STAT215: Solutions for Homework 1  
Due: Wednesday, Jan 30  

1. (10 pt) For $X \sim \text{Be}(\alpha, \beta)$,  
   (a) Evaluate $E[X^a(1 - X)^b]$ for all real numbers $a$ and $b$. For which $a, b$ is it finite?  
   (b) What is the MGF $M_{\log X}(t)$ for the random variable $Y := [\log X]$?  
      [HINT: Integration by parts is not needed for (a), and no new calculations at all are needed for (b)!

Solution:  
   (a)  
   
   
   $$E[X^a(1 - X)^b] = \int_0^1 x^a(1 - x)^b f(x) dx$$  
   $$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{(a+\alpha)-1}(1 - x)^{b+\beta-1} dx$$  
   $$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + a)\Gamma(\beta + b)}{\Gamma(\alpha + a + b + b)}$$  

When $a > -\alpha, b > -\beta$, it is finite.

(b)  

$$M_Y(t) = E(e^{tY}) = \int e^{ty} f(y) dy$$  
$$= \int x^t f(x) dx = E(X^t)$$  

This is just a special case with $a = t, b = 0$. So $M_{\log X}(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+t)}{\Gamma(\alpha+\beta+t)}$,  

$t > -\alpha$

2. (10 pt) Let $X \sim \text{Ge}(p)$.  
   (a) Find the moment generating function $M_X(t) := E[e^{tX}]$.  
   (b) Use this MGF to find the mean and variance of $X$.  

1
Solution:

(a)

\[ M_X(t) = E[e^{tX}] \]
\[ = \sum_{x=0}^{\infty} e^{tx} p(1-p)^x \]
\[ = \frac{p}{1 - (1-p)e^t}, \text{ for } t < -\log(1-p) \]

(b) The cumulant generating function \( c_X(t) = \log(M_X(t)) = \log p - \log(1 - qe^t) \), so

\[ EX = c'_X(0) = \frac{1-p}{p} \]

\[ VarX = c''_X(0) - (EX)^2 = \frac{qe^t}{(1-qe^t)^2} \bigg|_{t=0} = \frac{1-p}{p^2} \]

Note: If you use \( f(y) = pq^{y-1} \), then \( M_Y(t) = \frac{pe^t}{1-qe^t}, EY = 1/p, VarY = q/p^2 \).

3. (10 pt) The double-exponential distribution has pdf \( f(x) = \frac{1}{2}e^{-\lambda|x|} \), for fixed \( \lambda > 0 \).

(a) Find the MGF \( M_X(t) \) of a double exponential. For which \( t \) is it finite?

(b) Let \( U, V \overset{\text{iid}}{\sim} \text{Ex}(1) \), and find the MGF \( M_Y(t) \) of \( Y := U - V \). What is the distribution of \( Y \)?

(c) Find the mean and variance of a double exponential. If \( X_1, \ldots, X_n \) are i.i.d. double exponentials, find the MGF \( M_n(t) \) of the standardized mean,

\[ W_n := \frac{\bar{X} - E(X)}{\sqrt{Var(X)/n}}. \]

(d) What is the limit of \( M_n(t) \) as \( n \to \infty \)? What distribution has this function for its MGF?

Solution:

(a)

\[ M_X(t) = \int e^{tx} \frac{\lambda}{2} e^{-\lambda|x|} dx \]
\[ = \frac{\lambda}{2} \int_0^{\infty} e^{(t-\lambda)x} dx + \frac{\lambda}{2} \int_{-\infty}^{0} e^{(t+\lambda)x} dx \]
\[ = \frac{\lambda}{2} \frac{1}{-\lambda + t} e^{(t-\lambda)x}\bigg|_0^{\infty} + \frac{1}{\lambda + t} e^{(t+\lambda)x}\bigg|_0^{\infty} \]
\[ = \frac{\lambda^2}{\lambda^2 - t^2} \text{ if } |t| < \lambda \]
(b) 
\[ M_Y(t) = E[e^{t(V - U)}] = E[e^{tU}] = E[e^{-tV}] = \frac{1}{1 - t} \frac{1}{1 + t} = \frac{1}{1 - t^2} \]

which is the MGF of a double exponential so \( Y \) is a double exponential with \( \lambda = 1 \).

