STAT 701 HW4 Solutions, 4/10/23

(#5.5.2) Here we are testing $H_0 : \mu \in [0, \Delta]$ versus $H_A : \mu > \Delta$.

(a). $P_\mu(\sqrt{n}\,(\bar{X} - \Delta) \geq t) = (1 - \Phi(\sqrt{n}\,(\Delta - \mu) + t))$, and at $t = \sqrt{n}\,(\bar{X} - \Delta)$ we obtain the p-value (which is the sup of the rejection probabilities and occurs at $\mu = \delta$) as $\hat{p} = \Phi(-\sqrt{n}\,(\bar{X} - \Delta)$. In any case, this p-value under the null at $\mu = \Delta$ is Unif$(0, 1)$ distributed because p-values always are when the test statistic is continuously distributed.

(b)–(c). With $\pi(\cdot) = \phi(\cdot)$ the standard normal density, by sufficiency $\pi(\mu \,|\, \underline{X}) = \pi(\mu \,|\, \bar{X})$ is proportional to $\exp(-\,((1 + n)/2)(\mu - \bar{X} a_n)^2)$, where $a_n \equiv n/(n + 1)$, and is therefore the $\mathcal{N}(a_n\bar{X},\, a_n/n)$ density. The posterior probability of H_0 is

$$\tilde{p} \,=\, \Phi(\frac{\Delta - a_n\,\bar{X}}{\sqrt{a_n/n}}) \,-\, \Phi(\frac{-a_n\,\bar{X}}{\sqrt{a_n/n}})$$

So when $\mu = \Delta$, the random variable $\sqrt{n/a_n}\,(a_n\bar{X} - \Delta)$ tends to standard normal and $a_n\,\bar{X}/\sqrt{a_n/n} \to \infty$, so that $\tilde{p}$ is asymptotically Unif$(0, 1)$ distributed, as desired.

(d). For $\mu > \Delta$, again the second term of $\tilde{p}$ is negligible, so that $\bar{X} \approx \mu$ and $\tilde{p}$ is asymptotic to $\Phi(\sqrt{n}(\bar{X} - \Delta) + O_P(1/\sqrt{n}))$, and the ratio $\tilde{p}/\hat{p}$ tends to 1.

```
> f = function(Del,n) {
      an = n/(n+1)
      Xb = 1.645/sqrt(n)+Del
      pnorm(-sqrt(n/an)*(an*Xb-Del)) - pnorm(-sqrt(n/an)*an*Xb) }
> array( round(outer(c(.1,.24),c(10,20,50,100), f), 4), c(2,4),
          dimnames=list(c("Delt=0.1","Delt=0.2"),
          c("n=10","n=20","n=50","n=100")))
          n=10   n=20   n=50  n=100
Delt=0.1 0.0312 0.0361 0.0432 0.0476
Delt=0.2 0.0564 0.0562 0.0549 0.0533
```

(#6.1.1) (a). First, if $n = r$ with η, σ^2 unknown, then in the loglikelihood (6.1.11) we can put $\eta_i = u_i$ and take positive $\sigma^2 \to 0$ to argue that the likelihood can be made arbitrarily large, with no maximum. (b). When $n \geq r + 1$, it is easy to check that the given $(\hat{\eta}_1, \ldots, \hat{\eta}_r,\, \hat{\sigma}^2)$ is the unique calculus maximum of (6.1.11).

(#6.1.4) (a)–(b). In this problem, the data values Y_i have covariance matrix Σ such that $\Sigma_{ij} = \mathrm{cov}(Y_i, Y_j) = \sigma^2$ if $i = j = 1$, $= (1 + c^2)\sigma^2$ if $i = j > 1$, $= c\sigma^2$ if

$i = j \pm 1$, and $= 0$ otherwise. Let $W = \Sigma^{-1}$. When the error distribution is Gaussian, the Gaussian likelihood immediately implies the maximum likelihood estimate of θ is the minimizer of the quadratic form $(\underline{Y} - \underline{1})' W (\underline{Y} - \underline{1})$, which is the Weighted Least Squares (WLS) estimate $\underline{1}' W \underline{Y} / \underline{1}' W \underline{1}$ and is clearly linear in $\underline{Y}$. However as explained in a class-wide email, the coefficients a_j given in the problem must have a misprint, because they are not correct when $n = 2$ or 3. Some formula like that must hold because the solution of the equation $\Sigma \underline{v} = \underline{1}$ can be seen to satisfy (for components 2 through n-1) an inhomogeneous difference equation with constant coeffcients. The verification of that solution with a_j proportional to v_j) might have been tractable with correctly given closed-form formula, but I did not expect you to find the closed-form formula yourselves, and you did not need to in order to solve the remaining parts of the problem.

(d)–(e). These problem parts can be shown directly through the Cauchy-Schwarz inequality:

$$n^2 = [\underline{1}' \underline{1}]^2 = [(\Sigma^{1/2}\underline{1})' (\Sigma^{-1/2}\underline{1})]^2 \le \|\Sigma^{1/2}\underline{1}\|^2 \cdot \|\Sigma^{-1/2}\underline{1}\|^2 = (\underline{1}' \Sigma \underline{1}) \cdot (\underline{1}' \Sigma^{-1} \underline{1})$$

Since the variance of the MLE is $1/(\underline{1}' \Sigma^{-1} \underline{1})$ and the variance of $\bar{Y} = n^{-2} \underline{1}' \Sigma \underline{1}$, the result of this inequality is that the variance of $\bar{Y}$ is at least as large as the variance of the WLS. Moreover, according to the Cauchy-Schwarz inequality, the inequality is strict unless $\Sigma^{1/2}\underline{1}$ is proportional to $\Sigma^{-1/2}\underline{1}$, which happens only if $\Sigma\underline{1}$ is proportional to $\underline{1}$, which happens only if $c = 0$.

