ECE 361 Probability for Engineers (Fall, 2016)
Lecture 10a

§5.1 Markov and Chebychev inequalities

Example. Let $X \sim \text{Uni}[0,4]$ with $\mathbb{E}[X] = 2$. Then $\bar{F}_X(x) = 1 - F_X(x) = \mathbb{P}(X > x) = 1 - \frac{x}{4}$, $\mathbb{E}[X] = 2$ and $\text{var}(X) = 4/3$. The Markov inequality asserts:

$$\mathbb{P}(X > x) = 1 - \frac{x}{4} \leq \frac{2}{x}, \ x > 0. \tag{1}$$

The bound here is loose and is in fact trivial for $x \leq 2$. The Chebychev inequality asserts:

$$\mathbb{P}(X \geq x) = \mathbb{P}(X - 2 \geq x - 2) \leq \mathbb{P}(|X - 2| \geq x - 2) \leq \frac{4/3}{(x-2)^2}, \tag{2}$$

which is trivial for $x < \frac{2}{3}(3 + \sqrt{3}) \approx 3.1547$.

Example. Let $X \sim \text{Exp}(1)$ so that $\mathbb{E}[X] = \text{var}(X) = 1$ and thus the Markov inequality is

$$\mathbb{P}(X \geq x) \leq \frac{1}{x}, \tag{3}$$

which is trivial for $x \leq 1$. The Chebychev inequality is

$$\mathbb{P}(X \geq x) = \mathbb{P}(X - 1 \geq x - 1) \leq \mathbb{P}(|X - 1| \geq x - 1) \leq \frac{1}{(x-1)^2}. \tag{4}$$

The Chebychev bound is trivial for $x \leq 2$. Note the actual probability is $\mathbb{P}(X \geq x) = \mathrm{e}^{-x}$.

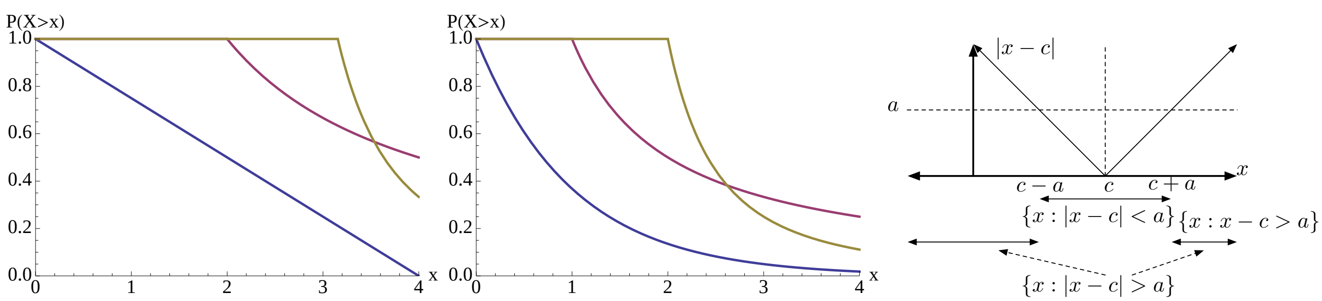

Figure 1: Left: the Markov inequality $2/x$ and Chebychev inequality $(4/3)/(x-2)^2$ for the first example. Center: the Markov inequality $1/x$ and the Chebychev inequality $1/(x-1)^2$ for the second example. Right: illustration of the key inequality $\{x : x - c > a\} \subset \{x : |x - c| > a\}$.

In the previous two examples the actual CDF was known; the value of these inequalities is more evident in cases where the CDF is unknown, as in the following example.

Example. Let $Z \sim N(0,1)$ and obtain bounds on $\mathbb{P}(|Z| > z)$. Observe

$$\begin{aligned} \mathbb{E}[|Z|] &= \int_{-\infty}^{+\infty} |z| f_Z(z) \mathrm{d}z = \cdots = \sqrt{\frac{2}{\pi}} \\ \mathbb{E}[|Z|^2] &= \mathbb{E}[Z^2] = \text{Var}(Z) = 1 \\ \text{Var}(|Z|) &= \mathbb{E}[|Z|^2] - \mathbb{E}[|Z|]^2 = 1 - \left(\sqrt{\frac{2}{\pi}}\right)^2 = 1 - \frac{2}{\pi} \end{aligned} \tag{5}$$

