Engr207b Final Exam

1. *Change point detection.* A sensor measures the state of a machine once per second, with the k'th measurement given by y_k. In normal operation, the y_k is normally distributed with mean m_1 and variance c_1. However, at some time the machine breaks, and consequently the sensor output changes to have mean m_2 with variance c_2.

   Sometime later the operator notices that the machine has broken, and a log of the sensor data is examined. We would like to estimate the time r at which the machine was last operating correctly. The prior on r is uniform over the time interval.

   The data is in the file `changedata.m`.

   (a) Give an expression for the posterior distribution of r given the data.

   (b) Given the measurement y, find the MAP estimate of r

   (c) Given the data y, what is the probability of error for your estimate?

   Solution.

   (a) We have $y \in \mathbb{R}^N$ where $N = 100$ and

   $$\begin{aligned} y_i &\sim \mathcal{N}(m_1, c_1) && \text{for } i = 1, \dots, r \\ y_i &\sim \mathcal{N}(m_2, c_2) && \text{for } i = r+1, \dots, N \end{aligned}$$

   hence the joint distribution is

   $$p(r, y) = \frac{1}{N} \left(\prod_{i=1}^{r} f_{c_1}^{\mathcal{N}}(y_i - m_1) \right) \left(\prod_{i=r+1}^{N} f_{c_2}^{\mathcal{N}}(y_i - m_2) \right)$$

   where

   $$f_c^{\mathcal{N}}(x) = \frac{1}{\sqrt{2\pi c}} \exp\left(\frac{-x^2}{2c} \right)$$

   Then the posterior of r is given by

   $$p^{|y}(r) = \frac{p(r, y)}{\sum_{k=1}^{N} p(k, y)}$$

   (b) The map estimate is $r = 40$

   (c) The probability of error is 0.4362.

2. *The t-test for linear regression.* Consider a model of the form

   $$y = ax + \epsilon,$$

   where $y \in \mathbb{R}^n$ is a vector of observations, $a \in \mathbb{R}^n$ is known, nonzero, and constant, $x \in \mathbb{R}$ is an unknown constant, and $\epsilon \in \mathbb{R}^n$ is a normal random vector with mean zero and variance $\sigma^2 I$.

(a) Let $\hat{x}$ be the least-squares estimate of x given y, and let $r = y - a\hat{x}$ be the least-squares residual. Compute the distribution of the vector $(\hat{x}, r)$.

(b) Let S be a d-dimensional subspace of $\mathbb{R}^n$, and let $P \in \mathbb{R}^{n\times n}$ be the matrix representing the orthogonal projection onto S. Show that if $z \in \mathbb{R}^n$ is normal with $z \sim \mathcal{N}(0, I)$ then $\|Pz\|^2$ has a chi-squared distribution with d degrees of freedom; that is, distribution χ^2_d. *Hint.* Let $q_1, \dots, q_d \in \mathbb{R}^n$ be an orthonormal basis for S. Then, $P = QQ^T$, where

$$Q = \begin{bmatrix} q_1 & \cdots & q_d \end{bmatrix}.$$

(c) If z is a normal random variable with $z \sim \mathcal{N}(0, I)$ and w is a chi-squared random variable with d degrees of freedom, and z and w are independent, then the random variable

$$\frac{z}{\sqrt{w/d}}$$

has a t-distribution with d degrees of freedom. Show that

$$\tau = \frac{(\hat{x} - x)\|a\|}{\|r\|/\sqrt{n-1}}$$

has a t distribution with $n - 1$ degrees of freedom.

(d) Now consider the specific parameter values

$$a = \begin{bmatrix} -2 \\ -1 \\ 1 \\ 2 \end{bmatrix}, \qquad x = 1, \qquad \sigma^2 = 2.2, \qquad \text{and} \qquad N = 1000.$$

Generate N samples of y. For each of these samples, compute the least-squares estimate $\hat{x}$, and the t-statistic τ. The p-value of τ is defined to be

$$p = F_{t_{n-1}}(|\tau|) - F_{t_{n-1}}(-|\tau|),$$

where $F_{t_{n-1}}$ is the cumulative distribution function of the t distribution with $n-1$ degrees of freedom. Intuitively, we expect τ to be close to zero, and p is the probability that τ is at least as extreme as the observed value. A small p-value is evidence that the model may not be correct. Compute the p-value for each τ, and plot the empirical cumulative distribution function of p. What appears to be the distribution of p?

