EECS502 Stochastic Processes Lecture 3 - 01/16/2020

Review of Probability Theory & Stochastic Processes Continued

Lecturer: Vijay G. Subramanian *Scribes: Kang Gong*

1 Probability Theory

Probability Space $(\Omega,\ \mathcal{F},\ \mathbb{P})$

Definition 1 (π-system)**.** *A collection S of subset of Ω is a π-system if it is closed under finite intersection.*

Theorem 2. *Given two probability distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ such that they are the same on a π-system S, then they are also the same on the $\sigma(S)$, i.e.,*

$$\begin{aligned} if \quad & \forall A \in S,\ \mathbb{P}_1(A) = \mathbb{P}_2(A), \\ then \quad & \forall B \in \sigma(S),\ \mathbb{P}_1(B) = \mathbb{P}_2(B). \end{aligned}$$

Remark This theorem means that specifying the values on a π-system uniquely specifies the probability distribution on the smallest $\sigma-$algebra containing it.

Example 3. For commonly used $\Omega = \mathbb{R}$, $S = \{(-\infty, x] :\ x \in \mathbb{R}\}$, S is a π-system since if $x_1 > x_2$, $(-\infty, x_1] \cap (-\infty, x_2] = (-\infty, x_2] \in S$. And $\mathbb{P}_X((-\infty, x]) = \mathbb{F}_X(x)$ is the cdf for random variable X.

Example 4. Random processes $\{X_t\}_{t\in\mathbb{R}}$ all takes values in $\mathbb{R}$.

Consider Borel σ-algebra $\mathcal{B}(\mathbb{R})$ and π-system $\{B_{t_i} : B_{t_i} \in \mathcal{B}(\mathbb{R})\}$. We can also use $\{B_{t_i} : B_{t_i} = (-\infty, x]\ \forall x \in \mathbb{R}\}$.

Let $\overline{B}_n = \prod_{i=1}^{n} B_{t_i}$, where $\prod$ means Cartesian Product, i.e., $x = (x_1, ..., x_n) \in \overline{B}_n \Rightarrow x_i \in B_{t_i}$ for any i.

The smallest σ-algebra containing such a π-system is called the Borel σ-algebra. And $\forall t_1 < t_2 < \cdots < t_n \in \mathbb{R}$ and $\forall n \in \mathbb{N}$, we have

$$\mathbb{P}(\overline{B}_n) = \mathbb{P}(X_{t_1} \in B_{t_1}, ..., X_{t_n} \in B_{t_n}),$$

i.e., specifying all finite dimensional marginals is equivalent to specifying the distribution of the random processes. The alternate specification specifies the joint CDF for all finite-dimensional collections.

Definition 5 (strictly stationary)**.** *A process $\{X_t : t \in \mathbb{R}\}$ is (strictly) stationary if for any finite set of indices $t_1 < t_2 < \cdots < t_n \in \mathbb{R}$ and any $\tau \in \mathbb{R}$ s.t.*

$$\begin{aligned} &\mathbb{P}(X_{t_1} \in B_1, X_{t_2} \in B_2, ..., X_{t_n} \in B_n) \\ =&\mathbb{P}(X_{t_1+\tau} \in B_1, X_{t_2+\tau} \in B_2, ..., X_{t_n+\tau} \in B_n). \end{aligned}$$

Remark The strict stationarity means that any time shifts don't change the distribution of the process.

Example 6. Let $\{X_n\}_{n\in\mathbb{Z}}$ be iid. Bernoulli($\frac{1}{2}$) random variables, which is a stationary process.

Consequence 1: $\mathbb{P}(X_t \in B)$ for all $B \in \mathcal{B}(\mathbb{R})$ is independent of t.
$\mathbb{F}_{X_t}(x)$ is independent of t.
$\mu_t = \mathbb{E}[X_t]$ is independent of t (if it exists).

Consequence 2: $\mathbb{P}(X_t \in B_1, X_s \in B_2) = \mathbb{P}(X_{t-s} \in B_1, X_0 \in B_2)$ is a function of $(t-s)$, so are the covariance $\mathbb{E}[X_t X_s]$ and the correlation function $\mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)] = R_X(t-s)$.

Definition 7 (wide-sense stationary)**.** *A process* $\{X_t\}_{t\in\mathbb{R}}$ *is wide-sense stationary (WSS) if*

$$\mu_t \triangleq \mathbb{E}[X_t] \equiv \mu, \quad \textit{(Not a function of } t\textit{)}$$
$$R_X(t,s) \triangleq \mathbb{E}[(X_t-\mu_t)(X_s-\mu_s)] = R_X(t-s).$$

Remark The strict stationary process is also WSS while the reverse is only true for Gaussian processes.

