Tutorial 4

1. Express the following families as exponential families, identifying the terms in the expression:

   (a) The beta family:
   $$p(x;\beta_1,\beta_2)=\frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}x^{\beta_1-1}(1-x)^{\beta_2-1}\qquad 0\le x\le 1\qquad \beta_1>0,\quad \beta_2>0.$$

   (b) The gamma family:
   $$p(x;\alpha,\lambda)=\frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{-\lambda x}\qquad x\ge 0,\quad \lambda>0,\alpha>0$$

2. Which of the following are exponential families? Prove or disprove.

   (a) The $U(0,\theta)$ family for $\theta>0$. That is, $X\sim U(0,\theta)$ if it has density
   $$p(x;\theta)=\begin{cases}\frac{1}{\theta} & 0\le x\le\theta\\ 0 & \text{other}\end{cases}$$

   (b) The family of densities:
   $$p(x;\theta)=\mathbf{1}_{[0,\theta]}(x)\exp\{-2\log\theta+\log(2x)\}$$
   where $\theta>0$.

   (c) The family of discrete probability mass functions
   $$p(x;\theta)=\frac{1}{9}\qquad x\in\{0.1+\theta,0.2+\theta,\dots,0.9+\theta\}\qquad \theta\in\mathbb{R}$$

   (d) The $N(\theta,\theta^2)$ family, $\theta>0$

   (e)
   $$p(x;\theta)=\frac{2(x+\theta)}{1+2\theta}\qquad 0<x<1,\quad \theta>0$$

   (f) $p(x,\theta)$ is the conditional probability mass function for a binomial(n,θ) variable, conditioned on $X>0$. (recall binomial has probability mass function $\binom{n}{k}\theta^k(1-\theta)^{n-k}$).

3. The *inverse Gaussian* density $IG(\mu,\lambda)$, is:
   $$f(x;\mu,\lambda)=\left(\frac{\lambda}{2\pi}\right)^{1/2}\frac{1}{x^{3/2}}\exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2x}\right\}\mathbf{1}_{\{x>0\}}\qquad \mu>0,\lambda>0.$$

   (a) Show that this is an exponential family generated by $T(X)=-\frac{1}{2}(X,\frac{1}{X})$ and $h(x)=\frac{1}{(2\pi)^{1/2}x^{3/2}}$.

(b) Show that the canonical parameters (η_1, η_2) are

$$\eta_1 = \frac{\lambda}{\mu^2}, \qquad \eta_2 = \lambda$$

and that the log partition function is:

$$A(\eta_1, \eta_2) = -\left(\frac{1}{2}\log(\eta_2) + \sqrt{\eta_1\eta_2}\right), \qquad \mathcal{E} = \mathbb{R}_+^2$$

(c) Find the moment generating function of T and show that

$$\mathbb{E}[X] = \mu, \quad \mathrm{Var}(X) = \frac{\mu^3}{\lambda}, \quad \mathbb{E}\left[\frac{1}{X}\right] = \frac{1}{\mu} + \frac{1}{\lambda}, \quad \mathrm{Var}\left(\frac{1}{X}\right) = \frac{1}{\lambda\mu} + \frac{2}{\lambda^2}.$$

4. Let $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ be a canonical exponential family generated by (T, h) and $\mathcal{E}^0 \neq \phi$. Show that T is minimal sufficient.

5. Let $p(x, \eta)$ be a one parameter canonical exponential family generated by $T(x) = x$ and $h(x)$: $x \in \mathcal{X} \subset \mathbb{R}$, such that $\mathbb{E}_\eta[X]$ is well defined for all $\eta \in \mathcal{E}^0$. Let $\psi(x)$ be a non-constant, non-decreasing function which satisfies $\int |x\psi(x)|p(x;\eta)dx < +\infty$ for all $\eta \in \mathcal{E}^0$. Show that $\mathbb{E}_\eta[\psi(X)]$ is strictly increasing in η.

   Hint: Let Cov denote covariance. Show that

   $$\mathrm{Cov}(X, Y) = \frac{1}{2}\mathbb{E}\left[(X - X')(Y - Y')\right]$$

   where (X, Y) and (X', Y') are independent, identically distributed. Compute $\frac{\partial}{\partial\eta}\mathbb{E}_\eta[\psi(X)]$ in terms of $\mathrm{Cov}_\eta(\psi(X), X)$.