(c) For a double-exponential, \( EX = M'_X(0) = 0, VarX = M''_X(0) = 2/\lambda^2 \). So

\[
M_n(t) = E(e^{tW_n}) = M_X(t/\sqrt{nV(x)}) \cdots M_X(t/\sqrt{nV(x)}) = M_X(t/\sqrt{nV(x)})^n = \left( \frac{1}{1 - t^2/2n} \right)^n
\]

(d) As \( n \to \infty \),

\[ M_n(t) \to \frac{1}{e^{-t^2/2}} = e^{t^2/2} \]

which is the MGF of a \( \text{No}(0, 1) \)

4. (10 pt) Bickel & Doksum pg 71, problem 1.2.1

Solution:

(a) 
\[
\pi(\theta = \theta_1|x = 0) = \frac{\pi(\theta = \theta_1)\pi(x = 0|\theta = \theta_1)}{\pi(\theta = \theta_1)\pi(x = 0|\theta = \theta_1) + \pi(\theta = \theta_2)\pi(x = 0|\theta = \theta_2)} = \frac{\frac{1}{2} \cdot 0.8}{\frac{1}{2} \cdot 0.8 + \frac{1}{2} \cdot 0.4} = \frac{2}{3}
\]

Similarly,
\[
\pi(\theta = \theta_2|x = 0) = \frac{\frac{1}{2} \cdot 0.4}{\frac{1}{2} \cdot 0.8 + \frac{1}{2} \cdot 0.4} = \frac{1}{3}
\]
\[
\pi(\theta = \theta_1|x = 1) = \frac{\frac{1}{2} \cdot 0.2}{\frac{1}{2} \cdot 0.2 + \frac{1}{2} \cdot 0.6} = \frac{1}{4}
\]
\[
\pi(\theta = \theta_2|x = 1) = \frac{\frac{1}{2} \cdot 0.6}{\frac{1}{2} \cdot 0.2 + \frac{1}{2} \cdot 0.6} = \frac{3}{4}
\]

We might summarize this posterior frequency function in the following way:

\[ \begin{array}{c|cc}
\downarrow \theta, x \to & 0 & 1 \\
\hline
\theta_1 & \frac{2}{3} & \frac{1}{3} \\
\theta_2 & \frac{1}{2} & \frac{1}{4} \\
\end{array} \]
(b) Note that \( \pi(x_1, x_2, \ldots, x_n|\theta = \theta_1) = \prod_{i=1}^{n}(0.2)^{x_i}(0.8)^{1-x_i} = (0.2)^{\sum x_i}(0.8)^{n-\sum x_i} \).

\[
\pi(\theta = \theta_1|x_1, x_2, \ldots, x_n) = \frac{\pi(\theta = \theta_1)\pi(x_1, x_2, \ldots, x_n|\theta = \theta_1)}{\pi(\theta = \theta_1)\pi(x_1, x_2, \ldots, x_n|\theta = \theta_1) + \pi(\theta = \theta_2)\pi(x_1, x_2, \ldots, x_n|\theta = \theta_2)} = \frac{\frac{1}{2}(0.2)^{\sum x_i}(0.8)^{n-\sum x_i}}{\frac{1}{2}(0.2)^{\sum x_i}(0.8)^{n-\sum x_i} + \frac{1}{2}(0.6)^{\sum x_i}(0.4)^{n-\sum x_i}} = \frac{3^{S_n} \cdot 2^{n-S_n}}{2^n + 3^{S_n} \cdot 2^{n-S_n}}
\]

where \( S_n = \sum x_i \).

Similarly,

\[
\pi(\theta = \theta_1|x_1, x_2, \ldots, x_n) = \frac{6^{S_n}}{2^n + 6^{S_n}}.
\]

(c) Similar to (b). We have:

\[
\pi(\theta = \theta_1|x_1, x_2, \ldots, x_n) = \frac{\pi(\theta = \theta_1)\pi(x_1, x_2, \ldots, x_n|\theta = \theta_1)}{\pi(\theta = \theta_1)\pi(x_1, x_2, \ldots, x_n|\theta = \theta_1) + \pi(\theta = \theta_2)\pi(x_1, x_2, \ldots, x_n|\theta = \theta_2)} = \frac{(0.25) \cdot (0.25)^{\sum x_i}(0.8)^{n-\sum x_i} + (0.75) \cdot (0.6)^{\sum x_i}(0.4)^{n-\sum x_i}}{(0.25) \cdot (0.2)^{\sum x_i}(0.8)^{n-\sum x_i} + (0.75) \cdot (0.6)^{\sum x_i}(0.4)^{n-\sum x_i}} = \frac{2^n}{2^n + 6^{S_n}}
\]

and

\[
\pi(\theta = \theta_2|x_1, x_2, \ldots, x_n) = \frac{(0.75) \cdot (0.6)^{\sum x_i}(0.4)^{n-\sum x_i}}{(0.25) \cdot (0.2)^{\sum x_i}(0.8)^{n-\sum x_i} + (0.75) \cdot (0.6)^{\sum x_i}(0.4)^{n-\sum x_i}} = \frac{6^{S_n}}{2^n + 3 \cdot 6^{S_n}}
\]

