PRINCETON UNIVERSITY
Department of Electrical Engineering
ELE 525: Random Processes in Information Systems
Final Examination Solutions (posted on January 21, 2014)
January 20 (Mon), 2014; 1:00 pm-4:00 pm
This is a **Closed Book Exam**.

Problem 1: Some definitions (20 points)
Give a brief definition of each term below (less than 100 words)

(a) A semi-Markov Process and a continuous-time Markov chain (CTMC)

(b) Karhunen-Loéve expansion

Solutions:

(a) A semi-Markov process (SMP) $X(t)$ is a right-continuous, piecewise constant process, which takes values in a finite or countably infinite set of states $\mathcal{S}$ and transitions at time $t_1, t_2, \ldots$. We define the nth sojourn time by $\tau_n = t_n - t_{n-1}, \quad n \geq 1$. We assume $t_0 = 0$. An SMP is characterized by

(i) the transition probability matrix (TPM) of the embedded Markov chain (EMC) $\boldsymbol{P} = [P_{ij}]$, where $P_{ij} = P[X(t_n) = j | X(t_{n-1} = i], \quad i, j \in \mathcal{S}$.

(ii) the set of distribution functions

$$F_{ij}(t) = P[\tau_n \leq t | X(t_n) = j, X(t_{n-1} = i], \quad i, j \in \mathcal{S}.$$

An SMP $X(t)$ is called a continuous-time Markov chain (CTMC) if the sojourn time distribution $F_{ij}(t)$ are all exponentially distributed according to

$$F_{ij}(t) = 1 - e^{-\nu_i t}, \quad t \geq 0, \ i, j \in \mathcal{S}.$$

Note that $F_{ij}(t)$ depends only on i, and not on j.

Remark: The infinitesimal generator $\boldsymbol{Q} = [Q_{ij}]$ defined in p. 460 can be derived from the above parameter, i.e., $Q_{ij} = \nu_i P_{ij}$

(b) The Karhunen-Loéve expansion is a generalized version of the Fourier series expansion that we can apply to a random process $X(t), \quad 0 \leq t \leq T$, which is not periodic, and not necessarily stationary. For the autocorrelation function of $X(t)$ denoted as $R(t,s) = E[X(t)X^*(s)]$, let λ_i and $u_i(t)$ be its eigenvalues and eigenfunctions such that

$$\int_0^T R(T,s) u_i(s)\, ds = \lambda_i u_i(t), \quad 0 \leq t \leq T, \quad i = 1, 2, \ldots,$$

and

$$\langle u_i, u_j \rangle = \int_0^T u_i^*(t) u_j(t)\, dt = \delta_{i,j}.$$

Then we can expand the process as

$$X(t) = l.i.m._{N\to\infty} \sum_{i=1}^{N} X_i u_i(t),$$

where X_i is the projection of $X(t)$ onto the ith coordinate $u_i(t)$:

$$X_i = \langle u_i(t), X(t) \rangle = \int_0^T X(t) u_i^*(t)\, dt.$$

Problem 2: Brownian motion (20 points)

Let $W(t)$ be Brownian motion (a.k.a. the Wiener process) with $W(0) = 0$ and $\mathrm{Var}[W(t)] = \alpha t$. In answering the following questions, you may refer to the five properties of the process $W(t)$: 1. Spatial homogeneity, 2. Temporal homogeneity, 3. Independent increments, 4. Markov property, and 5.Gaussian property.

(a) Show the following properties of $W(t)$

$$R_W(t, s) = \alpha \min\{s, t\}, \quad s, t > 0. \tag{1}$$
$$\mathrm{Var}[W(t) - W(s)] = \alpha |t - s|, \quad s, t > 0. \tag{2}$$

(b) Define a random process $Y(t)$ by

$$Y(t) = e^{W(t)}.$$

Find its mean $E[Y(t)$ and variance $\mathrm{Var}[Y(t)]$.

