1

z.z. $s \mapsto S_i(v)$ ist probability mass function.
bew.

$$\sum_{i=1}^{n} S_i(v) = \sum_{i=1}^{n} \frac{e^{v_i}}{\sum_{j=1}^{n} e^{v_j}} = \frac{\sum_{i=1}^{n} e^{v_i}}{\sum_{i=1}^{n} e^{v_i}} = 1$$

QED

z.z. $S_j(w) = S_j(v)$ für $w_i = v_i + c$
bew.

$$S_j(w) = \frac{e^{w_j}}{\sum_{i=1}^{n} e^{w_i}} = \frac{e^{x_j+c}}{\sum_{i=1}^{n} e^{v_i+c}} = \frac{e^c e^{x_j}}{\sum_{i=1}^{n} e^c e^{v_i}} = \frac{e^c e^{x_j}}{e^c \sum_{i=1}^{n} e^{v_i}} = \frac{e^{x_j}}{\sum_{i=1}^{n} e^{v_i}} = S_j(v)$$

QED

z.z. $\frac{\partial S_j(v)}{\partial v_i} = S_j(v)(\delta_{ij} - S_i(v))$
bew.

$$\begin{aligned}
\frac{\partial S_j(v)}{\partial v_i} &= \frac{\partial}{\partial v_i} \frac{e^{v_j}}{\sum_{k=1}^{n} e^{v_k}} \\
&= \frac{\left(\frac{\partial}{\partial v_i} e^{v_j}\right) \sum_{k=1}^{n} e^{v_k} - e^{v_j} \left(\frac{\partial}{\partial v_i} \sum_{k=1}^{n} e^{v_k}\right)}{\left(\sum_{k=1}^{n} e^{v_k}\right)^2} \\
&= \frac{\delta_{ij} e^{v_j} \sum_{k=1}^{n} e^{v_k} - e^{v_j} e^{v_i})}{\left(\sum_{k=1}^{n} e^{v_k}\right)^2} \\
&= \delta_{ij} S_j(v) - S_i(v) S_j(v) \\
&= S_j(v)(\delta_{ij} - S_i(v))
\end{aligned}$$

QED

2

Berechne $\frac{\partial H(t,q)}{\partial q_j} = \frac{\partial}{\partial q_j}\left(-\sum_{i=1}^{n} t_i log(q_i)\right) = -\frac{t_j}{q_j}$
z.z. $\frac{\partial H(t,\mathrm{softmax}(v))}{\partial v_i} = S_i(v) - t_i$
bew.

$$
\begin{aligned}
\frac{\partial H(t,\mathrm{softmax}(v))}{\partial v_i} &= \frac{\partial H(t,\mathrm{softmax}(v))}{\partial\,\mathrm{softmax}(v)}\frac{\partial\,\mathrm{softmax}(v)}{\partial v_i} \\
&= \left(-\frac{t_j}{S_j(v)}\right)_{j=1}^{n}{}^{\top}\left(S_j(v)(\delta_{ij} - S_i(v))\right)_{j=1}^{n} \\
&= \sum_{j=1}^{n} -\frac{t_j}{S_j(v)} S_j(v)(\delta_{ij} - S_i(v)) \\
&= \sum_{j=1}^{n} (t_j S_i(v) - t_j\delta_{ij}) \\
&= \sum_{j=1}^{n} t_j S_i(v) - t_i \\
&= \left(\sum_{j=1}^{n} t_j\right) S_i(v) - t_i \\
&= S_i(v) - t_i
\end{aligned}
$$

QED