A. Starting from the "standard" form of each PDF/PMF, show that the following distributions are in an exponential family, and find the corresponding b, c, θ, and $a(\phi)$.

(i) $Y \sim \mathbf{N}(\mu, \sigma^2)$ **for known** σ^2

Let's begin by writing the PDF for the normal distribution:

$$\begin{aligned}
f(y|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}(y^2 - 2y\mu + \mu^2)\right\} \cdot \exp\left\{\log\left((2\pi\sigma)^{-1/2}\right)\right\} \\
&= \exp\left\{\frac{y^2}{-2\sigma^2} + y\frac{\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma)\right\} \\
&= \exp\left\{\frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2} + \left(-\frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma)\right)\right\}
\end{aligned}$$

By letting $\theta = \mu, a(\phi) = \sigma^2, b(\theta) = \frac{1}{2}\mu^2$, and $c(y|\phi) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma)$, we can write the normal distribution with fixed variance in the form of an exponential family:

$$f(y|\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y|\phi)\right\}$$

(ii) $Y = Z/N$**, where** $Z \sim \mathbf{Binom}(N, P)$ **for known** N

We can easily obtain the PMF of Y through a transformation of random variables:

$$\begin{aligned}
P\left(Y = \frac{Z}{N}\right) &= P\left(Z = NY\right) \\
&= \binom{N}{NY} P^{NY}(1-P)^{N-NY} \\
&= \exp\left\{\log\left[\binom{N}{NY} P^{NY}(1-P)^{N-NY}\right]\right\} \\
&= \exp\left\{\log\left[\binom{N}{NY}\right] + NY\log(P) + N\log(1-P) - NY\log(1-P)\right\} \\
&= \exp\left\{Y\left[N\log\left(\frac{P}{1-P}\right)\right] - N\log\left(\frac{1}{1-P}\right) + \log\left[\binom{N}{NY}\right]\right\}
\end{aligned}$$

Let $\theta = N\log\left(\frac{P}{1-P}\right)$, $b(\theta) = N\log\left(\frac{1}{1-P}\right)$, $a(\phi) = 1$, and $c(y|\phi) = \log\left[\binom{N}{NY}\right]$ to get the form of an exponential family.

(iii) $\sim$ **Pois**(λ)
You know the drill!

$$\begin{aligned} f(y|\lambda) &= \frac{e^{-\lambda}\lambda^y}{y!} \\ &= \exp\left\{\log\left[\frac{e^{-\lambda}\lambda^y}{y!}\right]\right\} \\ &= \exp\left\{-\lambda + y\log(\lambda) - \log(y!)\right\} \\ &= \exp\left\{y\log(\lambda) - \lambda + (-\log(y!))\right\} \end{aligned}$$

The above is in the desired exponential family form since we can let $\theta = \log(\lambda), a(\phi) = 1, b(\theta) = \lambda$, and $c(y|\phi) = -\log(y!)$.

B. We want to characterize the mean and variance of a distribution in the exponential family. To do this, we'll take an unfamiliar route, involving a preliminary lemma. Define the score $s(\theta)$ as the gradient of the log-likelihood with respect to θ:

$$s(\theta) = \frac{\partial}{\partial\theta} \log L(\theta).$$

While we think of the score as a function of θ, clearly the score also depends on the data. So a natural question is: what can we say about the distribution of the score over different random realizations of the data under the true data-generating process, i.e., at the true θ? It turns out we can say the following, sometimes referred to as the score equations:

$$\begin{aligned} E\left[s(\theta)\right] &= 0 \\ \mathcal{I}(\theta) \equiv \mathbf{var}(s(\theta)) &= -E\left[H(\theta)\right], \end{aligned}$$

where the mean and variance are taken under the true θ. Prove these score equations.

First, we prove $E\left[s(\theta)\right] = 0$. Note that while the score is a function of θ, it's also dependent on the data y. Therefore, we can take the expected value of the score over the sample space $\mathcal{Y}$. Let's write out the form of $E\left[s(\theta)\right]$:

$$\begin{aligned} E\left[s(\theta)\right] &= \int_{\mathcal{Y}} s(\theta) f(y|\theta) dy \\ &= \int_{\mathcal{Y}} \frac{\partial}{\partial\theta} \log L(\theta) \cdot f(y|\theta) dy \end{aligned}$$

Now, we use the statistical trick that we can rewrite the likelihood function as a PDF, since we integrate over $\mathcal{Y}$ with PDF $f(y|\theta)$:

$$\begin{aligned} E\left[s(\theta)\right] &= \int_{\mathcal{Y}} \frac{\frac{\partial}{\partial\theta} f(y|\theta)}{f(y|\theta)} \cdot f(y|\theta) dy \\ &= \int_{\mathcal{Y}} \frac{\partial}{\partial\theta} f(y|\theta) dy \\ &= \frac{\partial}{\partial\theta} \int_{\mathcal{Y}} f(y|\theta) dy \\ &= \frac{\partial}{\partial\theta}(1) = 0, \end{aligned}$$

where we assume that any necessary technical conditions are met to switch the order of integration and differentiation.