Thus:

$$\begin{aligned}
\bar{F}_{|Z|}(z) &= \mathbb{P}(|Z|>z) \le \frac{\mathbb{E}[|Z|]}{z} = \sqrt{\frac{2}{\pi}} \times \frac{1}{z} \\
\bar{F}_{|Z|}(z) &= \mathbb{P}(|Z|>z) \le \mathbb{P}(||Z|-\mathbb{E}[|Z|]|) > z - \mathbb{E}[|Z|]) \le \frac{\mathrm{Var}(|Z|)}{(z-\mathbb{E}[|Z|])^2} = \frac{1-\frac{2}{\pi}}{\left(z-\sqrt{\frac{2}{\pi}}\right)^2}
\end{aligned} \tag{6}$$

Recall we can express the CDF for $|Z|$ in terms of the CDF for Z, denoted $F_Z(z) = \Phi(z) = \mathbb{P}(Z \le z)$:

$$\bar{F}_{|Z|}(z) = \mathbb{P}(|Z|>z) = \mathbb{P}(Z < -z \cup Z > z) = 2\mathbb{P}(Z>z) = 2(1-\mathbb{P}(Z \le z)) = 2(1-\Phi(z)). \tag{7}$$

All three expressions are shown in the figure below.

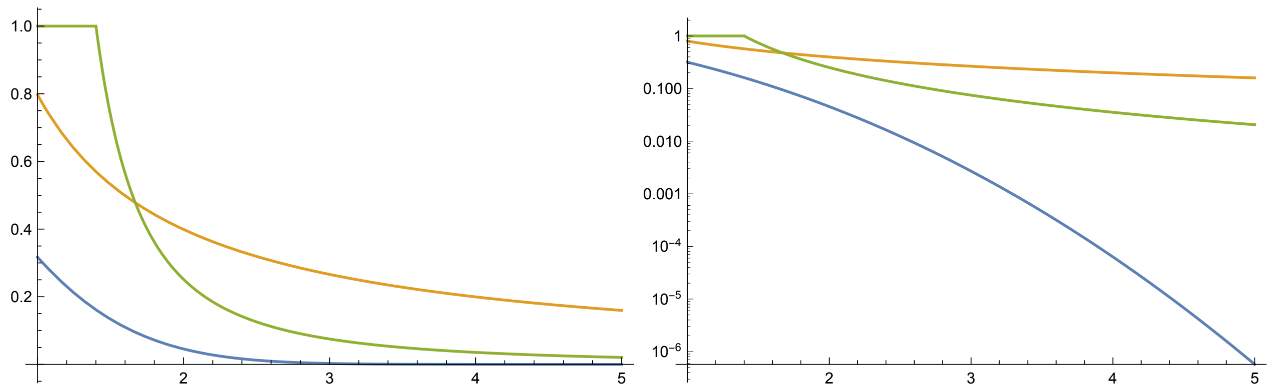

Figure 2: Normal tail probability $\mathbb{P}(|Z|>z)$ for $Z \sim N(0,1)$ (blue), with Markov (yellow) and Chebychev (green) upper bounds, for linear (left) and log (right) y-axes. Both bounds are polynomial decay in z while actual probability decays exponentially in z.

§5.2 The weak law of large numbers

The WLLN asserts the sample average of a collection of iid RVs, each with mean $\mathbb{E}[X]=\mu$ is close to μ with high probability. More precisely, the probability of the sample average deviating from μ by more than an arbitrarily small amount $\epsilon > 0$ goes to zero as the number of RVs, n, grows to infinity.

Theorem 1. *The weak law of large numbers. Let $X_1, X_2, \ldots$ be iid RVs with $\mathbb{E}[X]=\mu<\infty$. Define the sample mean of $\{X_1,\ldots,X_n\}$ as $M_n = \frac{1}{n}\sum_{i=1}^n X_i$. For every $\epsilon > 0$:*

$$\mathbb{P}(|M_n-\mu| \ge \epsilon) = \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| \ge \epsilon\right) \to 0, \quad \textit{as } n \to \infty. \tag{8}$$

Proof. Let $\mathrm{var}(X)=\sigma^2$. Note: $\mathbb{E}[M_n]=\mu$ and

$$\mathrm{var}(M_n) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\mathrm{var}\left(\sum_{i=1}^n X_i\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \tag{9}$$

Apply the Chebychev inequality:

$$\mathbb{P}(|M_n-\mu| \ge \epsilon) \le \frac{\mathrm{var}(M_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty, \ \forall \epsilon > 0. \tag{10}$$

