Exercise Sheet 4

Note: There are many competing notational standards for the exponential family, which can be used interchangeably, also for convenience. Here and in the rest of this sheet, we denote the natural parameter with $\gamma(\theta)$ instead of $c(\vartheta)$ as in the slides. Note that $g(\theta) = \exp(-A(\theta))$ in ex. 1 (b).

Exercise 4.1 - Exponential Family

(a) Show that the **Negative Binomial** distribution belongs to the exponential family. The density of a negative binomially distributed variable is:

$$f(x;p) = \binom{x+r_0-1}{x} p^{r_0}(1-p)^x, \quad x \in \mathbb{N}_0,$$

where $p \in [0,1]$ is a parameter and $r_0 \in \mathbb{N}$ <u>is known.</u>

Negative Binomial distribution with $p \in (0,1), r_0 \in \mathbb{N}$ (here: known and fixed) and $x \in \mathbb{N}_0$: We want to show that the density can be written as:

$$f_X(x;\theta) = h(x)\ g(\theta)\ \exp\left(\ \gamma(\theta) \cdot T(x)\ \right)$$

$$\begin{aligned} f(x;p) &= \binom{x+r_0-1}{x} p^{r_0}(1-p)^x \\ &= \underbrace{\binom{x+r_0-1}{x}}_{h(x)} \underbrace{p^{r_0}}_{g(p)} \exp(\underbrace{x}_{T(x)} \underbrace{\log(1-p)}_{\gamma(p)}) \end{aligned}$$

$\Rightarrow$ There is a one-parameter exponential family in $t(x) = x$ and $\gamma(p) = \log(1-p)$.

(b) **[Very hard]** Consider a simplified version of the **logistic distribution**, that has density:

$$f_\mu(x) = \frac{e^{(\mu-x)}}{(1+e^{(\mu-x)})^2}$$

with parameter $-\infty < \mu < \infty$, and support $\mathbb{R}$.

Show that this density does not belong to the exponential family.

Hint: It may be convenient to use $g(\mu, x) = \log(f_\mu(x))$.

To show that a distribution is part of the exponential family, we need to show that its density can be written as:

$$f(\mu, x) = h(x) \exp\left(\gamma(\mu)^\top T(x) - A(\mu)\right).$$

Here we assume μ is the parameter of the density. This is equivalent to writing the logarithm of the density as:

$$g(\mu, x) = \gamma(\mu)^\top T(x) - A(\mu) + \log(h(x)). \tag{1}$$

In our specific case:

$$g(\mu, x) = \log\left(\frac{e^{(\mu-x)}}{(1+e^{(\mu-x)})^2}\right) = \mu - x - 2\log\left(1+e^{(\mu-x)}\right). \tag{2}$$

Here the $\mu - x$ term doesn't pose any problem, since it can easily be assimilated into $A(\mu)$ and $\log(h(x))$. The real problem is the logarithm term for which we propose two different solutions, one being heuristic and the other one being fully rigorous.

Heuristic solution: For f to be exponential family, we need to be able to write $\log\left(1+e^{(\mu-x)}\right)$ as $\gamma(\mu)^\top T(x)$ which means a product of functions only containing μ and x. This looks very hard to do, and in fact it is not possible to split a logarithm in that way.

Rigorous proof: Let us suppose f belongs to the exponential family and look for a contradiction. Consider μ_0 a fixed parameter, $g(\mu, x)$ as defined above, and the following function:

$$G_{\mu_0}(\mu, x) = g(\mu, x) + g(\mu_0, 0) - g(\mu, 0) - g(\mu_0, x). \tag{3}$$

By substituting the exponential family expression from (1), we obtain:

$$G_{\mu_0}(\mu, x) = (\gamma(\mu) - \gamma(\mu_0))\,(T(x) - T(0)). \tag{4}$$

Now, the function G has the following property: for any pair of values x_1, x_2 and μ_1, μ_2

$$\underbrace{G_{\mu_0}(\mu_1, x_2) \cdot G_{\mu_0}(\mu_2, x_1)}_{a} = \underbrace{G_{\mu_0}(\mu_1, x_1) \cdot G_{\mu_0}(\mu_2, x_2)}_{b} \tag{5}$$

This can be verified easily by substituting (4) in it.

