MATH 557 - EXERCISES 4
SOLUTIONS

1. (a) To find the UMP test, consider

$$\begin{aligned} H_0 &: \theta = 1 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

for $\theta_1 > 1$. By Neyman-Pearson, the rejection region is constructed by looking at

$$\frac{f_{\mathbf{X}}(\mathbf{x};\theta_1)}{f_{\mathbf{X}}(\mathbf{x};1)} = \frac{\prod_{i=1}^{n} \theta_1 (1-x_i)^{\theta_1 - 1}}{1} = \theta_1^n \{T(\mathbf{x})\}^{\theta_1 - 1}$$

where $T(\mathbf{x}) = \prod_{i=1}^{n}(1-x_i)$. Hence the rejection region is defined by

$$\theta_1^n \{T(\mathbf{x})\}^{\theta_1-1} > c \qquad \text{or equivalently} \qquad T(\mathbf{x}) > c_1$$

where the requirement

$$\Pr[T(\mathbf{X}) \in \mathcal{R}_T; \theta = 1] = \Pr[T(\mathbf{X}) > c_1; \theta = 1] = \alpha$$

determines k_1 for any α. To simplify further

$$\prod_{i=1}^{n}(1-X_i) > c_1 \qquad \Longleftrightarrow \qquad -\sum_{i=1}^{n} \log(1-X_i) < -\log c_1 = c$$

say. Now, if $\theta = 1$, the data are uniformly distributed on (0,1). Also, if $X \sim Uniform(0,1)$, then $1 - X \sim Uniform(0,1)$, and

$$-\log(1-X) \sim Exponential(1)$$

Therefore the critical region is defined by $\Pr[T(\mathbf{X}) > c_1; \theta = 1] = \Pr[V < c; \theta = 1] = \alpha$, where

$$V = -\log T(\mathbf{X}) = -\sum_{i=1}^{n} \log(1-X_i) \sim Gamma(n,1).$$

Thus c is the α quantile of the $Gamma(n,1)$ distribution. This is the UMP test for any $\theta_1 > 1$, so it is the UMP test for the required hypotheses.

(b) Under H_1, the ML estimate of θ is

$$\widehat{\theta}_n = \underset{\theta \in \mathbb{R}^+}{\text{argmax}}\, \theta^n \{T(\mathbf{x})\}^{\theta-1} = -\frac{n}{\log T(\mathbf{x})} = -\frac{n}{\sum_{i=1}^{n} \log(1-X_i)} = -\frac{n}{\log T(\mathbf{x})}$$

Thus the LRT is based on the rejection region $\mathcal{R}_{\mathbf{X}}$ defined by

$$\lambda_{\mathbf{X}}(\mathbf{x}) = \frac{\mathscr{L}_n(1)}{\mathscr{L}_n(\widehat{\theta}_n)} = \frac{1}{\widehat{\theta}_n^n \{T(\mathbf{x})\}^{\widehat{\theta}_n - 1}} \leq c$$

which is equivalent to

$$n \log \widehat{\theta}_n + (\widehat{\theta}_n - 1) \log T(\mathbf{x}) \geq -\log c$$

or

$$-n \log(-\log T(\mathbf{x})) - \log T(\mathbf{x}) \geq -\log c - n \log n + n$$

which may be written

$$-n \log V + V \geq c$$

where $V \sim Gamma(n,1)$ as above. To solve this for c requires numerical steps.

2. Can use the Karlin-Rubin theorem in both cases.

   (a) The likelihood ratio for $\theta_1 < \theta_2$ for this model is

   $$\lambda(\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x};\theta_2)}{f_{\mathbf{X}}(\mathbf{x};\theta_1)} = \frac{\theta_1^n}{\theta_2^n}\exp\left\{T(\mathbf{x})\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)\right\}$$

   which is an increasing function of $T(\mathbf{x}) = \sum\limits_{i=1}^{n} x_i$. Thus the rejection region takes the form

   $$\mathcal{R} \equiv \left\{\mathbf{x} : T(\mathbf{x}) = \sum_{i=1}^{n} x_i > t_0\right\}$$

   To find t_0, we need to solve $\Pr[T(\mathbf{X}) > t_0 \,;\, \theta_0] = \alpha$. Here $T(\mathbf{X}) \sim \text{Gamma}(n, 1/\theta)$, so t_0 is the $1-\alpha$ quantile of this distribution.

