# PROBABILITY THEORY II

Matthew Joseph, notes by Ramdas Singh

Second Semester

# List of Symbols

$\Omega$, a sample space.

$\omega$, an element of a sample space.

$EX$, the expectation of the random variable $X$.

$\mathrm{Var}X$, the variance of the random variable $X$.

$N(\mu, \sigma^2)$, a normal distribution with expectation $\mu$ and variance $\sigma^2$.

# Contents

# Chapter 1

*January 3rd.*

Let $\Omega$ be a countable state space, and let each $\omega \in \Omega$ have a probabiltiy $P(\omega)$ associated with it.

**Lemma 1.1.** *For random variables $X, Y$ such that $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$. Then, $EX \leq EY$.*

*Proof.* This can easily be seen by summing over all terms via the alternate definition of the expectation,

$$EX = \sum_{\omega \in \Omega} X(\omega)P(\omega) \leq \sum_{\omega \in \Omega} Y(\omega)P(\omega) = EY. \tag{1.1}$$

■

We now state Markov's inequality.

**Theorem 1.2** (*Markov's inequality*)**.** *If $X$ is a non-negative randm variable, then for $a > 0$, we have*

$$P(X > a) \leq \frac{EX}{a}. \tag{1.2}$$

*Proof.* Define an indicator function $I_a(\omega)$ as 1 if $X(\omega) \geq a$, and 0 if otherwise. We then have

$$I_a(\omega) \leq \frac{X(\omega)}{a} \implies P(X \geq a) = EI_a \leq \frac{1}{a}EX. \tag{1.3}$$

■

**Remark 1.3.** A better upper bound here may be found by starting with $I_a(\omega)X(\omega)$ instead of just $X(\omega)$.

If we have $X \sim N(0,1)$, then we can find an upper bound for its probability density function.

$$P(X > a) = \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \leq \int_a^\infty \frac{1}{\sqrt{2\pi}} \frac{x}{a} e^{\frac{-x^2}{2}} dx = \frac{e^{\frac{-a^2}{2}}}{\sqrt{2\pi}a}. \tag{1.4}$$

Note that $X$ here is a random variable over a continuous state space; the previous lemma and Markov's inequality also work here. We are to show them for the continuous case instead of the discrete one.

*Proof.* Here, we have $0 \leq X(\omega) \leq Y(\omega)$ for all $\omega$ in our continuous state space $\Omega$. We see that $\{X > x\} \subseteq \{Y > x\} \implies P(X > x) \leq P(Y > x)$. Integrating both sides gives us $EX \leq EY$. ■

**Theorem 1.4** (*Chebyshev's inequality*)**.** *Let $X$ be a random variable with finite mean $\mu = EX$ and finite variance $\sigma^2 = Var(X)$. Then for $a > 0$,*

$$P(|X - \mu| > a) \leq \frac{Var(X)}{a^2}. \tag{1.5}$$

*Proof.* Start with the proof of Markov's inequality, replacing the indiciator function with one that's unity when $|X - \mu| \geq a$. ■

**Example 1.5.** Suppose $X_1, X_2, \ldots, X_n$ are $n$ independent and identically distributed random variables, with $EX_i = \mu$ and $\mathrm{Var}X_i = \sigma^2$. If $S_n = \sum X_i$, we then have

$$P(|S_n - n\mu| > a) \leq \frac{\mathrm{Var}S_n}{a^2} = \frac{n\sigma^2}{a^2}. \tag{1.6}$$

If we replace $a$ with $n^{\frac{1}{2}+\varepsilon}$, we then have

$$P(|S_n - n\mu| > n^{\frac{1}{2}+\varepsilon}) \leq \frac{\sigma^2}{n^{2\varepsilon}} \to 0 \text{ as } n \to \infty. \tag{1.7}$$

**Proposition 1.6.** *If $Var(X) = 0$, then $P(X = EX) = 1$.*

*Proof.* For all $\varepsilon > 0$, we have

$$P(|X - EX| > \varepsilon) \leq \frac{\mathrm{Var}X}{\varepsilon^2} = 0. \tag{1.8}$$

Define $A_n$ as $\{|X - EX| > \frac{1}{n}\}$. Taking $P(\bigcup A_n) = \lim_{n\to\infty} P(A_n)$, the proof follows. ■

## 1.1 The Law of Large Numbers

We start by stating the weak law of large numbers.

