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

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

Remark The key step is to establish that

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

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

$$\int_{\mathcal{S}_f} r_i(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = a_i \quad \text{for } 1 \le i \le m. \qquad (2)$$

Then f^* maximizes $h(f)$ over all pdf f defined on $\mathcal{S}$, subject to the constraints in (2).

Remark The key step is to establish that

$$\int_{\mathcal{S}} f^*(\mathbf{x}) \ln f^*(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{S}_f} f(\mathbf{x}) \ln f^*(\mathbf{x}) d\mathbf{x}. \qquad (3)$$

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

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

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

Then (3) and Theorem 10.45 together imply

Lemma 11.34 Let $\mathbf{Y}^* \sim \mathcal{N}(0, K)$ and $\mathbf{Y}$ be any random vector with correlation matrix K. Then

$$\int f_{\mathbf{Y}^*}(\mathbf{y}) \log f_{\mathbf{Y}^*}(\mathbf{y}) d\mathbf{y} = \int_{\mathcal{S}_{\mathbf{Y}}} f_{\mathbf{Y}}(\mathbf{y}) \log f_{\mathbf{Y}^*}(\mathbf{y}) d\mathbf{y}.$$