf{x}},$$

the joint pdf of $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

7. Hence, by Theorem 10.20, we have proved (1) with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \leq \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \leq i, j \leq n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \leq i, j \leq n$,

$$\mathrm{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

6. Accordingly, λ_0 and L have the unique solution given by

$$e^{-\lambda_0} = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}}$$

and

$$L = \frac{1}{2} \tilde{K}^{-1},$$

so that

$$f^*(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}} e^{-\frac{1}{2} \mathbf{x}^\top \tilde{K}^{-1} \mathbf{x}},$$

the joint pdf of $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

7. Hence, by Theorem 10.20, we have proved (1) with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

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

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

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top K^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

Theorem 10.45 Let $\mathbf{X}$ be a vector of n continuous random variables with correlation matrix $\tilde{K}$. Then

$$h(\mathbf{X}) \le \frac{1}{2} \log \left[(2\pi e)^n |\tilde{K}| \right], \qquad (1)$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

Proof

1. Let $r_{ij}(\mathbf{x}) = x_i x_j$ and $\tilde{K} = [\tilde{k}_{ij}]$.

2. Then the constraints on $f(\mathbf{x})$ are equivalent to

$$\begin{aligned} \tilde{k}_{ij} &= \int_{\mathcal{S}_f} r_{ij}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{S}_f} x_i x_j f(\mathbf{x}) d\mathbf{x} \\ &= E X_i X_j \end{aligned}$$

for $1 \le i, j \le n$.

3. By Theorem 10.41, the joint pdf that maximizes $h(\mathbf{X})$ has the form

$$f^*(\mathbf{x}) = e^{-\lambda_0 - \sum_{i,j} \lambda_{ij} x_i x_j} = e^{-\lambda_0 - \mathbf{x}^\top L \mathbf{x}},$$

where $L = [\lambda_{ij}]$.

4. Thus f^* is the joint pdf of a multivariate Gaussian distribution with zero mean.

5. Then for $1 \le i, j \le n$,

$$\mathrm{cov}(X_i, X_j) = E X_i X_j - (E X_i)(E X_j) = E X_i X_j.$$

Hence, $K_{f^*} = \tilde{K}$.

6. Accordingly, λ_0 and L have the unique solution given by

$$e^{-\lambda_0} = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}}$$

and

$$L = \frac{1}{2} \tilde{K}^{-1},$$

so that

$$f^*(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |\tilde{K}|^{1/2}} e^{-\frac{1}{2} \mathbf{x}^\top \tilde{K}^{-1} \mathbf{x}},$$

the joint pdf of $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

7. Hence, by Theorem 10.20, we have proved (1) with equality if and only if $\mathbf{X} \sim \mathcal{N}(0, \tilde{K})$.

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

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

$\mathcal{N}(\boldsymbol{\mu}, K)$:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2} (\mathbf{x}-\boldsymbol{\mu})^\top K^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$