a) Show that the Bernoulli distribution is a member of the exponential family, i.e. it can be written as $h(x)g(\eta)\exp(\eta u(x))$. What are $h(x)$, $g(\eta)$, and $u(x)$? Hint: use $\exp(\log())$.

$$p(x \mid \mu) = \mu^x (1 - \mu)^{1-x} = \exp(\log(\mu^x (1 - \mu)^{1-x})) = \exp(x \log \mu + (1 - x) \log(1 - \mu)) = \exp(x \log \mu + (1 - x) \log(1 - \mu)) = \exp(x \log \mu + \log(1 - \mu) - x \log(1 - \mu)) = (1 - \mu) \exp(x \log \mu - x \log(1 - \mu)) = (1 - \mu) \exp(x \log \left(\frac{\mu}{1 - \mu}\right))$$

$$h(x) = 1, \quad g(\eta) = 1 - \mu, \quad u(x) = x$$

b) From the result of a) for the natural parameter $\eta$, derive an expression for the parameter $\mu$. Try to draw a rough plot of $\mu$ as a function $f$ of $\eta$. What is the common name for that function?

$$\eta = \log\left(\frac{\mu}{1 - \mu}\right)$$

$$1 - \mu \exp(\eta) = \mu$$

$$\exp(\eta) - \mu \exp(\eta) = \mu$$

$$\exp(\eta) = \mu + \mu \exp(\eta)$$

$$\mu = \frac{\exp(\eta)}{1 + \exp(\eta)} = \frac{1}{1 + \exp(-\eta)} = \sigma(\eta)$$

This is the logistic sigmoid function.

c) From the expression for $g(\eta)$, compute the negative gradient of the logarithm, i.e. $-\frac{d \log g(\eta)}{d\eta}$. Use the function $f$ derived in b) to express $\mu$ and use the fact that the derivative of that function is $f(\eta)(1 - f(\eta))$. Interpret the result.
\[- \frac{d \log g(\eta)}{d \eta} = - \frac{d \log (1 - \sigma(\eta))}{d \eta} \quad (\mu = \sigma(\eta))\]
\[= \frac{1}{1 - \sigma(\eta)} \sigma(\eta)(1 - \sigma(\eta)) = \sigma(\eta) = \mu\]

In general, the negative gradient of the log of \(g(\eta)\) is the expected value over the sufficient statistics (the moments). This means that \(\mu\) is the sufficient statistics for the Bernoulli distribution.

d) The KL-divergence in bits between two discrete distributions \(p\) and \(q\) is defined as

\[KL(p\|q) = - \sum_x p(x) \log_2 \frac{q(x)}{p(x)}\]

Assume that both \(p(x \mid \mu)\) and \(q(x \mid \nu)\) are Bernoulli distributions where \(\mu = 1/2\) and \(\nu = 1/4\). Compute the KL-divergence \(KL(p\|q)\).

\[p(x = 1) = p(x = 0) = \frac{1}{2}\]
\[q(x = 1) = \frac{1}{4}, \quad q(x = 0) = \frac{3}{4}\]
\[\Rightarrow KL(p\|q) = - p(x) \log_2 \frac{q(x)}{p(x)} - p(1 - x) \log_2 \frac{q(1 - x)}{p(1 - x)}\]
\[= - \frac{1}{2} \log_2 \left(\frac{12}{41}\right) - \frac{1}{2} \log_2 \left(\frac{32}{41}\right)\]
\[= - \frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{3}{2}\]
\[= \frac{1}{2} - \frac{1}{2} (\log_2 (3) - 1)\]
\[= 1 - \frac{1}{2} \log_2 (3)\]