Practice Problems for Exam
Statistics 201B

1. Let $X_1, ..., X_n$ be iid exponential with density $f(x;\beta) = \beta e^{-\beta x}$ for $x > 0$ and $\beta > 0$.

   (a) Find the asymptotic (large sample) likelihood ratio test of size α for $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$.

   (b) Find a Wald test of size α for the same null hypothesis.

   Solution:

   (a) The MLE of β can be calculated as follows.

$$\mathcal{L}_n(\beta) = \prod_{i=1}^{n} f(X_i;\beta) = \beta^n e^{-\beta \sum_{i=1}^{n} X_i}.$$

$$\Rightarrow \ell_n(\beta) = \log \mathcal{L}_n(\beta) = n \log \beta - n\beta \bar{X}.$$

$$\Rightarrow \frac{d\ell_n}{d\beta}(\beta) = \frac{n}{\beta} - n\bar{X}.$$

$$\frac{d\ell_n}{d\beta}(\beta) = 0 \Rightarrow \hat{\beta} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}_n}.$$

$$\frac{d^2\ell_n}{d\beta^2}(\beta) = -\frac{n}{\beta^2} < 0.$$

   So the MLE is $\hat{\beta} = \bar{X}^{-1}$.
   The likelihood ratio statistic is

$$\begin{aligned}\lambda &= 2\log \frac{L_n(\hat{\beta})}{L_n(\beta_0)} = 2(\ell_n(\hat{\beta}) - \ell_n(\beta_0)) = 2\left[-n\log \bar{X}_n - n - \big(n \log \beta_0 - n\beta_0 \bar{X}_n\big)\right] \\ &= 2n\left[\beta_0 \bar{X}_n - \log(\beta_0 \bar{X}_n) - 1\right] \\ &= 2n\left[\frac{\beta_0}{\hat{\beta}} - \log\left(\frac{\beta_0}{\hat{\beta}}\right) - 1\right]\end{aligned}$$

   The asymptotic distribution of λ is χ^2_1. So the rejection region is

$$\lambda > \chi^2_{1,\alpha},$$

   where $\chi^2_{1,\alpha}$ is the $(1-\alpha)$th quantile of χ^2_1 distribution.

(b) We've already found the MLE as $\hat{\beta} = \frac{1}{\bar{X}}$. Therefore, we know that $\hat{\beta}$ is asymptotically normal with asymptotic variance given by $I_n(\beta)^{-1}$.

$$\begin{aligned} I_n(\beta) &= -E(\frac{d^2\ell_n}{d\beta^2}(\beta)) \\ &= -E(-\frac{n}{\beta^2}) = \frac{n}{\beta^2} \end{aligned}$$

We can estimate $I_n(\beta)$ with either $I_n(\hat{\beta})$ (the expected Fisher Information) or $J_n(\hat{\beta})$ (the observed Fisher's information). We'll use $I_n(\hat{\beta})$ to get a Wald's statistic of

$$W = \left|\frac{\hat{\beta} - \beta_0}{\hat{\beta}/\sqrt{n}}\right| = \sqrt{n}\left|1 - \frac{\beta_0}{\hat{\beta}}\right|$$

We will reject with W is greater than the $1 - \alpha/2$ quantile of a normal, $z_{1-\alpha/2}$

2. (Wasserman, Question 13) Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$.

(a) Find the asymptotic (large sample) likelihood ratio test of size α for $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

We know that the MLE for the full likelihood is

$$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

If $\mu = \mu_0$, then the log-liklihood is

$$\begin{aligned} \ell(\sigma^2) &= \sum_i \{-\frac{1}{2}log(2\pi\sigma^2) - \frac{(X_i - \mu_0)^2}{2\sigma^2}\} \\ &= \sum_i -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_i (X_i - \mu_0)^2\} \end{aligned}$$

Taking the derivative *with respect to* σ^2 gives us

$$\frac{d\ell}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_i (X_i - \mu_0)^2}{2(\sigma^2)^2} = \frac{1}{2\sigma^2}(\frac{\sum_i (X_i - \mu_0)^2}{\sigma^2} - n)$$

This is equal to 0 if $\sigma^2 = \frac{1}{n}\sum_i (X_i - \mu_0)^2$.

Checking if it is a maximum:

$$\frac{d^2\ell}{d(\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{\sum_i (X_i - \mu_0)^2}{(\sigma^2)^3} = \frac{1}{(\sigma^2)^2}(\frac{n}{2} - \frac{\sum_i (X_i - \mu_0)^2}{\sigma^2})$$

Plugging in $\sigma^2 = \frac{1}{n}\sum_i (X_i - \mu_0)^2$ gets us $\frac{n}{2} - n < 0$, so it is a maximum. Then our likelihood ratio test is given by

$$\lambda = -2(\ell(\hat{\mu}, \hat{\sigma}^2) - \ell(\mu_0, \hat{\sigma}_0^2))$$

where

$$\ell(\mu_0, \hat{\sigma}_0^2) = -\frac{n}{2}log(2\pi) - \frac{n}{2}\log(\hat{\sigma}_0^2) - \frac{n}{2}$$
$$\ell(\hat{\mu}, \hat{\sigma}^2) = -\frac{n}{2}log(2\pi) - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{n}{2}$$

