University of Illinois
ECE561

Spring 2008
Moulin

Final Exam
Wednesday, May 7, 2008

Name SOLUTION SET

Score ______________

Please do not turn this page until requested to do so.

You may use Poor's textbook and your class notes for this exam. Make sure to show all of your work on the longer problems for partial credit. **Box your answers, and write neatly.** Backs of pages may be used for scratch work if necessary.

There are 5 problems, for a total of 100 points. **No calculators are allowed.**

Problem 1 (20 pts). Let N_i, $1 \le i \le n$, be iid Bernoulli random variables with $Pr[N=1] = \frac{1}{1+e}$. Consider the following binary hypothesis test:

$$\begin{cases} H_0: & Y_i = N_i, \\ H_1: & Y_i = N_i \oplus 1, \quad 1 \le i \le n, \end{cases}$$

where $\oplus$ denotes modulo-2 addition.

(a) (10pts) Derive a test that maximizes $Pr[\text{say } H_0|H_0]$ subject to the constraint $Pr[\text{say } H_0|H_1] \le \beta$.

Let $\lambda = \frac{1}{1+e}$. The likelihood ratio is given by

$$L(\underline{y}) = \frac{p_1(\underline{y})}{p_0(\underline{y})} = \frac{(1-\lambda)^{n_1}\lambda^{n-n_1}}{\lambda^{n_1}(1-\lambda)^{n-n_1}} = \left(\frac{\lambda}{1-\lambda}\right)^{n-2n_1} = e^{2n_1-n}$$

Use threshold e^{τ} for LRT, get randomized test

$$\boxed{\tilde{\delta}(y) = \begin{cases} 1 \\ \gamma \\ 0 \end{cases} \text{if } n_1 \begin{matrix} > \\ = \\ < \end{matrix} \frac{\tau+n}{2}}$$

where τ is the largest value such that $\underbrace{P_1\left[N < \frac{\tau+n}{2}\right]}_{= P_0\left[N > \frac{n-\tau}{2}\right]} \le \beta$.

(b) (10pts) Let $\beta = e^{-n/10}$. Give a *simple* approximate value for the threshold of your test, and argue that the resulting test performs nearly as well as the optimal one.

$(a > \lambda)$. We have $P_0[N > na] \le e^{-n\mu^*(a)}$ where $\mu^*(a) = a\ln\frac{a}{\lambda} + (1-a)\ln\frac{1-a}{1-\lambda}$

$\downarrow$

tight on exponential scale as $n \to \infty$ (Cramer's theorem)

Use with $a = \frac{1-\tau/n}{2}$ and $\mu^*(a) = \frac{1}{10}$ $\Rightarrow a = (\mu^*)^{-1}\left(\frac{1}{10}\right)$

$$\Rightarrow \boxed{\frac{\tau}{n} = 1 - 2(\mu^*)^{-1}\left(\frac{1}{10}\right)}$$

(with $\frac{\tau}{n} = 1-2a$)

<u>Note</u>: Can further approximate $\mu^*(a) \sim \frac{(a-\lambda)^2}{2\lambda(1-\lambda)}$ for $a \downarrow \lambda$

$$\Rightarrow a \simeq \frac{2\lambda(1-\lambda)}{10} + \lambda = \frac{\lambda(6-\lambda)}{5}$$

$$\Rightarrow \boxed{\frac{\tau}{n} \simeq 1 - \frac{2\lambda(6-\lambda)}{5}}$$

Problem 2 (20 pts). Let Y_1 and Y_2 be independent Poisson random variables with respective parameters θ and θ^2.

(a) (10pts) Is there a one-dimensional sufficient statistic for θ given (Y_1, Y_2)? If so, identify it.

Poisson law : $e^{-\theta} \frac{\theta^y}{y!}$

Here $P_\theta(y_1, y_2) = P_\theta(y_1)\, P_{2\theta}(y_2) = e^{-\theta-\theta^2} \cdot \frac{\theta^{y_1+2y_2}}{y_1!\, y_2!}$

By Factorization theorem, $\boxed{T(y_1, y_2) = y_1 + 2y_2}$ is a sufficient statistic.

