Modern Statistics **Lecture 2 - 2/20/2025**

Probability and Random Variable

Lecturer:Xiangyu Chang *Scribe: Yijun Wang, PUA WENG YANG*

Edited by: Zhihong Liu

1 The Law of total probability

Definition 1 (Partition) *A collection of events* $\{B_1, B_2, ..., B_n\}$ *forms a* ***partition*** *of the sample space* Ω *if:*

- $B_i \cap B_j = \varnothing$ *for all* $i \neq j$.
- $\bigcup_{k=1}^{k} B_i = \Omega$.

Theorem 1 (Law of Total Probability) *Let* $\{B_1, B_2, ..., B_n\}$ *be a partition of* Ω *with* $P(B_i) > 0$ *for all* i*. Then for any event* A*:*

$$P(A) = \sum_{k=1}^{k} P(AB_k) = \sum_{k=1}^{k} P(A \mid B_k)P(B_k). \tag{1}$$

proof: Using the countable additivity axiom and conditional probability definition:

$$A = \bigcup_{k=1}^{k} AB_k$$

$$(AB_k) \bigcap (AB_l) = \varnothing$$

$$\begin{aligned} P(A) &= P\left(\bigcup_{k=1}^{k} AB_k\right) \\ &= \sum_{i=1}^{n} P(A \cap B_i) \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i) \end{aligned}$$

2 Bayesian Theorem

Theorem 2 (Bayes' Theorem) *For any events* $B_k (k = 1, 2, ...)$ *and* A *with* $P(A) > 0$*:*

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{k=1}^{k} P(A|B_k)P(B_k)}. \tag{2}$$

Example 1 (Naive Bayes Spam Filter) *Suppose:*

- *70% of emails are spam ($P(Spam) = 0.7$).*
- *20% of emails are spam ($P(Low\ Priority) = 0.2$).*
- *10% of emails are spam ($P(High\ Priority) = 0.1$).*
- *The word "essay writing" appears in:*
  - *90% of spam emails ($P(\text{"essay writing"}|Spam) = 0.9$).*
  - *1% of legitimate emails ($P(\text{"essay writing"}|Low\ Priority) = 0.01$).*
  - *1% of legitimate emails ($P(\text{"essay writing"}|High\ Priority) = 0.01$).*

Calculate the probability that an email containing "essay writing" is spam.

Solution:

$$\begin{aligned}
P(Spam|\text{"essay writing"}) &= \frac{P(\text{"essay writing"}|Spam)P(Spam)}{P(\text{"essay writing"})}, \\
P(\text{"essay writing"}) &= P(e.w.|Spam)P(Spam) + P(e.w.|Low\ Priority)P(Low\ Priority) \\
&\quad + P(e.w.|High\ Priority)P(High\ Priority) \\
&= 0.9 \times 0.7 + 0.01 \times 0.2 + 0.01 \times 0.1 \\
&= 0.633, \\
P(Spam|\text{"essay writing"}) &= \frac{0.9 \times 0.7}{0.633} = \frac{0.63}{0.633} \approx 0.9953.
\end{aligned}$$

3 Continuity of Probability

Definition 2 (Monotonic Sequences) *A sequence of events $\{A_n\}$ is:*

- ***Increasing*** *if $A_1 \subseteq A_2 \subseteq \cdots$ (denoted $A_n \uparrow A$).*
- ***Decreasing*** *if $A_1 \supseteq A_2 \supseteq \cdots$ (denoted $A_n \downarrow A$),*

where $A = \lim\limits_{n\to\infty} A_n$.

Theorem 3 (Continuity of Probability) *For any probability measure P:*

1. *If $A_n \uparrow A$, then $\lim_{n\to\infty} P(A_n) = P(A)$.*
2. *If $A_n \downarrow A$ and $P(A_1) < \infty$, then $\lim\limits_{n\to\infty} P(A_n) = P(A)$.*

Proof for Increasing Sequence: Let $B_1 = A_1,\ B_k = A_k \setminus \bigcup_{i=1}^{k-1} A_i$ for $n \geq 2$.
Then:

$$\begin{aligned}\lim_{n\to\infty} P(A_n) &= \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} B_n\right) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i)\\ &= P(\bigcup_{i=1}^{\infty} B_i) = P(\bigcup_{i=1}^{\infty} A_i)\\ &= P(A).\end{aligned}$$

