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 $\theta \in \Theta = \mathbb{R}_{++}$.
   Solution: $X_i \sim \text{Beta}(\theta, 1)$.

   $$\theta x^{\theta-1}\mathbf{1}(x \in (0,1)) = \mathbf{1}(x \in (0,1)) \cdot \theta \cdot \exp\left[(\theta-1)\log x\right] = h(x)c(\theta)\exp\left[w(\theta)t(x)\right]$$

   forms an exponential family.

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

   $$f_X(x_1, ..., x_n) = \frac{1}{\theta^n}\exp\left(-\frac{1}{\theta}\sum_{i=1}^n x_i\right)\mathbf{1}(x \in (0,\infty))$$

   forms an exponential family

   (c) $f_{X_i}(x;\theta) = \frac{1}{2}\exp(-|x-\theta|)$, with $\theta \in \Theta = \mathbb{R}$.
   Solution: $X_i \sim \text{DoubleExponential}(\theta, 1)$. Note that it is not possible to break the absolute value into $w(\theta)t(x)$ and thus this distribution does not form an exponential family.

   (d) $f_{X_i}(x;\theta) = \exp(-(x-\theta))\mathbf{1}(x \in (\theta,\infty))$, with $\theta \in \Theta = \mathbb{R}$.
   Solution: This distribution is the exponential location distribution. Its support is parameter-dependent so it doesn't form an exponential family.

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

   $$\frac{\theta^x}{x!}\exp(-\theta)\mathbf{1}(x \in \mathbb{Z}_+) = \frac{1}{x!}\mathbf{1}(x \in \mathbb{Z}_+)\exp(-\theta)\exp(x\log\theta) = h(x)c(\theta)\exp\left[w(\theta)t(x)\right]$$

   forms an exponential family.

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

   $$\sum_j (Y_j - \bar{Y})^2 = \sum_j (Y_j^2 - 2Y_j\bar{Y} + \bar{Y}^2) = \sum_j Y_j^2 - 2\bar{Y}\sum_j Y_j + n\bar{Y}^2 = \sum_j Y_j^2 - 2n\bar{Y}^2 + n\bar{Y}^2 = \sum_j Y_j^2 - n\bar{Y}^2$$

   as required.

   (b) Show that $\text{Var}(\bar{Y}) = \sigma^2/n$, 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 $\text{E}(S^2) = \sigma^2$.
   Solution: The variance of a sum of independent variables is the sum of the variances, so

   $$\text{Var}(\bar{Y}) = n^{-2}\text{Var}\left(\sum_j Y_j\right) = n^{-2}\sum_j \text{Var}(Y_j) = \sigma^2/n.$$

Moreover $\bar{Y} - \mu$ does not depend on j, so

$$\sum_{j=1}^{n}\{Y_j - \mu - (\bar{Y} - \mu)\}^2 = \sum_j (Y_j - \mu)^2 - 2(\bar{Y} - \mu)\sum_j (Y_j - \mu) + \sum_j (\bar{Y} - \mu)^2 = \sum_j (Y_j - \mu)^2 - n(\bar{Y} - \mu)^2,$$

and this has expectation $\sum_j \mathrm{Var}(Y_j) - n\mathrm{Var}(\bar{Y}) = n\sigma^2 - n\sigma^2/n = (n-1)\sigma^2$, as required.

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

Solution: Writing $Y_j - Y_k = Y_j - \bar{Y} - (Y_k - \bar{Y}) = U_j - U_k$, say, and noting that $\sum_{j=1}^{n} U_j = \sum_{j=1}^{n} Y_j - n\bar{Y} = 0$, we have

$$\sum_{j,k=1}^{n}(Y_j - Y_k)^2 = \sum_{j,k=1}^{n}(U_j^2 + U_k^2 - 2U_jU_k) = 2n\sum_{j=1}^{n}U_j^2 - 2\left(\sum_{j=1}^{n}U_j\right)^2 = 2n(n-1)S^2,$$

as required.