   (b) The likelihood ratio for $\theta_1 < \theta_2$ for this model is

   $$\lambda(\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x};\theta_2)}{f_{\mathbf{X}}(\mathbf{x};\theta_1)} = \frac{\theta_1^{n/2}}{\theta_2^{n/2}}\exp\left\{\frac{T(\mathbf{x})}{2}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)\right\}$$

   which is an increasing function of $T(\mathbf{x}) = \sum\limits_{i=1}^{n} (x_i-1)^2$. Thus the rejection region takes the form

   $$\mathcal{R} \equiv \left\{\mathbf{x} : T(\mathbf{x}) = \sum_{i=1}^{n} x_i > t_0\right\}$$

   To find t_0, we need to solve $\Pr[T(\mathbf{X}) > t_0 \,;\, \theta_0] = \alpha$. Here under the assumption $\theta = \theta_0$,

   $$\frac{T(\mathbf{X})}{\theta_0} \sim \chi^2_n \equiv \text{Gamma}(n/2, 1/2)$$

   so

   $$\Pr[T(\mathbf{X}) > t_0 \,;\, \theta_0] = \Pr[T(\mathbf{X})/\theta_0 > t_0/\theta_0 \,;\, \theta_0] = \alpha.$$

   implies that $t_0 = \theta_0 q_{n,1-\alpha}$, where $q_{n,1-\alpha}$ is the $1-\alpha$ quantile of the Chisquared distribution with n degrees of freedom.

3. Again using the Karlin-Rubin Theorem: The likelihood ratio for $\theta_1 < \theta_2$ for this model is

   $$\lambda(\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x};\theta_2)}{f_{\mathbf{X}}(\mathbf{x};\theta_1)} = \left(\frac{\theta_2}{\theta_1}\right)^{T(\mathbf{x})}\exp\left\{-n(\theta_2-\theta_1)\right\}$$

   where $T(\mathbf{x}) = \sum\limits_{i=1}^{n} x_i$. In this case, under $\theta = 2$,

   $$T(\mathbf{X}) = \sum_{i=1}^{n} X_i \sim Poisson(2n)$$

   Thus the distribution of $T(\mathbf{X})$ is discrete. A randomized test takes the form

   $$\phi^\star_{\mathcal{R}}(\mathbf{x}) = \begin{cases} 1 & T(\mathbf{x}) > c \\ \gamma & T(\mathbf{x}) = c \\ 0 & T(\mathbf{x}) \leq c \end{cases}$$

where c is the largest integer such that $\Pr[T(\mathbf{X}) > c; \theta] \le 0.05$, and γ is selected so that

$$\Pr[T(\mathbf{X}) > c; \theta] + \gamma \Pr[T(\mathbf{X}) = c; \theta = 2] = 0.05$$

In the example, $n = 6$, and $T(\mathbf{x}) = 18$, and by calculation $c = 18$

$$\Pr[T(\mathbf{X}) > 18; \theta = 2] = 0.0374 \qquad \Pr[T(\mathbf{X}) = 18; \theta = 2] = 0.0255$$

so that

$$\gamma = \frac{0.05 - \Pr[T(\mathbf{X}) > 18; \theta = 2]}{\Pr[T(\mathbf{X}) = 18; \theta = 2]} = \frac{0.05 - 0.0374}{0.0255} = 0.494$$

In this case, the hypothesis is rejected with probability $\gamma = 0.494$ as $T(\mathbf{x}) = 18$.

4. In this model, if $X = \sum_{i=1}^{n} X_i$, the MLE of θ is $\widehat{\theta}_n = \overline{X} = X/n$.

   (a) This is not a 1-1 mapping, so some care is needed. Inverting the transformation yields that

   $$\theta = \frac{1 \pm \sqrt{1 - 4\tau}}{2}$$

   so the likelihood needs to be worked out for both cases. In the first case

   $$L(\tau|\mathbf{x}) = \left(\frac{1 + \sqrt{1 - 4\tau}}{2}\right)^{x} \left(\frac{1 - \sqrt{1 - 4\tau}}{2}\right)^{n-x}$$

   and in the second

   $$L(\tau|\mathbf{x}) = \left(\frac{1 - \sqrt{1 - 4\tau}}{2}\right)^{x} \left(\frac{1 + \sqrt{1 - 4\tau}}{2}\right)^{n-x}$$

   where $0 < \tau < 1/4$. After some manipulation it follows that in both cases $x < n - x$ or $x > n - x$,

   $$\widehat{\tau} = \frac{1}{4} - \left(\frac{1}{2} - \frac{x}{n}\right)^2 = \overline{x}(1 - \overline{x})$$

   with the same result if $x = n - x$, and so the estimator is

   $$\widehat{\tau}_n(\mathbf{X}) = \overline{X}(1 - \overline{X})$$

   so in fact the non 1-1 nature of the reparameterization is not problematic.