Solution:

(a) For $t \geq s$,

$$R_W(t, s) = E[W(t)W(s)] = E[(W(t)-W(s)+W(s))W(s)] = E[(W(t)-W(s))W(s)]+E[W^2(s)].$$

Since $W(t) - W(s)$ and $W(s) = W(s) - W(0)$ are independent increments,the first term disappears. Hence

$$R_W(t, s) = \alpha s, \quad 0 \leq s \leq t.$$

Similarly, for $t \leq s$,

$$R_W(t, s) = \alpha t, \quad 0 \leq t \leq s.$$

Thus we obtain (1). To derive (2), we write for $t \geq s$,

$$\begin{aligned}\text{Var}[W(t) - W(s)] &= E[(W(t) - W(s))^2] = E[(W(t) - W(s))W(t)] - E[(W(t) - W(s))W(s)] \\ &= E[W^2(t)] - E[W(t)W(s)] = \alpha t - R_W(t, s) = \alpha(t - s).\end{aligned}$$

Similarly, for $t < s$, we find

$$\text{Var}[W(t) - W(s)] = \alpha(s - t).$$

Thus we derived (2).

Alternative derivation: For $t \geq s$, from the temporal homogeneity, $W(t) - W(s)$ and $W(t - s) - W(0) = W(t - 1)$ are identically distributed. Hence

$$\text{Var}[W(t) - W(s)] = \text{Var}[W(t - s)] = \alpha(t - s).$$

For $t < s$, $W(s) - W(t)$ and $W(s - t)$ are identically distributed, hence

$$\text{Var}[W(s) = W(t)] = \alpha(s - t).$$

Since $\text{Var}[-X] = \text{Var}[X]$ for any RV X, $\text{Var}[W(t) - W(s)] = \text{Var}[W(s) = W(t)]$. Hence,

$$\text{Var}[W(t) - W(s)] = \alpha|t - s|.$$

(b) Since $W(t)$ is a normal RV at given time t, $Y(t)$ is long-normally distributed (see. Textbook, pp. 165-167) Recall that the moment generating function (MGF) of a normal random variable X with zero mean and variance σ^2

$$M_X(\xi) = E[e^{\xi X}] = e^{\frac{(\sigma\xi)^2}{2}}.$$

The $W(t)$ is a normal RV with zero mean and variance αt, its MGF is readily obtained by setting $\sigma^2 = \alpha t$ in the above formula, i.e.,

$$M_{W(t)}(\xi) = E[e^{\xi W(t)}] = e^{\frac{\alpha t \xi^2}{2}}.$$

By setting $\xi = 1$ and $\xi = 2$, we have

$$M_{W(t)}(1) = E[e^{W(t)}] = e^{\frac{\alpha t}{2}}, \tag{3}$$

$$M_{W(t)}(2) = E[e^{2W(t)}] = e^{2\alpha t}. \tag{4}$$

From (3) we find

$$E[Y(t)] = E[e^{W(t)}] = e^{\frac{\alpha t}{2}}.$$

Similarly from (4) and the last equation,

$$\begin{aligned}\text{Var}[Y(t)] &= E[Y^2(t)] - E^2[Y(t)] = E[e^{2W(t)}] - e^{\alpha t} \\ &= e^{2\alpha t} - e^{\alpha t} = e^{\alpha t}(e^{\alpha t} - 1).\end{aligned}$$

Remarks:

(i) The process $Y(t) = e^{W(t))}$ is called a geometric Brownian motion and is often used to model the movement of a stock price, etc. See the textbook p. 509, Eq. (17.163).

(ii) Even if you don't remember the log-normal distribution or do not think of the MGF, you can compute $E[Y]$ directly: By setting $W(t) = X$, and $\alpha t = \sigma^2$, and $Y = e^X$, we have

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}\, dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\sigma^2)^2}{2\sigma^2} + \frac{\sigma^2}{2}}\, dx \\ &= e^{\frac{\sigma^2}{2}} = e^{\frac{\alpha t}{2}}. \end{aligned}$$

Similarly,

$$\begin{aligned} E[Y^2] &= \int_{-\infty}^{\infty} e^{2x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}\, dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-2\sigma^2)^2}{2\sigma^2} + 2\sigma^2}\, dx \\ &= e^{2\sigma^2} = e^{2\alpha t}. \end{aligned}$$

Problem 3: Estimation based on past values (20 points)

Consider estimating a time continuous process $X(t)$ in terms of two preceding values, $X(t-\tau_1)$ and $X(t-\tau_2)$, where $\tau_2 \geq \tau_1 \geq 0$, using the following linear estimation scheme:

$$\hat{X}(t) = \beta_1 X(t-\tau_1) + \beta_2 X(t-\tau_2). \tag{5}$$

Assume that $X(t)$ is a real-valued WSS (wide-sense stationary) process.

