MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Spring 2010)

Problem Set 11 (never due)

1. (a) Since $M_n = n$ when $K_n = 0$, we have $\hat{M}_n = n$ when $K_n = 0$. Consider now the case $K_n = k > 0$. The probability $q_{k,m}$ of having a total of $k > 0$ successes, with m trials before the first success (where $m \in \{0, 1, \ldots, n-k\}$) is given by

$$q_{k,m} = p^k(1-p)^{n-k}\binom{n-m-1}{k-1} = \frac{p^k(1-p)^{n-k}}{(k-1)!}(n-k+1-m)(n-k+2-m)\cdot\dots\cdot(n-m-1).$$

Since, for a given k, the value of $q_{k,m}$ is monotonically decreasing as $m \in \{0, 1, \ldots, n-k\}$ increases, the minimum of $1_{k,m}$ with respect to m is achieved at $m = 0$, i.e. $\hat{M}_n = 0$ when $K_n > 0$. Evidently, $\hat{M}_n$ is uniquely defined by n and K_n.

(b) Since the value $p = k/n$ maximizes $p^k(1-p)^{n-k}\binom{n}{k}$ over all $p \in [0, 1]$, the ML estimate of p is K_n/n. Since p^2 is a one-to-one function of $p \in [0, 1]$, the ML estimate of p^2 is $(K_n/n)^2$. Since K_n is not a constant unless $p \in \{0, 1\}$, and

$$\mathbf{E}[(K_n/n)^2] = \mathbf{E}[K_n/n]^2 + \text{var}(K_n/n) = p^2 + \frac{p(1-p)}{n},$$

we conclude that $(K_n/n)^2$ is not an unbiased estimate of p^2. However, for $n \to \infty$, $(K_n/n)^2$ is an asymptotically unbiased estimate of p^2.

Since, by the law of large numbers, K_n/n converges to p in probability, $(K_n/n)^2$ is a consistent estimate of p^2.

(c) This is a classical hypothesis testing setup, in which $g(K_n) = 1$ means rejecting the hypothesis H_0 that $p = 1/2$, in favor of the hypothesis H_1 that $p = 2/3$. We require the probability of false rejection of H_0 to be not larger than 0.5, and want to minimize the probability of false acceptance of H_0. According to the Neyman-Pearson Lemma, a likelihood ratio test $L(K_n) \leq \xi$, i.e. a function

$$g(k) = \begin{cases} 0, & L(k) \leq \xi, \\ 1, & L(k) > \xi, \end{cases}$$

where $L(k) = p_1(k)/p_0(k)$, while $p_0(k)$ equals $\mathbf{P}(K_n = k)$ when $p = 1/2$, and $p_0(k)$ equals $\mathbf{P}(K_n = k)$ when $p = 2/3$, will be optimal, as long as $\mathbf{P}(1 = g(K_n))$ equals 0.5 for $p = 1/2$. Since

$$p_0(k) = 2^{-n}\binom{n}{k}, \quad p_1(k) = 2^k 3^{-n}\binom{n}{k},$$

the likelihood ratio $L(k) = 2^k 2^n 3^{-n}$ is strictly monotonic with respect to k, i.e. every likelihood ratio test has the form $k \leq k_0$. Since selecting $k_0 = 4.5$ yields, for $p = 1/2$,

$$\mathbf{P}(K_n < 4.5) = 2^{-n}\sum_{k=0}^{4}\binom{n}{k} = 0.5,$$

an optimal g is given by

$$g(k) = \begin{cases} 0, & k \in \{0, 1, 2, 3, 4\}, \\ 1, & k \in \{5, 6, 7, 8, 9\}, \end{cases}$$

and achieves

$$\beta = \mathbf{P}(K_n \in \{0, 1, 2, 3, 4\}; p = 2/3) = 3^{-n}\sum_{k=0}^{4} 2^k\binom{n}{k} \approx 0.145.$$

2. We have

$$f_{Y|\Theta}(y|\theta) = \begin{cases} \frac{1}{|\theta|}, & |\theta| \in (0,1),\ y(\theta - y) > 0, \\ 0, & |\theta| \in (0,1),\ y(\theta - y) \le 0, \\ \text{not defined}, & |\theta| \notin (0,1) \end{cases}$$

Hence

$$f_{Y,\Theta}(y,\theta) = f_{Y|\Theta}(y|\theta) f_\Theta(\theta) = \begin{cases} \frac{1}{2|\theta|}, & |\theta| \in (0,1),\ y(\theta - y) \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

(a) The MAP estimator $\hat{\Theta}$ for Θ given Y is Y.

