ORF 526 Condensed Topic Notes
Miklos Racz

*Emily Walters**

Friend Center 008

September 12th, 2018 – December 31st, 2018

Contents

*ewalters@princeton.edu

1 Elementary Probability Theory

1.1 *Conditional Probability*

1.1 definition (Conditional Probability).

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

1.2 theorem (Baye's Theorem).

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\,\mathbb{P}(A)}{\mathbb{P}(B)}$$

1.2 *Statistical Independence*

1.3 definition (Statistical Independence). Two events A, B are statistically independent if

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

1.3 *Random Variables*

1.4 definition (Elementary Definition of Random Variables). Given a sample space Ω, a random variable is a numeric function on Ω.

The distribution of random variables can be defined by a probability distribution function. This can take multiple forms.

1.5 definition (Cumulative Density Function).

$$F_X(x) = \mathbb{P}(X \leq x)$$

1.6 definition (Probability Mass Function). The distribution of discrete random variables can be defined by a function of the form

$$f_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\})$$

1.7 definition (Probability Density Function). The distribution of a continuous random variable can be defined by a function f_X where

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$$

Notice that this also defines the cumulative density function as

$$F_X(x) = \int_{-\infty}^x f_X(t)dt$$

1.4 *Expected Value*

1.8 definition (Expected Value). The expected value of a discrete random variable X is defined as

$$\mathbb{E}[X] = \sum_{-\infty}^{\infty} x\mathbb{P}(X = x)$$

For a continuous random variable, we define it as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

where f_X is the probability density function of X.

1.9 remark (Properties of Expected Value).

- Linearity: For $a, b \in \mathbb{R}$, $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
- $\mathbb{E}[X]$ is finite if and only if $\mathbb{E}[|X|]$ is finite.
- $X \geq 0$ A.S., then $\mathbb{E}[X] \geq 0$.
- If $X \leq Y$ A.S. and both $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ exists (that is, $\min\{\mathbb{E}[X_+], \mathbb{E}[X_-]\} < \infty$ and $\min\{\mathbb{E}[Y_+], \mathbb{E}[Y_-]\} < \infty$), then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
- If $\mathbb{E}\left[|X^b|\right] < \infty$ and $0 < a \leq b$, then $\mathbb{E}[|X^a|] < \infty$.
- If X, Y are independent random variables then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

1.5 *Variance*

1.10 definition (Variance).

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

1.11 definition (Covariance).

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])]\,\mathbb{E}[(Y - \mathbb{E}[Y])]$$

1.12 remark (Properties of Variance).

- $\mathrm{Var}(X) \geq 0$.
- $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$.
- $\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y) + 2ab \mathrm{Cov}(X, y)$.

1.6 *Important Distributions*

Bernoulli Distribution

1.13 remark. The Bernoulli distribution is the distribution of the random variable that takes the value 1 with probability p and the value 0 with probability $(1-p)$.

1.14 fact (Bernoulli Distribution Properties).

- Expected Value: $\mathbb{E}[X] = p$.
- Variance: $\mathrm{Var}(X) = p(1-p)$.

Binomial Distribution

1.15 remark. The binomial distribution is the probability distribution of the number of successes in a sequence of n experiments, each with probability of success p. In other words, the sum of n independent random variables each with a Bernoulli distribution and probability p.

1.16 fact (Binomial Distribution Properties).

- PMF: $\mathbb{P}(X = k) = f(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}$.
- CDF: $\mathbb{P}(X \leq k) = F(k, n, p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$.
- Expected Value: $\mathbb{E}[X] = np$.
- Variance: $\mathrm{Var}(X) = np(1-p)$.

Geometric Distribution

1.17 remark. The geometric distribution describes the number of repeated Bernoulli trials (experiments with a probability of success p) needed to achieve one success. For example, how many times must a coin be flipped to get a heads.

1.18 fact (Geometric Distribution Properties).