(a) Find β_1 and β_2 such that the mean-square error (MSE) is minimized.

(b) Suppose that given $X(t-\tau_1)$, the knowledge of the older value $X(t-\tau_2)$ does not improve the estimation $\hat{X}(t)$. Show that the autocorrelation function must take the following form:

$$R_X(\tau) = R_X(0)e^{-\alpha|\tau|}.$$

Solution

(a) The mean square error is

$$\mathcal{E} = E[(\hat{X}(t) - X(t))^2] = E[(\beta_1 X(t-\tau_1) + \beta_2 X(t-\tau_2) - X(t))^2].$$

Take the partial derivative of $\mathcal{E}$ with respect to β_1 and β_2 and set them to zero:

$$\beta_1 R(0) + \beta_2 R(\tau_2 - \tau_1) - R(\tau_1) = 0 \tag{6}$$

$$\beta_1 R(\tau_2 - \tau_1) + \beta_2(0) - R(\tau_2) = 0. \tag{7}$$

Then we find

$$\beta_1 = \frac{R(0)R(\tau_1) - R(\tau_2)R(\tau_2 - \tau_1)}{R^2(0) - R^2(\tau_2 - \tau_1)}, \tag{8}$$

$$\beta_2 = \frac{R(0)R(\tau_2) - R(\tau_1)R(\tau_2 - \tau_1)}{R^2(0) - R^2(\tau_2 - \tau_1)}. \tag{9}$$

(b) The assumption means that $\beta_2 = 0$, i.e.,

$$R(0)R(\tau_2) - R(\tau_1)E(\tau_2 - \tau_1) = 0, \quad \text{for any} \quad \tau_2 \geq \tau_1 \geq 0, \tag{10}$$

or equivalently

$$R(0)R(s + t) = R(s)R(t), \quad \text{for any} \quad s \geq t \geq 0, \tag{11}$$

We can assume $R(0) = 1$ without loss of generality, by normalizing $X(t)$ so that $E[X^2(t)] = 1$. Then by taking the logarithm and defining $Q(t) = \ln R(t)$, the above equation becomes

$$Q(s + t) = Q(s) + Q(t) - Q(0) = Q(s) + Q(t), \quad s,\ t > 0.$$

Then it is clear that $Q(t)$ must take the form $Q(t) = at$, i.e., $R(\tau) = e^{a\tau}, \quad \tau > 0$. Since $R(\tau) \leq (0) = 1$, it is clear that $a \leq 0$. By setting $a = -\alpha$, we have $R(\tau) = e^{-\alpha\tau}, \quad \tau \geq 0$. Since an autocorrelation is a symmetric function we find $R(\tau) = e^{-\alpha|\tau|}$, or in general $R(\tau) = R(0)e^{-\alpha|\tau|}$.

Alternative proof: Let $s = h$ in (11). Then $R(0)R(t + h) = R(t)R(h)$. Thus,

$$\frac{R(t + h) - R(t)}{h} = R(t)\frac{R(h) - R(0)}{hR(0)}.$$

Then letting $h \to 0$, we have

$$R'(t) = R(t)R'(0)/R(0),$$

i.e.,

$$\frac{d \ln R(t)}{dt} = R'(0)/R(0) = a,$$

for some constant a. Thus,

$$\ln R(t) = at + b,$$

for some constant b. Thus

$$R(t) = e^b e^{at}.$$

By identifying $a = -\alpha$ and $e^b = R(0)$, we obtain the above result.

Problem 4: A hidden Markov model (HMM) and estimation algorithms (40 points)
Consider an information sequence $\{I_t\}$ $I_t \in \{+1, -1\}$, $t \in \mathcal{T} = \{0, 1, \ldots, T\}$. Assume the I_t's are i.i.d. with $P[I_t = +1] = P[I_t = -1] = 1/2$ for all $t \in \mathcal{T}$.