3. (a) If the estimators $T_1, \ldots, T_n$ are uncorrelated with common mean θ and known variances $v_1, \ldots, v_n$, find the unbiased estimator $\hat{\theta} = \sum_j a_j T_j$ that has minimum variance. **Solution**: If $\hat{\theta}$ is unbiased, then we must have $\mathrm{E}(\sum_j a_j T_j) = \sum_j a_j \theta = \theta$ for any possible θ, so $\sum a_j = 1$. Now $\mathrm{Var}(\sum_j a_j T_j) = \sum_j a_j^2 v_j$, and we seek to minimise this subject to $\sum a_j = 1$. The corresponding Lagrangian is

$$\sum_j a_j^2 v_j + \lambda\left(\sum a_j - 1\right),$$

and differentiation with respect to a_j and to λ gives

$$2v_j a_j + \lambda = 0, \quad j = 1, \ldots, n, \quad \sum_j a_j = 1,$$

resulting in $a_j = v_j^{-1}/\sum_i v_i^{-1}$ and $\mathrm{Var}(\hat{\theta}) = (\sum_j v_j^{-1})^{-1}$.

(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.

Solution: If the $T_j \perp\!\!\!\perp \mathcal{N}(\theta, \sigma^2 v_j)$ then the log likelihood function is

$$\ell(\theta, \sigma^2) = -\frac{1}{2}\sum_{j=1}^{n}\left\{\log(2\pi\sigma^2 v_j) + (t_j - \theta)^2/(\sigma^2 v_j)\right\} \equiv -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum(t_j - \theta)^2/v_j,$$

where the second expression ignores additive constants. Differentiation gives

$$\ell_\theta = \frac{1}{\sigma^2}\sum(t_j - \theta)/v_j, \quad \ell_{\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum(t_j - \theta)^2/v_j,$$

in an obvious notation, and it is easy to check that $\hat{\theta}$ is the sole solution of the first equation whatever the value of σ^2, and that $\hat{\sigma}^2 = n^{-1}\sum(t_j - \hat{\theta})^2/v_j$; the corresponding unbiased estimator uses the denominator $n-1$.

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

Solution: In this case we maximise not over $\sigma^2 \in (0, \infty)$ but over $\sigma^2 \geq 1$, so the estimator becomes $\tilde{\sigma}^2 = \max(\hat{\sigma}^2, 1)$.

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?

Solution: The probability of infection is $\Pr(Y > 0) = 1 - e^{-\mu}$, so if the infections are regarded as independent, the likelihood based on the number of infected eggs $R = \sum_{j=1}^m Y_j$, which is binomial with denominator m and probability $1 - e^{-\mu}$, is

$$\binom{m}{r}(1-e^{-\mu})^r(e^{-\mu})^{m-r} = \binom{m}{r}(e^{\mu}-1)^r(e^{-\mu})^m, \quad \mu > 0,$$

where r is the value of R.

The log likelihood $\ell(\mu) = r\log(e^{\mu} - 1) - m\mu$ has first and second derivatives

$$\ell'(\mu) = \frac{re^{\mu}}{e^{\mu}-1} - m, \quad \ell''(\mu) = -r\frac{e^{\mu}}{(e^{\mu}-1)^2},$$

so $e^{-\hat\mu} = (m-r)/m$, giving $\hat\mu = \log\{m/(m-r)\}$, and observed information

$$\jmath(\hat\mu) = -\ell''(\mu)|_{\mu=\hat\mu} = r^2(m-r)/m^2,$$

while the expected information is

$$\imath(\mu) = \mathrm{E}\{-\ell''(\mu)\} = \mathrm{E}(R)e^{\mu}/(e^{\mu}-1)^2 = m/(e^{\mu}-1),$$

because $\mathrm{E}(R) = m\Pr(Y = 1) = m(1 - e^{-\mu})$. The asymptotic distribution of $\hat\mu$ is $\mathbb{N}\{\mu, \imath(\mu)^{-1}\}$.

Alternatively we write $\hat\mu = g(r/m) = -\log(1 - r/m)$, with $g(u) = -\log(1-u)$ and $g'(u) = 1/(1-u)$, note that $R \sim B(m, 1-e^{-\mu})$, so for large m we have $R/m \sim \mathbb{N}\{1-e^{-\mu}, e^{-\mu}(1-e^{-\mu})/m\}$, and apply the delta method, which gives that $g(R/m)$ is approximately normal with mean and variance

$$g(1-e^{-\mu}) = \mu, \quad g'(1-e^{-\mu})^2\mathrm{Var}(R/m) = (e^{\mu}-1)/m.$$

The asymptotic variance is $(e^{\mu}-1)/m > 0$ for any $\mu > 0$, but the exact variance is infinite for any m, because

$$\mathrm{E}(\hat\mu) = \sum_{r=0}^m \log\{m/(m-r)\}\Pr(R = r) = \log m - \sum_{r=0}^m \log(m-r)\Pr(R = r) = \infty,$$

as $\Pr(R = m) = (1-e^{-\mu})^m > 0$ for any μ and m.