(d) When \( n = 2 \), then the observation gives \( \sum x_i = 1 \). The posterior distribution of \( \theta \) in that case is

\[
\pi(\theta = \theta_1|\sum x_i = 1) = \frac{2^2}{2^2 + 6^1} = 0.4, \text{ and } \pi(\theta = \theta_2|\sum x_i = 1) = 0.6 \text{ for the original prior.}
\]

For the prior \( \pi_1 \) given in part (c), the posterior distribution of \( \theta \) is

\[
\pi(\theta = \theta_1|\sum x_i = 1) = \frac{2^2}{2^2 + 3 \cdot 6^1} = \frac{2}{11} \approx 0.18, \text{ and } \pi(\theta = \theta_2|\sum x_i = 1) = \frac{9}{11} \approx 0.82.
\]

When \( n = 100 \), then the observation gives \( \sum x_i = 50 \). In that case, the original prior gives the following posterior distribution of \( \theta \):

\[
\pi(\theta = \theta_1|\sum x_i = 50) = \frac{100}{2^{100} + 6^{50}} \approx 1.57 \times 10^{-9}, \text{ and } \pi(\theta = \theta_2|\sum x_i = 1) \approx 1 \text{ for the original prior.}
\]

For the prior \( \pi_1 \) given in part (c), the posterior distribution of \( \theta \) is

\[
\pi(\theta = \theta_1|\sum x_i = 1) \approx 5.23 \times 10^{-10}, \text{ and } \pi(\theta = \theta_2|\sum x_i = 1) \approx 1.
\]
(e) Given that $\Sigma x_i = k$, we have, under the initial prior $\pi$,
\[
\pi(\theta = \theta_1 | \Sigma x_i = k) = \frac{\frac{1}{4} (0.2)^k (0.8)^{n-k}}{\frac{1}{4} (0.2)^k (0.8)^{n-k} + \frac{1}{4} (0.4)^k (0.6)^{n-k}},
\]
and
\[
\pi(\theta = \theta_2 | \Sigma x_i = k) = \frac{\frac{1}{4} (0.4)^k (0.6)^{n-k}}{\frac{1}{4} (0.2)^k (0.8)^{n-k} + \frac{1}{4} (0.4)^k (0.6)^{n-k}}.
\]
Since these denominators are the same, we may find the value of $\theta$ having the higher posterior probability by comparing numerators only:
\[
\pi(\theta = \theta_1 | \Sigma x_i = k) > \pi(\theta = \theta_2 | \Sigma x_i = k) \iff \frac{1}{4} (0.2)^k (0.8)^{n-k} > \frac{1}{4} (0.6)^k (0.4)^{n-k}
\]
\[
\iff \frac{0.2^k 0.8^{n-k}}{0.4^k} > 1
\]
\[
\iff -k \log 3 + (n - k) \log 2 > 0
\]
\[
\iff k/n < \frac{\log 2}{\log 3} \approx 0.39 \text{ (after a little algebra)}.
\]
This shows that under the original prior $\pi$, $\theta = \theta_1$ has a higher posterior probability given $\Sigma x_i$ than does $\theta = \theta_2$, so long as $k < 0.39n$.

For the prior $\pi_1$, similar algebra shows that $\theta = \theta_1$ has a higher posterior probability given $\Sigma x_i$ than does $\theta = \theta_2$, so long as $k < 0.39n - 0.61$, approximately.

Under the original prior, for $n = 2$, $\hat{\theta} = \theta_1$ if and only if $k = 0$. For $n = 100$, $\hat{\theta} = \theta_1$ if and only if $k \in \{0, 1, \ldots, 38\}$.

Under the prior $\pi_1$, for $n = 2$, $\hat{\theta} = \theta_1$ if and only if $k = 0$, as before. For $n = 100$, $\hat{\theta} = \theta_1$ if and only if $k \in \{0, 1, \ldots, 38\}$, exactly the same as under the other prior.

