STATISTICAL MODELLING Part IIC / Michaelmas 2025
Example Sheet 3 (of 4)

In questions that follow, by (p-parameter) exponential family we mean the collection of probability distributions with density function

$$f(y;\theta) = e^{\theta^T T(y) - K(\theta)} f_0(y),\ y \in \mathcal{Y},$$

where $\theta \in \Theta \subseteq \mathbb{R}^p$ is the *natural parameter*, $T : \mathcal{Y} \to \mathbb{R}^p$ gives the vector of *sufficient statistics*, and $K(\theta)$ is the *cumulant function*. A (canonical) generalized linear model assumes $Y_i \mid X_i \sim f(y;\theta_i)$ with sufficient statistic $T(y) = y$ and natural parameter $\theta_i = X_i^T \beta$ for $i = 1, \ldots, n$.

1. Show that the following families of distributions are (possibly multi-parameter) exponential families; all parameters are unknown unless stated otherwise. Find the corresponding natural parameters, sufficient statistics, and cumulant functions.

   (a) The normal distribution $\mathrm{N}(\mu, \sigma^2)$:

   $$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, y \in \mathbb{R}.$$

   (b) The negative binomial distribution $\mathrm{NegBin}(k,p)$ with known k:

   $$f(y;p) = \binom{y+k-1}{y} p^k (1-p)^y,\ y = 0, 1, 2, \ldots$$

   This describes the number of failures until k successes are reached in a sequence of independent Bernoulli trials.

   (c) The Gamma distribution $\mathrm{Gamma}(\alpha, \lambda)$:

   $$f(y;\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y},\ y > 0,$$

   where $\Gamma(\cdot)$ is the Gamma-function.

2. Let Y be a real-valued random variable whose moment generating function is finite on an open interval containing zero. Show that the first three cumulants are $\kappa_1 = \mathsf{E}(Y)$, $\kappa_2 = \mathsf{Var}(Y)$, $\kappa_3 = \mathsf{E}(Y - \kappa_1)^3$, respectively. Find the mean and variance of the negative binomial distribution $\mathrm{NegBin}(k,p)$ in terms of k and p.

3. Consider the multinomial distribution $\mathrm{Multinomial}(n, \pi)$ with probability mass function given by

   $$f(y;\pi) = \frac{n!}{y_1! \cdots y_L!} \pi_1^{y_1} \cdots \pi_L^{y_L}$$

   for any $0 \le y_1, \ldots, y_L \le n$ such that $y_1 + y_2 + \cdots + y_L = n$. Suppose n is known but $\pi = (\pi_1, \ldots, \pi_L)$ is unknown.

   (a) Write this as an exponential family.

   (b) We say an exponential family is *minimal* if the sufficient statistics are linearly independent. Is your answer in part (a) minimal? If not, can you write it as an exponential family that is minimal and give its cumulant function? *[Hint: Use* $\log(\pi_1/\pi_L), \ldots, \log(\pi_{L-1}/\pi_L)$ *as natural parameters.]*

4. Suppose $Y_1, \ldots, Y_n$ is an i.i.d. sample from $N(\mu, 1)$. What is the asymptotic distribution of the maximum likelihood estimator of $\mathsf{P}(Y_1 < 0)$?

5. Let $Y_1, \ldots, Y_n$ be independent Poisson random variables with mean μ. Compute the maximum likelihood estimator $\hat{\mu}$. By considering $n\hat{\mu}$, write down the distribution of $\hat{\mu}$ and deduce its asymptotic distribution directly. Verify that this asymptotic distribution agrees with that predicted by the general asymptotic theory for maximum likelihood estimators.

6. The *Box-Cox transformation* refers to the following function of a response variable $Y > 0$:

$$Y \mapsto Y^{(\lambda)} = \begin{cases} (Y^\lambda - 1)/\lambda, & \text{if } \lambda \neq 0, \\ \log Y, & \text{if } \lambda = 0. \end{cases}$$

   One can fit the usual normal linear model after this transformation.