Solution.

(a) The least-squares estimate of x is

$$\hat{x} = \frac{1}{\|a\|^2} a^T y,$$

and the corresponding residual is

$$r = y - a\hat{x} = \left(I - \frac{1}{\|a\|^2} aa^T \right) y.$$

Thus, the vector $(\hat{x}, r)$ is a linear function of y:

$$\begin{bmatrix} \hat{x} \\ r \end{bmatrix} = \begin{bmatrix} \frac{1}{\|a\|^2} a^T \\ I - \frac{1}{\|a\|^2} aa^T \end{bmatrix} y.$$

Since y is normal with mean ax and variance σ^2, and linear functions of normal random vectors are also normal random vectors, we see that $(\hat{x}, r)$ is a normal random vector with mean

$$\begin{bmatrix} \frac{1}{\|a\|^2} a^T \\ I - \frac{1}{\|a\|^2} aa^T \end{bmatrix} ax = \begin{bmatrix} x \\ 0 \end{bmatrix},$$

and variance

$$\sigma^2 \begin{bmatrix} \frac{1}{\|a\|^2} a^T \\ I - \frac{1}{\|a\|^2} aa^T \end{bmatrix} \begin{bmatrix} \frac{1}{\|a\|^2} a^T \\ I - \frac{1}{\|a\|^2} aa^T \end{bmatrix}^T = \sigma^2 \begin{bmatrix} \frac{1}{\|a\|^2} & 0 \\ 0 & I - \frac{1}{\|a\|^2} aa^T \end{bmatrix}.$$

(b) Since Q has orthonormal columns, we have that

$$\|Pz\|^2 = \|QQ^T z\|^2 = \|Q^T z\|^2.$$

Because z is a normal random vector with mean zero and variance I, $Q^T z \in \mathbb{R}^d$ is a normal random vector with mean zero and variance $Q^T Q = I$. Thus, we see that $Q^T z$ is a d-dimensional standard normal random vector. This implies that $\|Pz\|^2 = \|Q^T z\|^2$ has a chi-squared distribution with d degrees of freedom.

(c) We showed above that $\hat{x}$ is a normal random variable with mean x and variance $\frac{\sigma^2}{\|a\|^2}$. Therefore,

$$\frac{\hat{x} - x}{\sigma / \|a\|} = \frac{(\hat{x} - x)\|a\|}{\sigma}$$

is a standard normal random variable. We can write r as

$$\left(I - \frac{1}{\|a\|^2} aa^T\right) y = \left(I - \frac{1}{\|a\|^2} aa^T\right)(ax + \epsilon) = \left(I - \frac{1}{\|a\|^2} aa^T\right)\epsilon,$$

where $\frac{1}{\sigma}\epsilon$ is a standard normal random vector, and $I - \frac{1}{\|a\|^2} aa^T$ is the orthogonal projection matrix onto the $(n-1)$-dimensional subspace range$(a)^\perp$. Therefore,

$$\left\|\frac{1}{\sigma} r\right\|^2 = \left\|\frac{1}{\sigma}\left(I - \frac{1}{\|a\|^2} aa^T\right)\epsilon\right\|^2$$

has a chi-squared distribution with $n-1$ degrees of freedom. Moreover, $\hat{x}$ and r are uncorrelated normal random variables, and are therefore independent. This implies that $\hat{x}$ and $\|\frac{1}{\sigma^2} r\|^2$ are independent. Combining these results, we can conclude that

$$\tau = \frac{(\hat{x} - x)\|a\|}{\|r\| / \sqrt{n-1}} = \frac{(\hat{x} - x)\|a\| / \sigma}{\sqrt{\frac{1}{n-1}\|\frac{1}{\sigma} r\|^2}}$$

has a t-distribution with $n-1$ degrees of freedom.

(d) Based on the empirical cdf, the p-value appears to be uniformly distributed on $[0, 1]$.

The following code performs all of the calculations.

3. *Comparing estimators.* We have a sensor measurement y given by

$$y = x + w$$

The noise w is uniformly distributed on $[-b, b]$. We would like to estimate x, which has a prior which is uniform on $[-a, a]$.

(a) Let $f(x, y)$ be the joint pdf of (x, y). Sketch the region in the plane where f is nonzero.

