CSE 546: Machine Learning **Lecture 1**

The Central Limit Theorem

Instructor: Sham Kakade

1 The Central Limit Theorem

While true under more general conditions, the following is a rather simple proof of the central limit theorem. This proof provides some insight into our theory of large deviations (e.g. how far away a random variable is form its mean).

Recall that $M_X(\lambda) = \mathbb{E}e^{\lambda X}$ is the moment generating function of a random variable X.

Theorem 1.1. *Suppose $X_1, X_2, \dots X_n$ is a sequence of independent, identically distributed (i.i.d.) random variables with mean μ and variance σ_2. Suppose that the $M_X(\lambda)$ exists for all λ in a neighborhood of 0. Let $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Then for all x,*

$$\lim_{n\to\infty} \Pr\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z)$$

where $\Phi(\cdot)$ is the standard normal CDF.

Roughly, this says that, as $n \to \infty$, $\bar{X}_n$ is distributed according to a Gaussian with mean μ and variance σ^2/n.

Proof. Without loss of generality, assume $\mu = 0$. Define $\bar{Z}_n = \frac{\bar{X}_n}{\sigma/\sqrt{n}} = \frac{\sum_i X_i}{\sigma\sqrt{n}}$. By independence and properties of the MGF, we have:

$$M_{\bar{Z}_n}(\lambda) = \mathbb{E}e^{\lambda\frac{\sum_i X_i}{\sigma\sqrt{n}}} = \mathbb{E}e^{\lambda\frac{X_1}{\sigma\sqrt{n}}}\mathbb{E}e^{\lambda\frac{X_2}{\sigma\sqrt{n}}}\dots\mathbb{E}e^{\lambda\frac{X_n}{\sigma\sqrt{n}}} = (M_X(\frac{\lambda}{\sigma\sqrt{n}}))^n$$

where we have used independence of X_i in the first step.

As the moment generating function exists around 0 (and the derivatives of the moment generating function are the moments), Taylor's theorem implies:

$$\begin{aligned} M_X(s) &= M_X(0) + M_X'(0)s + \frac{1}{2}M_X''(0)s^2 + \frac{1}{3!}M_X'''(0)s^3 \dots \\ &= 1 + 0 + \frac{1}{2}M_X''(0)s^2 + o(s^2) \end{aligned}$$

where a function $g(s) = o(s^2)$ if $g(s)/s^2 \to 0$ as $s \to 0$. Hence,

$$M_X(\frac{\lambda}{\sigma\sqrt{n}}) = 1 + \frac{1}{2}\frac{\lambda^2}{n} + o(\frac{\lambda^2}{n})$$

where the last term is with respect to $n \to \infty$. Hence,

$$M_{\bar{Z}_n}(\lambda) = \left(1 + \frac{1}{2}\frac{\lambda^2}{n} + o(\frac{\lambda^2}{n})\right)^n \to \exp(\frac{\lambda^2}{2})$$

Thus the limiting moment generating function of $\bar{Z}_n$ is identical to that of a standard normal (in a neighborhood of 0 for λ). This proves they have identical CDFs (using properties of the MGF). ☐