Written solutions to the homework problems are due on Wednesday October 21, 2015 at the beginning of class.

The homework problems are divided into "regular" and "more involved" problems. In order to facilitate multiple graders, you should hand in these categories of problems separately. That is, hand in one paper that contains only the "regular" problems, and another paper that contains only the "more involved" problems.
As noted on the syllabus and discussed during the first class, copying of homework solutions is not allowed and, when detected, will be investigated as an infraction of the academy integrity policy of the University of Washington. While it is permissible to discuss problems with other students, TAs, or the instructor in order to learn how to solve a problem, your written solutions must be prepared without directly referencing any notes or solutions derived from other students or sources found on the internet.

REGULAR PROBLEMS

1. Suppose random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. Show that $Y = e^X$ has a log normal distribution.

2. Suppose we transform a random variable X according to $Y = aX + b$. Show that the following parametric distribution families are closed under such a transformation (i.e., that they are special cases of "location-scale families of distributions").

   (a) $X \sim \mathcal{U}(\alpha, \beta)$ has $Y \sim \mathcal{U}(a\alpha + b, a\beta + b)$

   (b) $X \sim \Gamma(\alpha, \beta, A)$ has $Y \sim \Gamma(\alpha, a\beta, aA + b)$

   (c) $X \sim \mathcal{N}(\mu, \sigma^2)$ has $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

3. Let $\vec{X} = (X_1, \ldots, X_n)$ be a random vector in which the X_i are independently distributed with an exponential family distribution having density (probability mass function) of the form

$$f_{\vec{X}}(\vec{x} \mid \vec{\theta}) = h(\vec{x}) \exp\left[\sum_{i=1}^{p} T_i(\vec{x})\eta_i(\vec{\theta}) + A(\vec{\theta})\right].$$

   For each of the following parametric distributions, show whether it does or does not belong to an exponential family by explicitly identifying $h(x)$, $\vec{\eta}(\vec{\theta})$, $\vec{T}(\vec{X})$, and $A(\vec{\theta})$

where possible. For every exponential family distribution, be sure to indicate the dimensionality of the statistic $\vec{T}(\vec{X})$ and whether the exponential family is curved. (Recall that the density (pmf) for a random vector of independent random variables is given by

$$f_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} f_{X_i}(x_i),$$

where f_{X_i} is the density (pmf) for the ith component of the random vector $\vec{X}$.

(a) Bernoulli distribution: $X_i \sim \mathcal{B}(1, p)$ with $\theta = p$.

(b) Binomial distribution with varying replications: $X_i \sim \mathcal{B}(m_i, p)$ with $\theta = p$.

(c) Geometric distribution: $X_i \sim Geom(p)$ with $\theta = p$.

(d) Negative binomial distribution with varying number of events: $X_i \sim NegB(r_i, p)$ with $\theta = p$.

(e) Poisson distribution: $X_i \sim \mathcal{P}(\lambda)$ with $\theta = \lambda$.

(f) Poisson distribution with varying time of follow-up: $X_i \sim \mathcal{P}(\lambda t_i)$ with $\theta = \lambda$ (t_i's are known constants).

(g) Uniform distribution: $X_i \sim \mathcal{U}(0, \theta)$.

(h) Exponential distribution (hazard parameterization): $X_i \sim \mathcal{E}(\lambda)$ with $\theta = \lambda$.

(i) Exponential distribution (mean parameterization): $X_i \sim \mathcal{E}(\mu)$ with $\theta = \mu$.

(j) Gamma distribution (unshifted): $X_i \sim \Gamma(\alpha, \beta, A = 0)$ with $\theta = (\alpha, \beta)$.

(k) Normal distribution with known variance: $X_i \sim \mathcal{N}(\mu, \sigma^2)$ with $\theta = \mu$.

(l) Normal distribution with unknown variance: $X_i \sim \mathcal{N}(\mu, \sigma^2)$ with $\vec{\theta} = (\mu, \sigma^2)$.

(m) Normal distribution with specified mean-variance relationship: $X_i \sim \mathcal{N}(\mu, \mu)$ with $\theta = \mu$.

(n) Normal distribution with specified mean-variance relationship: $X_i \sim \mathcal{N}(\mu, \mu^2)$ with $\theta = \mu$.

(o) Normal distribution with systematically varying means: $X_i \sim \mathcal{N}(c_i\mu, \sigma^2)$ with $\vec{\theta} = (\mu, \sigma^2)$ (c_i's are known constants).

(p) Normal distribution with systematically varying variances: $X_i \sim \mathcal{N}(\mu, c_i\sigma^2)$ with $\vec{\theta} = (\mu, \sigma^2)$ (c_i's are known constants).

(q) Lognormal distribution: $X_i \sim \mathcal{LN}(\mu, \sigma^2)$ with $\vec{\theta} = (\mu, \sigma^2)$.

(r) Weibull distribution: $X_i \sim Weib(p, \lambda)$ with $\vec{\theta} = (p, \lambda)$

MORE INVOLVED PROBLEMS

4. Let X be a continuous random variable with density $f_X(x\,|, \theta)$ for some real θ. Furthermore, suppose

   - $A = \{x : f_X(x\,|\,\theta) > 0\}$ (the "support" of the distribution of X) does not depend on θ, and
   - we can twice interchange the order of integration with respect to x and differentiation with respect to θ, so

$$\frac{\partial}{\partial\theta}\int f_X(x\,|\,\theta)\,dx = \int\left(\frac{\partial}{\partial\theta}f_X(x\,|\,\theta)\right)\,dx$$
$$\frac{\partial^2}{\partial\theta^2}\int f_X(x\,|\,\theta)\,dx = \int\left(\frac{\partial^2}{\partial\theta^2}f_X(x\,|\,\theta)\right)\,dx$$

   Consider the "efficient score" (transformation)

$$U(X) = \frac{\partial}{\partial\theta}\log\left(f_X(X\,|\,\theta)\right)$$

   (a) Find the expectation

$$\mu_U = E[U(X)] = \int_{-\infty}^{\infty} U(x)\,f_X(x\,|\,\theta)\,dx.$$

   (b) Show that

$$Var(U(X)) = \int_{-\infty}^{\infty}\left(U(x) - \mu_U\right)^2\;f_X(x\,|\,\theta)\,dx = -E\left[\frac{\partial^2}{\partial\theta^2}\log\left(f_X(X\,|\,\theta)\right)\right].$$

   (Note that this problem can be solved in the exact same way if X were a discrete random variable by substituting sums for integrals.)

5. Let X be a normally distributed random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ with σ^2 known and unknown $\theta = \mu$.

   (a) Derive the efficient score U (as given in the previous problem) for this distribution.

   (b) Now suppose that the efficient score derived from the normal distribution were used to transform some other random variable Y having mean μ and variance τ^2. What is are the mean and variance of $U(Y)$?