1. This problem concerns the Inverse Gaussian distribution. Let $\Phi$ denote the standard normal CDF and consider the function

$$F(y) = \begin{cases} 0, & y \leq 0, \\ \Phi\left(\sqrt{\frac{\lambda}{y}}(-1 + \frac{y}{\mu})\right) + e^{2\lambda/\mu}\Phi\left(-\sqrt{\frac{\lambda}{y}}\left(1 + \frac{y}{\mu}\right)\right), & y > 0. \end{cases}$$

(a) Show that $F$ has density $f$ given by

$$f(y) = \begin{cases} 0, & y \leq 0, \\ \left(\frac{\lambda}{2\pi y^{3}}\right)^{1/2}\exp\left\{-\frac{\lambda(y-\mu)^{2}}{2\mu^{2}y}\right\}, & y > 0. \end{cases}$$

(b) Show that the density can be written in exponential dispersion form. Identify: the canonical parameter $\theta$ and the dispersion parameter $\phi$ (in terms of $\lambda$ and $\mu$); the cumulant function, $b(\theta)$; the canonical link; and the variance function for this model.

2. The idea behind this exercise is to walk you through a way to construct an exponential dispersion family and to derive some basic properties of such a family. The properties that you will derive (and particularly the results of parts 2e through 2g) can be derived directly from the form of the exponential dispersion distribution given in class, so the construction given here is arguably just a curiosity, though it is at least a fairly interesting curiosity. We begin with a review of some basic terminology and results.

Let $Q$ represent a probability distribution (i.e., a probability measure) on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ represents the Borel subsets of $\mathbb{R}$, and let $Y$ be a random variable with distribution $Q$, i.e., $P(Y \in A) = Q(A)$, $A \in \mathcal{B}$. (Don’t worry too much if you don’t know about the Borel sets: suffice it to say that not all subsets $A \subset \mathbb{R}$ are Borel sets, but you’re unlikely to come up with one that isn’t without working pretty hard at it.) The moment generating function (mgf) of $Q$ (or equivalently of $Y$) is defined to be $M(t) = E(e^{tY}) = \int e^{ty} Q(dy)$. Because $e^{ty} > 0$, this integral is well-defined for all $t$, but it may be infinite. Let $\Theta = \{t : M(t) < \infty\}$. Of course $M(0) = 1$ and thus $0 \in \Theta$ is assured.

(a) Show that $\Theta$ is a convex subset of $\mathbb{R}$. ○

The only convex, nonempty subsets of $\mathbb{R}$ are intervals (here we count a singleton $\{t\}$ as an interval through the representation $\{t\} = [t, t]$), and hence $\Theta$ is an interval. It can be shown (see Billingsley, Probability and Measure, 3rd edn) that $M(t)$ is differentiable at all $t \in \Theta^o$ (the interior of $\theta$) and that its derivatives can be computed by differentiating across the expectation, i.e., $M^{(j)}(t) = E(Y^j e^{tY})$ for all $t \in \Theta^o$. If
0 ∈ Θ°, then Y has finite moments of all orders given by \( E(X^j) = M^{(j)}(0) \) and its distribution \( Q \) is determined (i.e., uniquely identified) by either of its moment generating function or its sequence of moments.

**Henceforth we will assume that 0 ∈ Θ°.**

The cumulant generating function (cgf) of \( Q \) (or \( Y \)) is defined to be \( K(t) = \log M(t) \). Since \( M(t) > 0 \) for all \( t \), the set of values \( t \) for which \( K(t) \) is finite is exactly \( \Theta \). Of course since \( M(t) = e^{K(t)} \) determines the distribution, so does \( K(t) \).

The \( j \)th cumulant of the distribution \( Q \) is most easily defined as \( \kappa_j = K^{(j)}(0) \) (this is guaranteed to work only because we have assumed \( 0 ∈ Θ° \)). If \( \alpha_j = E Y^j \) represents the \( j \)th moment of \( Y \) and \( \mu_j = E[(Y − μ)^j] \) its \( j \)th central moment (where \( μ = \alpha_1 \) represents the mean), then it is easily shown that the first few cumulants are given by

\[
\begin{align*}
\kappa_1 &= \alpha_1 = \mu_1 = μ \quad \text{(the mean)} \\
\kappa_2 &= \alpha_2 - \alpha_1^2 = \mu_2 = \sigma^2 \quad \text{(the variance)} \\
\kappa_3 &= \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 = \mu_3 \\
\kappa_4 &= \alpha_4 - 4\alpha_1\alpha_3 - 3\alpha_2^2 + 12\alpha_1^2\alpha_2 - 6\alpha_1^4 = \mu_4 - 3\mu_2^2.
\end{align*}
\]

Obviously we could similarly express the moments in terms of the cumulants, so the distribution \( Q \) is determined by its cumulants. In particular, the normal distribution is characterized by the fact that all its cumulants of order three and higher are zero. For what its worth, the coefficients of skewness and kurtosis are defined to be \( \kappa_3/\kappa_2^{3/2} = \mu_3/\sigma^3 \) and \( \kappa_4/\kappa_2^2 = \mu_4/\sigma^4 - 3 \), respectively, and of course these are both zero for a normal distribution.