Definition 8 (independence)**.** *X_1 and X_2 are independent if*

- *Joint Distribution:* $\mathbb{P}(X_1 \in B_1, X_2 \in B_2) = \mathbb{P}(X_1 \in B_1)\,\mathbb{P}(X_2 \in B_2)\ \forall\ B_1, B_2 \in \mathcal{B}(\mathbb{R})$.
- *Moments: For all separable functions* $f(x_1, x_2) = f_1(x_1)f_2(x_2)$, $\mathbb{E}[f(X_1, X_2)] = \mathbb{E}[f_1(X_1)]\,\mathbb{E}[f_2(X_2)]$. *The Gaussian random variables* X_1, X_2 *are independent iff* $\mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\,\mathbb{E}[X_2]$.
- *Characteristic function:* $\varphi_X(s) \triangleq \mathbb{E}[e^{isX}]$
  $\mathbb{E}[e^{is_1X_1+is_2X_2}] = \mathbb{E}[e^{is_1X_1}]\,\mathbb{E}[e^{is_2X_2}] \quad \forall\ s_1, s_2$. *If the characteristic function has a product form, then independence, and hence, the reverse is also true.*
  Moment generating function: $MGF_X(\theta) \triangleq \mathbb{E}[e^{\theta X}]$, $\theta \in \mathbb{R}$.
  $\mathbb{E}[e^{\theta_1X_1+\theta_2X_2}] = \mathbb{E}[e^{\theta_1X_1}]\,\mathbb{E}[e^{\theta_2X_2}] \quad \forall\ \theta_1, \theta_2$.
- *σ-algebra: Suppose X_1, X_2 take values in probability space $(E_1, \mathcal{E}_1, \mathbb{P})$ and $(E_2, \mathcal{E}_2, \mathbb{P})$ respectively. $\sigma(X_1)$ and $\sigma(X_2)$ are independent if* $\forall A_1 \in \mathcal{E}_1,\ A_2 \in \mathcal{E}_2$

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}_{A_1}(X_1)\mathbb{1}_{A_2}(X_2)]\\
=&\,\mathbb{P}(X_1 \in A_1, X_2 \in A_2)\\
=&\,\mathbb{P}(X_1 \in A_1)\,\mathbb{P}(X_2 \in A_2)\\
=&\,\mathbb{E}[\mathbb{1}_{A_1}(X_1)]\,\mathbb{E}[\mathbb{1}_{A_2}(X_2)].
\end{aligned}$$

Random variables are independent iff *the smallest σ-algebras $\sigma(X_1)$ and $\sigma(X_2)$ are independent.*

Definition 9 (Bayes's rule)**.** *The conditional probability is defined as*

$$\mathbb{P}(A|B) = \begin{cases} \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} & \textit{if } \mathbb{P}(B) > 0, \\ \textit{undefined otherwise.} \end{cases}$$

Remark Conditioning on a set or a random variable is actually equivalent to conditioning on a σ-algebra.

Example 10. $\mathbb{P}(A|B) = \mathbb{E}[\mathbb{1}_A(X) \mid \mathbb{1}_B(X) = 1]$ with $\mathbb{P}_X(X \in B) = \mathbb{P}(B)$ and $\mathbb{P}_X(X \in A \cap B) = \mathbb{P}(A \cap B)$. Given a set B, conditioning on $\sigma(B)$ is equivalent to conditioning on $\mathbb{1}_B(X)$ in that $\sigma(B) = \{\emptyset,\ B,\ B^c,\ \Omega\} = \sigma(\mathbb{1}_B(X))$.
Example 11. Given random variable X on a probability space $(\Omega,\ \mathcal{F},\ \mathbb{P})$ and σ-algebra $\mathcal{G} \subset \mathcal{F}$, we define the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ as follows. Let Y be another random variable, and $(X,\ Y)$ have some joint distribution. Then, we have $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$.