(f) The only time the two $\hat{\theta}$’s will disagree is if $k \in (0.39n - 0.61, 0.39n) \iff k = [0.39n]$ (approximately). For example, if $n = 1000$, then the two different priors would yield different estimates of $\theta$ only if $k$ happened to be 390. As $n$ goes to infinity, the probability of $x$ taking any one particular value $k$ goes to zero, so the region on which the two $\hat{\theta}$’s disagree has a probability going to zero.

5. (10 pt) Bickel & Doksum pg 74, problem 1.2.14

Solution:

(a) The solution to this problem is worked out in Gelman, Carlin, Stern, and Rubin (“Bayesian Data Analysis”, aka “The Red Book”), pp. 46-49. It involves more than one rather unpretty completing of a square. The result is as follows:

Posterior predictive distribution of $x_{n+1}$:
\[
x_{n+1} | x_1, x_2, \ldots, x_n \sim \mathcal{N}(\mu_n, \sigma_n^2 + \tau_n^2),
\]
where
\[
\mu_n = \frac{\left(\frac{1}{\sigma_0^2}\right)\theta_0 + \left(\frac{n}{\sigma_0^2}\right)\bar{x}}{\left(\frac{1}{\sigma_0^2}\right) + \left(\frac{n}{\sigma_0^2}\right)}, 
\]
and
\[
\tau_n^2 = \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_0^2}\right)^{-1}.
\]
Here, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

The predictive distribution (marginal) of $x_{n+1}$ is just a special case of the predictive distribution when $n = 0$, so we have $x_{n+1} \sim \mathcal{N}(\theta_0, \sigma_0^2 + \tau_0^2)$. 

5
(b) As \( n \) goes to infinity we note that the predictive distribution of \( x_{n+1} \) never changes. This is to be expected, since by not taking into account the data we’re gaining no information. Note that the mean of \( x_{n+1} \) is always the prior mean on \( \theta \), \( \theta_0 \), and the variance of \( x_{n+1} \) comes from two sources: the uncertainty in the observation (\( \sigma_0^2 \)) and the uncertainty in the prior estimate of \( \theta \) (\( \tau_0^2 \)).

However, as \( n \) goes to infinity the posterior predictive distribution of \( x_{n+1} \) given \( x_1, x_2, \ldots, x_n \) behaves differently. The posterior mean, which is a weighted average of the prior mean and the sample mean, gets a heavier and heavier weight on the sample mean. And the variance has less and less of its uncertainty coming from the uncertainty in \( \theta \), so that nearly all the uncertainty comes from the uncertainty in the observation.

In the limit, \( x_{n+1}|x_1, x_2, \ldots, x_n \sim \text{No}(\bar{x}, \sigma_0^2) \).

6. (10 pt) Bickel & Doksum pg 84, problem 1.5.3

**Solution:** The factorization theorem will be useful to determine a sufficient statistic for each example.

(a)

\[
p(x_1, x_2, \ldots, x_n|\theta) = \prod (\theta x^{\theta-1}) = \theta^n \prod x_i^{\theta-1} = \theta^n \exp((\theta - 1)\sum \log x_i)
\]

By the factorization theorem, \( \sum \log x_i \) is a sufficient statistic for \( \theta \). Note that taking the log was not necessary. Another sufficient statistic would be \( \prod x_i \). The latter is more cumbersome computationally, so it is common in a situation like this to take the logs.

(b)

\[
p(x_1, x_2, \ldots, x_n|\theta) = \prod (\theta ax^{a-1}\exp(-\theta x^a)) = \exp(n \log \theta + n \log a + (a - 1)\sum \log x_i - \theta \Sigma (x_i^a))
\]

A sufficient statistic for \( \theta \) is \( \Sigma x_i^a \).

(c)

\[
p(x_1, x_2, \ldots, x_n|\theta) = \prod \frac{\theta a^\theta}{x_i^{\theta+1}} = \exp(\Sigma (\log \theta + \theta \log a - (\theta + 1) \log x_i))
\]

A sufficient statistic for \( \theta \) is \( \Sigma \log x_i \).
7. (10 pt) Bickel & Doksum pg 85, problem 1.5.4
Solution:
(a) Suppose a bijection \( f \) exists between \( T_1 \) and \( T_2 \). Then \( T_1(x) = T_1(y) \Rightarrow f(T_1(x)) = f(T_1(y)) \Rightarrow T_2(x) = T_2(y) \). And the reverse is also true because \( f \), being a bijection, has an inverse.

(b) Let \( f(x) = \log(x) \). Then \( f(\prod x_i) = \log(\prod x_i) = \Sigma \log x_i \). \( f \) is a bijection, so these are equivalent statistics.