We now outline Jørgensen’s construction of an exponential dispersion family. Let \( Q^* \) be a probability distribution on \( \mathbb{R} \) with cumulant generating function \( b(t) \), let \( \Theta = \{ t : b(t) < ∞ \} \), and let

\[ Λ = \{ λ > 0 : \exists \text{ a probability distribution } Q^*_λ \text{ with cgf } λb(t) \}. \]

The set \( Λ \) of allowable values of \( λ \) is called the index set. Of course \( Q^*_1 = Q^* \).

(b) Prove that if \( λ_i \in Λ, i = 1, \ldots, n \), then \( λ = \sum_{i=1}^n λ_i \in Λ \), i.e., \( Λ \) is closed under the formation of (finite) sums. In particular, since \( 1 \in Λ \), this implies that \( \{1, 2, \ldots\} \subset Λ \). **Hint:** Consider \( X = \sum_{i=1}^n X_i \), where \( X_1, \ldots, X_n \) are independent with \( X_i \sim Q^*_λ \).

Now for \( λ ∈ Λ \) and \( θ ∈ Θ \), let \( Q^*_λ,θ \) be the distribution having density \( e^{xθ − λb(θ)} \) with respect to \( Q^*_λ \), i.e.,

\[ Q^*_λ,θ(A) = \int_A e^{xθ − λb(θ)} Q^*_λ(dx), \quad A ∈ \mathcal{A}. \]

(1)

(c) Verify that \( Q^*_λ,θ \) is a probability distribution for all \( θ ∈ Θ \) and \( λ ∈ Λ \) and find its cumulant generating function. **Hint:** Since \( e^{xθ − λb(θ)} \) is a nonnegative, measurable function of \( x \), (1) defines a measure on \( (\mathbb{R}, \mathcal{A}) \). Thus, to verify that \( Q^*_λ,θ \) is a probability distribution you need only show that \( Q^*_λ,θ(\mathbb{R}) = 1 \).
The method exemplified by (1), of generating a family of distributions $Q^*_{\lambda, \theta}$ from a single distribution $Q^*_{\lambda}$, is called (exponential dispersion) tilting. Of course in the present case, because $\lambda$ also varies over $\Lambda$, we have defined a two-parameter family of distributions, $\{Q^*_{\lambda, \theta} : \lambda \in \Lambda, \theta \in \Theta \}$. Jørgensen calls this an ED* family and writes $X \sim ED^*(\theta, \lambda)$ (ED is for “exponential dispersion” and the star suggests that we’re not done yet).

(d) Suppose that $X_1, \ldots, X_n$ are independent, with $X_i \sim Q^*_{\lambda_i, \theta}$, $\lambda_1, \ldots, \lambda_n \in \Lambda$, $\theta \in \Theta^\circ$. Prove that $X = \sum_{i=1}^n X_i \sim Q^*_{\lambda, \theta}$ where $\lambda = \sum_{i=1}^n \lambda_i \circ$. 

Now suppose that $X \sim Q^*_{\lambda, \theta}$ and let $Y = X/\lambda$ and $\phi = 1/\lambda$, and let $Q_{\phi, \theta}$ represent the distribution of $Y$. Then, by the general change of variable formula (see, e.g., Theorem 16.13 of Billingsley, Probability and Measure (3rd edition), p. 216),

$$Q_{\phi, \theta}(A) = \int_A \exp \left\{ \frac{y \theta - b(\theta)}{\phi} \right\} Q_\phi(A),$$

(2)

where $Q_{\phi}$ is the probability distribution defined by

$$Q_{\phi}(A) = Q^*_{1/\phi}(A/\phi), \quad A \in \mathcal{R},$$

(3)

with $A/\phi = \{x/\phi : x \in A\}$.