6. **Logistic Regression** In the following, $Y_1, \ldots, Y_n$ are the outcomes of random experiments $i = 1, \ldots, n$. For experiment i, you fix the values of covariates $z_{i1}, \ldots, z_{id}$. For example, suppose you are trying to find a cure for Coronavirus. For trial i, you choose the quantities of d different chemicals; these quantities are $z_{i1}, \ldots, z_{id}$. There are unknown parameters $\beta_1, \ldots, \beta_d$. You run experiment i on n_i individuals (who are independent of each other) and Y_i represents the number who are successfully cured and $n_i - Y_i$ the number for whom the treatement is not successful.

   If the model is correct, then you would like estimates $\widehat{\beta}_1, \ldots, \widehat{\beta}_p$ of the unknown parameters and, from this, estimate the success rate of the treatment for a given covariate vector $(z_1, \ldots, z_d)$.

   $(z_{1,.}, Y_1), \ldots, (z_{n,.}, Y_n)$ are observed, where $z_{1,.}, \ldots, z_{n,.}$ are d-row vectors and $Y_1, \ldots, Y_n$ are independent and $Y_j \sim \mathrm{Binomial}(n_j, \lambda_j)$. The success probability λ_j depends on the vector $z_{j,.}$. The function

   $$l(u) = \log\frac{u}{1-u}$$

   is called the *logit* function. In logistic regression, it is assumed that

   $$l(\lambda_i) = z_{i,.}\beta$$

where $\beta = (\beta_1, \ldots, \beta_d)^t$ is the parameter vector.

Show that $Y = (Y_1, \ldots, Y_n)$ is an exponential family of rank d if and only if $z_{.,1} \ldots, z_{.,d}$ are linearly independent.

Note The family is of rank k if and only if

$$\mathbb{P}\left(\sum_{j=1}^{k} c_j T_j(X) = c_0\right) < 1$$

for all $(c_0, c_1, \ldots, c_k) \not\equiv 0$.

7. Let $(X_1, X_2, \ldots X_n)$ be a *stationary Markov chain* with two states, 0 and 1. That is,

$$\mathbb{P}(X_i = x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) = \mathbb{P}(X_i = x_i | X_{i-1} = x_{i-1}) = p_{x_{i-1}, x_i}$$

where $\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$ is the matrix of *transition probabilities*. Suppose, furthermore, that

- $p_{00} = p_{11} = p$, so that $p_{10} = p_{01} = 1 - p$,
- $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$.

(a) Show that if $0 < p < 1$ is unknown, this is a full-rank one-parameter exponential family with $T = N_{00} + N_{11}$, where N_{ij} denotes the number of transitions from i to j. For example, the sequence 01011 has $N_{01} = 2$, $N_{11} = 1$, $N_{00} = 0$ and $N_{10} = 1$.

(b) Show that $\mathbb{E}[T] = (n-1)p$.

8. Let $X = (Z, Y)$ where $Y = Z + \theta W$, $\theta > 0$, Z and W are independent $N(0,1)$ variables. Let $X_1, \ldots, X_n$ be i.i.d. as X. Write the density of $X_1, \ldots, X_n$ as a canonical exponential family and identify T, h, η, A and $\mathcal{E}$. Find the expected value and variance of the sufficient statistic.

9. The *entropy* $h(p)$ of a random variable X with density p is defined by:

$$h(p) = \mathbb{E}\left[-\log p(X)\right] = -\int_S p(x) \log p(x) dx.$$

where $S = \{x : p(x) > 0\}$.

(a) Show that the canonical k parameter exponential family density

$$p(x, \eta) = \exp\left\{\sum_{j=1}^{k} \eta_j r_j(x) - A(\eta)\right\} \qquad x \in S$$

maximises $h(p)$ subject to the constraints

$$p(x) \geq 0, \quad \int_S p(x) dx = 1, \quad \int_S p(x) r_j(x) dx = \alpha_j, \quad 1 \leq j \leq k$$

for given $\alpha_1, \ldots, \alpha_k$ for which a solution exists, where $\eta_1, \ldots, \eta_k$ are chosen so that p satisfies the constraints.

