Problem 3. [50 pts] We have access to a data set $X_1, X_2, \cdots, X_n$ where X_i's are generated i.i.d. according to a distribution with the following pdf:

$$f(x|\sigma) = \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}}. \tag{1}$$

For this distribution, we know that $\mathbb{E}[X_i] = 0$, $\mathbb{E}[|X_i|] = \sigma$, and $\mathrm{Var}(X_i) = 2\sigma^2$. We consider a hypothesis testing problem with $H_0 : \sigma = \sigma_0$ and $H_a : \sigma = \sigma_1$. You may assume that $\sigma_0 < \sigma_1$.

(a) [10 pts] Write down the likelihood-ratio test-statistic for this setting. (You answer should be a function $T(X_1, \cdots, X_n)$)

Likelihood Function: $\mathcal{L}(\sigma) = \prod_{i=1}^{n} f(x|\sigma) = \prod_{i=1}^{n} \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}} = \left(\frac{1}{2\sigma}\right)^n \exp\left\{-\sum_{i=1}^{n} \frac{|x_i|}{\sigma}\right\}$

Likelihood Ratio Test Statistic:

$$\Lambda = \frac{\mathcal{L}(\sigma_0)}{\mathcal{L}(\sigma_1)} = \frac{\left(\frac{1}{2\sigma_0}\right)^n \exp\left\{-\frac{1}{\sigma_0}\sum_{i=1}^{n}|x_i|\right\}}{\left(\frac{1}{2\sigma_1}\right)^n \exp\left\{-\frac{1}{\sigma_1}\sum_{i=1}^{n}|x_i|\right\}} = \boxed{\left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left\{\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_0}\right)\sum_{i=1}^{n}|x_i|\right\}}$$

(b) [25 pts] Given a significance level α, specify the acceptance/rejection regions for the test-statistic obtained in part (a). [Hint: (i) By simplifying the likelihood ratio, show that the acceptance/rejection regions can be written in terms of the quantity $\frac{1}{n}\sum_{i=1}^{n}|X_i|$. (ii) Use the central limit theorem to argue that $\frac{1}{n}\sum_{i=1}^{n}|X_i|$ has a Gaussian distribution. (iii) Use this Gaussian distribution to derive the acceptance/rejection regions based on the significance level α.]

(i) Simplify The Likelihood Ratio

From the top, our rejection criteria in terms of Λ is $\Lambda < K$.
Reducing this to $\frac{1}{n}\sum_{i=1}^{n}|x_i|$ through algebraically equivalent statements:

$$\Lambda < K \Rightarrow \underbrace{\left(\frac{\sigma_1}{\sigma_0}\right)^n}_{>0 \text{ since } 0<\sigma_0<\sigma_1} \exp\left\{\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_0}\right)\sum_{i=1}^{n}|x_i|\right\} < K$$

$$\Rightarrow \ln\left[\left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left\{\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_0}\right)\sum_{i=1}^{n}|x_i|\right\}\right] < \ln(K)$$

$$\Rightarrow n[\ln(\sigma_1) - \ln(\sigma_0)] + \underbrace{\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_0}\right)}_{<0 \text{ since } \sigma_1>\sigma_0 \text{ so } \frac{1}{\sigma_1}<\frac{1}{\sigma_0}}\sum_{i=1}^{n}|x_i| < \ln(K)$$

$$\Rightarrow \sum_{i=1}^{n}|x_i| > \frac{\ln(K) - n[\ln(\sigma_1)-\ln(\sigma_0)]}{\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_0}\right)}$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}|x_i| > \left.\frac{\ln(K) - n[\ln(\sigma_1)-\ln(\sigma_0)]}{n\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_0}\right)}\right\} K'$$

(ii) Applying CLT to argue $Y = \frac{1}{n}\sum_{i=1}^{n}|x_i|$ has a Gaussian Distr.