   (b) Despite the non 1-1 nature of the reparameterization, the Delta method can still be used. If $g(t) = t(1 - t)$, then $\dot{g}(t) = 1 - 2t$, which is non zero if $t \neq 1/2$. Thus for $\theta \neq 1/2$, by the CLT

   $$\sqrt{n}(\overline{X} - \theta) \xrightarrow{d} Z \sim Normal(0, \theta(1 - \theta))$$

   and thus by the Delta method

   $$\sqrt{n}(\widehat{\tau}_n(\mathbf{X}) - \tau) \xrightarrow{d} Z \sim Normal(0, \tau(1 - 2\theta)^2)$$

   or, for large n

   $$\widehat{\tau}_n(\mathbf{X}) \stackrel{.}{\sim} Normal(\tau, \tau(1 - 2\theta)^2/n)$$

   If $\theta = 1/2$, this approximation yields a degenerate limiting distribution, so a second order Delta method must be used. We have that $\ddot{g}(t) = -2$, so

   $$n(\widehat{\tau}_n(\mathbf{X}) - 1/4) \xrightarrow{d} -\frac{1}{4}Q$$

   where $Q \sim \chi^2_1$, or, for large n

   $$\widehat{\tau}_n(\mathbf{X}) \stackrel{.}{\sim} \frac{1}{4} - \frac{1}{4}Gamma(1/4, 2n)$$

5. Note first that

$$\mathbb{E}_{T_{1n}}[T_{1n}] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathbf{X}}[X_i^2] - 1 = \frac{1}{n}\sum_{i=1}^n (\mu^2+1) - 1 = \mu^2$$

and

$$\begin{aligned}\mathbb{E}_{T_{2n}}[T_{2n}] &= \frac{1}{n^2}\mathbb{E}_{\mathbf{X}}\left[\left(\sum_{i=1}^n X_i\right)^2\right] - \frac{1}{n} = \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}_{\mathbf{X}}[X_i^2] + \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}_{\mathbf{X}}[X_iX_j] - \frac{1}{n} \\ &= \frac{1}{n^2}n(\mu^2+1) + 0 - \frac{1}{n} = \mu^2.\end{aligned}$$

so both statistics are unbiased (and hence asymptotically unbiased). Thus the ARE is the ratio of the asymptotic variances. For T_{1n}, we know by the CLT that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \theta\right) \xrightarrow{d} Z \sim Normal(0,\gamma)$$

where $\theta = \mu^2+1$, and $\gamma = \text{Var}_{f_X}[X_i^2]$, so the asymptotic variance of T_{1n} is γ. For T_{2n}, we have by the CLT

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} Z \sim Normal(0,1)$$

so by the Delta Method, for $\mu \neq 0$

$$\sqrt{n}(\overline{X}^2 - \mu^2) \xrightarrow{d} Z \sim Normal(0, 4\mu^2)$$

and for $\mu = 0$, as $\sqrt{n}\overline{X} \xrightarrow{d} Normal(0,1)$,

$$n\overline{X}^2 \xrightarrow{d} Q \sim \chi_1^2$$

where Q has variance 2. Hence

$$ARE_\mu(T_{1n}, T_{2n}) = \begin{cases} \dfrac{4\mu^2}{\gamma} & \mu \neq 0 \\[2ex] \dfrac{2}{\gamma} & \mu = 0 \end{cases}$$

Note however that there is a different rescaling in the $\mu = 0$ case. So in terms of large sample comparison,

$$T_{1n} \mathrel{\dot{\sim}} Normal(0, \gamma/n) \qquad T_{2n} \mathrel{\dot{\sim}} \text{Gamma}(1/2, n/2)$$

yielding a large sample variance ratio of $2/n\gamma$.