(b) The least mean squares estimator $\hat{\Theta}$ for Θ given Y is defined by $\hat{\Theta} = g(Y)$, where, for $y \in (0,1)$,

$$g(y) = \frac{\int_y^1 \frac{\theta d\theta}{2\theta}}{\int_y^1 \frac{d\theta}{2\theta}} = \frac{y-1}{\log(y)},$$

and, by symmetry,

$$g(-y) = -g(y) = \frac{y+1}{\log(|y|)}$$

for $y \in (-1,0)$. Since $\mathbf{P}(Y=0) = 0$, it does not really matter how $g(0)$ is defined.

(c) Since

$$\mathbf{E}[\Theta] = 0,\ \text{var}(\Theta) = \frac{2^2}{12} = \frac{1}{3},\ \mathbf{E}[W] = \frac{1}{2},\ \ \text{var}(W) = \frac{1}{12},$$

we have

$$\begin{aligned} \mathbf{E}[\Theta^2] &= \text{var}(Th) = \frac{1}{3}, \\ \mathbf{E}[W^2] &= \text{var}(W) + \mathbf{E}[W]^2 = \frac{1}{3}, \\ \mathbf{E}[Y] &= \mathbf{E}[\Theta W] = \mathbf{E}[\Theta]\mathbf{E}[W] = 0, \\ \text{var}(Y) &= \mathbf{E}[Y^2] = \mathbf{E}[\Theta^2 W^2] = \mathbf{E}[\Theta^2]\mathbf{E}[W^2] = \frac{1}{9}, \\ \mathbf{E}[Y\Theta] &= \mathbf{E}[\Theta^2 W] = \mathbf{E}[\Theta^2]\mathbf{E}[W] = \frac{1}{6}, \end{aligned}$$

the least mean squares linear estimator $\hat{\Theta}$ for Θ given Y is defined by $\hat{\Theta} = kY$, where

$$k = \frac{\text{cov}(Y,\Theta)}{\text{var}(Y)} = \frac{3}{2}.$$

(d) The least mean absolute error estimator $\hat{\Theta}$ for Θ given Y is defined by $\hat{\Theta} = h(Y)$, where, for $y > 0$, $h(y) = v$ is a number $v \in (y,1)$ such that

$$\int_y^v \frac{d\theta}{2\theta} = \int_v^1 \frac{d\theta}{2\theta},$$

i.e. $v = \sqrt{y}$. By symmetry, we have

$$h(y) = \frac{y}{\sqrt{|y|}}$$

for $|y| \in (0,1)$.

When $\Theta > 0$ is a constant, we have

$$f_Y(y;\Theta) = \begin{cases} 1/\Theta, & y \le \Theta, \\ 0, & y > \Theta. \end{cases}$$

(e) Given that the observed value of Y is y, $\hat{\Theta}$ is the argument of maximum, over $\theta > 0$, of $f_Y(y,\theta)$, i.e. $\hat{\Theta} = Y$.

(f) Since

$$\mathbf{P}(\theta X < \theta < \rho\theta X) = \mathbf{P}(X > 1/\rho),$$

we need $1/rho = 1 - 0.95 = 1/20$, hence $\rho = 20$.

3. Since, for $k \in \{0, 1, 2, \dots\}$,

$$\mathbf{P}(k-1 < T \le t) = e^{1-k} - e^{-t},$$

for $k - 1 < t \le k$ we have

$$F_{T|K}(t|k) = \frac{\mathbf{P}(k-1 < T \le t)}{\mathbf{P}(k-1 < T \le k)} = \frac{e^{1-k} - e^{-t}}{e^{1-k} - e^{-k}}.$$

Hence, for $k - 1 < t \le k$,

$$f_{T|K}(t|k) = \frac{e^{-t}}{e^{-k+1} - e^{-k}}.$$

(a) Since the maximum (with respect to t) of $f_{T|K}(t|k)$ is achieved at $t = k - 1$, we have $g_0(k) = k - 1$.

