1. Estimation in the Gamma distribution

   (a) The asymptotic covariance matrix of $(\hat{\alpha}, \hat{\beta})$ is equal to the inverse of the Fisher information matrix for a single observation times n^{-1}.

   As this family is a canonical exponential family, the information matrix is determined by differentiation of the log-normalising constant

$$c(\alpha, \beta) = \log \Gamma(\alpha) - \alpha \log \beta$$

   and we get

$$i(\theta) = \begin{pmatrix} \psi'(\alpha) & -1/\beta \\ -1/\beta & \alpha/\beta^2 \end{pmatrix},$$

   so taking inverses and dividing by n we get

$$\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) \stackrel{a}{=} \frac{1}{n\{\alpha\psi'(\alpha) - 1\}} \begin{pmatrix} \alpha & \beta \\ \beta & \beta^2\psi'(\alpha) \end{pmatrix},$$

   where $\psi'(\alpha)$ is the Trigamma function.

   The division with n was missing in the original problem sheet.

   (b) Using the delta method on the function

$$g(\alpha, \beta) = (\alpha, \gamma) = (\alpha, \alpha/\beta)$$

   yields

$$\frac{\partial g}{\partial \alpha} = (1, 1/\beta), \quad \frac{\partial g}{\partial \beta} = (0, -\alpha/\beta^2)$$

   so the delta method yields

$$\begin{aligned} \mathrm{Cov}(\hat{\alpha}, \hat{\gamma}) &\stackrel{a}{=} \frac{1}{n\{\alpha\psi'(\alpha) - 1\}} \begin{pmatrix} 1 & 0 \\ 1/\beta & -\alpha/\beta^2 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \beta & \beta^2\psi'(\alpha) \end{pmatrix} \begin{pmatrix} 1 & 1/\beta \\ 0 & -\alpha/\beta^2 \end{pmatrix} \\ &= \frac{1}{n} \begin{pmatrix} \frac{\alpha}{\alpha\psi'(\alpha)-1} & 0 \\ 0 & \alpha/\beta^2 \end{pmatrix}, \end{aligned}$$

   so $\hat{\alpha}$ and $\hat{\gamma}$ are asymptotically independent.

2. Estimation in the normal distribution with known coefficient of variation.

   (a) We get by differentiation of

$$\log f(x_i; \mu) = -\frac{1}{2}\log(2\pi) - \log \mu - \frac{x_i^2}{2\mu^2} + \frac{x_i}{\mu} - \frac{1}{2}$$

and summing over i that

$$S(\mu) = -\frac{n}{\mu} + \frac{\sum_i X_i^2}{\mu^3} - \frac{\sum_i X_i}{\mu^2}.$$

If we let $S = \sum X_i$ and $SS = \sum X_i^2$ this can be rewritten as

$$S(\mu) = -\frac{n}{\mu} + \frac{SS}{\mu^3} - \frac{S}{\mu^2}.$$

(b) The likelihood equation is obtained by equating the score statistic to 0. Doing this and multiplying with μ^3 yields the equation

$$n\mu^2 + \mu S - SS = 0$$

which has exactly one positive root

$$\hat{\mu} = \frac{-S + \sqrt{S^2 + 4nSS}}{2n}$$

unless $SS = 0$, which implies all X_i are equal to zero.

(c) We get by further differentiation that

$$j(\mu) = -\frac{n}{\mu^2} + \frac{3SS}{\mu^4} - \frac{2S}{\mu^3}.$$

Since $\hat{\mu}$ satisfies the likelihood equation we have

$$\frac{S}{\hat{\mu}^2} = \frac{SS}{\hat{\mu}^3} - \frac{n}{\hat{\mu}}.$$

Inserting this into the expression for $j(\hat{\mu})$ we further get

$$j(\hat{\mu}) = \frac{n}{\hat{\mu}^2} + \frac{SS}{\hat{\mu}^4} > 0,$$

so the root of the likelihood equation is the unique local (and therefore global) maximum.