(b) (10pts) Is there a MVUE for θ or for some invertible function of θ?

$E[T(Y)] = E[Y_1 + 2Y_2] = \theta + 2\theta^2$

$\Rightarrow T(\underline{Y})$ is an unbiased estimator of $\theta + 2\theta^2$

Moreover $T(\underline{Y})$ is a sufficient statistic in a 1-dim. exponential family

$\Rightarrow T(\underline{Y})$ is a complete sufficient statistic

and is $\boxed{\text{MVUE for } \theta + 2\theta^2}$

Problem 3 (20 pts).

(a) (10pts) Derive the ML estimator of $\theta \geq 0$ given the data

$$Y_i = \frac{\theta}{i^2} + N_i, \quad 1 \leq i \leq n$$

where $\{N_i,\ 1 \leq i \leq n\}$ are iid $\mathcal{N}(0, 1)$.

This is a variation (due to the constraint $\theta \geq 0$) of the problem $\underline{Y} = \theta \underline{S} + \underline{N}$ studied in class. The signal here is $S_i = \frac{1}{i^2}$ for $1 \leq i \leq n$. The ML estimator is

$$\hat{\theta}_{ML} = \arg\max_{\theta \geq 0} \ln p(\underline{y}|\theta) = \arg\min_{\theta \geq 0} \underbrace{\|\underline{y} - \theta \underline{s}\|^2}_{\text{quadratic in } \theta \text{, with global max at } \hat{\theta}_u = \frac{\underline{y}^T \underline{s}}{\underline{s}^T \underline{s}}}$$

$$\Rightarrow \hat{\theta}_{ML} = \max\left\{0, \hat{\theta}_u\right\} = \boxed{\max\left\{0, \frac{\sum_{i=1}^{n} y_i/i^2}{\sum_{i=1}^{n} 1/i^4}\right\}}$$

(b) (5pts) Is the ML estimator unbiased?

We have seen that the unconstrained estimator $\hat{\theta}_u$ is unbiased: $E[\hat{\theta}_u] = \theta$.

However $E[\hat{\theta}_{ML}] = E[\max\{0, \hat{\theta}_u\}] \neq \theta$

$\Rightarrow \hat{\theta}_{ML}$ is <u>biased</u>.

(c) (5pts) Is the ML estimator consistent as $n \to \infty$? Why?

- First we analyze $\hat{\theta}_u$. We have

$$\mathrm{Var}[\hat{\theta}_u] \underset{\substack{\downarrow \\ \text{since } \hat{\theta}_u \text{ is unbiased}}}{=} E(\hat{\theta}_u - \theta)^2 = E\left[\frac{\sum_{i=1}^{n} N_i/i^2}{\sum_{i=1}^{n} 1/i^4}\right]^2 = \frac{\sigma^2}{\sum_{i=1}^{n} 1/i^4} > \frac{\sigma^2}{\underset{\substack{\downarrow \\ \text{Riemann-zeta function}}}{\zeta(4)}} \quad \forall n.$$

Hence $\hat{\theta}_u$ is not consistent. The signal $\underline{s}$ is too weak, its total energy $< \zeta(4)$ does not tend to ∞ as $n \to \infty$.

- Returning to $\hat{\theta}_{ML}$, we have

$$\Pr\left[|\hat{\theta}_{ML} - \theta| \geq \theta\right] \geq \Pr\left[\hat{\theta}_{ML} = 0\right] = \Pr\left[\hat{\theta}_u \leq 0\right] = Q\left(\frac{\theta}{\sqrt{\mathrm{Var}(\hat{\theta}_{ML})}}\right)$$ does not $\to 0$ as $n \to \infty$

$\Rightarrow \hat{\theta}_{ML}$ is not consistent

Problem 4 (20 pts). Let Y be a binary random variable with $Pr[Y=1] = \theta = 1 - Pr[Y=0]$. Consider the randomized estimator

$$\hat{\theta} = \begin{cases} 1/2 & \text{with prob. } \gamma \\ Y & \text{with prob. } 1-\gamma \end{cases}$$

whose performance will be assessed using the cost function $C(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$.