- PMF: $\mathbb{P}(X = k) = f(k, p) = (1-p)^{k-1} p$.
- CDF: $\mathbb{P}(X \leq k) = F(k, p) = 1 - (1-p)^k$.
- Expected Value: $\mathbb{E}[X] = \frac{1}{p}$.
- Variance: $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

Poisson Distribution

1.19 remark. The Poisson distribution describes the probability that a given number of events will occur in a fixed length of time if the events occur at a constant rate and the chance of one occurring is independent of the time since the last event. For example, the number of meteors that strike the earth in one year. λ denotes the average number of events in a unit time.

1.20 fact (Poisson Distribution Properties).

- PMF: $\mathbb{P}(X = k) = f(k, \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$.
- CDF: $\mathbb{P}(X \leq k) = F(k, \lambda) = e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}$.
- Expected Value: $\mathbb{E}[X] = \lambda$.
- Variance: $\text{Var}(X) = \lambda$.

Exponential Distribution

1.21 remark. The exponential distribution describes the time between events in a Poisson point process. That is, a process in which events occur continuously and independently at a constant average rate. λ denotes the average number of events in a unit time.

1.22 fact (Poisson Distribution Properties).

- PDF: $f(x, \lambda) = \lambda e^{-\lambda x}$.
- CDF: $\mathbb{P}(X \leq k) = F(x, \lambda) = \int_{-\infty}^{x} f(t, \lambda) dt = 1 - e^{-\lambda x}$.
- Expected Value: $\mathbb{E}[X] = \frac{1}{\lambda}$.
- Variance: $\text{Var}(X) = \frac{1}{\lambda^2}$.

Normal Distribution

1.23 remark. The normal distribution is a continuous probability distribution with a number of unique properties. It is particularly important due to its relation to the central limit theorem.

1.24 fact (Normal Distribution Properties).

- PDF: $f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
- CDF: $\mathbb{P}(X \leq k) = F(x, \lambda) = \int_{-\infty}^{x} f(t, \lambda) dt = \frac{1}{2}[1 + erf(\frac{x-\mu}{\sigma\sqrt{2}})]$, where erf, the error function, is non-elementary.
- Expected Value: $\mathbb{E}[X] = \mu$.
- Variance: $\text{Var}(X) = \sigma^2$.

1.7 *Important Inequality Theorems*

1.25 theorem (Markov's Inequality). *Suppose X is a non-negative random variable and $a > 0$. Then*

$$\mathbb{P}\left(X \geq a\right) \leq \frac{\mathbb{E}\left[X\right]}{a}$$

1.26 theorem (Chebyshev's Inequality). *Let X be a random variable with $\mathbb{E}\left[X\right] = \mu$ finite and non-zero variance $\mathrm{Var}(x) = \sigma^2$. Then for an real $k > 0$,*

$$\mathbb{P}\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

1.8 *Important Limit Theorems*

1.27 theorem (Weak Law of Large Numbers). *Let $\{X_i\}$ be i.i.d. random variables with mean μ, and let $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$. Then S_n converges to μ in probability. That is, for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left(|S_n - \mu| > \epsilon\right) = 0$$

That is, for any ϵ, we can guarantee that S_n is within ϵ of μ with arbitrary probability given a sufficient n.

1.28 theorem (Strong Law of Large Numbers). *Let $\{X_i\}$ be i.i.d. random variables with mean μ, and let $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$. Then S_n converges to μ almost surely. That is,*

$$\mathbb{P}\left(\lim_{n\to\infty} S_n = \mu\right) = 1$$

1.29 theorem (Central Limit Theorem). *Let $\{X_i\}$ be i.i.d. random variables with mean μ and finite variance σ^2, and let $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$. Then*

$$\sqrt{n}(S_n - \mu) \to \mathcal{N}(0, \sigma^2)$$

as $n \to \infty$. That is, as $n \to \infty$, the cumulative distribution function of $\sqrt{n}(S_n - \mu)$ converges pointwise to the CDF of the normal distribution centered at 0 with variance σ^2.

2 Measure Theory

2.1 *(Sigma) Algebras*

2.1 definition. A collection C of subsets of E (the universe) is called an algebra if:

- $\emptyset \in C$.
- $A \in C \implies A^c \in C$.