The information sequence is sent over a linear and time-invariant channel, which is dispersive so that a "+1" signal sent at time t appears at the channel output at times $t, t+1, \ldots, t+d$ with amplitudes $h_0, h_1, \ldots, h_d$, respectively. Thus, the channel output at time t is given by

$$O_t = \sum_{i=0}^{d} I_{t-i} h_i, \quad t \in \mathcal{T} \tag{12}$$

The observation sequence Y_t is given by

$$Y_t = O_t + N_t, \quad t \in \mathcal{T} \tag{13}$$

where the noise $\{N_t\}$ are i.i.d. Gaussian variables with zero mean and variance σ^2.

Consider the case $d = 1$, hence the parameters of interest are $\boldsymbol{\theta} = (h_0, h_1, \sigma)$. In the questions (a) through (e), assume that the parameters $\boldsymbol{\theta}$ are fixed and known.

(a) *Formulation of a hidden Markov model*
Formulate the system in terms of a hidden Markov model (HMM)[1]

Define a set of Markov states, denoted $\mathcal{S}$, and draw the state transition diagram.

(b) *Conditional joint probabilities* $p(j, y|i)$

Find the following conditional joint probabilities for all state pairs $i, j \in \mathcal{S}$, and $y \in \mathcal{R}$.

$$p(j, y|i)\, dy = P[S_t = j, y < Y_t \leq y + dy | S_{t-1} = i], \quad i, j \in \mathcal{S}, -\infty < y < \infty. \tag{14}$$

(c) *Posterior probability and joint probability*

When the observation sequence $\boldsymbol{Y}_0^T = \boldsymbol{y}$ is given, the posterior probability of the state sequence $\boldsymbol{S}_0^T = \boldsymbol{s}$ is $\pi(\boldsymbol{s}|\boldsymbol{y}) = \frac{p(\boldsymbol{s},\boldsymbol{y})}{p(\boldsymbol{y})}$, and the initial probability is denoted by $\pi(s_0, y_0)$, i.e.,

$$\pi(s_0, y_0) dy = P[S_0 = s_0, y_0 < Y_0 \leq y_0 + dy].$$

Express the joint probability $p(\boldsymbol{s}, \boldsymbol{y})$ in terms of $\pi(s_0, y_0)$ and the conditional joint probabilities defined in part (b).

[1] *Definition (Hidden Markov Model).* A Markov process (S_t, Y_t) is called a partially observable Markov process or HMM, if its state transition probability does not depend on $Y_{t-1} = y'$, i.e.,

$$p_{S_t, Y_t | S_{t-1}, Y_{t-1}}(j, y | i, y') = p(j, y|i).$$

(d) *Auxiliary variables and a recursion formula*

Consider the following auxiliary variables:

$$\alpha_t(j, \boldsymbol{y}_0^t) = \max_{\boldsymbol{s}_0^{t-1}} P[\boldsymbol{S}_0^{t-1} = \boldsymbol{s}_0^{t-1}, S_t = j.\boldsymbol{Y}_0^t = \boldsymbol{y}_0^t] \tag{15}$$

Find the recursion formula for the auxiliary variables.

(e) *Simplify the algorithm for a MAP sequence estimation*

Simplify the above recursion formula and describe the resultant algorithm to find a MAP (maximum a posteriori probability) state sequence estimate $\hat{\boldsymbol{s}}^*$ from the observation $\boldsymbol{Y}_0^T = \boldsymbol{y}$. Then find the MAP information sequence $\hat{\boldsymbol{i}}^* = (\hat{i}_0^*, \hat{i}_1^*, \ldots, \hat{i}_T^*)$.

(f) *A maximum likelihood estimate (MLE) of the channel parameters*

Now assume that the information sequence $\boldsymbol{i}_0^T$ sent over the channel is known to the receiver, but now the channel parameters $\boldsymbol{\theta}$ are unknown and have to be estimated. Obtain a maximum likelihood estimate (MLE) of the parameters $\boldsymbol{\theta}$ from the observation $\boldsymbol{y}$ and the given information sequence $\boldsymbol{i}$.

Solution:

(a) Most of you defined the state process $S_t = (I_t, I_{t-1})$, which is a state-based HMM, and the number of states is four, i.e., $\mathcal{S} = \{(-1, -1), (-1, +1), (+1, -1), (+1, +1)\}$.

I choose to adopt, instead, a transition-based HMM, similar to Example 20.1 of pp. 578-579 (where the state is defined by $S_=(I_t, I_{t-1})$ instead of $S_t = (I_t, I_{t-1}, I_{t-2})$, which would be the case in a state-based HMM).

