**STAT 701 HW6 Solutions, 5/11/23**

**(# 1).** In this problem, the score and information involve the digamma and trigamma functions, $D(a_0) = d \log \Gamma(a_0)/da = \Gamma'(a_0)/\Gamma(a_0)$ and trigamma function $T(a_0) = d^2 \log \Gamma(a_0)/da^2$). At $a_0 = 2$, `digamma(2) = 0.4227843, trigamma(2) = 0.6449341`. In this Problem, the Neyman-Pearson test would be fully optimal if there were no nuisance parameter $b$. Since $b$ is unknown, it must be estimated. In part (c), you are asked to approximate (using the CLT) the probability of rejection of your tests when $a = 3$ (but $b$ is still unknown). The answer is based on the asymptotic distributions under the alternative, which depend on $\alpha, b, n$.

(a) 'Locally optimal' signals the Rao Score approach. The (one-sided, un-squared) score statistic for fixed $b$ is

$$n^{-1/2} \, \nabla_a \, \log L(\mathbf{X}, a, b) \Big|_{a=2} \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \log X_i \;+\; n\,(\log b - 0.4228)$$

and the restricted MLE $\hat{b}^{(r)} = 2/\bar{X}$, so

$$R_n \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \log X_i \;+\; \sqrt{n}\,\{\log(2/\bar{X}) \;-\; 0.4228\}$$

and the asymptotic variance is $(I(\theta)^{-1})_{aa} = T(2) - 1/2 = 0.1449$. Thus the one-sided test rejects when

$$(0.1449\, n)^{-1/2} \sum_{i=1}^{n} (\log(2X_i/\bar{X}) - 0.4228) \;>\; z_\alpha \;=\; \Phi^{-1}(1-\alpha) \tag{1}$$

(b). Since the likelihood ratio at $a = 3$ over $a = 2$ for each fixed $b$ is $(2b)^n\,(X_1 \cdots X_n)$, the optimal Neyman-Pearson test statistic **for any known b** would be $\sum_{i=1}^n \log X_i$. But the Neyman-Pearson Lemma does not quite apply here, and the "locally optimal" rationale is not immediately persuasive versus the alternative $a = 3$. No test we have studied is exactly optimal here, but a reasonable test statistic would be the LRT, although the Wilks Theorem does not apply to it because our alternative does not consist of all possible values for the parameter $a$. Nevertheless the LRT statistic is

$$\log \frac{\left(\prod_{i=1}^{n} X_i\right)^2 (3/\bar{X})^{3n}\, e^{-3n}}{\left(\prod_{i=1}^{n} X_i\right) (2/\bar{X})^{3n}\, e^{-2n}} \;=\; Const. + \sum_{i=1}^{n} \log(X_i) \;-\; n\, \log(\bar{X})$$

After centering under $H_0$ and scaling by $1/\sqrt{n}$, this statistic is exactly the same as the Rao Score statistic. It is definitely a sensible statistic. Another choice, which is asymptotically the same under the null hypothesis, is the Wald test based on the MLE $\hat{a}$. This statistic $\hat{a}$

standardized using large-sample theory at $a = 2$ has mean 2 and asymptotic variance (that turns out not to depend on $b$) obtained from the (per-observation) Information matrix as

$$(I(a,b))^{-1}_{11} \Big|_{a=2,\, b=\hat{b}^{(r)}} \;=\; \begin{pmatrix} T(a) & -1/b \\ -1/b & a/b^2 \end{pmatrix}^{-1}_{11} \Big|_{a=2,\, \hat{b}^{(r)}} \;=\; (T(2) - 1/2)^{-1} \;=\; 6.8997$$

So the Wald test rejects when $\sqrt{n/6.8997}\,(\hat{a} \,-\, 2) \;>\; \Phi^{-1}(\alpha)$.

**(c).** The asymptotic equivalence mentioned in (b) between Wald and Rao-Score does not persist under distant alternatives. So it makes sense to approximate the power of each versus $H_A: \;\; a = 3$. We use $n = 40$, $\alpha = 0.05$. The Rao-score statistic at $a = 3$ has mean and variance not depending on $b$ and found by numerical integration, respectively equal to

$$(40 \cdot 0.1449)^{-1/2}(40) \cdot (0.0693) \;=\; 1.1514$$

and

$$(40 \cdot 0.1449)^{-1}(40) \cdot (T(3) - 1/3) \;=\; 0.4250$$

So the approximate Normal(1.1514, 0.4250) probability of $[1.96, \infty)$ is $1 - \Phi((1.96 - 1.1514)/\sqrt{0.4250}) \;=\; 0.1074$. The corresponding approximate mean and variance for the Wald Test Statistic under $H_A: \; a = 3$ are respectively

$$(40/6.8997)^{1/2} \;=\; 2.4078 \qquad \text{and} \qquad (T(3) - 1/3)^{-1}/6.8997 \;=\; 2.3528$$

So the approximate power for the Wald test is

$$1 \;-\; \Phi((1.96 - 2.4078)/\sqrt{2.3528}) \;=\; 0.615$$

This calculation suggests that the Wald test is much more powerful for the distant alternative $H_A: \; a = 3$ in this example than the Rao-Score test.