Hint This is very easy using Lagrange multipliers; maximise the integrand

(b) Find the maximum entropy densities when $r_j(x) = x^j$ in the following cases:

i. $S = (0, +\infty), \quad \alpha_1 > 0$

ii. $S = \mathbb{R}$, $k = 2$, $\alpha_1 \in \mathbb{R}, \quad \alpha_2 \in \mathbb{R}_+$

iii. $S = \mathbb{R}$, $k = 3$, $\alpha_1 \in \mathbb{R}$, $\alpha_2 > 0$, $\alpha_3 \in \mathbb{R}$.

10. Suppose that $p(x, \theta)$ is a positive density on the real line, which is continuous in x for each θ and such that if X_1, X_2 is a sample of size 2 from $p(., \theta)$ then $X_1 + X_2$ is sufficient for θ. Show that $p(., \theta)$ corresponds to a one-parameter exponential family of distributions with $T(x) = x$.

Answers

1. (a)

$$p(x;\beta_1,\beta_2) = \exp\left\{(\beta_1 - 1)\log x + (\beta_2 - 1)\log(1-x) - \log\frac{\Gamma(\beta_1)\Gamma(\beta_2)}{\Gamma(\beta_1+\beta_2)}\right\}$$

$$h(x) = 1, \quad T_1(x) = \log x, \quad \eta_1(\beta) = \beta_1 - 1, \quad T_2(x) = \log(1-x), \quad \eta_2(\beta) = \beta_2 - 1,$$

$$B(\beta_1,\beta_2) = \log\frac{\Gamma(\beta_1)\Gamma(\beta_2)}{\Gamma(\beta_1+\beta_2)}$$

(b)

$$p(x;\alpha,\lambda) = \exp\{(\alpha-1)\log(x) - \lambda x - (\log\Gamma(\alpha) - \alpha\log(\lambda)\}$$

$$h(x) \equiv 1, \quad T_1(x) = \log(x), \quad \eta_1(\theta) = (\alpha - 1), \quad T_2(x) = x, \quad \eta_2(\theta) = -\lambda$$

$$B(\theta) = \log\Gamma(\alpha) - \alpha\log(\lambda)$$

2. (a) no: $p(x;\theta) = \theta^{-1}\mathbf{1}_{[0,\theta]}(x)$. For an exponential family:

$$p(x;\theta) = h(x)\exp\{\eta(\theta)T(x) - B(\theta)\}$$

so that

$$h(x) = \exp\{-\eta(\theta)T(x) + B(\theta) - \log\theta\}\mathbf{1}_{[0,\theta]}(x)$$

so that $h(x) = 0$ for all $x > \theta$. Since h does *not* depend on θ, $x \in [0,1]$ and $\Theta = (0,+\infty)$, hence $h(x) = 0$ for all $x > \theta$ for all $\theta > 0$, hence $h(x) = 0$ for all $x > 0$, so that $p(x;\theta) \equiv 0$, which is a contradiction.

(b) no: same as for (a): assume it is exponential family then:

$$h(x)e^{(T(x),\eta(\theta)) - B(\theta)} = \mathbf{1}_{[0,\theta]}(x)\exp\{-2\log\theta + \log(2x)\}$$

so that

$$h(x) = \mathbf{1}_{[0,\theta]}(x)\exp\{B(\theta) - 2\log\theta + \log(2x) - (T(x),\eta(\theta))\}$$

Here $\Theta = (0,+\infty)$ and $h(x) = 0$ for all $x > \theta$. This holds for all $\theta > 0$ hence $h(x) \equiv 0$ so that $p(x;\theta) \equiv 0$ which is a contradiction.

(c) no;

$$p(x;\theta) = \frac{1}{9}\sum_{j=1}^{9}\mathbf{1}_{0.1j}(x-\theta) = p(x-\theta;0)$$

so that

$$h(x)\exp\{\eta(\theta)T(x) - B(\theta)\} = h(x-\theta)\exp\{\eta(0)T(x-\theta) - B(0)\}.$$

If we take $\theta = 0$, we see that $h(x)$ has support (i.e. is non-zero for) $x \in \{0.1, 0.2, \dots, 0.9\}$. That is, $h(x) = 0$ for any x which does not belong to this set of values. Now, let us consider *arbitrary* θ, we see that $h(x)$ has support $\{0.1 + \theta, \dots, 0.9 + \theta\}$; $h(x) = 0$ for any x which does not belong to this set of values. Since $\Theta = \mathbb{R}$, therefore so that $h \equiv 0$, which gives a contradiction.