**Theorem 1.7** (*Weak law of large numbers*)**.** *Let $\{X_k\}_{k\geq 1}$ be a sequence of independent and identically distributed random variables with $E|X_i| < \infty$. Let $\mu = EX_i$. Then for any $a > 0$,*

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > a\right) = 0. \tag{1.9}$$

*Proof.* For now, let us assume that $\Omega$ is countable. We begin with the case where the variance of $X_i$, $\sigma^2$, is finite. Fix $a > 0$, and let $S_n = X_1 + X_2 + \ldots + X_n$. Then,

$$P\left(\left|\frac{S_n}{n} - \mu\right| > a\right) = P(|S_n - n\mu| > na) \leq \frac{\mathrm{Var}S_n}{n^2a^2} = \frac{n\sigma^2}{n^2a^2} \to 0 \text{ as } n \to \infty. \tag{1.10}$$

We now focus the case when the variance, $\sigma^2$, is infinite. Assume that the expected value, $\mu$, is 0; if it were non-zero, we would then instead work with $X_i - \mu$. Let $\delta > 0$; we shall choose a particular $\delta$ later. For each $n$, define $n$ pairs of random variables, $U_1, V_1, \ldots, U_n, V_n$, as $U_k = X_k, V_k = 0$ if $|X_k| \leq \delta n$, and $U_k = 0, V_k = X_k$ if $|X_k| > \delta n$. $X_k$ can be rewritten as $U_k + V_k$. We then have

$$\{|X_1 + \ldots + X_n| \geq na\} \subseteq \{|U_1 + \ldots + U_n| \geq \frac{na}{2}\} \cup \{|V_1 + \ldots + V_n| \geq \frac{na}{2}\} \tag{1.11}$$

$$\implies P\left(|X_1 + \ldots + X_n| \geq na\right) \leq P\left(|U_1 + \ldots + U_n| \geq \frac{na}{2}\right) + P\left(|V_1 + \ldots + V_n| \geq \frac{na}{2}\right). \tag{1.12}$$

We focus on the first term on the right hand side. The $U_i$'s are independently and identically distributed, so

$$P\left(|U_1 + \ldots + U_n| \geq \frac{na}{2}\right) \leq \frac{4E[|U_1 + \ldots + U_n|^2]}{a^2n^2} = \frac{4}{a^2n^2}\left(\mathrm{Var}(U_1 + \ldots + U_n) + (nEU_i)^2\right). \tag{1.13}$$

For the variance, we have

$$\mathrm{Var}(U_1 + \ldots + U_n) = n\mathrm{Var}U_i \leq nEU_i^2 \leq nE[|U_i|\,|U_i|] \leq \delta n^2 E[|U_i|] \tag{1.14}$$

which transforms the previous equation as

$$P\left(|U_1 + \ldots + U_n| \geq \frac{na}{2}\right) \leq \frac{4}{a^2n^2}\left(\delta n^2 E[|U_i|] + (nEU_i)^2\right). \tag{1.15}$$

A lemma (to be proven later) states that $E[|U_i|] = E[|X_i|]$ as $n \to \infty$, and $EU_i = EX_i = 0$ too. So,

$$P\left(|U_1 + \ldots + U_n| \geq \frac{na}{2}\right) \leq \frac{4}{a^2n^2}\left(\delta n^2 E[|U_i|] + (nEU_i)^2\right) \leq \frac{4\delta E[|U_i|]}{a^2} + \frac{4}{a^2}(EU_i)^2. \tag{1.16}$$

For the second term on the right hand side, begin with

$$P(V_1 + \ldots + V_n \neq 0) \leq P(\{V_1 \neq 0\} \cup \ldots \cup \{V_n \neq 0\}) \leq nP(V_i \neq 0) = n \sum_{|x|>\delta n} P(X_i = x)$$
$$\leq n \sum_{|x|>\delta n} \frac{|x|}{\delta n} P(X_i = x) = \frac{1}{\delta} E[|V_i|]. \tag{1.17}$$

The rightmost term here tends to 0 as $n \to \infty$. Now choose $\delta$ to be $\frac{\varepsilon a^2}{|6E|X_i||}$, and then choose $N$ to be large enough such that for all $n > N$, both the terms are smaller than $\frac{\varepsilon}{2}$. ∎

# Appendices

## Chapter A

# Appendix

Extra content goes here.

Appendix

# Index