This gives us

$$\lambda = 2(-\frac{n}{2}\log(\hat{\sigma}^2) + \frac{n}{2}\log(\hat{\sigma}_0^2)) = n\log(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}) = n\log(\frac{\sum_i (X_i - \mu_0)^2}{\sum_i (X_i - \bar{X})^2})$$

Asymptotically, λ will be χ^2_{k-q}, where k is the dimensions of the full space (R^2) and q is the dimension of the smaller space (a line in R^2). So $k = 2$ and $q = 1$, so asymptotically λ will follow a χ^2_1 distribution. This gives us an asymptotic likelihood ratio test that rejects if λ is greater than the $1 - \alpha$ quantile of a χ^2_1 distribution.
As a sanity check, we can see

$$\frac{\sum_i (X_i - \mu_0)^2}{\sum_i (X_i - \bar{X})^2} = \frac{\sum_i (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2}{\sum_i (X_i - \bar{X})^2}$$

(the cross-product term cancels) so if $\bar{X}$ is close to μ_0, λ will be close to 0 and as $\bar{X}$ gets far away from μ, λ will be large.

(b) Find a Wald test of size α for the same null hypothesis.

Under the null hypothesis,

$$T = \frac{\bar{X} - \mu_0}{\hat{\sigma}}$$

where

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}(X - \bar{X})^2$$

is asymptotically $N(0,1)$. Indeed, we could pick different estimators for σ^2 so long as they were all consistent for σ^2 under the null.

We would reject when $|T|$ is larger than $z_{1-\alpha/2}$, the $1-\alpha/2$ quantile of a standard normal distribution.

3. Suppose that instead of 0-1 loss, you had a loss that was not symmetric, so that different mistakes counted differently,

$$L(Y,\hat{Y}) = \begin{cases} 0 & \hat{Y} = Y \\ a & \hat{Y} = 0 \text{ and } Y = 1 \\ b & \hat{Y} = 1 \text{ and } Y = 0 \end{cases}$$

Show that the following is true for the optimal (Bayes) classification rule $h^*(x)$ in this case:

(a)

$$h^*(x) = \begin{cases} 1 & \text{if } r(x) > \frac{b}{a+b} \\ 0 & \text{otherwise} \end{cases}$$

(b) The decision boundary is given by

$$\{x : P(Y=1|X) = \frac{b}{a}P(Y=0|X)\}$$

where $r(x) = P(Y=1|X) = \frac{\pi_1 f_1(x)}{\pi_1 f_1(x) + (1-\pi_1) f_0(x)}$

Solution: We can write our loss as a single function using indicator functions:

$$L(Y,\hat{Y}) = aI(Y=0,\hat{Y}=0) + bI(Y=0,\hat{Y}=1).$$

$$\begin{aligned} E(L(Y,\hat{Y})) &= aI(\hat{Y}=0)E(I(Y=1)|X) + bI(\hat{Y}=1)E(I(Y=0)|X) \\ &= aI(\hat{Y}=0)P(Y=1|X) + bI(\hat{Y}=1)P(Y=0|X) \end{aligned}$$

In otherwords, if we choose $\hat{Y}=0$, our loss is $aP(Y=1|X)$ and if we choose $\hat{Y}=1$, our loss is $bP(Y=0|X)$. Therefore, to minimize this quantity for a given X, we want to choose $\hat{Y}=1$ if $bP(Y=0|X) < aP(Y=1|X)$, and otherwise choose $\hat{Y}=0$. Since $P(Y=0|X) = 1 - P(Y=1|X)$, this gives that we want to choose $\hat{Y}=1$ if $r(x) > \frac{a}{a+b}$. And our decision boundary will be when these are equal, $\{x : aP(Y=1|X=x) = bP(Y=0|X=x)\}$

4. Show that decision boundary for logistic regression is linear.

Solution:

Our logistic classification rule is

$$h(x) = \begin{cases} 1 & \text{if } r(x) > 1/2 \\ 0 & \text{otherwise} \end{cases}$$

where

$$r(x) = P(Y = 1|X) = \frac{e^{\beta_0 + \sum_j \beta_j x^j}}{1 + e^{\beta_0 + \sum_j \beta_j x^{(j)}}}$$

Thus the decision boundary is $\{x : r(x) = 1/2\}$.

$$\begin{aligned} e^{\beta_0 + \sum_j \beta_j x^j} &= 1/2 + 1/2 e^{\beta_0 + \sum_j \beta_j x^{(j)}} \\ e^{\beta_0 + \sum_j \beta_j x^j} &= 1 \\ \beta_0 + \sum_j \beta_j x^j &= 0 \end{aligned}$$

So the decision boundary is

$$\{x : \beta_0 + \sum_j \beta_j x^j = 0\}$$

which defines a linear boundary in R^p (a hyperplane)

Logistic regression will estimate the quantities β_j, so the final rule will replace β_j with $\hat{\beta}_j$, but this does not change the fact that the decision boundary is linear.