(d)

$$p(x;\theta) = \exp\left\{-\frac{x^2}{2\theta^2} + \frac{x}{\theta} - \frac{1}{2} - \frac{1}{2}\log(2\pi) - \log\theta\right\}$$

This does (technically) satisfy the definition of an exponential family, so the answer is YES. Note, however, that Θ is one-dimensional, yet we need a two-dimensional sufficient statistic $(T_1(x), T_2(x)) = (-x^2, x)$ and a two functions $\eta_1(\theta) = \frac{1}{2\theta^2}$ and $\eta_2(\theta) = \theta$. This is known as a *curved* exponential family.

(e)

$$p(x;\theta) = 2\exp\{\log(x+\theta) - \log(1+2\theta)\}\mathbf{1}_{[0,1]}(x)$$

no; the canonical parameter is infinite dimensional.

$$p(x;\theta) = 2x\exp\{\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}\frac{\theta^n}{x^n} - \log(1+2\theta)\}\mathbf{1}_{[0,1]}(x)$$

giving a sufficient statistic of $T(x) = (\frac{-1^{n-1}}{x^n})_{n\geq 1}$ and a canonical parameter vector of $\eta = (\theta^n)_{n\geq 1}$. For an exponential family, these have to be finite dimensional.

(f)

$$\mathbb{P}_\theta(X > 0) = 1 - (1-\theta)^n$$

$$p(x,\theta) = \frac{1}{1-(1-\theta)^n}\binom{n}{x}\theta^x(1-\theta)^{n-x} = \binom{n}{x}\exp\left\{x\log\frac{\theta}{1-\theta} + n\log(1-\theta) - \log(1-(1-\theta)^n)\right\}$$

yes

3. (a) Comes from expanding

$$-\frac{\lambda(x-\mu)^2}{2\mu^2 x} = -\frac{\lambda x}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2x}$$

which gives sufficient statistic $-\frac{1}{2}(x, \frac{1}{x})$ and canonical coordinates $(\eta_1, \eta_2) = (\frac{\lambda}{\mu^2}, \lambda)$. The $h(x) = \frac{1}{(2\pi)^{1/2}x^{3/2}}$ comes directly from the first part of the expression for the density and the log partition function is:

$$B(\mu, \lambda) = -\frac{\lambda}{\mu} - \frac{1}{2}\log\lambda.$$

(b) For $T(x) = -\frac{1}{2}(x, \frac{1}{x})$, the above expansion also gives $\eta_1 = \frac{\lambda}{\mu^2}$, $\eta_2 = \lambda$ and

$$A(\eta_1, \eta_2) = -\frac{\lambda}{\mu} - \frac{1}{2}\log\lambda = -\sqrt{\eta_1\eta_2} - \frac{1}{2}\log\eta_2.$$

(c) Using $M_T(s) = \exp\{A(\eta+s) - A(\eta)\}$ we have:

$$M_{T;\eta}(s_1, s_2) = \left(\frac{\eta_2}{\eta_2 + s}\right)^{1/2} \exp\left\{\sqrt{\eta_1\eta_2} - \sqrt{(\eta_1+s_1)(\eta_2+s_2)}\right\}$$

To compute expectations and variances, use $\dot{A}(\eta) = \mathbb{E}_\eta[T]$ and $\ddot{A}(\eta) = \Sigma_T$.