$$E(Y) = E\left(\frac{1}{n}\sum_{i=1}^{n}|x_i|\right) = \frac{1}{n}\sum_{i=1}^{n}E(|x_i|) = \frac{1}{\cancel{n}}\cdot\cancel{n}\cdot\sigma = \sigma$$

$$Var(Y) = Var\left(\frac{1}{n}\sum_{i=1}^{n}|x_i|\right) \overset{iid}{=} \frac{1}{n^2}\cdot n\cdot Var(|x_i|) = \frac{\sigma^2}{n}$$

$$\left(\begin{aligned} \text{where } Var(|x_i|) &= E(|x_i|^2) - [E(|x_i|)]^2 = (E(x_i^2)) - [E(|x_i|)^2] \\ &= (Var(x_i) + [E(x_i)]^2) - [E[|x_i|]]^2 \\ &= 2\sigma^2 + 0 - \sigma^2 = \sigma^2 \end{aligned}\right)$$

Then $Y \sim N\left(\sigma, \frac{\sigma^2}{n}\right)$

cont. on final pg. →

[part (b) continued]

(iii) <u>Defining Rejection Region</u>

$$\alpha = \Pr(\text{Type I Error})$$
$$= \Pr(\text{Reject } H_0 \mid H_0 \text{ is true})$$
$$= \Pr(\Lambda < k \mid \sigma = \sigma_0)$$
$$= \Pr\Bigg(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{Y} > \underbrace{\frac{\ln(k) - n[\ln(\sigma_1) - \ln(\sigma_0)]}{n\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_0}\right)}}_{k'} \;\Bigg|\; \sigma = \sigma_0\Bigg)$$
$$= \Pr(Y > k' \mid \sigma = \sigma_0)$$
$$= \Pr\Bigg(\underbrace{\frac{Y - E(Y)}{\sqrt{Var(Y)}}}_{Z \sim N(0,1)} > \frac{k' - E(Y)}{\sqrt{Var(Y)}} \;\Bigg|\; \sigma = \sigma_0\Bigg)$$
$$= \Pr\Bigg(Z > \frac{k' - \sigma}{\sqrt{\frac{\sigma^2}{n}}} \;\Bigg|\; \sigma = \sigma_0\Bigg)$$
$$= \Pr\left(Z > \frac{\sqrt{n}(k' - \sigma_0)}{\sigma_0}\right)$$
$$= \boxed{1 - \Phi\left(\frac{\sqrt{n}(k' - \sigma_0)}{\sigma_0}\right)} \quad \text{where } k' = \frac{\ln(k) - n[\ln(\sigma_1) - \ln(\sigma_0)]}{n\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_0}\right)}$$

is our rejection region.

$$\therefore k' = \Phi^{-1}(1-\alpha)\cdot\frac{\sigma_0}{\sqrt{n}} + \sigma_0$$

$$\Rightarrow n\left[\Phi^{-1}(1-\alpha)\cdot\frac{\sigma_0}{\sqrt{n}} + \sigma_0\right]\cdot\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_0}\right) + n[\ln(\sigma_1) - \ln(\sigma_0)] = \ln(k)$$

$$\Rightarrow k = \exp\left\{n\left[\ln(\sigma_1) - \ln(\sigma_0) + \left(\frac{1}{\sigma_1} - \frac{1}{\sigma_0}\right)\left(\sigma_0 + \Phi^{-1}(1-\alpha)\cdot\frac{\sigma_0}{\sqrt{n}}\right)\right]\right\}$$

Statistic

(c) [15 pts] Consider the likelihood-ratio test in part (a) with the acceptance/rejection regions derived in part (b). Let t_{data} denote the realized value of the test on the sample data; i.e., if the sample data is $x_1, x_2, \cdots, x_n$, then $t_{\text{data}} = T(x_1, x_2, \cdots, x_n)$. Calculate the p-value for t_{data} and simplify your answer as much as you can.

$$\text{Let } t' = \frac{\ln(t_{data}) - n[\ln(\sigma_1) - \ln(\sigma_0)]}{n\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_0}\right)}$$

By similar algebra to part(b):

$$\text{p-value} = \Pr(\Lambda < t_{data})$$

b/c find prob/area bounded by test statistic instead of critical value this time.

$$= \Pr\left(\frac{1}{n}\sum_{i=1}^{n}|X_i| > t'\right)$$

$$= \Pr\left(Z > \frac{\sqrt{n}\,(t' - \sigma_0)}{\sigma_0}\right)$$

$$= \boxed{1 - \Phi\left(\frac{\sqrt{n}(t' - \sigma_0)}{\sigma_0}\right)}$$