6. In this case, $X \sim Exp(\phi)$ and $Y \sim Exp(\theta\phi)$, so if $\lambda_1 = \phi$ and $\lambda_2 = \theta\phi$, the MLEs are $\widehat{\lambda}_1 = 1/\overline{X}$ and $\widehat{\lambda}_2 = 1/\overline{Y}$, so that by invariance

$$\widehat{\phi}_n = \widehat{\lambda}_1 = 1/\overline{X} \qquad \widehat{\theta}_n = \widehat{\lambda}_2/\widehat{\lambda}_1 = \overline{X}/\overline{Y}$$

Thus

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathbf{Z} \sim Normal(\mathbf{0}_2, \{\mathcal{I}_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)\}^{-1})$$

where $\mathcal{I}_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)$ is the Fisher Information. We have joint density

$$\phi^2\theta\exp\{-[\phi x + \theta\phi y]\} \qquad x, y > 0$$

so that

$$\ell(\theta,\phi) = 2\log\phi + \log\theta - (\phi x + \theta\phi y)$$

and

$$\frac{\partial \ell}{\partial \phi} = \frac{2}{\phi} - x - \theta y \qquad\qquad \frac{\partial \ell}{\partial \theta} = \frac{1}{\theta} - \phi y$$

$$\frac{\partial^2 \ell}{\partial \phi^2} = -\frac{2}{\phi^2} \qquad \frac{\partial^2 \ell}{\partial \theta^2} = -\frac{1}{\theta^2} \qquad \frac{\partial^2 \ell}{\partial \phi \partial \theta} = -y$$

yielding the matrix $\Psi(X,Y;\theta,\phi)$ and Fisher information

$$\Psi(X,Y;\theta,\phi) = \begin{bmatrix} \dfrac{2}{\phi^2} & Y \\ Y & \dfrac{1}{\theta^2} \end{bmatrix} \qquad \mathcal{I}_{(\theta,\phi)}(\theta,\phi) = \begin{bmatrix} \dfrac{1}{\theta^2} & \dfrac{1}{\phi\theta} \\ \dfrac{1}{\phi\theta} & \dfrac{2}{\phi^2} \end{bmatrix}$$

as $\mathbb{E}_Y[Y;\theta,\phi] = 1/(\phi\theta)$. Thus

$$\mathcal{I}_{(\theta,\phi)}\{(\theta,\phi)\}^{-1} = \begin{bmatrix} 2\theta^2 & -\phi\theta \\ -\phi\theta & \phi^2 \end{bmatrix}$$

Thus

$$\sqrt{n}\begin{pmatrix} \widehat{\theta}_n - \theta_0 \\ \widehat{\phi}_n - \phi_0 \end{pmatrix} \xrightarrow{d} \mathbf{Z} \sim Normal\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 2\theta_0^2 & -\phi_0\theta_0 \\ -\phi_0\theta_0 & \phi_0^2 \end{bmatrix}\right)$$

or, for large n,

$$\begin{pmatrix} \widehat{\theta}_n \\ \widehat{\phi}_0 \end{pmatrix} \mathrel{\dot\sim} Normal\left(\begin{pmatrix} \theta_0 \\ \phi_0 \end{pmatrix}, \begin{bmatrix} 2\theta_0^2 & -\phi_0\theta_0 \\ -\phi_0\theta_0 & \phi_0^2 \end{bmatrix}\right)$$

7. For the Poisson case, for $\lambda > 0$

$$\ell_n(\lambda) = -n\lambda + t_n \log\lambda - \sum_{i=1}^n \log x_i! \qquad \text{where } t_n = \sum_{i=1}^n x_i$$

$$\dot{\ell}_n(\lambda) = -n + \frac{t_n}{\lambda} \qquad \ddot{\ell}_n(\lambda) = -\frac{t_n}{\lambda^2}$$

and hence the MLE, from $\dot{\ell}_n(\widehat{\lambda}_n) = 0$, is $\widehat{\lambda}_n = t_n/n = \overline{x}$, with estimator $T_n/n = \overline{X}$. Then

$$W_n = n(\widehat{\theta}_n - \theta_0)^\top \widehat{I}_n(\widehat{\theta}_n)(\widehat{\theta}_n - \theta_0) \qquad \text{(WALD)}$$

