Handout 22: Chebyshev, Markov, CLT

We finish the semester today by establishing some convergence results. These all have important applications if you continue on in statistics.

Inequalities

Chebyshev's inequality is a simple but quite useful tool in probability, as it applies to any random variable whose variance is finite, and it leads to an immediate proof of a version of a famous result called the "weak law of large numbers" which gives conditions under which the mean of a random sample will converge to the mean of the sampled population as the sample size grows large. Another reason for the impact of Chebyshev's inequality on Statistics is that it showed quite dramatically the utility and relevance of the *variance* as a measure of dispersion. At first view, the quantity σ^2 seems like a reasonable but somewhat arbitrary measure of the "distance" of a random variable from its expected value. Chebyshev's inequality leaves no doubt that this measure says something quite meaningful about the dispersion of a distribution. The inequality applies to both theoretical and empirical distributions, that is, it can be applied when discussing the probability mass function or density of a random variable or a histogram based on discrete or continuous data.

Theorem 1 (Chebyshev's Inequality) *For any random variable X with finite mean μ and variance σ^2, and for any $\epsilon > 0$:*

$$\mathbb{P}\left(|X - \mu| > \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2}$$

Proof. We give a proof in the continuous case. The proof of the discrete case is similar.

$$
\begin{aligned}
\sigma^2 &= \mathbb{E}(X-\mu)^2 \\
&= \int_{\infty}^{\infty} (x-\mu)^2 f(x)dx \\
&\geq \int_{\{x:|x-\mu|\geq\epsilon\}} (x-\mu)^2 f(x)dx \\
&= \int_{\{x:|x-\mu|\geq\epsilon\}} \epsilon^2 f(x)dx \\
&= \epsilon^2 \cdot \int_{\{x:|x-\mu|\geq\epsilon\}} f(x)dx \\
&= \epsilon^2 \cdot \mathbb{P}\left(|X-\mu| > \epsilon\right)
\end{aligned}
$$

Dividing the first and last term of the sequence above by ϵ^2 yields the result ■.

A related result that is often useful in its own right is Markov's Inequality.

Theorem 2 (Markov's Inequality) *For any random variable X and for any $a > 0$,*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}|X|}{a}.$$

Proof. Consider the indicator function of the event $\{|X| \geq a\}$, that is, the function $I_{[a,\infty)}(|X|)$. For any value of a, including $a > 0$ (the case of interest), we have that:

$$a \cdot I_{[a,\infty)}(|X|) \leq |X|.$$

Taking the expected value of both sides, it follows that:

$$a \cdot \mathbb{P}(|X| \geq a) \leq \mathbb{E}|X|.$$

An inequality that can be rewritten, by dividing by a, as the desired result ■.

Chebyshev's inequality turns out to be a very useful tool in studying questions concerning the convergence of a sequence of random variables $\{X_n\}$ to a fixed constant. Such questions are of particular interest in applications in which an unknown parameter is estimated from data. To formalize this, we need a definition of convergence in the context of random variables.

Definition 1 (Convergence in Probability) *Let $\{X_i, i = 1, 2, \ldots\}$ be a sequence of random variables. Then X_n is said to converge in probability to a constant c, denoted $X_n \xrightarrow{p} c$, if for any $\epsilon > 0$:*

$$P\left(|X_n - c| \geq \epsilon\right) \to 0 \, as \, n \to \infty.$$

From this definition, we can now prove a version of the law of large numbers.

Theorem 3 (The Weak Law of Large Numbers) *Let $X_1, X_2, \ldots \overset{iid}{\sim} F$, where F is a distribution with finite mean mu and variance σ^2. Let $\bar{X}_n$ be the mean of the first n random variables. Then $\bar{X}_n \xrightarrow{p} \mu$ as $n \to \infty$.*

Proof. By Chebyshev's inequality:

$$\begin{aligned} P\left(|X_n - c| \geq \epsilon\right) &\leq \frac{\sigma^2_{\bar{X}}}{\epsilon^2} \\ &= \frac{\sigma^2}{n\epsilon^2} \end{aligned}$$

And clearly the last term limits to 0 as $n \to \infty$, completing the result.