# Massachusetts Institute of Technology

Department of Electrical Engineering & Computer Science

**6.041/6.431: Probabilistic Systems Analysis**

(Fall 2011)

**Tutorial 11 Solutions**
**December 9, 2011**

1. (a) We first find the posterior distribution for $\Theta$:

$$\begin{aligned}
f_{\Theta|X}(\theta \mid k) &= \frac{p_{X|\Theta}(k \mid \theta) \cdot f_\Theta(\theta)}{p_X(k)} \\
&= \frac{p_{X|\Theta}(k \mid \theta) \cdot f_\Theta(\theta)}{\int_0^1 p_{X|\Theta}(k \mid t) \cdot f_\Theta(t)dt} \\
&= \frac{\binom{n}{k} \cdot \theta^k \cdot (1-\theta)^{(n-k)} \cdot 1}{\int_0^1 \binom{n}{k} \cdot t^k \cdot (1-t)^{(n-k)} \cdot 1dt} \\
&= \frac{\binom{n}{k} \cdot \theta^k \cdot (1-\theta)^{(n-k)}}{\binom{n}{k} \int_0^1 \cdot t^k \cdot (1-t)^{(n-k)}dt} \\
&= \frac{\theta^k \cdot (1-\theta)^{(n-k)}}{\frac{k!(n-k)!}{(k+n-k+1)!}} \\
&= \frac{(n+1)!}{k!(n-k)!}\, \theta^k \,(1-\theta)^{(n-k)}
\end{aligned}$$

To find the MAP estimate, we need to find the value $\hat{\theta}$ that maximizes the posterior. We differentiate the posterior PDF and set the derivative to 0 then solve for $\theta$, obtaining,

$$k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1} = 0$$

which yields

$$\hat{\theta}_{\text{MAP}}(k) = \frac{k}{n}.$$

(b) The linear LMS estimator is given by:

$$\begin{aligned}
\hat{\Theta} &= \mathbf{E}[\Theta] + \frac{\text{cov}(\Theta, X)}{\text{var}(X)}\,(X - \mathbf{E}[X]) \\
\mathbf{E}[\Theta] &= 1/2 \\
\mathbf{E}[X] &= \mathbf{E}\left[\mathbf{E}[X|\Theta]\right] = \mathbf{E}[n\Theta] = n/2 \\
\text{var}(X) &= \mathbf{E}\left[\text{var}(X|\Theta)\right] + \text{var}\left(\mathbf{E}[X|\Theta]\right) \\
&= \mathbf{E}[n\Theta(1-\Theta)] + \text{var}(n\Theta) \\
&= n\mathbf{E}[\Theta] - n(\text{var}(\Theta) + (E[\Theta])^2) + n^2\text{var}(\Theta) \\
&= n/6 + n^2/12 \\
\mathbf{E}[\Theta X] &= \mathbf{E}\left[\mathbf{E}[\Theta X|\Theta]\right] = \mathbf{E}[n\Theta^2] = \frac{n}{3} \\
\text{cov}(\Theta, X) &= \mathbf{E}[\Theta X] - \mathbf{E}[\Theta]\mathbf{E}[X] = \frac{n}{3} - \frac{1}{2}\frac{n}{2} = \frac{n}{12} \\
\hat{\Theta} &= \frac{1}{2} + \frac{n/12}{n/6 + n^2/12}\,(X - n/2) = \frac{X+1}{n+2}
\end{aligned}$$

(c) The LMS estimator can be found by integrating

$$\begin{aligned} E[\Theta \mid X = k] &= \int_0^1 \theta \frac{(n+1)!}{k!(n-k)!} \theta^k (1-\theta)^{(n-k)} \, d\theta \\ &= \frac{(n+1)}{k!(n-k)!} \int_0^1 \theta^{k+1} \cdot (1-\theta)^{(n-k)} d\theta \\ &= \frac{(n+1)}{k!(n-k)!} \cdot \frac{(k+1)!(n-k)!}{(n+2)!} \\ &= \frac{k+1}{n+2} \end{aligned}$$

and

$$\hat{\theta}_{\text{LMS}}(k) = E[\Theta \mid X = k] = \frac{k+1}{n+2}.$$

(d) The mean square error is minimized by the conditional mean estimator. Therefore, $MSE_{MAP} \geq MSE_{LMS}$.