**(# 2).** Denote the log-likelihoods for $X_1, \ldots, X_m$ and for $X_{m+1}, \ldots, X_n$ respectively as $\log L_{f,m}(\theta)$, $\log L_{g,n-m}(\theta)$, so that the overall log-likelihood $\log L(\theta)$ is just the sum of these separate log-likelihoods. Let the per-observation information for $\theta$ with density $f$ as $I_f(\theta)$ and with density $g$ as $I_g(\theta)$. Then our large-sample MLE theory under the regularity conditions in Bickel-Doksum chapter 6 assure us that there exists $\epsilon > 0$ (not shrinking to 0 as $m, n-m \to \infty$) such that with probability converging to 1 as $n \to \infty$, both log-likelihoods are strictly concave on $B_\epsilon(\theta)$ and with a unique maximum, and that

$$\frac{-1}{m}\, \nabla_\theta^{\otimes 2} \log L_{f,m}(\theta) \;\overset{P}{\approx}\; I_f(\theta) \;\;, \qquad \frac{-1}{n-m}\, \nabla_\theta^{\otimes 2} \log L_{g,n-m}(\theta) \;\overset{P}{\approx}\; I_g(\theta)$$

so that $\log L(\cdot)$ is strictly concave on $B_\epsilon(\theta)$ and

$$\frac{-1}{n}\, \nabla_\theta^{\otimes 2} \log L(\theta) \;\overset{P}{\approx}\; \lambda\, I_f(\theta) \,+\, (1-\lambda) I_g(\theta) \;\equiv\; I(\theta)$$

Moreover,

$$\sqrt{m}\,(\theta^{(1)} - \theta) \;\overset{P}{\approx}\; (I_f(\theta))^{-1} m^{-1/2} \nabla \,\log L_{f,m}(\theta)$$

and

$$\sqrt{n-m}\,(\theta^{(2)} - \theta) \;\overset{P}{\approx}\; (I_g(\theta))^{-1} (n-m)^{-1/2} \nabla \,\log L_{g,n-m}(\theta)$$

The standard Taylor Series expansion of $0 = \nabla \log L(\hat{\theta})$ around the base-point $\theta$ shows that

$$\sqrt{n}\,(\hat{\theta} - \theta) \;\overset{P}{\approx}\; [I(\theta)]^{-1}\, n^{-1/2}\, \nabla \log L(\theta) \;\overset{\mathcal{D}}{\longrightarrow}\; \mathcal{N}(\underline{0},\, (I(\theta))^{-1})$$

The existence and consistency of $\hat{\theta}$ can also be shown by checking that

$$\sqrt{n}\,(\hat{\theta} \;-\; (I(\theta)^{-1} \lambda I_f(\theta)\, \theta^{(1)} \;-\; (I(\theta)^{-1}(1-\lambda) I_g(\theta)\, \theta^{(2)}) \;\overset{P}{\longrightarrow}\; 0$$

**(#2.4.18)**. The problem requires a lot of careful algebra. But the most economical approach is to use the conditional density (coming from prediction theory or projections)

$$\mathcal{L}(Z_i \,|\, Y_i) \;=\; \mathcal{N}(\mu_1 \;+\; \beta_2 \gamma\, (Y_i - \alpha),\; \gamma \sigma^2) \;\;, \qquad \gamma \;\equiv\; \frac{\sigma_1^2}{\beta_2^2 \sigma_1^2 + \sigma^2} \;, \quad \alpha \;\equiv\; \beta_1 + \beta_2 \mu_1$$

Then the complete-data log-likelihood has the form $\sum_{i=1}^n \,[\log f_{Y_i}(Y_i) \;+\; \log f_{Z_i|Y_i}(Z_i \,|\, Y_i)]$

$$=\; -\frac{n}{2}\, \log(2\pi\, \sigma_1^2\, \sigma^2) \;-\; \frac{1-\gamma}{2\sigma^2} \sum_{i=1}^n (Y_i - \alpha)^2 \;-\; \frac{1}{2\gamma\, \sigma^2} \sum_{i=1}^n (Z_i - \mu_1 - \beta_2 \gamma\, (Yi - \alpha))^2$$

