1. Hidden Markov Models

Definition 1.1 (Hidden Markov Models)**.** A HMM is a Markov process (X, Y) on a $(\mathbb{X} \times \mathbb{Y}, \mathcal{X} \times \mathcal{Y})$ with transition kernel $K(x', dy')P(x, dx')$ and initial distribution $K(x, dy)\mu_0(dx)$, where P is a probability kernel from $\mathbb{X}$ to $\mathbb{X}$, K is a probability kernel from $\mathbb{X}$ to $\mathbb{Y}$, and μ_0 is a probability measure on $\mathbb{X}$.

Definition 1.2 (Non-degeneracy)**.** A HMM has non-degenerate observations if $K(x, dy) = \lambda(x, y)\phi(dy)$ for some measurable positive function λ on $\mathbb{X} \times \mathbb{Y}$ and a probability measure ϕ on $\mathbb{Y}$.

Definition 1.3 (Notation)**.** We set $\pi_{k|n} = P_{X_k|Y_{0:n}}$, $\pi_k = \pi_{k|k}$, $\lambda_{k:n}(x_{k:n}, y_{k:n}) = \prod_{j=k}^{n} \lambda(x_j, y_j)$, $P_{k:n}(x_{k-1}, dx_{k:n}) = \prod_{j=k}^{n} P(x_{j-1}, dx_j)$ for $k \geq 1$, and $P_{0:n}(dx_{0:n}) = P_{1:n}(x_0, dx_{1:n})\mu_0(dx_0)$. We let f denote an arbitrary bounded measurable function on $\mathbb{X}$.

Theorem 1.4 (Filtering, smoothing, prediction)**.** *Let (X, Y) be a HMM with non-degenerate observations as in Definitions 1.1 and 1.2. Then*

$$\pi_{k|n}(y_{0:n}, f) = \frac{\rho_{k|n}(y_{0:n}, f)}{\rho_{k|n}(y_{0:n}, 1)},$$

$$\rho_{k|n}(y_{0:n}, f) = \int f(x_k)\lambda_{0:n}(x_{0:n}, y_{0:n})P_{0:k\vee n}(dx_{0:k\vee n}).$$

The smoothing densities $\alpha_{k|n} = \pi_{k|n}/\pi_k$ and $\beta_{k|n} = \rho_{k|n}/\rho_k$, $k \leq n$, satisfy

$$\alpha_{k|n}(y_{0:n}, x_k) = \frac{\beta_{k|n}(y_{k+1:n}, x_k)}{\int \beta_{k|n}(y_{k+1:n}, x_k)\pi_k(y_{0:k}, dx_k)},$$

$$\beta_{k|n}(y_{k+1:n}, x_k) = \int \lambda_{k+1:n}(x_{k+1:n}, y_{k+1:n})P_{k+1:n}(x_k, dx_{k+1:n}).$$

Theorem 1.5 (Recursions)**.** *The prediction step for $k \geq n$ is*

$$\pi_{k+1|n}(y_{0:n}, f) = \int f(x_{k+1})P(x_k, dx_{k+1})\pi_{k|n}(y_{0:n}, dx_k),$$

$$\rho_{k+1|n}(y_{0:n}, f) = \int f(x_{k+1})P(x_k, dx_{k+1})\rho_{k|n}(y_{0:n}, dx_k).$$

The correction step for $k \geq 0$ is

$$\pi_{k+1}(y_{0:k+1}, f) = \frac{\int f(x_{k+1})\lambda(x_{k+1}, y_{k+1})\pi_{k+1|k}(y_{0:k}, dx_{k+1})}{\int \lambda(x_{k+1}, y_{k+1})\pi_{k+1|k}(y_{0:k}, dx_{k+1})},$$

$$\rho_{k+1}(y_{0:k+1}, f) = \int f(x_{k+1})\lambda(x_{k+1}, y_{k+1})\rho_{k+1|k}(y_{0:k}, dx_{k+1}).$$

The filtering step is a prediction followed by a correction step. The smoothing step for $k \leq n$ is

$$\begin{aligned}\alpha_{k-1|n}(y_{0:n}, x_{k-1}) &= \frac{\int \lambda(x_k, y_k)\alpha_{k|n}(y_{0:n}, dx_k)P(x_{k-1}, dx_k)}{\int \lambda(x_k, y_k)\alpha_{k|n}(y_{0:n}, dx_k)P(x_{k-1}, dx_k)\pi_k(y_{0:k}, dx_k)} \\ &= \frac{\int \lambda(x_k, y_k)\alpha_{k|n}(y_{0:n}, dx_k)P(x_{k-1}, dx_k)}{\int \lambda(x_k, y_k)P(x_{k-1}, dx_k)\pi_{k-1}(y_{0:k-1}, dx_{k-1})},\end{aligned}$$

$$\beta_{k-1|n}(y_{k:n}, x_{k-1}) = \int \lambda(x_k, y_k)\beta_{k|n}(y_{k+1:n}, x_k)P(x_{k-1}, dx_k).$$