   (a) Read `?MASS::boxcox` and find out how the tuning parameter λ is selected in the implmentation of Box-Cox transformation for linear models in the `MASS` package.

   (b) Look at the `cabbages` data in the `MASS` package (use `?cabbages` to find out about the dataset). Investigate whether the planting date has a significant effect on the weight of the cabbage head. Write out the models you have fitted and explain any conclusions you come to.

7. Suppose $Y_1, \ldots, Y_n$ is an i.i.d. sample from a one-parameter exponential family with natural parameter $\theta \in \Theta \subseteq \mathbb{R}$ and sufficient statistic $T(y) = y$.

   (a) Show that the distribution of $\bar{Y} = \sum_{i=1}^n Y_i/n$ is in an exponential family with natural parameter $\theta^{(n)} = n\theta$. What is the cumulant function $K^{(n)}(\theta^{(n)})$?

   (b) The deviance of θ_1 from θ_2 is defined as

$$D(\theta_1, \theta_2) = 2\mathsf{E}_{\theta_1}\left\{\log \frac{f(Y;\theta_1)}{f(Y;\theta_2)}\right\}.$$

   Let $D^{(n)}(\theta_1, \theta_2)$ denote that the deviance in the exponential family for $\bar{Y}$ with natural parameter $\theta_1^{(n)} = n\theta_1$ from that with $\theta_2^{(n)} = n\theta_2$. Show that $D^{(n)}(\theta_1, \theta_2) = nD(\theta_1, \theta_2)$.

   (c) Show that the likelihood ratio statistic for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, after a monotone transformation, is given by $D^{(n)}(\hat{\theta}, \theta_0)$. What is the limiting distribution of this statistic when $n \to \infty$? Justify your answer using the approximation

$$D(\theta_1, \theta_2) \approx I^{(1)}(\theta_2)(\theta_1 - \theta_2)^2 \text{ when } \theta_1 \approx \theta_2,$$

   where $I^{(1)}(\theta_2)$ is the Fisher information of one observation from the the distribution $f(\cdot; \theta_2)$.

8. Consider a canonical generalized linear model and write $\hat{\mu}_i = g^{-1}(X_i^T\hat{\beta})$ where $\hat{\beta}$ is the maximum likelihood estimate of β. Show that if the model includes the intercept term (a column of the model matrix X is a vector of ones), then

$$\sum_{i=1}^n \hat{\mu}_i = \sum_{i=1}^n Y_i.$$

9. We say a probability distribution is in an *exponential dispersion family* if its density can be written as

$$f(y;\theta,\sigma^2) = e^{\{\theta y - K(\theta)\}/\sigma^2} f_0(y;\sigma^2),$$

where $\theta \in \Theta \subseteq \mathbb{R}$ is called the natural parameter, $\sigma^2 > 0$ is called the dispersion parameter, and $f_0(y;\sigma^2)$ is some density function. Suppose $Y_1,\ldots,Y_n$ is an i.i.d. sample from such a distribution and $\mathsf{Var}(Y_1) > 0$.

   (a) Compute the cumulant generating function of Y_1 and use it to show that $\mathsf{E}(Y_1) = K'(\theta)$ and $\mathsf{Var}(Y_1) = \sigma^2 K''(\theta)$.

   (b) Show that the MLE of $\mathsf{E}(Y_1)$ is given by the sample mean $\bar{Y} = \sum_{i=1}^n Y_i/n$.

   (c) Show that the Gamma distribution is an exponential dispersion family with natural parameter $\theta = -\lambda/\alpha$ and dispersion parameter $\sigma^2 = 1/\alpha$.

10. Let $Y_1,\ldots,Y_n \sim \text{Gamma}(\alpha,\lambda)$ be i.i.d.

   (a) Using the function `optim` in R with option `hessian = TRUE` (or otherwise), write a function `gammaMLE` that returns the maximum likelihood estimator of (α,λ) and its approximate covariance matrix.

   (b) Simulate a dataset with $n = 1000$, $\alpha = 2$, $\lambda = 3$ and use

```
fit <- glm(Y ~ 1, family = Gamma)
```

   to find an estimate of the natural and dispersion parameters of the Gamma distribution (see Question 9c). Do these estimates agree with those obtained by your `gammaMLE`?