and

$$R_n = Z_n^\top \mathcal{I}_{\theta_0}(\theta_0)^{-1} Z_n \qquad \text{with} \quad Z_n = \frac{1}{\sqrt{n}}\dot{\ell}_n(\theta_0). \qquad \text{(SCORE)}$$

we have, in the 1-d case

$$W_n = (\widehat{\theta}_n - \theta_0)^2 (n\widehat{I}_n(\widehat{\theta}_n)) = -(\widehat{\theta}_n - \theta_0)^2 \ddot{\ell}_n(\widehat{\theta}_n) \qquad R_n = \frac{\{Z_n(\theta_0)\}^2}{\mathcal{I}_{\theta_0}(\theta_0)}$$

Thus we have

- **Wald Statistic:**

$$W_n = -(\widehat{\theta}_n - \theta_0)^2 \ddot{\ell}_n(\widehat{\theta}_n) = -(\overline{X} - \lambda_0)^2 \left(\frac{-S_n}{(\overline{X})^2}\right) = n\frac{(\overline{X} - \lambda_0)^2}{\overline{X}} = \frac{n(\widehat{\lambda}_n - \lambda_0)^2}{\widehat{\lambda}_n}$$

- **Rao Statistic:** in this case, we can compute the Fisher Information exactly - we have

$$\mathcal{I}_{\lambda_0}(\lambda_0) = \mathbb{E}_X\left[-\Psi\left(X;\lambda_0\right)\right] = \mathbb{E}_X\left[\frac{X}{\lambda_0^2};\lambda_0\right] = \frac{1}{\lambda_0^2}\mathbb{E}_X\left[X;\lambda_0\right] = \frac{\lambda_0}{\lambda_0^2} = \frac{1}{\lambda_0}$$

so therefore

$$R_n = \frac{\{Z_n\}^2}{\mathcal{I}_{\lambda_0}(\lambda_0)} = \frac{\lambda_0}{n}\left(\frac{T_n}{\lambda_0} - n\right)^2 = \frac{n(\overline{X} - \lambda_0)^2}{\lambda_0}$$

If the Fisher Information can be computed exactly, then the exact version should be used for the Score statistic rather than an estimated version. Here that would imply that

$$R_n = -\left\{\dot{\ell}_n(\theta_0)\right\}^2\left\{\ddot{\ell}_n(\theta_0)\right\}^{-1} = \frac{-\left(\dfrac{T_n}{\lambda_0} - n\right)^2}{-T_n/\lambda_0^2} = \frac{(T_n - n\lambda_0)^2}{T_n} = \frac{n(\overline{X} - \lambda_0)^2}{\overline{X}}$$

that is, identical to Wald.

- **Likelihood Ratio Statistic**: the likelihood ratio is

$$\lambda_{\mathbf{X}}(\mathbf{x}) = \frac{\mathscr{L}_n(\lambda_0)}{\mathscr{L}_n(\widehat{\lambda}_n)} = \frac{e^{-n\lambda_0}\lambda_0^{T_n}}{e^{-n\widehat{\lambda}_n}\widehat{\lambda}_n^{T_n}} = \exp\left\{n(\widehat{\lambda}_n - \lambda_0) - T_n(\log\widehat{\lambda}_n - \log\lambda_0)\right\}$$

or equivalently

$$-2\lambda_{\mathbf{X}}(\mathbf{x}) = -2n(\widehat{\lambda}_n - \lambda_0) + 2T_n(\log\widehat{\lambda}_n - \log\lambda_0)$$

For a $1-\alpha$ confidence interval, we utilize the result that each of the test statistics has an approximate χ_1^2 distribution as $n \longrightarrow \infty$. For W_n and R_n, we have

$$\left\{\lambda : n(\widehat{\lambda}_n - \lambda)^2/\widehat{\lambda}_n \leq c_{1-\alpha}\right\} \qquad \text{and} \qquad \left\{\lambda : n(\widehat{\lambda}_n - \lambda)^2/\lambda \leq c_{1-\alpha}\right\}$$

respectively, where $c_{1-\alpha}$ is the $1-\alpha$ quantile of the χ_1^2 distribution. For the LRT, we have

$$\left\{\lambda : -2n(\widehat{\lambda}_n - \lambda) + 2t_n(\log\widehat{\lambda}_n - \log\lambda) \leq c_{1-\alpha}\right\}$$