Exercise: We derive a numerical trick for computing the Lagrange multipliers during maximum entropy inference.

(Source: Physics of Statistical Inference Course)

The maximum entropy probability distribution associated with fixed observables $\{\langle f_\mu \rangle\}_{\mu=1}^{K}$ and overall normalization $\sum_x P(x)=1$ is given by

$$p(x) = \frac{1}{Z} \exp\left(-\sum_{\mu=1}^{K} \lambda_\mu f_\mu(x)\right)$$

where $Z = e^{1+\lambda_0}$, and $\{\lambda_\mu\}_{\mu=0}^{K}$ are Lagrange multipliers.

We find a formula for these Lagrange multipliers.

Let $F_\mu \equiv \langle f_\mu \rangle$. Then, the entropy is given by

$$S(\{F_\mu\}) = -\sum_x p(x) \log p(x) =$$

$$\sum_x p(x)\left[1+\lambda_0+\sum_{\mu=1}^{K}\lambda_\mu f_\mu(x)\right] =$$

$$1+\lambda_0+\sum_{\mu=1}^{K}\lambda_\mu \sum_x p(x) f_\mu(x) = \underbrace{1+\lambda_0}_{\log Z}+\sum_{\mu=1}^{K}\lambda_\mu F_\mu.$$

We now compute

$$\frac{\partial S}{\partial F_\ell} = \frac{\partial \log Z}{\partial F_\ell} + \sum_\mu \left(\frac{\partial \lambda_\mu}{\partial F_\ell} F_\mu + \lambda_\mu \frac{\partial F_\mu}{\partial F_\ell} \right)$$

$$= \sum_\mu \left(\frac{\partial \log Z}{\partial \lambda_\mu} \frac{\partial \lambda_\mu}{\partial F_\ell} + F_\mu \frac{\partial \lambda_\mu}{\partial F_\ell} + \lambda_\mu \delta_{\mu\ell} \right).$$

We remark that for Z to be a normalization for $p(x)$ above, it must obey

$$Z = \sum_x \exp\left(-\sum_\mu \lambda_\mu f_\mu(x)\right) \Longrightarrow$$

$$\frac{\partial \log Z}{\partial \lambda_\mu} = -\frac{1}{Z} \sum_x f_\mu(x) \exp\left(-\sum_\mu \lambda_\mu f_\mu(x)\right) =$$

$$-\langle f_\mu \rangle = -F_\mu \Longrightarrow$$

$$\frac{\partial S}{\partial F_\ell} = \sum_\mu \left(-F_\mu \frac{\partial \lambda_\mu}{\partial F_\ell} + F_\mu \frac{\partial \lambda_\mu}{\partial F_\ell} \right) + \lambda_\ell$$

$$\Longrightarrow \boxed{\frac{\partial S}{\partial F_\ell} = \lambda_\ell}$$