(e) The ML estimator is the same as the MAP estimator derived in part (a) because the prior is uniform.

2. (a) $X$ is a binomial random variable with parameters $n = 3$ and given the probability $p$ that a single bit is flipped in a transmission over the noisy channel:

$$p_X(k; p) = \begin{cases} \binom{3}{k} p^k (1-p)^{3-k}, & k = 0, 1, 2, 3 \\ 0 & \text{o.w.} \end{cases}$$

(b) To derive the ML estimator for $p$ based on $X_1, \ldots, X_n$, the numbers of bits flipped in the first $n$ three-bit messages, we need to find the value of $p$ that maximizes the likelihood function:

$$\hat{p}_n = \arg\max_p \, p_{X_1,\ldots,X_n}(k_1, k_2, \ldots, k_n; p)$$

Since the $X_i$'s are independent, the likelihood function simplifies to:

$$p_{X_1,\ldots,X_n}(k_1, k_2, \ldots, k_n; p) = \Pi_{i=1}^n p_{X_i}(k_i; p) = \Pi_{i=1}^n \binom{3}{k_i} p^{k_i} (1-p)^{3-k_i}$$

The log-likelihood function is given by

$$\log(p_{X_1,\ldots,X_n}(k_1, k_2, \ldots, k_n; p)) = \sum_{i=1}^n \left( k_i \log(p) + (3 - k_i)\log(1-p) + \log\binom{3}{k_i} \right)$$

We then maximize the log-likelihood function with respect to p:

$$\begin{aligned}
\frac{1}{p}\left(\sum_{i=1}^{n} k_i\right) - \frac{1}{1-p}\left(\sum_{i=1}^{n}(3-k_i)\right) &= 0 \\
\left(3n - \sum_{i=1}^{n} k_i\right) p &= \left(\sum_{i=1}^{n} k_i\right)(1-p) \\
\hat{p}_n = \frac{1}{3n}\sum_{i=1}^{n} k_i &
\end{aligned}$$

This yields the ML estimator:

$$\hat{P}_n = \frac{1}{3n}\sum_{i=1}^{n} X_i$$

(c) The estimator is unbiased since:

$$\begin{aligned}
\mathbf{E}_p[\hat{P}_n] &= \frac{1}{3n}\sum_{i=1}^{n}\mathbf{E}_p[X_i] \\
&= \frac{1}{3n}\sum_{i=1}^{n} 3p \\
&= p
\end{aligned}$$

(d) By the weak law of large numbers, $\frac{1}{n}\sum_{i=1}^{n} X_i$ converges in probability to $\mathbf{E}_p[X_i] = 3p$, and therefore $\hat{P}_n = \frac{1}{3n}\sum_{i=1}^{n} X_i$ converges in probability to $p$. Thus $\hat{P}_n$ is consistent.

(e) Sending 3 bit messages instead of 1 bit messages does not affect the ML estimate of $p$. To see this, let $Y_i$ be a Bernoulli RV which takes the value 1 if the $i$th bit is flipped (with probability $p$), and let $m = 3n$ be the total number of bits sent over the channel. The ML estimate of $p$ is then

$$\hat{P}_n = \frac{1}{3n}\sum_{i=1}^{n} X_i = \frac{1}{m}\sum_{i=1}^{m} Y_i.$$

Using the central limit theorem, $\hat{P}_n$ is approximately a normal RV for large $n$. An approximate 95% confidence interval for $p$ is then,

$$\left[\hat{P}_n - 1.96\sqrt{\frac{v}{m}}, \hat{P}_n + 1.96\sqrt{\frac{v}{m}}\right]$$

where $v$ is the variance of $Y_i$.
As suggested by the question, we estimate the unknown variance $v$ by the convervative upper bound of 1/4. We are also give that $n = 100$ and the number of bits flipped is 20, yielding $\hat{P}_n = \frac{2}{30}$. Thus, an approximate 95% confidence interval is $[0.01, 0.123]$.

(f) Other estimates for the variance are the sample variance and the estimate $\hat{P}_n(1-\hat{P}_n)$. They potentially result in narrower confidence intervals than the conservative variance estimate used in part (e).