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 $\theta_g = \theta$.

Solution: The minimum value of zero is attained when $\nabla_y \log f(y;\theta') - \nabla_y \log f(y;\theta) \equiv 0$, and this clearly occurs when $\theta' = \theta_g$, say. Now suppose that $\nabla_y \log f(y;\theta') - \nabla_y \log g(y) \equiv 0$, and integrate with respect to y to obtain

$$\log f(y;\theta') - \log g(y) = a(\theta')$$

where $a(\cdot)$ is an arbitrary function of θ alone. Hence $g(y)b(\theta') = f(y;\theta')$ for all y, which implies that as both f and g are densities, they must be identical.

(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. **Solution**: (i) In this case the log likelihood is $-\frac{1}{2}\log(2\pi\tau^2) - (y-\eta)^2/(2\tau^2)$, and the two derivatives are $-(y-\eta)/\tau^2$ and $-1/\tau^2$, so the objective function is

$$\mathrm{E}\left\{(Y-\eta)^2/\tau^4 - 2/\tau^2\right\} = \tau^{-4}\left\{\mathrm{E}\left\{(Y-\mu+\mu-\eta)^2\right\} - 2\tau^2\right\} = \tau^{-4}\left\{\sigma^2 + (\mu-\eta)^2 - 2\tau^2\right\},$$

which is clearly minimised by taking $\eta = \mu$ and then setting $\tau^2 = \sigma^2$. The empirical estimators minimise

$$\tau^{-4}\sum_{j=1}^{n}(Y_j - \eta)^2 - 2n/\tau^2 = \tau^{-4}\sum_{j=1}^{n}(Y_j - \bar{Y})^2 + \tau^{-4}n(\bar{Y}-\eta)^2 - 2n/\tau^2,$$

so $\tilde{\eta} = \bar{Y}$ and $\tilde{\tau}^2 = n^{-1}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$; these are the maximum likelihood estimators.

(ii) Here the log likelihood is $\log\lambda' - y\lambda'$ for $\lambda' > 0$, so the two derivatives are $-\lambda'$ and 0 and thus the objective function is $(\lambda')^2$, which is minimised by taking $\tilde{\lambda} = 0$. This is not a sensible solution, and it arises because the argument from the original expression to that above fails (the integration by parts involves another, non-zero, term). The original expression would be $\mathrm{E}\{[-\lambda' - (-\lambda)\}^2] = (\lambda - \lambda')^2$, which is minimised by taking $\lambda' = \lambda$.

(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

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

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

Solution: On writing

$$\{\nabla_y \log f(y;\theta) - \nabla_y \log g(y)\}^2 = \{\nabla_y \log f(y;\theta)\}^2 - 2\nabla_y \log f(y;\theta)\nabla_y \log g(y) + \{\nabla_y \log g(y)\}^2,$$

we see that the population version of the estimator is

$$\theta_g = \arg\min_\theta \int \{\nabla_y \log f(y;\theta)\}^2 w(y)g(y)\ \mathrm{d}y - 2\int \{\nabla_y \log f(y;\theta)\nabla_y \log g(y)\} w(y)g(y)\ \mathrm{d}y,$$

and the previous argument and integration by parts shows that the second integral here is

$$[w(y)\nabla_y \log f(y;\theta)g(y)]_{y_-}^{y_+} - \int \nabla_y\{w(y)\nabla_y \log f(y;\theta)\}g(y)\ \mathrm{d}y,$$

which leads to (1) if $w(y)$ ensures that the first term here equals zero. This is the case with $w(y) = y$, $g(y) = \lambda e^{-\lambda y}$ and $\nabla_y \log f(y;\theta) = -\lambda'$, and then as $\mathrm{E}(Y) = 1/\lambda$, (1) becomes $(\lambda')^2/\lambda - 2\lambda'$, minimisation of which with respect to λ' gives $\lambda' = \lambda$, as expected.