E1 245: Online Prediction & Learning **Fall 2014**

Lecture 2 — August 6

Lecturer: Aditya Gopalan *Scribe: Shreyas S*

2.1 Review of Math Fundamentals

2.1.1 Probability

Probability Space: A probability space is defined to be a triple $(\Omega, \mathscr{F}, \mathbb{P})$. The Ω is the set of all possible outcomes. The $\mathscr{F}$ is a σ-algebra, i.e., a collection of events $\mathscr{F} \subseteq 2^{\Omega}$ such that it satisfies the following properties:

1. $\Omega \in \mathscr{F}$.
2. $\forall E \in \mathscr{F},\ E^c \in \mathscr{F}$.
3. If $E_1,\ E_2\ \cdots$ is a sequence of sets in $\mathscr{F}$, then $\cup_{i=1}^{\infty} E_i \in \mathscr{F}$.

The function $\mathbb{P} : \mathscr{F} \to \mathbb{R}$ is called probability measure and satisfies following properties:

1. $\forall E \in \mathscr{F},\ \mathbb{P}(E) \geq 0$.
2. $\mathbb{P}(\Omega) = 1$.
3. If $E_1,\ E_2\ \cdots$ is a sequence of disjoint sets (mutually exclusive events) in $\mathscr{F}$, then $\mathbb{P}(\sum_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$.

Random Variables: The function $\mathbf{X} : \Omega \to \mathbb{R}$ is called a Random variable if $\mathbf{X}$ is measurable i.e., $\forall x \in \mathbb{R}$ following holds

$$\{\omega : \mathbf{X}(\omega) \leq x\} \in \mathscr{F}$$

Expectation of a Random Variable: The function $\mathbb{E}[\mathbf{X}] : \mathbf{X} \to \mathbb{R}$ called Expectation of a Random variable $\mathbf{X}$ is defined as follows:

$$\mathbb{E}[\mathbf{X}] = \int_{\Omega} \mathbf{X}(\omega)\mathbb{P}(d\omega)$$

The following definitions of $\mathbb{E}[\mathbf{X}]$ hold depending on the nature of sample space:

1. If Ω is discrete, then $\mathscr{F} = 2^{\Omega}$ and $\mathbb{P}[w \in \Omega : \mathbf{X}(w) = x] = p_x$. The expectation is given by $\mathbb{E}[\mathbf{X}] = \sum_x x p_x$.

2. If $\mathbf{X}$ is continuous random variable, with probability density function $f_{\mathbf{X}} : \mathbb{R} \to \mathbb{R}$ then expectation is given by $\mathbb{E}[\mathbf{X}] = \int_{\mathbb{R}} x f_{\mathbf{X}}(x) dx$.

Some of the properties of expectation are given below:

1. If $\mathbf{X}$ and $\mathbf{Y}$ are random variables, then $\mathbb{E}[\mathbf{X}+\mathbf{Y}] = \mathbb{E}[\mathbf{X}] + \mathbb{E}[\mathbf{Y}]$.
2. If $\mathbf{X}$ is non-negative random variable i.e., $\mathbb{P}[\mathbf{X} < 0] = 0$ then expectation of $\mathbf{X}$ is given by $\mathbb{E}[\mathbf{X}] = \int_0^\infty \mathbb{P}[\mathbf{X} \geq x] dx$.
3. Corresponding to any event $E \in \mathscr{F}$, we can define its Indicator random variable $1_{\mathbf{E}}$ is defined as
$$1_{\mathbf{E}}(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & otherwise \end{cases}$$
then we have $\mathbb{E}[1_{\mathbf{E}}] = \mathbb{P}(E), \forall E \in \mathscr{F}$.

Variance of a Random Variable: The variance of a random variable $\mathbf{X}$ is given

$$Var(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])^{\mathbf{2}}]$$

The variance $Var(\mathbf{X}) = 0$ iff $\mathbb{P}[\mathbf{X} = \mathbb{E}[\mathbf{X}]] = 1$.

Conditional Probability and Expectation: Let $\mathbf{X}$ and $\mathbf{Y}$ be two discrete random variables on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ then conditional Expectation of $\mathbf{X}$ given $\mathbf{Y}$.

$$\mathbb{E}[\mathbf{X}/\mathbf{Y} = \mathbf{y}] = \sum_x x \, \mathbb{P}[\mathbf{X} = x/\mathbf{Y} = y]$$

where $\mathbb{P}[\mathbf{X} = x/\mathbf{Y} = y] = \frac{\mathbb{P}[\mathbf{X}=x, \mathbf{Y}=y]}{\mathbb{P}[\mathbf{Y}=y]}$ is called the conditional probability of $\mathbf{X}$ given $\mathbf{Y}$ with $\mathbb{P}[\mathbf{Y} = y] > 0$. Let $\mathbf{X}$ and $\mathbf{Y}$ be two continuous random variables on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ then conditional Expectation of $\mathbf{X}$ given $\mathbf{Y}$.

$$\mathbb{E}[\mathbf{X}/\mathbf{Y} = \mathbf{y}] = \int_{\mathbb{R}} x \, f_{\mathbf{X}/\mathbf{Y}}(x/y) dx$$

where $f_{\mathbf{X}/\mathbf{Y}}(x/y) = \frac{f_{\mathbf{X},\mathbf{Y}}(x,y)}{f_{\mathbf{Y}}(y)}$ is called the conditional probability of $\mathbf{X}$ given $\mathbf{Y}$.