(b) Find the MMSE estimator for x given measurement $y = y_{\text{meas}}$, i.e. find ϕ such that

$$x_{\text{mmse}} = \phi(y_{\text{meas}})$$

(c) What is the mean square error $\mathrm{E}(x - \phi(y))^2$?

(d) What is the conditional mean square error $\mathrm{E}(x - \phi(y)^2 \mid y = y_{\text{meas}})$.

(e) Find the linear estimator ϕ_l that minimizes the mean square error.

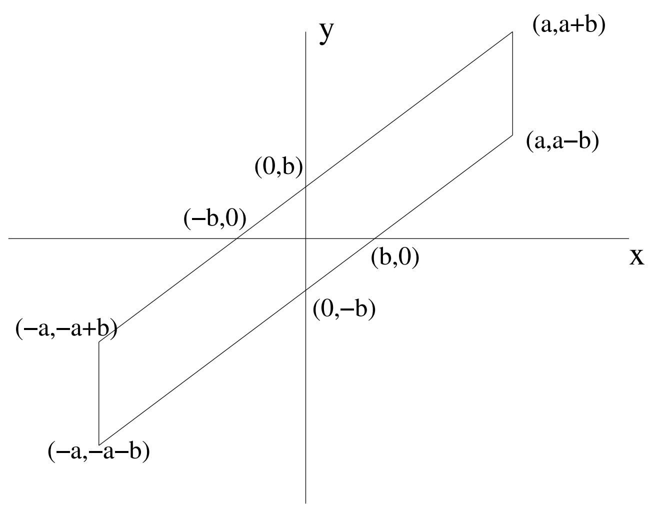
(f) What is the mean square error of ϕ_l?

(g) Suppose $a = 1$ and $b = 0.5$. Plot $\phi(y)$ and $\phi_l(y)$.

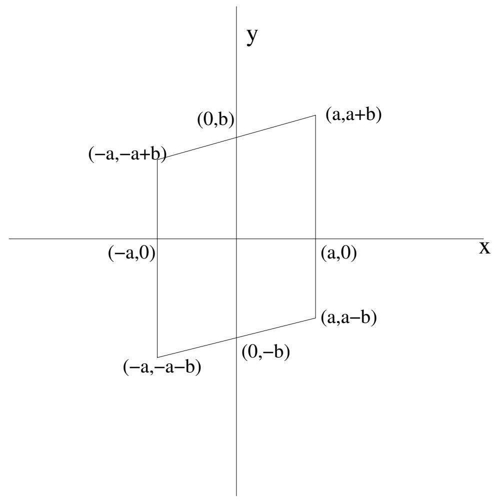
(h) Suppose $a = 2$ and $b = 3$. Plot $\phi(y)$ and $\phi_l(y)$.

Solution.

(a) For $a \geq b$, the plot of the region in $\mathbb{R}^2$ for which $f(x, y)$ is nonzero is shown below.



The plot for the case $a < b$ is shown below.



(b) The MMSE estimate for a given y is given by $x_{\mathrm{mmse}} = \mathrm{E}(x \mid y)$. We will consider two cases, $a \geq b$ and $a < b$.

- If $a \geq b$, then the MMSE estimator is

$$\phi(y) = \begin{cases} \frac{1}{2}(y + b - a) & y \in [-a-b, b-a] \\ y & y \in [b-a, a-b] \\ \frac{1}{2}(y - b + a) & y \in [a-b, a+b] \end{cases}$$

- If $a < b$ then the MMSE estimator is

$$\phi(y) = \begin{cases} \frac{1}{2}(y + b - a) & y \in [-a-b, a-b] \\ 0 & y \in [a-b, b-a] \\ \frac{1}{2}(y - b + a) & y \in [b-a, a+b] \end{cases}$$

(c) The MSE in the two cases is:

- If $a \geq b$, then the mean square error is

$$\mathrm{E}(x - \phi(y))^2 = \frac{b^2(2a - b)}{6a}$$

- If $a < b$ then the mean square error is

$$\mathrm{E}(x - \phi(y))^2 = \frac{a^2(2b - a)}{6b}$$

(d) The conditional MSE in the two cases is:

- If $a \geq b$, then the conditional MSE is

$$\mathrm{E}((x - \phi(y))^2 | y = d) = \begin{cases} \frac{(d+a+b)^2}{12} & d \in [-a-b, b-a] \\ \frac{b^2}{3} & d \in [b-a, a-b] \\ \frac{(d-a-b)^2}{12} & d \in [a-b, a+b] \end{cases}$$

