1 Derivation 2 of Bernoulli distributions as an exponential family

Let X be your sample space:

$$X \in \{'heads', 'tails'\} \tag{1}$$

Let us define a sufficient statistic, and induce the natural parameters.
We define the sufficient statistic by an indicator function

$$u(x) = I(x =' heads') \tag{2}$$

Note for multinoulli, we can specify indicator functions for $K - 1$ classes, and obtain a sufficient statistic vector:

$$\begin{aligned} u_1(x) &= I(x =' cat') \\ u_2(x) &= I(x =' dog') \\ u_3(x) &= I(x =' hippo') \\ &\dots \end{aligned} \tag{3}$$

Now we define our intrinsic measure to restrict our samples to our discrete set of samples:

$$h(x) = \begin{cases} 1^1 & x =' heads' \, or \, x =' tails' \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The partition/normalization function can be calculated by integration/summing over X.

$$\begin{aligned} Z(\eta) &= \sum_X e^{u(x) \cdot \eta} \\ &= e^{0 \cdot \eta} + e^{1 \cdot \eta} \\ &= 1 + e^{\eta} \end{aligned} \tag{5}$$

Now we know for exponential families,

$$\nabla_\eta \log Z(\eta) = E_{x \sim p(x|\eta)} \left[u(x) \right] \tag{6}$$

Using calculus to simplify the left side

$$\begin{aligned} \frac{d}{d\eta} \log(1 + e^{\eta}) &= \frac{e^{\eta}}{1 + e^{\eta}} \\ &= \sigma(\eta) \end{aligned} \tag{7}$$

The right side is a sufficient statistic of the indicator function for heads

$$E_{x \sim p(x|\eta)} \left[I(x =' heads') \right] = Pr\left(X =' heads' \right) \tag{8}$$

Thus, we have usual relation for bernoulli that we use for logistic regression

$$Pr\left(X =' heads' \right) = \sigma(\eta) \tag{9}$$