You can just trust me on equations (2) and (3) if you want to, but here is a (perhaps overly) detailed derivation. Let $h_{\lambda}(x) = x/\lambda$, so that $h^{-1}_{\lambda}(y) = \lambda y$ and, for $A \subset \mathbb{R}$, $h^{-1}_{\lambda}(A) = \lambda A = \{\lambda x : x \in A\}$. Then

$$P(Y \in A) = P(X \in \lambda A) = Q^*_{\lambda, \theta}(\lambda A) = \int_{\lambda A} e^{x\theta - \lambda b(\theta)} Q^*_\lambda(dx)$$

$$= \int_{h^{-1}_{\lambda}(A)} e^{\lambda h_{\lambda}(x)\theta - \lambda h_{\lambda}(b(\theta))} Q^*_{\lambda}(dx) = \int_A e^{\lambda y\theta - \lambda h_{\lambda}(b(\theta))} (Q^*_\lambda \circ h^{-1}_{\lambda})(dy)$$

$$= \int_A e^{[y\theta - b(\theta)]/\phi} (Q^*_{1/\phi} \circ h^{-1}_{1/\phi})(dy) = \int_A e^{[y \theta - b(\theta)]/\phi} Q_{\phi}(dy),$$

where $Q_{\phi}(A) = (Q^*_{1/\phi} \circ h^{-1}_{1/\phi})(A) = Q^*_{1/\phi}(h^{-1}_{1/\phi}(A)) = Q^*_{1/\phi}(A/\phi)$. (The equality in the middle of the second line in the display above is where the change of variables theorem is applied; the rest is really just notation. The change of variables theorem itself is a general version of the familiar “Jacobian of the transformation” stuff for deriving the density of a transformed random vector, or more precisely, the familiar Jacobian stuff is a special case of the general change of variables theorem.)

(e) Find the cumulant generating function of $Q_{\phi, \theta}$ and show that the $j$th cumulant is given by

$$\kappa_j = \phi^{j-1} b^{(j)}(\theta)$$

as long as $\theta \in \Theta^\circ$. In particular, this implies that if $Y \sim Q_{\phi, \theta}$, then $\mu := E(Y) = b'(\theta)$ and $\sigma^2 := \text{Var}(Y) = \phi b''(\theta) = \phi b'(b^{-1}(\mu)) = \phi V(\mu)$, as derived in class by other means. Hint: You can save some work by using your answer to part 2c and the transformation $Y = X/\lambda$. Do express your final answer in terms of $\phi$ and $\theta$, though, not $\lambda$. \(\circ\)
The family of distributions so defined, \( \{ Q_{\phi, \theta} : 1/\phi \in \Lambda, \theta \in \Theta \} \), is called an exponential dispersion (ED) family, and Jørgensen writes \( Y \sim ED(\theta, \phi) \). Again, \( \Lambda \) is called the index set and \( \Theta \) is called the canonical parameter domain. Equation (2) says that \( Q_{\phi, \theta} \) has density \( \exp\{ [y\theta - b(\theta)]/\phi \} \) with respect to \( Q_\phi \). To see more clearly the connection with the expression given in class for a density of the exponential dispersion form, we consider the two most important special cases:

- If \( Q_x^\lambda \) has density \( a^\ast(x; \lambda) \) with respect to Lebesgue measure, then \( Q_\phi \) has density \( a(y; \phi) = (1/\phi)a^\ast(y/\phi; 1/\phi) \) with respect to Lebesgue measure (you can check this using the usual Jacobian stuff applied to the transformation \( x \mapsto y = \phi x \) with \( \phi = 1/\lambda \)). Thus in this case,

  \[
  Q_{\phi, \lambda}(A) = \int_A \exp\left\{ \frac{y\theta - b(\theta)}{\phi} \right\} a(y; \phi) \, dx = \int_A \exp\left\{ \frac{y\theta - b(\theta)}{\phi} + c(y; \phi) \right\} \, dx
  \]

  where \( c(y; \phi) = \log a(y; \phi) \), i.e., \( Q_{\phi, \lambda} \) has density

  \[
  f(y; \theta, \phi) = \exp\left\{ \frac{y\theta - b(\theta)}{\phi} + c(y; \phi) \right\}, \quad x \in \phi D,
  \]

  with respect to Lebesgue measure.