Definition 12 (conditional expectation)**.** *For random variable X and σ-algebra $\mathcal{G} \subset \mathcal{F}$, $\mathbb{E}[X|\mathcal{G}]$ is measurable with respect to $\mathcal{G}$. And for all $A \in \mathcal{G}$, $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}] \cdot \mathbb{1}_A]$. Specifically, $\mathbb{E}[X|Y]$ is a measurable function of Y, i.e., $g(Y)$. For every function of Y, $f(Y)$, we have*

$$\begin{aligned}
\mathbb{E}[Xf(Y)] &= \mathbb{E}[\mathbb{E}[X|Y] \cdot f(Y)]\\
&= \mathbb{E}[g(Y) \cdot f(Y)].
\end{aligned}$$

Example 13. $\mathbb{E}[X] = \mathbb{E}[X|\{\emptyset,\ \Omega\}]$ is constant, since the only functions measurable under $\{\emptyset,\ \Omega\}$ are constants.
Example 14. X and Y have a joint distribution (discrete) $\mathbb{P}_{X,Y}(x, y) \quad \forall x, y \in \mathbb{Z}$.

$$g(y) = \begin{cases} \sum_{x\in\mathbb{Z}} x\,\mathbb{P}_{X|Y}(x|y) = \sum_{x\in\mathbb{Z}} x \cdot \dfrac{\mathbb{P}_{X,Y}(x,y)}{\mathbb{P}_Y(y)} & \text{if } \mathbb{P}_Y(y) > 0, \\ 0 \quad \text{otherwise.} \end{cases}$$

$$\mathbb{E}[X|Y] = g(Y).$$

Example 15. X and Y have a joint density (continuous) $f_{X,Y}(x,y) \quad \forall x,y \in \mathbb{R}$.

$$
\begin{aligned}
g(y) &= \begin{cases} \int x f_{X|Y}(x|y)dx = \int x \cdot \frac{f_{X,Y}(x,y)}{f_Y(y)}dx & \text{if } f_Y(y) > 0, \\ 0 \quad \text{otherwise.} \end{cases} \\
\mathbb{E}[X|Y] &= g(Y).
\end{aligned}
$$

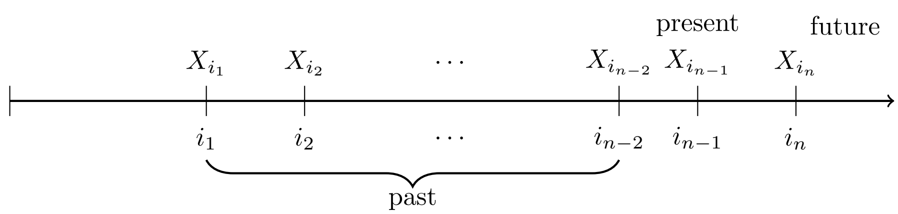
If X and Y are jointly Gaussian, $\mathbb{E}[X|Y] = a + bY$, a linear function of Y.

2 Stochastic Process

Definition 16 (Markov process)**.** *Given an index set I and probability space $(E,\ \mathcal{E},\ \mathbb{P})$, a random process $\{X_i\}_{i\in I}$ taking values in E is Markov if for any finite sequence of indices $i_1 < i_2 < \cdots < i_n \in I$,*

$$
\begin{aligned}
\forall A \in \mathcal{E}, \quad & \mathbb{P}(X_{i_n} \in A | X_{i_1}, X_{i_2}, ..., X_{i_{n-1}}) \\
= & \mathbb{P}(X_{i_n} \in A | X_{i_{n-1}}).
\end{aligned}
$$

Remark For a Markov process $\{X_i\}_{i\in I}$, the future is conditionally independent of the past given the present.



Example 17. Suppose the discrete Markov process $\{X_i\}_{i\in\mathbb{Z}}$ has all positive conditional probabilities, then we have

$$
\begin{aligned}
& \mathbb{P}(X_{i_n} = a_n, X_{i_{n-1}} = a_{n-1}, ..., X_{i_1} = a_1) \\
= & \mathbb{P}(X_{i_n} = a_n \mid X_{i_{n-1}} = a_{n-1}, ..., X_{i_1} = a_1)\, \mathbb{P}(X_{i_{n-1}} = a_{n-1}, ..., X_{i_1} = a_1) \\
= & \mathbb{P}(X_{i_n} = a_n \mid X_{i_{n-1}} = a_{n-1})\, \mathbb{P}(X_{i_{n-1}} = a_{n-1}, ..., X_{i_1} = a_1) \\
= & \mathbb{P}(X_{i_n} = a_n \mid X_{i_{n-1}} = a_{n-1})\, \mathbb{P}(X_{i_{n-1}} = a_{n-1} \mid X_{i_{n-2}} = a_{n-2}) \cdots \mathbb{P}(X_{i_2} = a_2 \mid X_{i_1} = a_1)\, \mathbb{P}(X_{i_1} = a_1).
\end{aligned}
$$