Finally, to find a contradiction we just need to find concrete values for μ_0, μ_1, μ_2 and x_1, x_2 such that (5) is not satisfied in our specific case. This means we calculate a and b in our concrete case by plugging (2) into (3) and check whether we obtain the same number. This is not the case for most arbitrarily chosen values, for example we can take $\mu_0 = 0, \mu_1 = -1, \mu_2 = 1$, and $x_1 = -1, x_2 = 1$:

$$\begin{aligned} G_{\mu_0}(\mu, x) = {} & (\mu - x - 2\log\left(1+e^{(\mu-x)}\right)) + (\mu_0 - 0 - 2\log\left(1+e^{(\mu_0-0)}\right)) - \\ & (\mu - 0 - 2\log\left(1+e^{(\mu-0)}\right)) - (\mu_0 - x - 2\log\left(1+e^{(\mu_0-x)}\right)) \\ = {} & 2\cdot\left(-\log\left(1+e^{(\mu-x)}\right) - \log\left(1+e^{\mu_0}\right) + \log\left(1+e^{\mu}\right) + \log\left(1+e^{(\mu_0-x)}\right)\right) \\ \overset{\mu_0=0}{=} {} & 2\cdot\left(-\log\left(1+e^{(\mu-x)}\right) - \log(2) + \log\left(1+e^{\mu}\right) + \log\left(1+e^{-x}\right)\right) \end{aligned}$$

$$\begin{aligned} G_{\mu_0=0}(\mu_1, x_1) &= G_{\mu_0=0}(-1,-1) = 2\cdot\left(-\log 2 - \log 2 + \log\left(1+e^{-1}\right) + \log(1+e)\right) \approx 0.47 \\ G_{\mu_0=0}(\mu_2, x_2) &= G_{\mu_0=0}(1,1) = 2\cdot\left(-\log(2) - \log(2) + \log(1+e) + \log\left(1+e^{-1}\right)\right) \approx 0.47 \\ G_{\mu_0=0}(\mu_1, x_2) &= G_{\mu_0=0}(-1,1) = 2\cdot\left(-\log\left(1+e^{-2}\right) - \log(2) + 2\log(1+e^{-1})\right) \approx -0.38 \\ G_{\mu_0=0}(\mu_2, x_1) &= G_{\mu_0=0}(1,-1) = 2\cdot\left(-\log\left(1+e^{2}\right) - \log(2) + 2\log(1+e)\right) \approx -0.38 \end{aligned}$$

By multiplying the values we immediately can see that $0.47^2 \neq 0.38^2$.

$\square$

Morale: showing *nonexistence* is hard, since we can't just check all possibilities and verify they're all wrong. Proof by contradiction helps in these cases.

Exercise 4.2 - Exponential Family

(a) Let X be a real random variable, with distribution belonging to the exponential family. Let $\boldsymbol{\theta} \in \mathbb{R}^p$ be its parameter vector and $u : \mathbb{R} \to \mathbb{R}$ a differentiable function with differentiable inverse u^{-1}.

Show that $Z = u(X)$ also follows an exponential family distribution.

Hint: remember the transformation rule for densities (see exercise sheet 1).

The distribution family of X is an exponential family and thus has a density of the form

$$f_X(x;\boldsymbol{\theta}) = h(x)g(\boldsymbol{\theta})\exp(\gamma(\boldsymbol{\theta})^\top T(x))$$

with suitable functions h, g, γ and T. This means that the density of Z is

$$\begin{aligned} f_Z(z;\boldsymbol{\theta}) &= (u^{-1})'(z) \cdot f_X(u^{-1}(z);\boldsymbol{\theta}) \\ &= \underbrace{(u^{-1})'(z)\ h(u^{-1}(z))}_{=:\tilde{h}(z)}\ g(\boldsymbol{\theta})\ \exp(\gamma(\boldsymbol{\theta})^\top \underbrace{T(u^{-1}(z))}_{=:\widetilde{T}(z)}) \end{aligned}$$

and as such, f_Z again has the desired form.

(b) Write down the joint density of an i.i.d. sample $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ where $X_i \sim Geom(p)$ (p unknown parameter). Show that it is an exponential family distribution and determine the natural parameter.

We are in a setting where a Bernoulli trial with success p is repeated as often as necessary, until a success occurs for the first time, the total number of trials performed (up to and including the first success) is modelled. For the density of one X_i we have:

$$f(x_i; p) = p(1-p)^{x_i - 1}$$

and thus

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x};p) &= \prod_{i=1}^n p(1-p)^{x_i-1} \\ &= p^n(1-p)^{\sum_{i=1}^n (x_i-1)} \\ &= p^n \cdot \exp\left(\left(\sum_{i=1}^n (x_i - 1)\right) \cdot \log(1-p)\right) \quad \text{Split sum} \\ &= p^n \cdot \exp\left(\left(\sum_{i=1}^n x_i - n\right) \cdot \log(1-p)\right) \quad \text{multiply} \\ &= p^n \cdot \exp\left((\sum_{i=1}^n x_i)\log(1-p) - n\log(1-p)\right) \quad \text{pull n into log} \\ &= p^n \cdot \exp\left((\sum_{i=1}^n x_i)\log(1-p)\right) \cdot \cancel{\exp(\log}(((1-p)^{-n})) \\ &= \underbrace{1}_{h(x)} \cdot \underbrace{\left(\frac{p}{1-p}\right)^n}_{g(p)} \exp\left(\underbrace{(\sum_{i=1}^n x_i)}_{T(x)} \underbrace{\log(1-p)}_{\gamma(p)} \right) \end{aligned}$$

The natural parameter of this exponential family is $\gamma(p) = \log(1-p)$.