- If $a < b$ then the conditional MSE is

$$\mathrm{E}((x-\phi(y))^2|y=d) = \begin{cases} \frac{(d+a+b)^2}{12} & d \in [-a-b, a-b] \\ \frac{a^2}{3} & d \in [a-b, b-a] \\ \frac{(d-a-b)^2}{12} & d \in [b-a, a+b] \end{cases}$$

(e) Since x and w are uniformly distributed over intervals of length $2a$ and $2b$ respectively, we have $\Sigma_x = \frac{a^2}{3}$, and $\Sigma_w = \frac{b^2}{3}$. Also

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$

Hence

$$\begin{aligned} \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \Sigma_x & 0 \\ 0 & \Sigma_w \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\mathsf{T}} \\ &= \begin{bmatrix} \frac{a^2}{3} & \frac{a^2}{3} \\ \frac{a^2}{3} & \frac{a^2+b^2}{3} \end{bmatrix} \end{aligned}$$

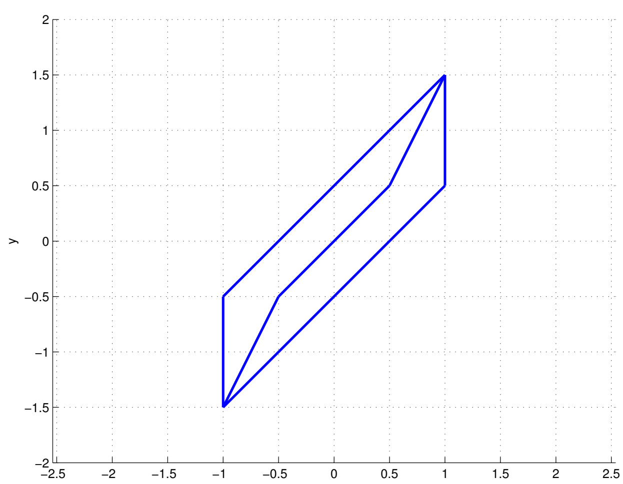
The LMMSE is given by

$$\begin{aligned} \hat{x}_{\text{lmmse}} &= \mu_x + \Sigma_{xy}\Sigma_y^{-1}(y-\mu_y) \\ &= \frac{a^2}{a^2+b^2} y \end{aligned}$$

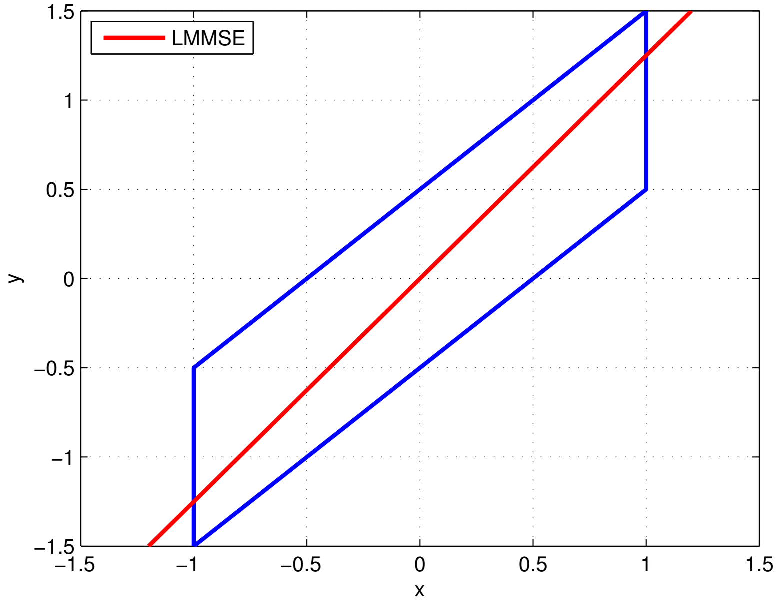
(f) The mean square error of the LMMSE is

$$\begin{aligned} \mathbf{E}(\|x-\hat{x}_{\text{lmmse}}\|^2) &= \mathbf{trace}(\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}) \\ &= \frac{a^2b^2}{3(a^2+b^2)} \end{aligned}$$