- If \( Q_x^\lambda \) has probability mass function \( a^\ast(x; \lambda) \) defined on some countable subset \( D \) of \( \mathbb{R} \) (or to put it more abstractly, if \( Q_x^\lambda \) has density \( a^\ast(x; \lambda) \) with respect to counting measure on \( D \)), then \( Q_\phi \) has probability mass function \( a(y; \phi) = a^\ast(y/\phi; 1/\phi) \) for \( y \) in the set \( \phi D = \{ \phi x : x \in D \} \) (you can check this directly). Thus in this case \( Q_{\phi, \theta} \) has probability mass function

  \[
  f(y; \theta, \phi) = \exp\left\{ \frac{y\theta - b(\theta)}{\phi} \right\} a(y; \phi) = \exp\left\{ \frac{y\theta - b(\theta)}{\phi} + c(y; \phi) \right\}, \quad x \in \phi D,
  \]

  where again, \( c(y; \phi) = \log a(y; \phi) \).

(f) Suppose that \( Y_1, \ldots, Y_n \) are independent, with \( Y_i \sim Q_{\phi_i, \theta} \), where \( \phi_i = \phi/w_i \) and \( \theta \in \Theta^0 \). Let \( w_i = \sum_{i=1}^n w_i \) and let \( \bar{Y} = (1/w) \sum_{i=1}^n w_i Y_i \) be the weighted average of the \( Y_i \)'s. Show that \( \bar{Y} \sim Q_{\phi/w, \theta} \). \( \text{Hint: } \text{Use the result of part 2d.} \)

(g) Part 2f implies in particular that if \( Y_1, \ldots, Y_n \sim \text{i.i.d.} Q_{\phi, \theta}, \theta \in \Theta^0, \) then \( \bar{Y}_n = (1/n) \sum_{i=1}^n Y_i \sim Q_{\phi/n, \theta} \). Use this to argue that if \( Y \sim Q_{\phi, \theta}, \theta \in \Theta^0, \) then

  \[
  \frac{Y - \mu}{\sqrt{\phi V(\mu)}} \xrightarrow{d} N(0, 1) \quad \text{as } \phi \downarrow 0.
  \]  \tag{4}

In other words, for small \( \phi \), an exponential dispersion distribution can be approximated by a normal distribution with mean \( \mu \) and variance \( \phi V(\mu) \). \( \text{Hint: } \text{Appeal to the central limit theorem. The resulting argument does not provide a completely rigorous proof of (4), so you should not worry too much about that. A rigorous proof can be constructed using the mgf.} \).
3. Suppose $Y \sim \text{Gamma}(\mu, \phi)$ (in terms of the usual $(\alpha, \beta) = (\text{shape}, \text{scale})$ parametrization, $\mu = \alpha \beta$ and $\phi = 1/\alpha$). Note that $Y$ has an exponential distribution when $\phi = 1$.

In this problem you are asked in several places to plot the density of $X = h(Y)$ for some smooth, strictly monotone function $h$. Recall that

$$f_X(x) = f_Y(y) \left| \frac{dy}{dx} \right| = f_Y(h^{-1}(x)) \left| \frac{d}{dx} h^{-1}(x) \right| = f_Y(h^{-1}(x)) \frac{1}{h'(h^{-1}(x))}. $$

For plotting purposes, there is no need to actually work with $h^{-1}(x)$. Rather we plot $f_Y(y) / |h'(y)|$ on the vertical axis against $h(y)$ on the horizontal for an appropriate range of $y$ values. Note that for a given $y$ value, if $x = h(y)$, then $y = h^{-1}(x)$ and thus

$$(h(y), f_Y(y) / |h'(y)|) = (x, f_Y(h^{-1}(x)) / |h'(h^{-1}(x))|),$$

which justifies our method of plotting.

(a) Show that the density of $Y/\mu$ does not depend on the value of $\mu$. This will be useful in parts (b), (d), and (e) because it implies that the graphs there do not depend on the value of $\mu$.

(b) Evaluate the normal approximation to the distribution of $Y$ by plotting the density of $(Y - \mu) / \sqrt{\phi \mu}$ for various values of $\phi$.

(c) Derive the moment generating function of $X = \log Y$ and hence find formulas for $E(X)$ and $\text{var}(X)$.

(d) Plot the density of $(X - E(X)) / \sqrt{\text{var}(X)}$ for the values of the dispersion parameter considered in part (a). Hence evaluate the normal approximation to the distribution of $X$.

(e) Plot the density of the signed (and scaled) likelihood root statistic

$$r = \text{sign}(Y - \mu) \sqrt{\frac{2}{\phi} \left\{ -\log \left( \frac{Y}{\mu} \right) + \frac{Y - \mu}{\mu} \right\}}$$

for the values of the dispersion parameter considered in part (a). Hence evaluate the normal approximation to the distribution of $r$. (Note that $r$ is the scaled deviance residual for a single observation, but using the true value of $\mu$ rather than $\hat{\mu}$.)