(c) These are not equivalent statistics. Here’s a counterexample: If \( x_1 = 1 \) and \( x_2 = 4 \) then \( \Sigma x_i = 5 \). And if \( y_1 = 2 \) and \( y_2 = 3 \), then we also have \( \Sigma y_i = 5 \). But
\[
\log 1 + \log 4 \neq \log 2 + \log 3
\]
So the information contained in these two statistics is not identical.

(d) We can uniquely express the pair \([\Sigma x_i, \Sigma(x_i - \bar{x})^2]\) in terms of the pair \([\Sigma x_i, \Sigma x_i^2]\).
A little algebra shows that the reverse is also uniquely possible, with \( \Sigma x_i^2 = \Sigma (x_i - \bar{x})^2 + \frac{1}{n}(\Sigma x_i)^2 \). Thus, a bijection between these two statistics exist, and they are therefore equivalent.

(e) These are not equivalent statistics. Here’s an example, let’s begin with \( x_1 = 3, x_2 = 0, x_3 = 0 \), then we have \( T_1 = (3,27) \) and \( T_2 = (3,6) \); now another set \( 3,1,-1 \), then again \( T_1 = (3,27) \) but \( T_2 = (3,0) \). More generally, to obtain a counterexample, let’s begin with \( x_1 = 1, x_2 = 2, \) and \( x_3 = 3 \). We have \( \Sigma x_i = 6 \) and \( \Sigma x_i^3 = 36 \). That system has two equations but three variables, and therefore more than one distinct solution to construct a counter example.

8. (10 pt) Bickel & Doksum pg 86, problem 1.5.14
Solution:
(a) By definition, the posterior distribution of \( \theta \) given \( X \) is \( \pi(\theta|X) = \frac{\pi(\theta)\pi(X|\theta)}{\int_\theta \pi(\theta)\pi(X|\theta)\,d\theta} \)
If the posterior distribution depends upon \( X \) only through \( T(X) \), then the likelihood function \( \pi(X|\theta) \) must also depend upon \( X \) only through \( T(X) \), for it is only in the likelihood function that \( X \) appears in the expression for the posterior distribution of \( \theta \) given \( X \). And if the likelihood function \( \pi(X|\theta) \) depends upon \( X \) only through \( T(X) \), then all \( X \) having a common \( T(X) \) belong to the same information class, and their distribution given \( T(X) \) will not depend upon \( \theta \). Thus, in that case, \( T(X) \) is a sufficient statistic.
(b) If $T(X)$ is a sufficient statistic for $\theta$, then by the factorization theorem, the
distribution of $X$ given $\theta$ can be factored thus: $\pi(X|\theta) = g(T(X), \theta)h(X)$. But
then the posterior distribution of $\theta$ given $X$ can be expressed in a similar form,
by substitution: $\pi(\theta|X) = \int g(T(X), \theta)h(X)d\theta$. Thus, the posterior distribution
of $\theta$ given $X$ depends upon $X$ only through $T(X)$.

9. (10 pt) Bickel & Doksum pg 87, problem 1.6.2
Solution:

(a) 
\[
\begin{align*}
\pi(x|\theta) & = \prod_{i=1}^{n} \theta x_i^{\theta-1}I_{(1,\infty)}(x_{1:n}) \\
& = \exp(n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i)I_{(1,\infty)}(x_{1:n}) \\
\eta(\theta) & = \theta - 1 \\
B(\theta) & = -n \log \theta \\
T(x) & = \sum_{i=1}^{n} \log x_i \\
h(x) & = I_{(0,1)}(x_{1:n})
\end{align*}
\]

(b) 
\[
\begin{align*}
\pi(x|\theta) & = \prod_{i=1}^{n} ax_i^{a-1} \exp(-\theta x_i^a)I_{(0,\infty)}(x_{1:n}), (a \text{ is a constant}) \\
& = \exp(n \log \theta - \theta \sum_{i=1}^{n} x_i^a)ax^n_{\prod_{i=1}^{n} x_i^{a-1}}I_{(0,\infty)}(x_{1:n}) \\
\eta(\theta) & = -\theta \\
B(\theta) & = -n \log \theta \\
T(x) & = \sum_{i=1}^{n} x_i^a \\
h(x) & = a^n x^n_{\prod_{i=1}^{n} x_i^{a-1}}I_{(0,\infty)}(x_{1:n})
\end{align*}
\]