4 Random Variable

Definition 3 (Random Variable) *A **random variable** X is a measurable function from a probability space $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:*

$$X : \Omega \to \mathbb{R}$$

- *X: is a mapping, measurable function.*
- $\Omega : outcomes.$
- $\omega \in \Omega, X(\omega) \in \mathbb{R}.$

Example 2 (Coin Toss) *Consider the probability space $(\Omega, \mathcal{F}, P)$ where:*

- *Sample space $\Omega = \{H, T\}$ represents the elementary outcomes.*
- *Define the **indicator random variable** $X : \Omega \to \{0, 1\}$ by:*

$$X(\omega) = \begin{cases} 1, & \text{if } \omega = H \text{ (Head)}. \\ 0, & \text{if } \omega = T \text{ (Tail)}. \end{cases}$$

Example 3 (Stock Price of Today is Higher than Yesterday) *Consider a financial asset's daily closing prices over time. Define:*

- ***Sample space**: $\Omega = \{\omega\}_{t\in\mathbb{N}}$ where ω represents wether the stock price of today is higher than yesterday.*
- ***Event space**: $\mathcal{F} = \sigma(\{H, L\})$ where:*
  - *H, price increase.*
  - *L , price decrease.*

- ***Random variable**: $X : \Omega \to \{0, 1\}$ defined as:*

$$X(\omega) = \begin{cases} 1, & \text{if } \omega \in H. \\ 0, & \text{if } \omega \in L. \end{cases}$$

Example 4 (Five Coin Tosses) *Consider the canonical experiment of five independent coin tosses:*

- ***Sample space**: $\Omega = \big\{\omega = (\omega_1, \ldots, \omega_5) \big| \omega_i \in \{H, T\}\big\}$ where H denotes Heads and T denotes Tails.*

- ***Random variable**: $X : \Omega \to \{0, 1, 2, 3, 4, 5\}$ defined as:*

$$X(\omega) = \sum_{i=1}^{5} \mathbf{1}_{\{\omega_i = H\}}.$$

  counting the total number of Heads in the sequence

- ***Realizations**:*

$$X(H, H, H, H, H) = 5.$$

5 the Relationship Between Random Variable and Probability

- $X : \Omega \to \mathbb{R}$.

  An outcome in the sample space is defined as a real number.

- Define a set $A = [a, b] \subseteq \mathbb{R}$, then $P(X \in A) = P(\omega | \omega \in \Omega, X(\omega) \in A)$.

Example 5 (Binomial Probability of Coin Tosses) *Consider a sequence of n independent trials representing fair coin tosses. Let X be the random variable counting the number of heads observed.*

where:

- $\Omega = \{HH...H, HH...T, ..., TT...T\}$ *(total $n + 1$ element).*

- *if $\omega = \overbrace{HH...H}^{k} TT...T$, then $X(\omega) = k$.*

$$P(X(\omega) = k) = P(\omega = \overbrace{HH...H}^{k} TT...T) = \binom{n}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{n-k}.$$

***Key Observations**:*

- *The distribution is symmetric when $p = \frac{1}{2}$.*

- *Sum of probabilities:*

$$P(\Omega) = 1 \Rightarrow \sum_{k=0}^{n} P(X(\omega) = k) = 1 \Rightarrow \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^n = 2^n \left(\frac{1}{2}\right)^n = 1.$$

6 Cumulative Distribution Function(CDF)

Definition 4 (Cumulative Distribution Function) *For a random variable X, its **Cumulative Distribution Function (CDF)** is the function $F_X : \mathbb{R} \to [0, 1]$ defined by:*

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}).$$

Theorem 4 (CDF Characterization) *Any valid CDF must satisfy:*

1. ***boundedness**: $0 \leq F_X(x) \leq 1$.*
2. ***Monotonic increasing**: $x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$.*
3. ***Normalized**:*

$$\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to +\infty} F_X(x) = 1.$$

4. ***Right-Continuity**: $\lim_{y \to x^+} F_X(y) = F_X(x^+) = F_X(x)$.*

Theorem 5 (the Property of CDF) *1. $P(X = x) = F_X(x) - F_X(x^-)$.*

2. *if $F_X(x)$ is continuous, then $P(X = x) = 0$, $P(a \leq x \leq b) = P(a < x \leq b) = P(a < x < b) = P(a \leq x < b)$.*
3. *$P(x < X \leq y) = F_X(y) - F_X(x)$.*
4. *$P(X > x) = 1 - F_X(x)$.*