MATH562 – Fall 2025

Problem Set: Week 3

1. (**Exponential families**) Let $X = (X_1, \cdots, X_n)$ be a random sample from a distribution with density $f_{X_i}(x;\theta)$, where $\theta \in \Theta$ is an unknown parameter. For each of the five densities below, identify the parametric family and determine whether it forms an exponential family.

   (a) $f_{X_i}(x;\theta) = \theta x^{\theta-1}\mathbf{1}(x \in (0,1))$, with $\theta \in \Theta = \mathbb{R}_{++}$.

   (b) $f_{X_i}(x;\theta) = \frac{1}{\theta}\exp(-\frac{x}{\theta})\mathbf{1}(x \in (0,\infty))$, with $\theta \in \Theta = \mathbb{R}_{++}$.

   (c) $f_{X_i}(x;\theta) = \frac{1}{2}\exp(-|x-\theta|)$, with $\theta \in \Theta = \mathbb{R}$.

   (d) $f_{X_i}(x;\theta) = \exp(-(x-\theta))\mathbf{1}(x \in (\theta,\infty))$, with $\theta \in \Theta = \mathbb{R}$.

   (e) $f_{X_i}(x;\theta) = \frac{\theta^x}{x!}\exp(-\theta)\mathbf{1}(x \in \mathbb{Z}_+)$, with $\theta \in \Theta = \mathbb{R}_+$.

2. Let $Y_1, \dots, Y_n \overset{i.i.d.}{\sim} (\mu, \sigma^2)$, let $\bar{Y} = n^{-1}\sum_{j=1}^n Y_j$ and $S^2 = (n-1)^{-1}\sum_{j=1}^n (Y_j - \bar{Y})^2$.

   (a) Verify that $(n-1)S^2 = \sum_{j=1}^n Y_j^2 - n\bar{Y}^2$.

   (b) Show that $\mathrm{Var}(\bar{Y}) = \sigma^2/n$, and by writing $\sum_{j=1}^n (Y_j - \bar{Y})^2 = \sum_{j=1}^n \{Y_j - \mu - (\bar{Y} - \mu)\}^2$ and expanding, show that $\mathrm{E}(S^2) = \sigma^2$.

   (c) Show that an alternative form for S^2 is $\{2n(n-1)\}^{-1}\sum_{j,k=1}^n (Y_j - Y_k)^2$.

3. (a) If the estimators $T_1, \dots, T_n$ are uncorrelated with common mean θ and known variances $v_1, \dots, v_n$, find the unbiased estimator $\hat{\theta} = \sum_j a_j T_j$ that has minimum variance.

   (b) Show that if the T_j are normally distributed then $\hat{\theta}$ is the maximum likelihood estimator, and discuss how it should be modified if $\mathrm{Var}(T_j) = \sigma^2 v_j$ for each j, with σ^2 unknown.

   (c) How should the estimator of σ^2 in (b) be modified if it is believed that $\sigma^2 \geq 1$?

4. Eggs are thought to be infected with a bacterium *salmonella enteriditis*, so that the number of organisms, Y, in each egg has a Poisson distribution with mean μ. The value of Y cannot be observed directly, but after a period it becomes certain whether the egg is infected ($Y > 0$) or not ($Y = 0$). Out of m such eggs, r are found to be infected. Find the maximum likelihood estimator $\hat{\mu}$ of μ and its asymptotic variance. Is the exact variance of $\hat{\mu}$ defined?

5. The score-matching estimator of a parameter θ corresponds to the population expression

$$\theta_g = \arg\min_{\theta'} \mathrm{E}\left[\{\nabla_y \log f(Y;\theta') - \nabla_y \log g(Y)\}^2 w(Y)\right],$$

   where $w(y)$ is a positive weight function, $\nabla_y \cdot$ denotes $\mathrm{d}\cdot/\mathrm{d}y$ and $Y \sim g$.

   (a) Show that if $g(y) = f(y;\theta)$ and the density f is identifiable, i.e., no two parameter values give the same density, then the minimum is achieved when the expression above equals zero, and then $\theta_g = \theta$.

   (b) If $w(y) \equiv 1$, find the parameters that minimise

$$\mathrm{E}\left[\{\nabla_y \log f(Y;\theta)\}^2 + 2\nabla_y^2 \log f(Y;\theta)\right],$$

   when (i) f is the $\mathbb{N}(\eta, \tau^2)$ density and and $Y \sim \mathbb{N}(\mu, \sigma^2)$, (ii) f is the $\exp(\lambda')$ density and $Y \sim \exp(\lambda)$. In the case of (i) also find the empirical estimators. Discuss.

(c) Show that if $w(y)$ is chosen so that $w(y)g(y)\nabla_y \log f(y;\theta) = 0$ at the limits of integration for y, then score-matching amounts to minimising

$$\mathrm{E}\left[w(Y)\left\{\nabla_y \log f(Y;\theta)\right\}^2 + 2w(Y)\nabla_y^2 \log f(Y;\theta) + 2\nabla_y w(Y)\nabla_y \log f(Y;\theta)\right] \tag{1}$$

and give the empirical version of this expression. Does setting $w(y) = y$ fix the problem in (b)(ii)?