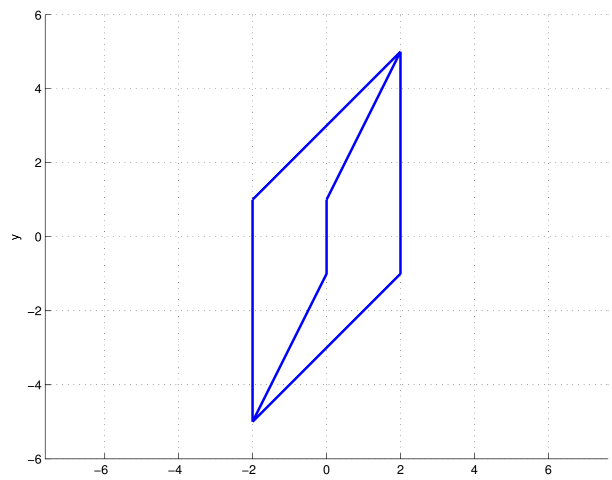
(g) The MMSE estimator is



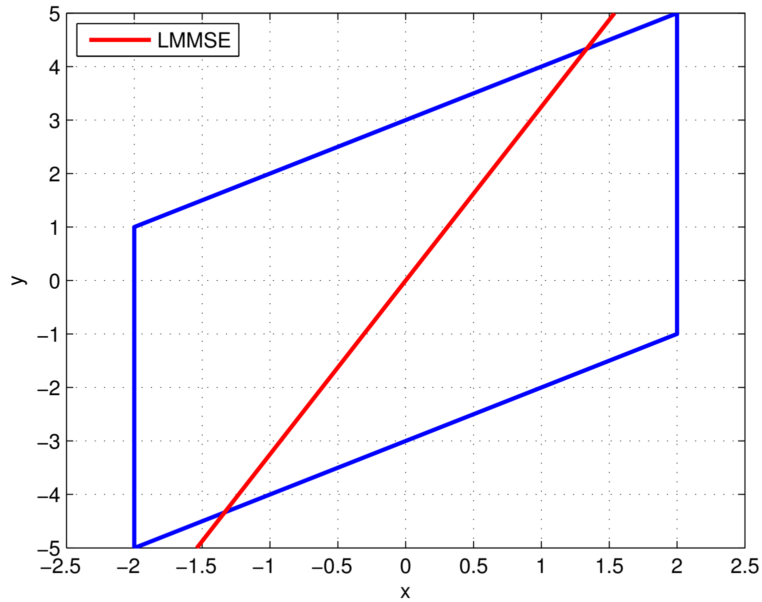
and the LMMSE is



(h) The MMSE estimator is



and the LMMSE is



4. *Chocolate.* A manufacturer makes large boxes of chocolate. Inside each box there are two kinds, one filled with coconut the other filled with hazelnut, but they are identical from the outside. The manufacturer makes five different kinds of boxes.

box type	contents
1	100% coconut
2	80% coconut, 20% hazelnut
3	50% coconut 50% hazelnut
4	25% coconut 75% hazelnut
5	100% hazelnut

The prior distribution of boxes is $0.1, 0.2, 0.4, 0.2, 0.1$. You buy a box, and start tasting them, one after the other. All of the n chocolates you've eaten so far are coconut. You'd like to guess which type of box you bought.

(a) Plot as a function of n the probability that the box is of type i given that all of the n chocolates so far were coconut.

(b) Plot as a function of n the probability that the next chocolate will be coconut, given that all of the n chocolates so far were coconut.

Solution.

(a) Let $x : \Omega \to \{1, 2, 3, 4, 5\}$ be the random variable indicating which box is chosen, and let $y_i : \Omega \to \{1, 2\}$ be the random variable for which type of chocolate, coconut (1) or hazelnut (2), is chosen at time i. Using Bayes' law, we have

$$\text{Prob}(x = a|y = b) = \frac{\text{Prob}(y = b|x = a)\,\text{Prob}(x = a)}{\sum_a \text{Prob}(y = b|x = a)\,\text{Prob}(x = a)}$$

Since the y_i are conditionally IID given x, then

$$\text{Prob}(y = b|x = a) = \text{Prob}(y_1 = b_1|x = a) \cdots \text{Prob}(y_n = b_n|x = a)$$

Since each $b_i = 1$, we have

$$\text{Prob}(y_i = 1|x = a) = \begin{cases} 1 & a = 1 \\ 0.8 & a = 2 \\ 0.5 & a = 3 \\ 0.25 & a = 4 \\ 0 & a = 5 \end{cases}$$