Example. Polling. Let p be the (unknown but fixed) fraction of voters supporting a given candidate. We interview n randomly selected candidates and record M_n, the fraction of voters polled that support the candidate. Viewing M_n as an estimate of p, we wish to identify the required number of voters to be polled to ensure the probability of substantial error in our estimate of p is acceptably small. To this end, first observe that each $X_i \sim \mathrm{Ber}(p)$ with $\mathbb{E}[X_i] = p$ and $\mathrm{var}(X_i) = p(1-p) \leq 1/4$. Then, M_n has $\mathbb{E}[M_n] = p$ and $\mathrm{var}(M_n) = \frac{p(1-p)}{n}$. We upper bound the probability of our estimate being in error exceeding ϵ using the Chebychev inequality and the variance bound:

$$\mathbb{P}(|M_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}. \tag{11}$$

If the probability of error exceeding ϵ had to be smaller than $\delta > 0$, then we require $n_{\min} = n(\epsilon, \delta) = \frac{1}{4\delta\epsilon^2}$ samples. For example, $\epsilon = 1\%$ and $\delta = 5\%$ gives that at least $n_{\min} = 50,000$ citizens must be polled. In fact this is a conservative number since the Chebychev and variance bounds are each conservative.

§5.3 Convergence in probability

In calculus we learn about the limits of sequences of numbers. We say $\{a_n\} \to a$ if for any $\epsilon > 0$ there exists n_ϵ such that $|a_n - a| < \epsilon$ for all $n \geq n_\epsilon$. Intuitively, arbitrarily small deviations ϵ from a never occur for sufficiently large n.

We wish to talk about convergence of RVs, and it turns out there are several natural definitions. The first is *convergence in probability* defined below.

Definition 1. *Let $\{Y_1, Y_2, \ldots\}$ be a sequence of RVs (not necessarily independent) and let $a \in \mathbb{R}$. Say Y_n converges in probability to a if for every $\epsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - a| \geq \epsilon) = 0. \tag{12}$$

Using this definition, the WLLN asserts $M_n \to \mu$ in probability.

Convergence in probability may also be thought of as asserting the probability of Y_n deviating from a by more than an arbitrarily small ϵ happens with an arbitrarily small probability δ for sufficiently large n:

$$\mathbb{P}(|Y_n - a| \geq \epsilon) \leq \delta, \ \forall n \geq n_0(\delta, \epsilon). \tag{13}$$

Call ϵ the accuracy level and δ the confidence level, so that $Y_n \to a$ means that for sufficiently large n, we have arbitrarily high confidence that Y_n is arbitrarily close to a (accuracy).

Example. Let $Y_n = \min\{X_1, \ldots, X_n\}$ where $\{X_1, X_2, \ldots\}$ are independent uniform RVs on $[0,1]$. Claim: $Y_n \to 0$ in probability. Proof:

$$\mathbb{P}(|Y_n - 0| \geq \epsilon) = \mathbb{P}(X_1 \geq \epsilon, \ldots X_n \geq \epsilon) = \mathbb{P}(X_1 \geq \epsilon) \cdots \mathbb{P}(X_n \geq \epsilon) = (1-\epsilon)^n. \tag{14}$$

Thus:

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - 0| \geq \epsilon) = \lim_{n\to\infty} (1-\epsilon)^n = 0. \tag{15}$$

Example. Let $Y \sim \mathrm{Exp}(1)$ and define $Y_n = Y/n$, so that $\{Y_1, Y_2, \ldots\}$ are a sequence of dependent RVs. Claim: $Y_n \to 0$ in probability. Proof:

$$\mathbb{P}(|Y_n - 0| \geq \epsilon) = \mathbb{P}(Y_n \geq \epsilon) = \mathbb{P}(Y \geq n\epsilon) = \mathrm{e}^{-n\epsilon}, \tag{16}$$

so

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - 0| \geq \epsilon) = \lim_{n\to\infty} \mathrm{e}^{-n\epsilon} = 0. \tag{17}$$

We now show that $Y_n \to 0$ in probability does not mean $\mathbb{E}[Y_n] \to 0$.

Example. Define a sequence of discrete RVs $\{Y_1, Y_2, \ldots\}$ where $Y_n = 0$ with probability $1 - 1/n$ or $Y_n = n^2$ with probability $1/n$. Then $\mathbb{E}[Y_n] = n \to \infty$ but $Y_n \to 0$ in probability:

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - 0| \geq \epsilon) = \lim_{n\to\infty} \frac{1}{n} = 0. \tag{18}$$

References

[1] *Introduction to Probability, 2nd Edition* by Dimitri P. Bertsekas and John N. Tsitsiklis, Athina Scientific Press, 2008.