Then the major E-step calculation based on an initial parameter $\theta_* = (\mu_{1*}, \alpha_{1*}, \beta_{2*}, \gamma_*, \sigma_*^2)$ and final parameter $\theta$ is for $i \geq m+1$ given by

$$E_{\theta_*}[(Z_i - \mu_1 - \beta_2 \gamma\, (Y_i - \alpha))^2 \,|\, Y_i] \;=\; (\mu_{1*} + \beta_{2*} \gamma_* (Y_i - \alpha_*) - \mu_1 \;-\; \beta_2 \gamma\, (Y_i - \alpha))^2 \;+\; \sigma_*^2\, \gamma_*$$

Like all EM problems, this one is impossible if you do not carefully distinguish the initial parameters (used to calculate conditional expectations) from the M-step free-and-then-maximized parameters. The M-step maximizes over $\theta$ in the expression

$$-\frac{n}{2} \log(\frac{\sigma^2 \gamma}{1-\gamma}) \;-\; \frac{1-\gamma}{2\sigma^2} \sum_{i=1}^n (Y_i - \alpha)^2 \;-\; \frac{1}{2\gamma\sigma^2} \Big\{ \sum_{i=1}^m (Z_i - \mu_1 - \beta_2 \gamma (Y_i - \alpha))^2$$

$$+ \sum_{i=m+1}^n (\mu_{1*} + \beta_{2*} \gamma_* (Y_i - \alpha_*) - \mu_1 \;-\; \beta_2 \gamma\, (Y_i - \alpha))^2 \;+\; (n-m)\, \sigma_*^2\, \gamma_* \Big\}$$

Maximization is still a bit laborious, but you can verify without too much trouble by substituting the expression for $\partial/\partial\mu_1 \;=\; 0$ into $\partial/\partial\alpha \;=\; 0$, that $\hat{\alpha} = \bar{Y}$ and then

$$\hat{\mu}_1 \;=\; \frac{1}{n} \Big\{ \sum_{i=1}^m Z_i \;+\; (\mu_{1*} - \beta_{2*} \gamma_* \alpha_*)(n-m) \;+\; \beta_{2*} \gamma_* \sum_{i=m+1}^n Y_i \Big\}$$

**(# 4)**. (*Chi-squared goodness of fit test problem.*)

```
  prob4dat = scan("hw6prob4.dat", sep=" ")   # numeric vector, length 80
         ## first step is to find the discrete dataset based on given cut-points
  cvec = c(0, 0.35, 0.5, 0.625, 0.8, Inf)
  count4 = hist(prob4dat, breaks=cvec, plot=F)$count
> count4
[1] 18 10 14 22 16
                     ## maximize Weibull logLik for these count data
   negWeib = function(x, dat=count4, cuts=cvec) {
        alph=x[1]; lam=x[2]
        probs = diff(pweibull(cuts, alph, lam^(-1/alph)))
        -sum(count4*log(probs)) }
   tmp = nlm(negWeib, c(2.5,3))
   tmp$estimate
[1] 2.387675 2.589567
> expec4 = 80*diff(pweibull(cvec, tmp$est[1], tmp$est[2]^(-1/tmp$est[1])))
> expec4
[1] 15.22838 15.99817 14.34189 16.93424 17.49733
   sum((count4-expec4)^2/expec4)
[1] 4.405005                                   # df = 5-1-2 =2
  1-pchisq(4.405,2)
[1] 0.1105265       ## so we would accept the null that the data are Weibull
```

**(# 5)**. Here the $\epsilon_i$ are the unobserved variables, and the complete-data log-likelihood is

$$n\log(\lambda(1-p)) \,+\, (n-\sum_{i=1}^{n}\epsilon_i)\log(2) \,+\, \sum_{i=1}^{n}\epsilon_i\,\log(p\lambda/(1-p)) \,-\, \lambda\sum_{i=1}^{n}(2-\epsilon_i)Y_i$$

The conditional expectation needed is $E(\epsilon_i\,|\,Y_i) \;=\; p\,e^{-\lambda Y_i}/(pe^{-\lambda Y_i}+2(1-p)e^{-2\lambda Y_i})$, and the M-steps are straightforward.

Define $\rho_* = p_*/(p_* + 2(1-p)\exp(-\lambda Y_i))$. Then the M-step maximizes

$$n\log(\lambda(1-p)) + n\rho_*\log(p/(\lambda(1-p))) - n\lambda\bar{Y}(2-\rho_*)$$

and max occurs at $p=\rho_*,\ \lambda = (1-\rho_*)/(\bar{Y}(2-\rho_*))$.