Exercise 4.3 - Sufficiency

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$ be an i.i.d. sample, with $X_i \sim Po(\lambda)$ (Poisson distributed) for $i = 1, \ldots, n$.

(a) By using the conditional distribution of $\boldsymbol{X}$ given $T(\boldsymbol{X}) = t$, show that the statistic $T(\boldsymbol{X}) = \sum\limits_{i=1}^{n} X_i$ is sufficient for λ.

$$
\begin{aligned}
f(x_i;\lambda) &= \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \\
\Rightarrow f_{\boldsymbol{X}}(\boldsymbol{x};\lambda) &\overset{X_i \text{ indep.}}{=} \prod_{i=1}^{n} f(x_i;\lambda) = \lambda^{\sum_{i=1}^{n} x_i} \cdot e^{-n\lambda} \cdot \prod_{i=1}^{n} \left[\frac{1}{x_i!}\right] \\
&= \frac{\lambda^{\sum_i x_i}}{\prod_i x_i!} \cdot e^{-n\lambda} \text{ , for } x_i \in \{0, 1, \ldots\} = \mathbb{N}_0, \forall\, i = 1, \ldots, n
\end{aligned}
$$

The conditional distibution:

$$
\begin{aligned}
f_{\boldsymbol{X}|T(\boldsymbol{X})}(\boldsymbol{x} \mid t\,;\lambda) &= \frac{f_{X,T(X)}(\boldsymbol{x}, t;\lambda)}{f_{T(X)}(t;\lambda)} \\
&= \frac{\mathbb{P}\left(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, \sum_{i=1}^{n} X_i = t\right)}{\mathbb{P}\left(\sum_{i=1}^{n} X_i = t\right)} \\
&= \begin{cases} 0 \text{ , if } \sum X_i \neq t \\ \frac{\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)}{\mathbb{P}(\sum X_i = t)} \text{, else} \quad (*) \end{cases}
\end{aligned}
$$

1. Numerator of $(*)$:

$$
\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f_{\boldsymbol{X}}(\boldsymbol{x};\lambda) = \frac{\lambda^{\sum x_i}}{\prod_i x_i!} e^{-n\lambda}
$$

2. Denominator of $(*)$:
   For Poisson-distributed and independent X_i with parameters λ_i, the following holds:

$$
X_i \sim Pois(\lambda_i) \text{ indep. } \Rightarrow \sum_{i=1}^{n} X_i \sim Pois\left(\sum_{i=1}^{n} \lambda_i\right) \overset{\forall i:\ \lambda_i = \lambda}{=} Pois(n\lambda)
$$

   Thus:

$$
\mathbb{P}\left(\sum X_i = t\right) = f_{T(X)}(t;\lambda) = \frac{(n\lambda)^t}{t!} e^{-n\lambda}
$$

$(*)$ together:

$$
\begin{aligned}
f_{\boldsymbol{X}|T(\boldsymbol{X})}(\boldsymbol{x} \mid t\,;\lambda) &= \frac{\lambda^{\sum x_i} \cancel{e^{-n\lambda}} t!}{(\prod_i x_i!)(n\lambda)^t \cancel{e^{-n\lambda}}} \\
&= \frac{t! \cancel{\lambda^{\sum x_i}}}{(\prod_i x_i!) n^t \cancel{\lambda^t}} \qquad\qquad \sum x_i = t \\
&= \left(\frac{1}{n}\right)^t \cdot \frac{t!}{\prod_i x_i!} \\
&= \left(\frac{1}{n}\right)^t \cdot \frac{t!}{x_1! \cdots x_n!} \\
&\widehat{=}\ Mult\left(t; \frac{1}{n}, \ldots, \frac{1}{n}\right) \quad \text{with } m = t \text{ and } p_1 = \ldots = p_n = \frac{1}{n}
\end{aligned}
$$

It can be seen that the conditional distribution of $\boldsymbol{X}$ given $T(\boldsymbol{X}) = t$ is independent of λ. We have shown that $T(\boldsymbol{X}) = \sum X_i$ is sufficient for λ.