(d) We use that $\mathbf{E}(X^2) = \mathbf{V}(X) + \{\mathbf{E}(X)\}^2 = 2\mu^2$ and take expectations in the expression for $j(\mu)$ to get

$$i(\mu) = \mathbf{E}\{j(\mu)\} = -\frac{n}{\mu^2} + \frac{6n\mu^2}{\mu^4} - \frac{2n\mu}{\mu^3} = \frac{3n}{\mu^2}.$$

The asymptotic variance of $\hat{\mu}$ is the inverse of this, thus equal to $\mu^2/(3n)$.

(e) SSD/μ^2 is distributed as $\chi^2(n-1)$, which is the distribution of

$$Y = \sum_1^{n-1} Z_i$$

where Z_i are i.i.d. and $Z_i \sim \chi^2(1)$. It follows that

$$\frac{SSD}{n-1} \overset{\text{a}}{\sim} \mathcal{N}\{\mu^2, 2\mu^4/n\}.$$

For $g(x) = 2\sqrt{x}$ we have $g'(x) = 1/\sqrt{x}$ so the delta method yields that

$$2\sqrt{SSD/(n-1)} \overset{\text{a}}{\sim} \mathcal{N}\{2\mu, 2\mu^2/n\}$$

and since $\bar{X}$ and SSD are independent we further get

$$\mathbf{V}(\tilde{\mu}) = \frac{1}{9}\frac{\mu^2 + 2\mu^2}{n} = \frac{\mu^2}{3n},$$

so this estimator is also asymptotically efficient.

3. Equivalent forms of the asymptotic distribution of the MLE

(a) As $\hat{\theta}$ is consistent, $\hat{\theta} \xrightarrow{P} \theta_0$. As i is continuous, this implies $i(\hat{\theta}) \xrightarrow{P} i(\theta_0)$. Since Cramér's conditions imply that

$$\sqrt{ni(\theta_0)}(\hat{\theta} - \theta_0) \overset{\text{a}}{\sim} \mathcal{N}(0,1),$$

Slutsky's theorem now yields

$$\sqrt{ni(\hat{\theta})}(\hat{\theta} - \theta_0) \overset{\text{a}}{\sim} \mathcal{N}(0,1).$$

(b) Taylor's theorem yields that

$$\begin{aligned} j_n(\hat{\theta})/n &= -\frac{1}{n}\sum_{k=1}^{n} \frac{\partial^2}{\partial\theta^2}\log f(X_k;\theta_0) - \frac{1}{n}\sum_{k=1}^{n} \frac{\partial^3}{\partial\theta^3}\log f(X_k;\theta^*)(\hat{\theta} - \theta_0) \\ &= -\frac{1}{n}\sum_{k=1}^{n} \frac{\partial^2}{\partial\theta^2}\log f(X_k;\theta_0) + R(X,\theta^*) \end{aligned}$$

for some θ^* between θ_0 and $\hat{\theta}$.
The first of these terms converges in probability to $i(\theta_0)$.
As $\hat{\theta}$ is consistent and the third derivative is bounded in a neighbourhood of 0, the absolute value of the second term satisfies

$$|R(X,\theta^*)| \leq |\hat{\theta} - \theta_0|\frac{1}{n}\sum_{k=1}^{n} M_k \xrightarrow{P} 0,$$

and hence

$$j_n(\hat{\theta})/n \xrightarrow{P} i(\theta_0).$$

(c) We have

$$\sqrt{j_n(\hat{\theta})}(\hat{\theta} - \theta_0) = \sqrt{\frac{j_n(\hat{\theta})/n}{i(\theta_0)}}\sqrt{ni(\theta_0)}(\hat{\theta} - \theta_0).$$

The result under (b) shows that the fraction converges in probability to 1. Slutsky's theorem now yields

$$\sqrt{j_n(\hat{\theta})}(\hat{\theta} - \theta_0) \overset{\text{a}}{\sim} \mathcal{N}(0, 1),$$

as required.