UNIVERSITY
OF OSLO

IN4310 Probability and Random Variables

ali@uio.no

Axioms for Events and Axioms of Probability [1, 2]

Sample space Ω the set of all sample points for a given experiment

Events are subsets of the sample space

A^c Complement of an event A is the set of points in Ω but not A

Axioms for events

Ω is an event

For every sequence of events $A_1, A_2, \cdots$, the $\bigcup_{i=1}^{\infty} A_i$ is an event

For every event A, the complement A^c is an event

Axiom of Probability: a probability rule $\Pr$ is a function mapping each event to a real number so that

$\Pr(\Omega) = 1$

For every event A, $\Pr(A) \geq 0$

For disjoint events $A_1, A_2, \cdots$, $\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{n=1}^{\infty} \Pr(A_i)$

Properties of Probability Rule $\Pr$

Range $0 \leq \Pr(A) \leq 1$ for any event A

Complement $\Pr(A^c) = 1 - \Pr(A)$

Events A and B are independent if $\Pr(A \cap B) = \Pr(A) \Pr(B)$

Events A and B are mutually exclusive if $\Pr(A \cap B) = 0$

$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

Conditional probability: probability of A given B under $\Pr(B) > 0$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Law of Total Probability and Bayes' Rule

$$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B^c) = \Pr(A|B)\Pr(B) + \Pr(A|B^c)\Pr(B^c)$$

Suppose $\bigcup_{i=1}^{\infty} B_i = \Omega$ where B_i's are mutually exclusive

$$\Pr(A) = \sum_{i=1}^{\infty} \Pr(A \cap B_i) = \sum_{i=1}^{\infty} \Pr(A|B_i)\Pr(B_i)$$

Bayes' rule: Given $\Pr(A|B_i)$, we obtain $\Pr(B_i|A)$

$$\Pr(B_i|A) = \frac{\Pr(A|B_i)\Pr(B_i)}{\Pr(A)} = \frac{\Pr(A|B_i)\Pr(B_i)}{\sum_{j=1}^{\infty} \Pr(A|B_j)\Pr(B_j)}$$

Random Variable and Distribution Function

Formally, a random variable (RV) is a function $X : \Omega \to \mathbb{R}$ s.t.
1) X is undefined or infinite for a subset of Ω that has 0 probability;
2) $\{\omega \in \Omega | X(\omega) \leq x\}$ is an event for each $x \in \mathbb{R}$;
3) for every finite set of RVs $X_1, \cdots, X_n$, we have $\{\omega \in \Omega | X_1(\omega) \leq x_1, \cdots, X_n(\omega) \leq x_n\}$ is an event for each $x_1 \in \mathbb{R}, \cdots, x_n \in \mathbb{R}$.

(Cumulative) Distribution Function (CDF) of X is a function $F_X : \mathbb{R} \to \mathbb{R}$ defined by

$$F_X(x) = \Pr(\{\omega \in \Omega | X(\omega) \leq x\})$$

We omit ω for simplicity $F_X(x) = \Pr(\{X \leq x\})$
F_X is non-decreasing, right-continuous, $F(-\infty) = 0$ and $F(\infty) = 1$

Probability Density Function and Probability Mass Function

X is a Continuous RV if its cdf F_X is differentiable e.g., Gaussian and exponential RVs

Probability Density Function (PDF)

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$$

$$\Pr(\{X \in [x, x + \Delta x]\}) \approx f_X(x)\Delta x \text{ for small } \Delta x$$

X is a Discrete RV if its cdf F_X is piece-wise constant with jumps at $x = x_i$ s.t. $\Pr(\{X = x_i\}) > 0$ e.g., Bernoulli and Poisson RVs

Probability Mass Function (PMF)

$$f_X(x) = \Pr(\{X = x\}) = \lim_{\delta \to 0^+} F_X(x) - F_X(x - \delta)$$

Let g denote a function of RV X

$\mathbb{E}[g(X)] = \int g(x) f_X(x)\,\mathrm{d}x$ if X is continuous
$\mathbb{E}[g(X)] = \sum_i g(x_i) f_X(x_i)$ if X is discrete