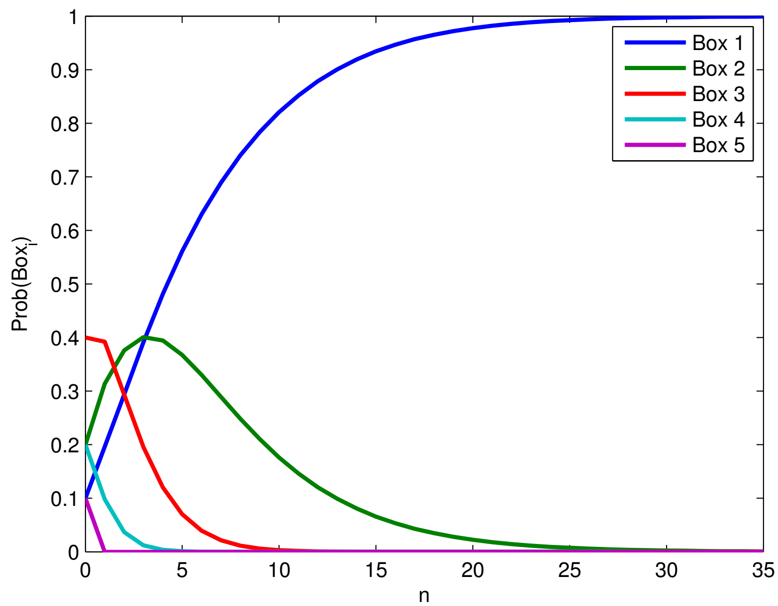
Let $g \in \mathbb{R}^5$, where $g_j = \text{Prob}(y_i = 1|x = j)$, and $p(t) \in \mathbb{R}^5$, for $t = 0, 1, \ldots$, where $p(t)_j = \text{Prob}(x = j|y_t = 1, \ldots, y_1 = 1)$. Our prior on x is

$$p(0) = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.4 \\ 0.2 \\ 0.1 \end{bmatrix}$$

Then, for any $t \geq 1$, we have

$$\begin{aligned} p(t)_j &= \text{Prob}(x = j|y_t = 1, \ldots, y_1 = 1) \\ &= \frac{g_j^t p(0)_j}{\sum_j g_j^t p(0)_j} \\ &= \frac{g_j p(t-1)_j}{\sum_j g_j p(t-1)_j} \\ &= \frac{g_j p(t-1)_j}{g^T p(t-1)} \end{aligned}$$

which allows us to compute this recursively. We can plot this posterior distribution for each t, assuming that $y_t = 1$ for all t, shown below.



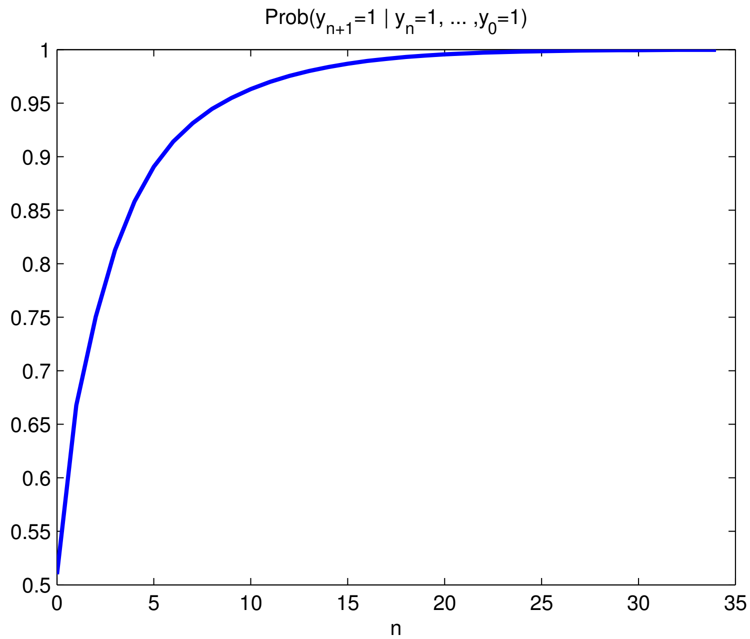
(b) Let $f(n) = \text{Prob}(y_{n+1} = 1 | y_n = 1, \ldots, y_1 = 1)$. Again, using Bayes' law, for each $n = 0, 1, \ldots$, we have

$$\begin{aligned} f(n) &= \text{Prob}(y_{n+1} = 1 | y_n = 1, \ldots, y_1 = 1) \\ &= \frac{\sum_j \text{Prob}(y_{n+1} = 1, \ldots, y_1 = 1 | x = j) \, \text{Prob}(x = j)}{\sum_k \text{Prob}(y_n = 1, \ldots, y_1 = 1 | x = k) \, \text{Prob}(x = k)} \\ &= \frac{\sum_j g_j^{n+1} p(0)_j}{\sum_k g_k^n p(0)_k} \end{aligned}$$