- $A, B \in C \implies A \cup B \in C$.

C is a σ-algebra if:

- $\emptyset \in C$.
- $A \in C \implies A^c \in C$.
- $A_1, A_2, \cdots \in C \implies \bigcup_{i=1}^{\infty} A_i \in C$.

Noticed the strengthened version of property three.

Examples:

- $\mathcal{E} = \{\emptyset, E\}$ trivial σ-algebra.
- $\mathcal{E} = 2^E$, the discrete σ-algebra.

2.2 THEOREM (Intersections and Unions of σ-Algebras).
- *Any (countable or uncountable) intersection of σ-algebras is a σ-algebra.*
- *The union of two σ-algebras is not necessarily a σ-algebra.*

2.3 DEFINITION. Generated σ-algebra] Let $\mathcal{C}$ be a collection of subsets of E. Take all σ-algebras that contain $\mathcal{C}$. Take their intersection. This σ-algebra is called the σ-algebra generated by $\mathcal{C}$, and is denoted by $\sigma(\mathcal{C})$.

2.4 DEFINITION (Borel Algebras and Borel Sets). If E is a topological space, and $\mathcal{C}$ is the collection of all open sets of E, then $\sigma(\mathcal{C})$ is called the Borel σ-algebra. Its elements are called Borel sets. The Borel σ-algebra is denoted by $\mathcal{B}_E$ or $\mathcal{B}(E)$.

2.5 DEFINITION (Measurable Spaces and Measurable Sets). A pair $(E, \mathcal{E})$ is a measurable space if $\mathcal{E}$ is a σ-algebra on E. The sets in $\mathcal{E}$ are called measurable sets.

2.6 DEFINITION (Measurable Rectangles). Let $(E, \mathcal{E})$, $(F, \mathcal{F})$ are two measurable spaces. If $A \subset E$ and $B \subset F$ are measurable sets, then $A \times B$ is called a measurable rectangle.

2.7 DEFINITION (Products of Measurable Spaces). The product $(E \times F, \mathcal{E} \otimes \mathcal{F})$ where $\mathcal{E} \otimes \mathcal{F} = \sigma(\{A \times B \mid A \in \mathcal{E}, B \in \mathcal{F}\})$, is a measurable space.

2.2 *Measures*

2.8 DEFINITION (Measure). $\mu : \mathcal{E} \to \mathbb{R}^+$ is a measure on $(E, \mathcal{E})$ if

1. $\mu(\emptyset) = 0$
2. If $A_1, A_2, \cdots \in \mathcal{E}$ are pairwise disjoint, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Property (2) is called countable additivity or σ-additivity.

2.9 definition (Probability Measure). A probability measure is a measure μ such that $\mu(E) = 1$.

2.10 definition. A probability space is a triple $(E, \mathcal{E}, \mu)$ such that $(E, \mathcal{E})$ is a measurable space, and μ is a probability measure.

Probability spaces are often denoted $(\Omega, \mathcal{F}, \mathbb{P})$.

Examples of Measures

1. The Dirac Measure. $x \in E$, $\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$
2. The Counting Measure. $D \in E$, $\mu(A) =$# of points in $A \cap D$. If D is countable, then $\mu(A) = \sum_{x \in D} \delta_x(A)$.
3. Discrete Measure. $D \subset E$ countable, $m(x)$ is some real value for every $x \in D$. $\mu(A) = \sum_{x \in D} m(x)\delta_x(A)$.
4. The Uniform Measure on $\{1, 2, \ldots, n\}$. The discrete measure with $m(x) = \frac{1}{n}$.
5. The Lebesgue Measure. $Leb(A) =$ length of A where A is an interval.

2.3 *Measurable Functions*

2.4 *Integration of Measurable Functions*

3 Probability Spaces

3.1 *Expected Value*

3.2 *Almost Sure and Almost Everywhere*

3.3 *Inequalities and Bounds*

3.4 *Borel-Cantelli Lemmas*

3.5 *Law of Large Numbers, Central Limit Theorem*

3.6 *Weak Convergence*

4 Markov Chains