(a) (12 pts) Give an expression for the conditional risk of this estimator, in terms of θ and γ.

Let $\hat{\theta}_1 = \frac{1}{2}$ (picked with prob. γ)

and $\hat{\theta}_2 = Y$ (——— $1-\gamma$)

We have $R_\theta(\hat{\theta}) = \gamma R_\theta(\hat{\theta}_1) + (1-\gamma) R_\theta(\hat{\theta}_2)$

where $R_\theta(\hat{\theta}_1) = E(\theta - \frac{1}{2})^2 = (\theta - \frac{1}{2})^2$

$R_\theta(\hat{\theta}_2) = E(\theta - Y)^2 = \theta(\theta-1)^2 + (1-\theta)\theta^2 = \theta(1-\theta)$

$$\Rightarrow \boxed{R_\theta(\hat{\theta}) = \gamma(\theta - \tfrac{1}{2})^2 + (1-\gamma)\theta(1-\theta)}$$

(b) (5 pts) Find the value of γ that minimizes the worst-case risk over $\theta \in [0,1]$.

$$R_\theta(\hat{\theta}) = \theta(1-\theta) + \gamma\left[(\theta - \tfrac{1}{2})^2 - \theta(1-\theta)\right]$$

$$= \theta(1-\theta) + \gamma\left[2\theta^2 - 2\theta + \tfrac{1}{4}\right]$$

$$= (1-2\gamma)\underbrace{\theta(1-\theta)}_{\leq 1/4} + \frac{\gamma}{4} \begin{cases} \leq \frac{1-2\gamma}{4} + \frac{\gamma}{4} = \frac{1-\gamma}{4} & \text{for } \gamma \leq 1/2 \\ \leq \gamma/4 & \text{for } \gamma \geq 1/2 \end{cases}$$

$\Rightarrow$ choosing $\boxed{\gamma = \frac{1}{2}}$ yields an equalizer rule (with risk $= \frac{1}{8}$ $\forall\theta$)

(c) (3 pts) Can you argue that the above test is minimax?

This is an equalizer rule but we need to check whether it is also a Bayes rule.

The minimax rule (as given in class) is $\hat{\theta}_{MX} = \frac{1}{2}Y + \frac{1}{4}$ and has risk $r_{MX} = \frac{1}{16} < \frac{1}{8}$

$\Rightarrow$ The test above is not minimax.

There was a typo in the statement of this problem! The problem I wanted you to solve had $\sum_{i=1}^{8}$ and $\sin^2$ switched, See next page.

Problem 5 (20 pts). Consider the following 8-dimensional integral, viewed as a function of $t \geq 0$:

$$f(t) = \int_{[-1,1]^8} \underbrace{\exp\left\{-t\sum_{i=1}^{8}\sin^2\left(\frac{\pi x_i}{2}\right)\right\}}_{e(\underline{x})} dx_1\, dx_2 \cdots dx_8, \quad t \geq 0.$$

(a) (4pts) Give the minimum and the maximum value of the integrand.

$$e^{-8t} \leq e(\underline{x}) \leq 1$$

achieved when $x_i = \pm 1$ for all $1 \leq i \leq 8$.

achieved when $\underline{x} = \underline{0}$

(b) (8 pts) Propose a computationally efficient discretization method for evaluating $f(0.1)$. How many points are needed to obtain a precision of 1%? (A ballpark estimate is sufficient.)

Since $\exp\{\sum_{i=1}^{8} \cdots\} = \prod_{i=1}^{8} \exp\{\cdots\}$, the 8-dimensional integral giving $f(t)$ reduces to $f(t) = (f_1(t))^8$, where

$$f_1(t) = \int_{-1}^{1} \exp\left\{-t \sin^2\left(\frac{\pi x}{2}\right)\right\} dx,$$

which can be evaluated using the standard trapezoidal rule. (approximation power $= O(1/n^2)$).