If we use $p(t)$ from part a, then for $n \geq 1$, we can compute $f(n)$ as

$$f(n) = \sum_j g_j p(n)_j = g^T p(n)$$

We can plot f against n below



5. *Recursive estimation.* We would like to estimate a random variable $x : \Omega \to \{1, 2, 3, \ldots, 30\}$. We have two sensors. Sensor 1 gives $y = x + w$, where $\text{Prob}(w = i) = d_i$ and d is

$$d = \begin{bmatrix} 1 & 3 & 8 & 8 & 3 & 1 \end{bmatrix}$$

Sensor 2 gives measurement $y = f(x) + v$ where

$$f(x) = \begin{cases} 1 & \text{if } 0 < x \leq 10 \\ 2 & \text{if } 10 < x \leq 20 \\ 3 & \text{if } 20 < x \leq 30 \end{cases}$$

and v satisfies

$$\text{Prob}(v = 0) = 0.25 \qquad \text{Prob}(v = 1) = 0.75$$

(a) Find the transition matrices G_1 and G_2 for the two sensors

(b) The prior on x is uniform. We take measurements in sequence, given by

$$A = \begin{bmatrix} 1 & 18 \\ 1 & 19 \\ 2 & 3 \\ 1 & 18 \\ 1 & 18 \\ 2 & 3 \\ 2 & 3 \\ 1 & 18 \\ 2 & 3 \\ 1 & 20 \end{bmatrix}$$

Here the first column is the sensor and the second is the corresponding reading. Compute, for each t, the posterior distribution of x given measurements 1 through t. Plot these distributions.

(c) Compute, for each t, the MAP estimate of x.

Solution.

(a) We have $G_1 = \frac{1}{24} \begin{bmatrix} G_{11} & G_{12} \end{bmatrix}$ where

$$G_{11} = \begin{bmatrix}
1 & 3 & 8 & 8 & 3 & 1 & & & & & & & & & & & & \\
& 1 & 3 & 8 & 8 & 3 & 1 & & & & & & & & & & & \\
& & 1 & 3 & 8 & 8 & 3 & 1 & & & & & & & & & & \\
& & & 1 & 3 & 8 & 8 & 3 & 1 & & & & & & & & & \\
& & & & 1 & 3 & 8 & 8 & 3 & 1 & & & & & & & & \\
& & & & & 1 & 3 & 8 & 8 & 3 & 1 & & & & & & & \\
& & & & & & 1 & 3 & 8 & 8 & 3 & 1 & & & & & & \\
& & & & & & & 1 & 3 & 8 & 8 & 3 & 1 & & & & & \\
& & & & & & & & 1 & 3 & 8 & 8 & 3 & 1 & & & & \\
& & & & & & & & & 1 & 3 & 8 & 8 & 3 & 1 & & & \\
& & & & & & & & & & 1 & 3 & 8 & 8 & 3 & 1 & & \\
& & & & & & & & & & & 1 & 3 & 8 & 8 & 3 & 1 & \\
& & & & & & & & & & & & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & & & & & & & & & & & 1 & 3 & 8 & 8 & 3 \\
& & & & & & & & & & & & & & 1 & 3 & 8 & 8 \\
& & & & & & & & & & & & & & & 1 & 3 & 8 \\
& & & & & & & & & & & & & & & & 1 & 3 \\
& & & & & & & & & & & & & & & & & 1 \\
\\
\\
\\
\\
\\
\\
\\
\\
\\
\\
\\
\\
\end{bmatrix}$$