(#6.1.14) Here we deal with the linear Gaussian regression model with $p = r = 2$ and nonrandom design matrix Z with first column $\underline{1}$. From the book's treatment of this topic, we know $\hat{\beta} - \beta = (Z'Z)^{-1} Z' \epsilon \sim \mathcal{N}(0, \sigma^2)$ is independent of $s^2 = (n/(n-2))\, \hat{\sigma}^2 = (n-2)^{-1} \sum_{i=1}^n (Y_i - Z\hat{\beta})^2$. The estimate of $\beta_1 + \beta_2 z$ is $(1, z)\hat{\beta}$, and $(1, z)(\hat{\beta} - \beta) \sim \mathcal{N}(0, (1, z)(Z'Z)^{-1}(1, z)'\sigma^2)$.

(a). The confidence interval for $\beta_1 + \beta_2 z$ is $(1, z)\hat{\beta} \pm \{(1, z)(Z'Z)^{-1}(1, z)'\}^{1/2} s \cdot t_{n-2,\alpha/2}$.

(b). For a new observation $Y_0 \sim \mathcal{N}(\beta_1 + \beta_2 z,\, (1, z)(Z'Z)^{-1}(1, z)'\sigma^2)$ independent of the original sample $\underline{Y}$, the point-predictor is again $(1, z)\hat{\beta}$, and the prediction interval is $(1, z)\hat{\beta} \pm \{1 + (1, z)(Z'Z)^{-1}(1, z)'\}^{1/2} s \cdot t_{n-2,\alpha/2}$.

(#6.2.9) In this problem $(Y_i - \mu)/\sigma = \epsilon_i$ are *iid* random variables with density $f(\cdot)$, and $\rho(x) = -\log(f(x))$ a strictly convex twice differentiable function. The log-likelihood for the parameters (μ, σ) is $-n \log(\sigma) - \sum_{i=1}^n \rho((Y_i - \mu)/\sigma)$. (a). When $\sigma = \sigma_0$ is known, then the derivative of the log-likelihood with respect to μ is $-\sigma_0^{-1} \sum_{i=1}^n \rho'((Y_i - \mu)/\sigma_0)$, so the likelihood equation solving for the root $\mu = \hat{\mu}$ in this equation is as given in the problem, and the solution is unique if it exists, because this function is strictly increasing and continuous in μ. The solution exists by the intermediate value theorem if $\rho(-\infty) < 0 < \rho(\infty)$. Because f is assumed strictly positive and log-concave, this must be true. (For example, if $\rho(infty) \le 0$, then $f'(x) \ge 0$ everywhere and strictly increases and is not integrable.

(b). The likelihood in terms of $\theta_1 = 1/\sigma$ for fixed known $\theta_2^o = \mu/\sigma$, differentiated with respect to θ_1, yields the function $h(\theta_1) = n/\theta_1 - \sum_{i=1}^n$ that is strictly decreasing with respect to θ_1. Moreover, by the reasoning in (a) saying that $\rho(-\infty) < 0 < \rho(\infty)$, it follows that $h(0+) = \infty$ and $h(\infty) < 0$. Therefore, by the Intermediate Value Theorem there is a root $\hat{\theta}_1$ solving $h(\theta_1) = 0$, and the strict decrease of $h(\cdot)$ implies that root is unique.

(#6.2.10) We did the univariate version of part (a) of this problem a while ago in class.

(a). If $\theta_n^* = \theta_0 + O_P(n^{-1/2}$, then

$$\frac{1}{n} \sum_{i=1}^n J_\Psi(X_i, \theta^* \xrightarrow{P} E(J_\Psi(X_1, \theta_0)$$

and by the hint,

$$\Big[\frac{1}{n} \sum_{i=1}^n J_\Psi(X_i, \theta^*)\Big]^{-1} \frac{1}{n} \Psi(X_i, \theta_n^* =$$

$$\Big[E(J_\Psi(X_1, \theta_0) + o_P(1)\Big]^{-1} \sum_{i=1}^n \Psi(X_i, \theta_0 \; - \; (1 + o_P(1)) \, (\theta_n^* - \theta_0)$$

which by the definition of $\bar{\theta}_n$ in the problem implies property (6.2.3).

(b). Under A0-A4, the uniform version of the uniform inverse function theorem applies to the functions $g_n(\theta) = n^{-1} \sum_{i=1}^n \Psi(X_i, \theta)$ and $g(\theta) = E(\Psi(X_1, \theta))$. The conclusion of the uniform inverse function is just what we want for the assertion of this problem part. The tricky step of this problem part is to show or assume condition (i). I think we may need an assumption stronger than (A4) to prove this rigorously. An example of such as assumption is: $E[\sup_{\theta \in S_\epsilon(\theta_0)} \|J_\Psi(X_1, \theta)\|] < \infty$ for some sufficiently small $\epsilon > 0$.

(c). The point is that under the assumptions and hint, using (a), all Newton-Raphson iterates to find the root of the estimation equation starting from θ_n^* fall within the ϵ ball found in (b).