$$\dot{A}(\eta_1,\eta_2) = -\begin{pmatrix} \frac{1}{2}\frac{\eta_2^{1/2}}{\eta_1^{1/2}} \\ \frac{1}{2}\frac{\eta_1^{1/2}}{\eta_2^{1/2}} + \frac{1}{2\eta_2} \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} \mathbb{E}[X] \\ \mathbb{E}\left[\frac{1}{X}\right] \end{pmatrix}$$

$$\mathbb{E}[X] = \mu, \qquad \mathbb{E}\left[\frac{1}{X}\right] = \frac{1}{\mu} + \frac{1}{\lambda}$$

$$\ddot{A}(\eta_1,\eta_2) = \frac{1}{4}\begin{pmatrix} \eta_1^{-3/2}\eta_2^{1/2} & -\eta_1^{-1/2}\eta_2^{-1/2} \\ -\eta_1^{-1/2}\eta_2^{-1/2} & \eta_1^{1/2}\eta_2^{-3/2} + \frac{2}{\eta_2} \end{pmatrix}$$

$$\mathrm{Var}(X) = \frac{\mu^3}{4\lambda} \qquad \mathrm{Var}\left(\frac{1}{X}\right) = \frac{1}{\mu\lambda} + \frac{2}{\lambda^2}.$$

4.

$$p(x,\theta) = h(x)\exp\left\{\sum_{j=1}^k T_j(x)\theta_j - A(\theta)\right\}$$

$$\log L(\theta,x) - \log L(\theta,y) = (\log h(x) - \log h(y)) + \sum_{j=1}^k (T_j(x) - T_j(y))\theta_j$$

clearly does not depend on θ if and only if $T(x) = T(y)$.

5. For a one-parameter exponential family,

$$p(x;\eta) = h(x)\exp\{\eta T(x) - A(\eta)\}$$

and here $T(x) = x$ so that

$$\begin{aligned}
\frac{\partial}{\partial\eta}\mathbb{E}_\eta[\psi(X)] &= \frac{\partial}{\partial\eta}\int h(x)e^{\eta x - A(\eta)}\psi(x)dx = \int h(x)\left(\frac{d}{d\eta}e^{\eta x - A(\eta)}\right)\psi(x)dx \\
&= \int h(x)e^{\eta x - A(\eta)}(x - \dot{A}(\eta))\psi(x)dx \\
&= \int p(x;\eta)x\psi(x)dx - \dot{A}(\eta)\int p(x;\eta)\psi(x)dx.
\end{aligned}$$

To exchange $\frac{\partial}{\partial\eta}$ with $\int$ for all $\eta \in \mathcal{E}^0$, then for each $\eta \in \mathcal{E}^0$ we need an $\epsilon > 0$ and an integrable function $H(x)$ such that $\sup_{\xi\in(\eta-\epsilon,\eta+\epsilon)}\left|\frac{\partial}{\partial\xi}h(x)e^{\xi x - A(\xi)}\psi(x)\right| \leq H(x)$ for all $\eta \in \mathcal{E}$ and such that $\int H(x)dx < +\infty$. Now,

$$\frac{\partial}{\partial \xi} h(x) e^{\xi x - A(\xi)} \psi(x) = (x - \dot{A}(\xi)) \psi(x) h(x) e^{\xi x - \dot{A}(\xi)}$$

so, for a given $\eta \in \mathcal{E}^0$, choose an ϵ such that $(\eta - \epsilon, \eta + \epsilon) \subset \mathcal{E}^0$, let $C = \sup_{\xi \in (\eta - \epsilon, \eta + \epsilon)} |\dot{A}(\xi)|$ and let:

$$H(x) = h(x) e^{(\eta + \epsilon)x} (|x| + C) |\psi(x)| \mathbf{1}_{[0, +\infty)}(x) + h(x) e^{(\eta - \epsilon)x} (|x| + C) |\psi(x)| \mathbf{1}_{(-\infty, 0]}(x)$$

Then $\int H(x) dx < +\infty$ and satisfies the criterion, so that the derivative and integral may be exchanged.

Therefore, using $\dot{A}(\eta) = \mathbb{E}_\eta[T(X)]$:

$$\frac{\partial}{\partial \eta} \mathbb{E}_\eta[\psi(X)] = \mathbb{E}_\eta[X \psi(X)] - \mathbb{E}_\eta[\psi(X)] \mathbb{E}_\eta[X] = \mathrm{Cov}(X, \psi(X)).$$

Under the conditions placed on ψ, $(x - y)(\psi(x) - \psi(y))$ is non negative and positive with positive probability. The result follows.