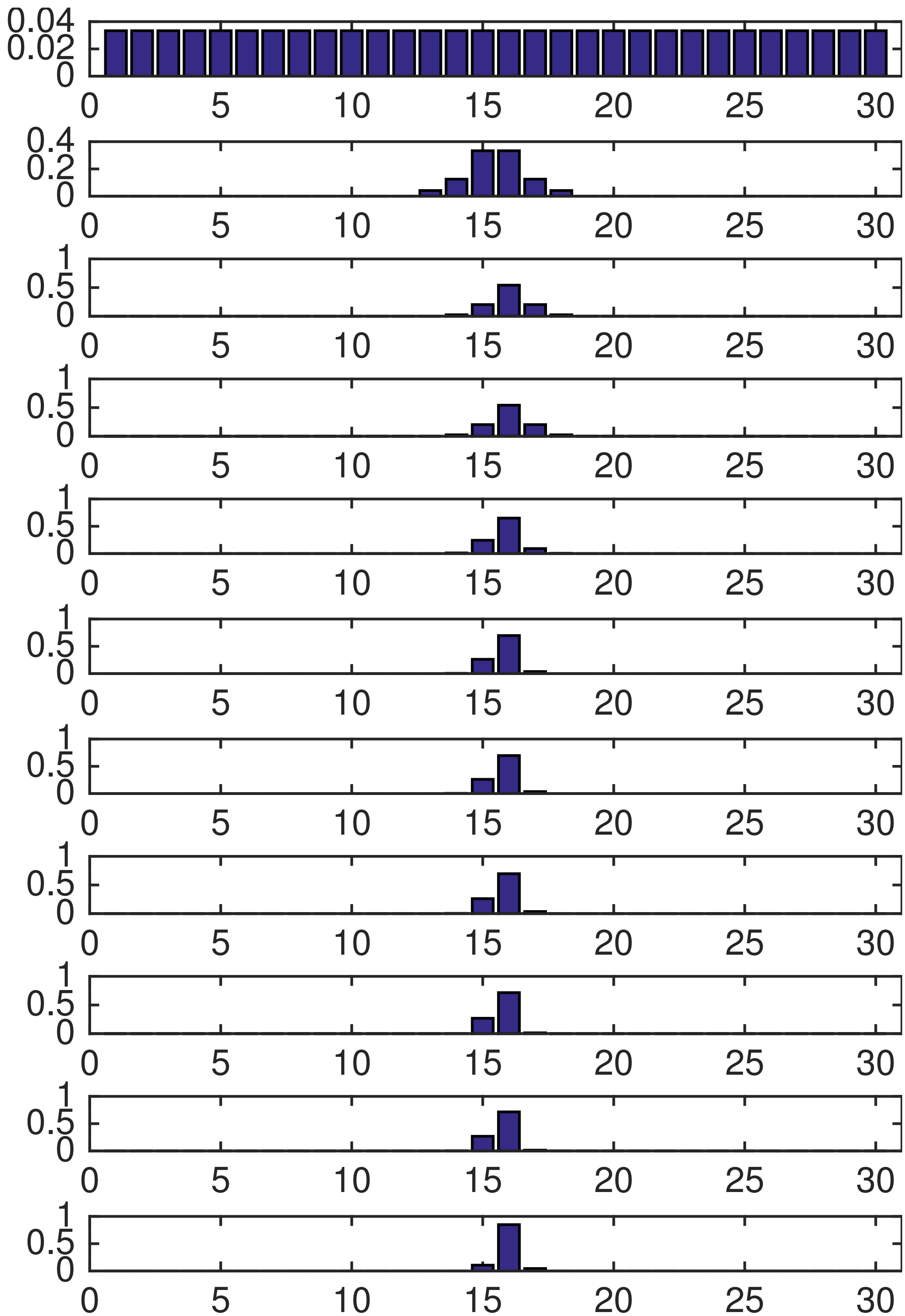
and

$$G_{12} = \begin{bmatrix}
\\
\\
\\
\\
\\
\\
\\
\\
\\
\\
\\
\\
\\
1 \\
3 & 1 \\
8 & 3 & 1 \\
8 & 8 & 3 & 1 \\
3 & 8 & 8 & 3 & 1 \\
1 & 3 & 8 & 8 & 3 & 1 \\
& 1 & 3 & 8 & 8 & 3 & 1 \\
& & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & & & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & & & & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & & & & & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & & & & & & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & & & & & & & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & & & & & & & & 1 & 3 & 8 & 8 & 3 & 1 \\
& & & & & & & & & & & 1 & 3 & 8 & 8 & 3 & 1 &
\end{bmatrix}$$

The second sensor is

$$G_2 = \frac{1}{4}\begin{bmatrix} 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

(b) The posterior distributions are below.



(c) The MAP estimates are $15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16$.