(b) Show again that $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ is a sufficient statistic for λ, this time by using the Neyman-Fisher factorization.

The following applies to the density of $\boldsymbol{X}$:

$$f_{\boldsymbol{X}}(\boldsymbol{x}; \lambda) = \underbrace{\lambda^{\sum x_i} e^{-n\lambda}}_{g(T(\boldsymbol{x});\, \lambda)} \underbrace{\prod_{i=1}^{n} \frac{1}{x_i!}}_{h(\boldsymbol{x})}$$

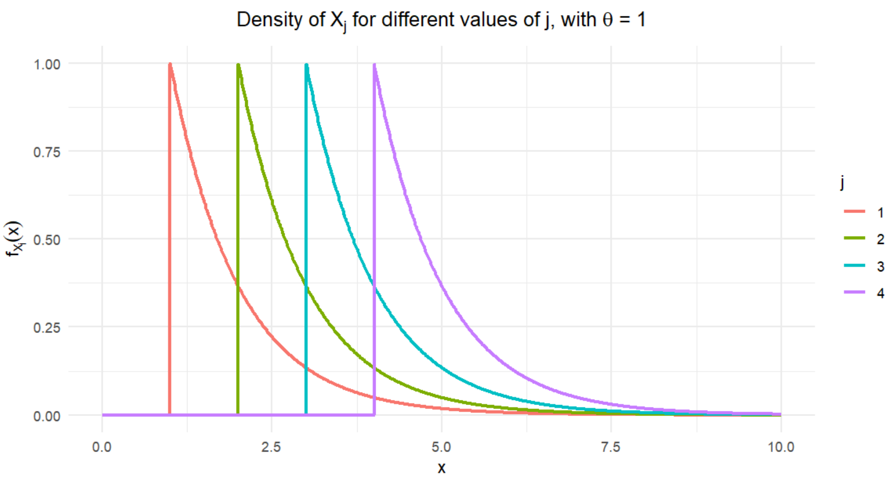
Sufficiency follows from this representation of the density.

Exercise 4.4 - Sufficiency

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ for $j = 1 \ldots n$ be an independent sample from the pdf

$$f_{X_j}(x; \theta) = \begin{cases} e^{j\theta - x} & x \geq j\theta \\ 0 & x < j\theta. \end{cases}$$

Find a sufficient statistic for θ and prove that it is sufficient.



To find a sufficient statistic for θ, note that that all $j > 0$ and all X_j values must satisfy $X_j \geq j\theta$ for the joint density to be positive. Rearranging, we arrive at $\frac{X_j}{j} \geq \theta$. This suggests that the minimum of the ratios $\frac{X_j}{j}$ could be sufficient since:

- If $\min_j \left(\frac{X_j}{j}\right) \geq \theta$, then all conditions $X_j \geq j\theta$ are satisfied.
- The minimum captures the threshold above which θ must lie.

Thus, consider the statistic $T(\mathbf{X}) = \min_j \frac{X_j}{j}$.

Since the X_j are independent, the joint pdf of $X_1, \ldots, X_n$ is:

$$\begin{aligned} f(x_1, \ldots, x_n; \theta) &= \prod_{j=1}^{n} f_{X_j}(x_j; \theta) \\ &= \prod_{j=1}^{n} e^{j\theta - x_j} \mathbb{1}_{[j\theta, \infty)}(x_j) \\ &= e^{\theta \sum_{j=1}^{n} j - \sum_{j=1}^{n} x_j} \prod_{j=1}^{n} \mathbb{1}_{[j\theta, \infty)}(x_j). \end{aligned}$$

The indicator term $\prod_{j=1}^{n} \mathbb{1}_{[j\theta,\infty)}(x_j)$ is equivalent to $\mathbb{1}_{[\theta,\infty)}\left(\min_j \frac{x_j}{j}\right)$, so:

$$f(x_1,\ldots,x_n;\theta) = \underbrace{\left(e^{-\sum_{j=1}^{n} x_j}\right)}_{h(\boldsymbol{x})} \cdot \underbrace{\left(e^{\theta \sum_{j=1}^{n} j} \mathbb{1}_{[\theta,\infty)}\left(\min_j \frac{x_j}{j}\right)\right)}_{g(T(\boldsymbol{x});\theta)}.$$

This is a factorization into a term $h(\boldsymbol{x})$ that does not depend on θ and a term $g(T(\boldsymbol{x});\theta)$ that depends on θ through $T(\mathbf{X}) = \min_j \frac{X_j}{j}$. $T(\mathbf{X})$ captures all information about θ while giving the tightest bound on θ that is consistent with all observations.