So in this problem, I define the state $S(t)$ simply by

$$S_t = I_t,$$

Hence, $\mathcal{S} = \{+1, -1\}$. The transition diagram of the Markov chain has two states $+1$ and -1. The transition from $+1$ to -1 happens with probability $1/2$, and to $+1$ itself with probability $1/2$. Similarly, the transition from -1 to $+1$ happens with probability $1/2$, and to itself, with probability $1/2$. (The diagram is skipped here, since the above description should suffice)

(b) Noting $S_t = I_t$, we have

$$\begin{aligned} p(s_t, y_t | s_{t-1}) &= \frac{1}{2} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(y_t - o_t)^2}{2\sigma^2} \right\} \\ &= \frac{1}{2} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{[y_t - h_0 s_t - h_1 s_{t-1}]^2}{2\sigma^2} \right\}, \quad s_{t-1}, s_t \in \mathcal{S} = \{+1, -1\} \end{aligned}$$

for $t \in \mathcal{T}$, where we assume $s_{-1} = 0$.

(c) The initial probability is given by

$$\pi(s_0, y_0) = p(s_0, y_0|s_{-1} = 0) = \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{[y_0 - h_0 s_0]^2}{2\sigma^2}\right\}$$

Thus,

$$\begin{aligned} p(\boldsymbol{s}, \boldsymbol{y}) &= \pi(s_0, y_0) \prod_{t=1}^{T} p(s_t, y_t|s_{t-1}) \\ &= \frac{1}{(2\sigma)^{T+1}(2\pi)^{\frac{T+1}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=0}^{T}[y_t - h_0 s_t - h_1 s_{t-1}]^2\right\} \end{aligned} \quad (16)$$

(d) Since

$$\begin{aligned} \alpha_t(j, \boldsymbol{y}_0^t) &= \max_{\boldsymbol{s}_0^{t-1}} P[\boldsymbol{s}_0^{t-1}, j, \boldsymbol{y}_0^t] \\ &= \max_{i\in\mathcal{S}} \max_{\boldsymbol{s}_0^{t-2}} P[\boldsymbol{s}_0^{t-2}, S_{t-1} = i, \boldsymbol{y}_0^{t-1}] P[j, y_t|\boldsymbol{s}_0^{t-2}, S_{t-1} = i, \boldsymbol{y}_0^{t-1}] \\ &= \max_{i\in\mathcal{S}} \max_{\boldsymbol{s}_0^{t-2}} P[\boldsymbol{s}_0^{t-2}, S_{t-1} = i, \boldsymbol{y}_0^{t-1}] P[j, y_t|S_{t-1} = i] \\ &= \max_{i\in\mathcal{S}} \left(\max_{\boldsymbol{s}_0^{t-2}} P[\boldsymbol{s}_0^{t-2}, S_{t-1} = i, \boldsymbol{y}_0^{t-1}]\right) P[j, y_t|S_{t-1} = i] \\ &= \max_{i\in\mathcal{S}} \left\{\alpha_{t-1}(i, \boldsymbol{y}_0^{t-1}) p(j, y_t|i)\right\}. \end{aligned}$$

where $i, j \in \mathcal{S} = \{+1, -1\}$.

(e) Take the logarithm of the auxiliary variable

$$\ln \alpha_t(j, \boldsymbol{y}_0^t) = \max_{i\in\mathcal{S}} \left\{\ln \alpha_{t-1}(i, \boldsymbol{y}_0^{t-1}) + \ln p(j, y_t|i)\right\}.$$

Here

$$\begin{aligned} \ln p(j, y_t|i) &= -c - \frac{1}{2\sigma^2}[y_t - h_0 j - h_1 i]^2 \\ &= y_t(h_0 j + h_1 i) - \frac{1}{2}(h_0 j + h_1 i)^2 + y_t^2 - c, \end{aligned}$$

where

$$c = -\ln(2\sqrt{2\pi}\sigma),$$

Thus, we define a new auxiliary variable or metric:

$$m_t(j, \boldsymbol{y}_0^t) = \max_{i\in\mathcal{S}} \left\{m_{t-1}(i, \boldsymbol{y}_0^{t-1}) + y_t(h_0 j + h_1 i) - \frac{1}{2}(h_0 j + h_1 i)^2\right\} \quad (17)$$

Thus, the Viterbi algorithm reduces to the following simple operation. $m_{t-1}(i, \boldsymbol{y}_0^{t-1}$ is the metric or score given to the surviving path that has ended at state i at time $t-1$, where $i = +1$ or $i = -1$. The surviving path that enters state j at time t is greater of the two competing paths determined by (17). The initial value of the metric is $m_0(j, y_0) = y_0 h_0 j - \frac{1}{2}(h_0 j)^2$.