Now, we prove that $\text{var}(s(\theta)) = -E\left[H(\theta)\right]$. Using the provided hint, suppose we differentiate the first equation with respect to θ^T:

$$\begin{aligned}
\frac{\partial}{\partial \theta^T} E(s(\theta)) &= \frac{\partial}{\partial \theta^T} \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \log L(\theta) f(y|\theta) dy \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta^T} \left[\frac{\partial}{\partial \theta} \log L(\theta) f(y|\theta) \right] dy \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \log L(\theta) \cdot \frac{\partial}{\partial \theta^T} f(y|\theta) + f(y|\theta) \frac{\partial^2}{\partial \theta^T \theta} \log L(\theta) dy \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \log L(\theta) \cdot \frac{\partial}{\partial \theta^T} L(\theta) dy + \int_{\mathcal{Y}} f(y|\theta) \frac{\partial^2}{\partial \theta^T \theta} \log L(\theta) dy \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \log L(\theta) \cdot \frac{\partial}{\partial \theta^T} \log L(\theta) \cdot f(y|\theta) dy + E\left[H(\theta)\right] \\
&= E\left[\frac{\partial}{\partial \theta} \log L(\theta) \cdot \frac{\partial}{\partial \theta^T} \log L(\theta) \right] + E\left[H(\theta)\right] \\
&= E\left[s(\theta) s(\theta)^T \right] + E\left[H(\theta)\right] \\
&\overset{set}{=} \frac{\partial}{\partial \theta^T}(0) = 0
\end{aligned}$$

Before we get to the big reveal, let's acknowledge the nice property that we used to obtain the fifth equality:

$$\begin{aligned}
\frac{\partial}{\partial \theta} \log L(\theta) &= \frac{\frac{\partial}{\partial \theta} L(\theta)}{f(y|\theta)} \\
\implies \frac{\partial}{\partial \theta} L(\theta) &= \frac{\partial}{\partial \theta} \log L(\theta) \cdot f(y|\theta)
\end{aligned}$$

Now, we see from the above that

$$\begin{aligned}
\text{var}\left[s(\theta)\right] &= E[s(\theta)s(\theta)^T] - (E[s(\theta)])^2 \\
&= E[s(\theta)s(\theta)^T] \\
&= -E[H(\theta)]
\end{aligned}$$

C. Use the score equations you just proved to show that, if $Y \sim f(y|\theta, \phi)$ is an exponential family, then

$$E(Y) = b'(\theta)$$
$$\mathbf{var}(Y) = a(\phi)b''(\theta)$$

Thus, the variance of Y is a product of two terms: $b''(\theta)$ depends only on the canonical parameter θ, and hence on the mean, since we showed that $E(Y) = b'(\theta)$; $a(\phi)$ is independent of θ. Note that the most common form of a is $a(\phi) = \phi/w$ where ϕ is called a dispersion parameter and w is a known prior weight that can vary from one observation to another.

Recall that $E[s(\theta)] = E\left[\frac{\partial}{\partial\theta}\log L(\theta)\right] = 0$. For exponential families, we know that the log-likelihood is

$$\begin{aligned}
\log L(\theta) &= \log\left[\prod_{i=1}^{n}\exp\left\{\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i|\phi)\right\}\right] \\
&= \sum_{i=1}^{n}\left[\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i|\phi)\right] \\
&= \frac{\theta}{a(\phi)}\sum_{i=1}^{n} y_i - \frac{nb(\theta)}{a(\phi)} + \sum_{i=1}^{n} c(y_i|\phi)
\end{aligned}$$

By taking the expectation of the gradient of the log-likelihood with respect to θ, we obtain the following:

$$\begin{aligned}
E[s(\theta)] &= \int_{\mathcal{Y}} \frac{\partial}{\partial\theta}\left[\frac{\theta}{a(\phi)}\sum_{i=1}^{n} y_i - \frac{nb(\theta)}{a(\phi)} + \sum_{i=1}^{n} c(y_i|\phi)\right] f(y|\theta)dy \\
&= \int_{\mathcal{Y}}\left[\frac{1}{a(\phi)}\sum_{i=1}^{n} y_i - \frac{nb'(\theta)}{a(\phi)}\right] f(y|\theta)dy \\
&= E\left[\frac{\sum_{i=1}^{n} y_i}{a(\phi)} - \frac{nb'(\theta)}{a(\phi)}\right] \\
&= \frac{1}{a(\phi)}\sum_{i=1}^{n} E(Y) - \frac{nb'(\theta)}{a(\phi)} \\
&\stackrel{set}{=} 0
\end{aligned}$$