Independence: For $A, B \in \mathscr{F}$, then events A and B are said to be independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Similarly for random variables $\mathbf{X}$ and $\mathbf{Y}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ are said to be independent if $\forall x, y \in \mathbb{R}$ the following holds

$$\mathbb{P}(\mathbf{X} \leq x, \mathbf{Y} \leq y) = \mathbb{P}[\mathbf{X} \leq x] \cdot \mathbb{P}[\mathbf{Y} \leq y]$$

or alternatively,

$$\mathbb{E}[\mathbf{X}/\mathbf{Y}=\mathbf{y}] = \sum_x x\,\mathbb{P}[\mathbf{X}=x/\mathbf{Y}=y]$$

where $\mathbb{P}[\mathbf{X}=x/\mathbf{Y}=y] = \frac{\mathbb{P}[\mathbf{X}=x,\mathbf{Y}=y]}{\mathbb{P}[\mathbf{Y}=y]}$ where $\mathbb{P}[\mathbf{Y}=y] > 0$.

Strong Law of Large Numbers: Let $X_1,\ X_2,\ X_3,\cdots$ be a sequence of i.i.d random variables. Suppose that $\mathbb{E}[|\mathbf{X_1}|] < \infty$ and $\mathbb{E}[\mathbf{X_1}] = \mu$. Then

$$\mathbb{P}[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i = \mu] = 1$$

Central Limit of Theorem: Let $X_1, X_2, X_3,\cdots$ be a sequence of i.i.d random variables. Suppose that $\mathbb{E}[\mathbf{X_1}] = \mu$ and $Var(\mathbf{X}_1) = \sigma^2$. Then

$$\mathbb{P}[\ (\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i - \mu) \geq \frac{\varepsilon}{\sqrt{n}}] \to \Phi(\varepsilon/\sigma)\ \ as\ \ n\to\infty$$

where Φ is the cumulative distribution of standard normal random variable.

Markov Inequality: For a non-negative random variable $\mathbf{X}$, $\forall z > 0$,

$$\mathbb{P}[\ \mathbf{X} \geq z] \leq \frac{\mathbb{E}[\mathbf{X}]}{z}.$$

Proof:

$$\begin{aligned}
\mathbb{E}[\mathbf{X}] &= \mathbb{E}[\mathbf{X}\ \mathbf{1}_{\{\mathbf{X}\geq\mathbf{z}\}} + \mathbf{X}\ \mathbf{1}_{\{\mathbf{X}<\mathbf{z}\}}] \\
&= \mathbb{E}[\mathbf{X}\ \mathbf{1}_{\{\mathbf{X}\geq\mathbf{z}\}}] + \mathbb{E}[\mathbf{X}\ \mathbf{1}_{\{\mathbf{X}<\mathbf{z}\}}] \\
&\geq \mathbb{E}[\mathbf{X}\ \mathbf{1}_{\{\mathbf{X}\geq\mathbf{z}\}}]\ \text{, since } \mathbb{E}[\mathbf{X}\ \mathbf{1}_{\{\mathbf{X}<\mathbf{z}\}}] \geq 0 \\
&\geq \mathbb{E}[\mathbf{X}\ \mathbf{1}_{\{\mathbf{X}\geq\mathbf{z}\}}] \\
&\geq \mathbb{E}[\mathbf{z}\ \mathbf{1}_{\{\mathbf{X}\geq\mathbf{z}\}}] \\
&\geq z\ \mathbb{P}[\mathbf{X} \geq z]
\end{aligned}$$

□

Chebyshev Inequality:

$$\mathbb{P}[\ |\mathbf{X} - \mathbb{E}[\mathbf{X}]| \geq z] \leq \frac{Var(\mathbf{X})}{z^2}.$$

Hoeffding's Inequality: Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be independent random variables such that $\mathbf{X}_i \in [a_i,\ b_i]$ for $a_i,\ b_i \in \mathbb{R}, \forall i$.

$$\mathbb{P}[\ \frac{1}{n}(\sum_{i=1}^{n} \mathbf{X}_i - \sum_{i=1}^{n} \mathbb{E}[\mathbf{X_i}]) > \varepsilon] \leq \exp\Big(-\frac{2n\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\Big)$$