(c) 
\[
\begin{align*}
\pi(x|\theta) & = \prod_{i=1}^{n} \frac{\theta a^\theta}{x_i^{\theta+1}}I_{(a,\infty)}(x_{1:n}) \\
& = \exp(n(\log \theta + \theta \log a) - (\theta + 1) \sum_{i=1}^{n} \log x_i)I_{(a,\infty)}(x_{1:n}) \\
\eta(\theta) & = -(\theta + 1) \\
B(\theta) & = -n \log \theta - n \theta \log a \\
T(x) & = \sum_{i=1}^{n} \log x_i \\
h(x) & = I_{(a,\infty)}(x_{1:n})
\end{align*}
\]
10. (10 pt) Bickel & Doksum pg 87, problem 1.6.4
Solution:

(a) The family of uniform distributions is not an exponential family. If $X$ has pdf $\pi(x|\theta) = \frac{1}{\theta}1_{[0,\theta]}(x)$, then $\pi(x|\theta)$ takes the value 0 over a part of its domain that depends upon $\theta$. But if a function $f$ is in an exponential family expressed as $f(x|\theta) = h(x)e^{T(x)\eta(\theta) - B(\theta)}$, then the only way the function can take the value 0 is if $h(x) = 0$, which cannot depend upon $\theta$.

(b) Rewriting this in a different form shows that it is a triangular function: $\pi(x|\theta) = \frac{2}{\theta^2} x$ over $x \in (0, \theta)$ and $\pi(x|\theta) = 0$ elsewhere. This is not an exponential family, for the same reason as the uniform distribution in (a). The locations where $\pi(x|\theta) = 0$ depends upon $\theta$, but in an exponential family, such a set of locations could only depend upon $x$.

(c) Once again, this function has probability mass equal to zero at lots of locations, but they depend upon $\theta$. So it cannot be in any exponential family.

(d) WHOOPS! Sorry we didn’t catch this earlier—this is kind of a trick question. In §1.6.3 Bickel & Doksum define “curved exponential families,” where (as in this example) the parameter space $\Theta$ has lower dimension ($l = 1$) than the natural parameter $\eta(\theta)$ or natural sufficient statistic $T(X)$ (both $k = 2$). The authors would say this is not an exponential family, since $l < k$. I feel the distinction is subtle and misleading—so we will treat either “yes” (with $k = 2$ and $\eta, T$ as below) or “no” (with the explanation that $l < k$) as correct answers. If you answered “no” and did not receive credit for this problem, please bring your homework to either Zhenglei or to me for us to correct our grading error. Meanwhile, the “yes” answer we’ll accept is:

- Yes, this is an exponential family.

$$
\pi(x|\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp \left( -\frac{1}{2\theta^2}(x - \theta)^2 \right)
$$

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

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

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

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

$$
= \exp \left( -\frac{1}{2} \log(2\pi) - \frac{1}{2} \right) \exp \left( [x, x^2] \cdot \left[ \frac{1}{\theta}, -\frac{1}{2\theta^2} \right]' - \log(\theta) \right)
$$

Here we have $\eta(\theta) = \left[ \frac{1}{\theta}, -\frac{1}{2\theta^2} \right]'$, $T(x) = [x, x^2]$, $B(\theta) = \log \theta$, and $h(x) = \exp \left( -\frac{1}{2} \log(2\pi) - \frac{1}{2} \right)$.
(e) \( \pi(x|\theta) = \frac{2(x+\theta)}{1+2\theta} = \exp(\log 2 + \log(x+\theta) - \log(1+2\theta)) \)

The \( \log(x+\theta) \) part of this expression cannot be factored into a \( T(x) \) part and an \( \eta(\theta) \) part. So this family is not an exponential family.

(f) \[
\pi(x|\theta) = \frac{{n \choose x} \theta^x (1-\theta)^{n-x}}{1-(1-\theta)^n} = \frac{{n \choose x} \cdot (1-\theta)^n \cdot \left( \frac{\theta}{1-\theta} \right)^x}{1-(1-\theta)^n} = \frac{{n \choose x} \exp \left( \log \left( \frac{(1-\theta)^n}{1-(1-\theta)^n} \right) + x(\log \theta - \log(1-\theta)) \right)}{1-(1-\theta)^n} \]

This is an exponential family. \( \eta(\theta) = \log \theta - \log(1-\theta) \), \( T(x) = x \), \( B(\theta) = -\log \left( \frac{(1-\theta)^n}{1-(1-\theta)^n} \right) \), and \( h(x) = \frac{n \choose x} \).