For $t = 0.1$, $f(0.1) \approx \int_{[-1,1]^8} d\underline{x} = 2^8 = 256$ (ballpark estimate)

So we need a relative accuracy of $\frac{1}{256}$ on $f(0.1)$ and $\frac{1}{256 * 8} = \frac{1}{2048}$ on $f_1(0.1)$ $\Rightarrow \frac{1}{n^2} \simeq \frac{1}{2048} \Rightarrow \boxed{n \approx 64}$

(c) (8 pts) Propose a computationally efficient discretization method for evaluating $f(2)$. Explain why your method from (b) would be inefficient here.

For $t = 2$, the integrand $e_1(x) = \exp\{-t \sin^2(\frac{\pi x}{2})\}$ has a strong peak at $x = 0$ $\Rightarrow$ need finer discretization. Trapezoidal rule is still appropriate.

Problem 5 (20 pts). Consider the following 8-dimensional integral, viewed as a function of $t \geq 0$:

$$f(t) = \int_{[-1,1]^8} \underbrace{\exp\left\{-t\sum_{i=1}^{8} \sin^2\left(\frac{\pi x_i}{2}\right)\right\}}_{e(\underline{x})} dx_1\, dx_2 \cdots dx_8, \quad t \geq 0.$$

If $\sum_{i=1}^{8}$ and $\sin^2$ are switched, the 8-dim. integral does not factor into a product of 1-D integrals

(a) (4pts) Give the minimum and the maximum value of the integrand.

$$\exp\{-8t\} \leq \underbrace{\text{Integrand}}_{=e(\underline{x})} \leq 1$$

$\exp\{-8t\}$ achieved when $x_i = \pm 1\ \forall i$

1 achieved when $x_i = 0\ \forall i$

(b) (8 pts) Propose a computationally efficient discretization method for evaluating $f(0.1)$. How many points are needed to obtain a precision of 1? (A ballpark estimate is sufficient.)

Since the dimension of the integral is > 4, a stochastic method is appropriate. For $t=0.1$, we obtain from Part (a): $e^{-0.8} \leq e(\underline{x}) \leq 1$

$\Rightarrow$ the integrand varies slowly.

Draw points $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ from uniform distribution over $[-1,1]^8$ and approximate $f(t)$ by

$$\boxed{\hat{f}_n(t) = \frac{2^8}{n} \sum_{j=1}^{n} e(\underline{x}_j)}$$

where normalization constant 2^8 ensures that

$$E[\hat{f}_n(t)] = 2^8\, E[e(\underline{X})] = f(t).$$

Standard Deviation $\text{s.d.}[\hat{f}_n(t)] = \frac{2^8\, \text{s.d.}[e(\underline{X})]}{\sqrt{n}} \simeq \frac{2^8}{\sqrt{n}} \Rightarrow$ choose $\sqrt{n} \simeq 2^8$

$\Rightarrow \boxed{n \simeq 2^{16}}$

(c) (8 pts) Propose a computationally efficient discretization method for evaluating $f(2)$. Explain why your method from (b) would be inefficient here.

For $t=2$, we obtain from Part (a): $e^{-16} \leq e(\underline{x}) \leq 1$

Most of the mass of $e(\underline{x})$ is highly concentrated near $\underline{x} = \underline{0}$.

$\Rightarrow$ Sampling from uniform distribution as in Part (b) would be very ineffective

$\Rightarrow$ sample from pdf that resembles $e(\underline{x})$ in the vicinity of $\underline{0}$

e.g., use $g(\underline{x}) = \left(\frac{4}{\pi t}\right)^{-4} \exp\left\{-t\left(\frac{\pi}{2}\right)^2 \sum_{i=1}^{8} x_i^2\right\}$

(iid $\mathcal{N}(0,\sigma^2)$ with $\sigma^2 = \frac{2}{t\pi^2} = \frac{1}{\pi^2}$)

and obtain

$$\boxed{\hat{f}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{e(\underline{x}_i)}{g(\underline{x}_i)}}$$

(importance sampling)