By manipulating the above equations, we get

$$E(Y) = b'(\theta)$$

Now, let's obtain the variance of Y:

$$\begin{aligned}
\text{var}(s(\theta)) &= \text{var}\left[\frac{1}{a(\phi)}\sum_{i=1}^{n} y_i - \frac{nb'(\theta)}{a(\phi)}\right] \\
&= \frac{1}{a(\phi)^2}\sum_{i=1}^{n}\text{var}(Y) \\
&\overset{set}{=} -E\left[H(\theta)\right]
\end{aligned}$$

Note that

$$\begin{aligned}
-E[H(\theta)] &= -E\left[\frac{\partial}{\partial\theta^T}\left(\frac{1}{a(\phi)}\sum_{i=1}^{n} y_i - \frac{nb'(\theta)}{a(\phi)}\right)\right] \\
&= -\int_{\mathcal{Y}}\frac{\partial}{\partial\theta^T}\left(\frac{1}{a(\phi)}\sum_{i=1}^{n} y_i - \frac{nb'(\theta)}{a(\phi)}\right) f(y|\theta)dy \\
&= \int_{\mathcal{Y}}\frac{nb''(\theta)}{a(\phi)} f(y|\theta)dy \\
&= E\left[\frac{nb''(\theta)}{a(\phi)}\right] \\
&= \frac{nb''(\theta)}{a(\phi)}
\end{aligned}$$

By combining the two above derivations, we see that

$$\begin{aligned}
\frac{1}{a(\phi)^2}\sum_{i=1}^{n}\text{var}(Y) &= \frac{nb''(\theta)}{a(\phi)} \\
\implies \text{var}(Y) &= a(\phi)b''(\theta)
\end{aligned}$$

D. To convince yourself that your result in (C) is correct, use these results to compute the mean and variance of the $\mathbf{N}(\mu, \sigma^2)$ distribution.

A.

A.

Solving for Beta

Michael R. Schwob*

January 27, 2022

The results below are generated from an R script.

```
library(matlib) # for inv() function
library(matrixcalc) # for LU decomposition
library(microbenchmark) # for comparing the two methods

###
### Inversion Method
###

invertFun <- function(X, y, W){
  beta <- inv(t(X)%*%W%*%X)%*%t(X)%*%W%*%y
  return(beta)
}

###
### Using LU Decomposition
###

solveFun <- function(X, y, W){

  ## Pseudo-code step 1
  decomp <- lu.decomposition(t(X)%*%W%*%X)
  L <- decomp$L
  U <- decomp$U

  ## Pseudo-code step 2
  y <- forwardsolve(L, t(X)%*%W%*%y)

  ## Pseudo-code step 3
  beta <- backsolve(U, y)
  return(beta)
}

###
### Simulate Data and Implement
###

N <- c(800)
P <- c(200)
```

*This report is automatically generated with the R package **knitr** (version 1.37).

```r
for(i in 1:length(N)){

  ## Initialize variables
  n <- N[i]
  p <- P[i]
  W <- diag(n) # identity matrix for W for
  X <- matrix(rnorm(n*p), n, p) # design matrix
  y <- rnorm(n, 0.3*X[,1]+0.5*X[,2], 1)

  ## Implementation
  assign(paste0("benchmark",i),microbenchmark(invertFun(X, y, W), solveFun(X, y, W), times=10)) # save t
}

###
### Print Results
###

for(i in 1:length(N)){
  print(paste0("Benchmark when N=",N[i]," and P=",P[i]))
  print(get(paste0("benchmark",i)))
}
```

```
## [1] "Benchmark when N=800 and P=200"
## Unit: milliseconds
##                expr        min         lq       mean     median         uq        max
##  invertFun(X, y, W) 30428.7280 31680.0271 31664.8722 31772.3117 31918.7712 32158.4157
##   solveFun(X, y, W)   793.4299   797.2425   800.1262   800.1762   803.7212   805.5329
##  neval
##     10
##     10
```

The R session information (including the OS info, R version and all packages used):

```r
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.3 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.9.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.9.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C               LC_TIME=en_US.UTF-8
##  [4] LC_COLLATE=en_US.UTF-8     LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C                  LC_ADDRESS=C
## [10] LC_TELEPHONE=C             LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] microbenchmark_1.4.9 matrixcalc_1.0-5     matlib_0.9.5
```

```
##
## loaded via a namespace (and not attached):
##  [1] rgl_0.108.3       digest_0.6.29     MASS_7.3-55       R6_2.5.1
##  [5] xtable_1.8-4      jsonlite_1.7.3    magrittr_2.0.1    evaluate_0.14
##  [9] highr_0.9         stringi_1.7.6     rlang_0.4.12      carData_3.0-5
## [13] car_3.0-12        tools_4.1.2       stringr_1.4.0     htmlwidgets_1.5.4
## [17] xfun_0.29         abind_1.4-5       fastmap_1.1.0     compiler_4.1.2
## [21] htmltools_0.5.2   knitr_1.37
```

```
Sys.time()
```

```
## [1] "2022-01-27 18:14:39 CST"
```