(b) $v = g_1(k)$ is the number from $(k-1, k)$ such that

$$\int_{k-1}^{v} e^{-t} dt = \int_{v}^{k} e^{-t} dt,$$

i.e.

$$g_1(k) = \log\left(\frac{e^{-k} + e^{-k+1}}{2}\right).$$

(c)

$$g_2(k) = \int_{k-1}^{k} t f_{T|K}(t|k) dt = k - \frac{1}{e-1}.$$

4. (a) The joint probability distribution of $Y \in [20, 110]$ and $U \in \{A, B\}$ is given by

$$f_{Y,U}(y,u) = \begin{cases} \frac{1}{3}a(y), & u = A, \\ \frac{2}{3}b(y), & u = B, \end{cases}$$

where

$$\begin{aligned} a(y) &= \max\left\{0, \frac{1}{30} - \frac{|y-80|}{30 \cdot 30}\right\}, \\ b(y) &= \max\left\{0, \frac{1}{30} - \frac{|y-50|}{30 \cdot 30}\right\}. \end{aligned}$$

According to the MAP principle, $g_{mpe}(y) = A$ when $f_{Y,U}(y, A) > f_{Y,U}(y, B)$ (i.e. $y \in (70, 110)$), and $g_{mpe}(y) = B$ when $f_{Y,U}(y, A) < f_{Y,U}(y, B)$ (i.e. $y \in (20, 70)$, while the value of g for other arguments is irrelevant, because Y is either in $(70, 110)$ or in $(20, 70)$ with probability 1. In other words,

$$g_{mpe}(y) = \begin{cases} A, & y \geq 70, \\ B, & y < 70. \end{cases}$$

(b) For a type A duck, $f_Y(y) = f_Y(y; A) = a(y)$, while $f_Y(y) = f_Y(y; B) = b(y)$ for a type B duck. The ML principle calls for $g(y) = A$ whenever $f_Y(y; A) \geq f_Y(y; B)$, i.e.

$$g_{ML}(y) = \begin{cases} A, & y \geq 65, \\ B, & y < 65. \end{cases}$$

(c) According to the Neyman-Pearson Lemma, an optimal test g_* is given by the likelihood ratio criterion $f_Y(y; A) \geq \xi f_Y(y; B)$ for some $\xi > 0$, which will be equivalent to $y \geq y_0$ for some $y_0 \in (50, 80)$. For this ratio test to reject A falsely with probability 1/18, we need $y_0 = 60$, i.e. $\xi = a(y_0)/b(y_0) = 0.5$. Hence

$$g_*(y) = \begin{cases} A, & y \geq 60, \\ B, & y < 60, \end{cases}$$

and the probability of false acceptance of A is 2/9.

5. For $y \in (0, 1)$, $n \in \{1, 2, \dots\}$ we have

$$F_{Y,N}(y, n) = \mathbf{P}(Y \leq y, N = n) = \mathbf{P}(X^n \leq y, N = n) = \mathbf{P}(X \leq y^{1/n}, N = n) = y^{1/n} p(1-p)^{n-1},$$

hence

$$f_{Y,N}(y, n) = \frac{dF_{Y,N}(y, n)}{dy} = \frac{1}{n} y^{1/n-1} p(1-p)^{n-1}.$$

The most likely value $\hat{n}$ of N, given that $Y = y$, is any argument of maximum of $f_{Y,N}(y, n)$ with respect to positive integer n.

To find this maximum (for the case $y = e^{-12}$, $p = 1 - e^{-1}$), let us first maximize the function

$$h(r) = \frac{1}{r} y^{1/r} (1-p)^r$$

with respect to $r \in (0, \infty)$. Note that

$$\frac{d \log(h(r))}{d} r = \log(1-p) - \frac{\log(y)}{r^2} - \frac{1}{r}.$$

When $y = e^{-12}$ and $p = 1 - e^{-1}$, this yields

$$\frac{d \log(h(r))}{d} r = -1 - \frac{1}{r} + \frac{12}{r^2},$$

which is positive for $r \in (0, 3)$, and negative for $r \in (3, \infty)$. Hence $r = 3$ is the argument of maximum. Since $r = 3$ is readily integer, we have $\hat{N} = 3$ for $Y = e^{-12}$ and $p = 1 - e^{-1}$.