   (c) Use the output of your `gammaMLE` to find an approximate confidence interval for the natural parameter of the Gamma distribution. Does your result agree with `summary(fit)`? *Hint: You may find the following (vector version) delta method useful: suppose* $\sqrt{n}(\hat{\eta}_n - \eta) \xrightarrow{d} Z$ *where* $\hat{\eta}_n$ *is a random vector in* $\mathbb{R}^d$ *and* $g : \mathbb{R}^d \to \mathbb{R}^m$ *is continuously differentiable at* η*, then* $\sqrt{n}(g(\hat{\eta}_n) - g(\eta)) \xrightarrow{d} \nabla g(\eta)^T Z$.

11. In the general form of a generalized linearl model, it is assumed that $Y_1,\ldots,Y_n$ are independent and $Y_i \sim f(y;\theta_i,\sigma_i^2)$ follows a distribution from a exponential dispersion family (see Question 9) with the mean and dispersion parameters modelled by

$$g(\mu_i) = X_i^T\beta \quad \text{and} \quad \sigma_i^2 = \sigma^2/w_i,$$

where g is a strictly increasing and twice differentiable function, $\sigma^2 > 0$ is possibly unknown, and $w_i, i = 1,\ldots,n$ are some known weights.

   (a) Show that the β-score is given by

$$U_\beta(\beta,\sigma^2) = \nabla_\beta\, l(\beta,\sigma^2) = \frac{1}{\sigma^2} X^T W R, \;\; R = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix} \text{ with } R_i = g'(\mu_i)(Y_i - \mu_i),$$

   for some diagonal matrix W that you must specify.

   (b) When σ^2 is unknown, explain why the following might be a reasonable estimator of σ^2:

$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^n w_i \frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}.$$

(c) Show that the Fisher information matrix for (β, σ^2) is block-diagonal

$$I^{(n)}(\beta, \sigma^2) = \begin{pmatrix} I^{(n)}_{\beta\beta}(\beta, \sigma^2) & 0 \\ 0 & I^{(n)}_{\sigma^2\sigma^2}(\beta, \sigma^2) \end{pmatrix}, \quad \text{where } I^{(n)}_{\beta\beta}(\beta, \sigma^2) = \frac{1}{\sigma^2} X^T W X.$$

Use this to construct an asymptotic $(1-\alpha)$ confidence interval for $\beta_j, j \in \{1, \ldots, p\}$.

12. The next table contains counts N_y of the number of claims y made in a single year by $n = 9461$ car insurance policy holders.

Claims y	0	1	2	3	4	5	6	7
Counts N_y	7840	1317	239	42	14	4	4	1

Suppose Y_k, the number of claims to be made in a single year by policy holider $k = 1, \ldots, n$, follows a Poisson distribution with parameter $\theta_k > 0$, and $Y_1, \ldots, Y_n$ are independent given $\theta_1, \ldots, \theta_n$.

(a) Suppose $\theta_1 = \cdots = \theta_n = \theta$. Find a conjugate prior for θ and describes the posterior update for θ.

(b) Now suppose $\theta_1, \ldots, \theta_n$ are i.i.d. random variables with density function g. Show that

$$\mathsf{E}(\theta_k \mid Y_k) = (Y_k + 1) f(Y_k + 1)/f(Y_k), \ k = 1, \ldots, n,$$

where $f(y) = \mathsf{P}(Y_1 = y), y = 0, 1, \ldots$ is the marginal probability mass function of Y_1.

(c) The insurance company is interested in finding an estimator $\hat{N}_y$ of the expected number of claims in a succeeding year for its customers who has made y claims this year. By using an empirical estimate of $f(y)$ in the above formula, find an estimator $\hat{N}_y$ for $y = 0, 1, \ldots, 6$ and apply it to the dataset above. What can you say about your estimates? If your estimates do not behave as expected for larger values of y, what can you do to improve them?