6. Using the notations of the question, and setting $z_{j.} = (z_{j1}, \ldots, z_{jd})^t$,

$$p(y_1, \ldots, y_n, \underline{\beta}) = \left(\prod_{j=1}^n \binom{n_j}{y_j} \right) \exp \left\{ \sum_{i=1}^d \beta_i \left(\sum_{j=1}^n y_j z_{ji} \right) - \sum_{j=1}^n n_j \log(1 - \lambda_j) \right\}$$

The family is of rank k if and only if

$$\mathbb{P}(\sum_{j=1}^k c_j T_j(X) = c_0) < 1$$

for all $(c_0, c_1, \ldots, c_k)$. Here

$$\mathbb{P}\left(\sum_{i=1}^d c_i T_i(Y) = c_0 \right) = \mathbb{P}\left(\sum_{j=1}^n \left(\sum_{i=1}^d c_i z_{ji} \right) Y_j = c_0 \right)$$

If the (column) vectors $z_{.1}, \ldots, z_{.d}$ are not linearly independent, then (by definition) $c_1, \ldots, c_d$ may be found so that $\sum_{i=1}^d c_i z_{.,i} = 0$.

If they are linearly independent, then clearly the family is of rank d.

7. (a)

$$\mathbb{P}((X_1, \ldots, X_n) = (x_1, \ldots, x_n)) = \frac{1}{2} p_{00}^{n_{00}} p_{01}^{n_{01}} p_{10}^{n_{10}} p_{11}^{n_{11}}$$

where $n_{00} + n_{01} + n_{10} + n_{11} = n - 1$, the total number of transitions. It follows that

$$\begin{aligned}
\mathbb{P}((X_1, \ldots, X_n) = (x_1, \ldots, x_n)) &= \frac{1}{2} \exp \left\{ (n_{00} + n_{11}) \log p + (n_{01} + n_{10}) \log(1 - p) \right\} \\
&= \frac{1}{2} \exp \left\{ (n_{00} + n_{11}) \log \left(\frac{p}{1 - p} \right) + (n - 1) \log(1 - p) \right\}
\end{aligned}$$

The result now follows from the formula for an exponential family; $h(x) = \frac{1}{2}$, $T(x) = n_{00} + n_{11}$, $\eta(p) = \log\left(\frac{p}{1-p}\right)$, $B(p) = -(n-1)\log(1-p)$.

(b) Let $Y_i = 1$ if transition i is either $0 \mapsto 0$ or $1 \mapsto 1$ and let $Y_i = 0$ otherwise. Then

$$T = Y_1 + \ldots Y_{n-1}.$$

Since $\mathbb{E}[Y_j] = p$, the result follows.

8. $X = \binom{Z}{Y} \sim N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & 1 \\ 1 & 1+\theta^2 \end{array}\right)\right)$. Covariance matrix is $\Sigma = \left(\begin{array}{cc} 1 & 1 \\ 1 & 1+\theta^2 \end{array}\right)$ so $|\Sigma| = \theta^2$ and $\Sigma^{-1} = \frac{1}{\theta^2}\left(\begin{array}{cc} 1+\theta^2 & -1 \\ -1 & 1 \end{array}\right)$. It follows that

$$f_{(Z,Y)}(z,y) = \frac{1}{2\pi|\theta|}\exp\left\{-\frac{1}{2\theta^2}(z^2 + (1+\theta^2)y^2 - 2zy)\right\}$$

giving:

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{1}{(2\pi)^n|\theta|^n}\exp\left\{-\frac{1}{2\theta^2}\sum_{j=1}^n (z_j - y_j)^2 - \frac{1}{2}\sum_{j=1}^n y_j^2\right\}$$

so that

$$T(x_1,\ldots,x_n) = \sum_{j=1}^n (z_j - y_j)^2, \qquad h(x_1,\ldots,x_n) = \frac{1}{(2\pi)^n}e^{-\sum_{j=1}^n y_j^2}, \qquad \eta = -\frac{1}{2\theta^2}$$

$$A(\eta) = -\frac{n}{2}\log\frac{1}{\theta^2} = -\frac{n}{2}\log(-2\eta), \qquad \mathcal{E} = (0, +\infty)$$