Note that the metric $m_t(j, \boldsymbol{y}_0^t)$ is equivalent to the correlator (or equivalently matched filter) output of the received sequence y_0^t and the surviving state sequence $\boldsymbol{s}_0^t$, which is equal to the plausible information sequence $\boldsymbol{I}_0^t = \boldsymbol{i}_0^t$, since the state S_t is defined as I_t in this HMM.

(f) Maximum likelihood estimation of the model parameters in this problem can be simply solved analytically, therefore we do not need an iterative algorithm such as the EM algorithm discussed in the text, which may be required when the state sequence is hidden. The likelihood function is

$$L_{\boldsymbol{y}}(\boldsymbol{\theta}) = p(\boldsymbol{s}, \boldsymbol{t}; \boldsymbol{\theta}),$$

where $p(\boldsymbol{s}, \boldsymbol{y}; \boldsymbol{\theta})$ is the same as $p(\boldsymbol{s}, \boldsymbol{y})$ obtained in (16) in part (c). We now write the model parameters $\boldsymbol{\theta} = (h_0, h_1, \sigma)$ explicitly in the equation.

$$\begin{aligned} L_{\boldsymbol{y}}(\boldsymbol{\theta}) = p(\boldsymbol{s}, \boldsymbol{y}; \boldsymbol{\theta}) &= \pi(s_0, y_0) \prod_{t=1}^{T} p(s_t, y_t | s_{t-1}) \\ &= \frac{1}{(2\sigma)^{T+1}(2\pi)^{\frac{T+1}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=0}^{T} [y_t - h_0 s_t - h_1 s_{t-1}]^2 \right\} \end{aligned} \tag{18}$$

The log-likelihood function is given be

$$J(\boldsymbol{\theta}) = \log L_{\boldsymbol{y}}(\boldsymbol{\theta}) = -(T+1)\ln(2\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{t=0}^{T} [y_t - h_0 s_t - h_1 s_{t-1}]^2$$

Take the partial derivatives of J with respect to h_0, h_1 and σ and set them to zero.

$$\begin{aligned} \frac{\partial J}{\partial h_0} &= -\frac{1}{\sigma^2} \sum_{t=0}^{T} s_t (y_t - h_0 s_t - h_1 s_{t-1}) = 0, \\ \frac{\partial J}{\partial h_1} &= -\frac{1}{\sigma^2} \sum_{t=0}^{T} s_{t-1} (y_t - h_0 s_t - h_1 s_{t-1}) = 0, \\ \frac{\partial J}{\partial \sigma} &= -\frac{T+1}{\sigma} + \frac{1}{\sigma^3} \sum_{t=0}^{T} [y_t - h_0 s_t - h_1 s_{t-1}]^2 = 0. \end{aligned}$$

Then we find

$$\hat{h}_0 = \frac{\sum_t s_t y_t \sum_t s_t^2 - \sum_t s_{t-1} y_t \sum_t s_t s_{t-1}}{,} \sum_t s_t^2 \sum_t s_{t-1}^2 - (\sum_t s_t s_{t-1})^2$$

$$\hat{h}_1 = -\frac{\sum_t s_t y_t \sum_t s_t s_{t-1} - \sum_t s_{t-1} y_t \sum_t s_t^2}{\sum_t s_t^2 \sum_t s_{t-1}^2 - (\sum_t s_t s_{t-1})^2}$$

$$\hat{\sigma}^2 = \frac{\sum_{t=0}^{T} [y_t - h_0 s_t - h_1 s_{t-1}]^2}{T+1}.$$

In the above formulae, $s_t = i_t$ for all $t \in \mathcal{T} = [0, 1, \ldots, T]$ and $s_{-1} = i_{-1} = 0$.