Mean of X is $\mathbb{E}[X]$ and Variance of X is $\mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$

Exponential RV: X has exponential distribution with parameter λ if $F_X(x) = 1 - \exp(-\lambda x)$ for $x \geq 0$

$$f_X(x) = \lambda \exp(-\lambda x)$$

$\mathbb{E}[X] = \mathrm{Var}(X) = 1/\lambda$

Joint Distribution

Let $\mathbf{X} = [X_1, \cdots, X_n]^\top$ where RV X_i's are defined on a common probability space

Joint CDF $\quad F_{\mathbf{X}}(\mathbf{x}) = \Pr(\{X_1 \leq x_1, \cdots, X_n \leq x_n\})$

Joint PDF $\quad f_{\mathbf{X}}(\mathbf{x})$ obtained by taking $\partial/\partial x_i$ for all i

$$\Pr(\{X_1 \in [x_1, x_1 + \Delta x_1], \cdots, X_n \in [x_n, x_n + \Delta x_n]\}) \approx P_{\mathbf{X}}(\mathbf{x})\Delta x_1 \cdots \Delta x_n$$

Joint PMF $\quad f_{\mathbf{X}}(\mathbf{x}) = \Pr(\{X_1 = x_1, \cdots, X_n = x_n\})$

Conditional Density and Bayes' Rule

Conditional density of Y given X

$$f_Y(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Bayes' rule: given conditional density of Y given X the conditional density for X given Y

$$f_X(x|y) = \frac{f_Y(y|x) f_X(x)}{f_Y(y)} = \frac{f_Y(y|x) f_X(x)}{\int f_Y(y|x) f_X(x) \, \mathrm{d}x}$$

$$f_X(x|y) = \frac{f_Y(y|x) f_X(x)}{f_Y(y)} = \frac{f_Y(y|x) f_X(x)}{\sum_x f_Y(y|x) f_X(x)}$$

Markov's Inequality

If $X \geq 0$ and $\epsilon > 0$, then

$$\Pr(\{X \geq \epsilon\}) \leq \frac{\mathbb{E}[X]}{\epsilon}$$

Proof:

$$\begin{aligned}
\Pr(\{x \geq t\mathbb{E}[X]\}) &= \sum_{x \geq t\mathbb{E}[X]} \Pr(\{X = x\}) \\
&\leq \sum_{x \geq t\mathbb{E}[X]} \Pr(\{X = x\}) \frac{x}{t\mathbb{E}[X]} \\
&\leq \sum_{x} \Pr(\{X = x\}) \frac{x}{t\mathbb{E}[X]} \\
&= \mathbb{E}\left[\frac{X}{t\mathbb{E}[X]}\right] = \frac{1}{t}
\end{aligned}$$

Conditional density of Y given X

$$f_Y(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Bayes' rule: given conditional density of Y given X the conditional density for X given Y

$$f_X(x|y) = \frac{f_Y(y|x)f_X(x)}{f_Y(y)} = \frac{f_Y(y|x)f_X(x)}{\int f_Y(y|x)f_X(x)\,\mathrm{d}x}$$
$$f_X(x|y) = \frac{f_Y(y|x)f_X(x)}{f_Y(y)} = \frac{f_Y(y|x)f_X(x)}{\sum_x f_Y(y|x)f_X(x)}$$

Hoeffding's Theorem

Let $X_1, \cdots, X_n$ be independent RVs and $X_i \in [a_i, b_i]$. Then for any ϵ, $S_n = \sum_{i=1}^n X_i$ satisfies

$$\Pr(\{S_n - \mathbb{E}[S_n] \geq \epsilon\}) \leq \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

$$\Pr(\{S_n - \mathbb{E}[S_n] \leq -\epsilon\}) \leq \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

References

[1] R. Gallager. Introduction and Review of Probability. https://ocw.mit.edu/courses/6-262-discrete-stochastic-processes-spring-2011/.

[2] M. Mohri. Foundations of Machine Learning Introduction to ML. https://cs.nyu.edu/~mohri/mls/ml_introduction.pdf.