Hence

$$\mathbb{E}_\eta[T] = \frac{dA}{d\eta} = -\frac{n}{2\eta} = n\theta^2$$

$$\mathrm{Var}_\eta(T) = \frac{d^2A}{d\eta^2} = \frac{n}{2\eta^2} = 2n\theta^4$$

9. (a) Lagrange method of multipliers: if we maximise the integrand pointwise, then this maximises the integral. Maximise

$$-p(x)\log p(x) - \lambda_0 p(x) - \sum_{j=1}^k p(x) r_j(x)\lambda_j$$

then choose $\lambda_0, \lambda_1, \ldots, \lambda_k$ to satisfy constraints. Taking derivative w.r.t. $p(x)$, maximum satisfies:

$$-\log p(x) - 1 - \lambda_0 - \sum_{j=1}^k r_j(x)\lambda_j = 0$$

so that p is of the form:

$$p(x) = \exp\left\{-(1+\lambda_0) - \sum_{j=1}^{k} \lambda_j r_j(x)\right\}$$

Choose $\lambda_0, \lambda_1, \dots, \lambda_k$ so that the constraints are satisfied. For an exponential family, this is clearly the case if $\lambda_j = -\eta_j$ for $j = 1, \dots, k$ and $A(\eta) = 1 + \lambda_0$.

(b) i. $p(x) = \frac{1}{\alpha_1} \exp\{-x/\alpha_1\} \qquad x \in (0, +\infty)$

ii.

$$p(x) = \exp\left\{\eta_1 x + \eta_2 x^2 - A(\eta)\right\}$$

No solution for $\alpha_2 < \alpha_1^2$; this would require random variables which satisfy: $\mathbb{E}[X^2] < \mathbb{E}[X]^2$. It follows that α_2 satisfies $\alpha_2 > \alpha_1^2$. Set $\sigma^2 = \alpha_2 - \alpha_1^2$, then

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\alpha_1)^2}{2\sigma^2}\right\} \qquad -\infty < x < +\infty$$

iii.

$$p(x) = \exp\left\{\eta_1 x + \eta_2 x^2 + \eta_3 x^3 - A(\eta)\right\} \qquad -\infty < x < +\infty$$

Clearly it doesn't exist!

10. It follows from the factorisation theorem that

$$p(x_1, \theta) p(x_2, \theta) = h(x_1, x_2) g(x_1 + x_2, \theta).$$

Fix a point θ_0 and let $r(x, \theta) = \log p(x, \theta) - \log p(x, \theta_0)$. Let $q(z, \theta) = \log g(z, \theta) - \log g(z, \theta_0)$. Then

$$r(x_1, \theta) + r(x_2, \theta) = q(x_1 + x_2, \theta)$$

so that $r(., \theta)$ and $q(., \theta)$ are linear in x;

$$r(x, \theta) = a(\theta) + b(\theta)x.$$

It follows that

$$p(x, \theta) = p(x, \theta_0) \exp\left\{a(\theta) + b(\theta)x\right\}$$

Let $h(x) = p(x, \theta_0)$, then this density is an exponential family with $T(x) = x$.

Establishing linearity in x The density is continuous and positive, hence so are r and q. Since $q(x_1 + x_2) = r(x_1) + r(x_2)$, it follows that $q(x) = r(x) + r(0)$ so that $q(0) = 2r(0)$ and $q(x_1 + x_2) = q(x_1) + q(x_2) - q(0)$. Now set $f(x) = q(x) - q(0)$ so that

$$f(x_1 + x_2) = f(x_1) + f(x_2).$$

It follows that for any $x_1, \dots, x_n$,

$$f(x_1 + \ldots + x_n) = f(x_1) + \ldots + f(x_n).$$

In particular,

$$f(1) = nf\left(\frac{1}{n}\right) \Rightarrow f\left(\frac{1}{n}\right) = \frac{1}{n}f(1)$$

and

$$f\left(\frac{k}{n}\right) = \frac{k}{n}f(1).$$

It follows that for x rational, $f(x) = xf(1)$ and hence, by continuity, it follows that $f(x) = xf(1)$ for all x. It follows that $q(x) = a + bx$ for constants a and b and hence that $r(x) = \frac{a}{2} + bx$.