W13 Notes

Deviations

01 Theory - Tail estimation

Every distribution must trail off to zero for large enough $|X|$. The regions where X trails off to zero (large magnitude of X) are informally called 'tails'.

Tail probabilities

A **tail probability** is a probability with one of these forms, for any $c > 0$:

$$P[X \geq c] \qquad P[X \leq -c] \qquad P[|X - \mu_X| \geq c]$$

Markov's inequality

Assume that $X \geq 0$. Take any $c > 0$.

Then the **Markov's inequality** states:

$$P[X \geq c] \quad \leq \quad \frac{\mu_X}{c}$$

Chebyshev's inequality

Take any X, and $c > 0$.

Then the **Chebyshev's inequality** states:

$$P[|X - \mu_X| \geq c] \quad \leq \quad \frac{\sigma_X^2}{c^2}$$

Markov vs. Chebyshev

Chebyshev's inequality works for any X, and it usually gives a better estimate than Markov's inequality.

The main value of Markov's inequality is that it only requires knowledge of μ_X.

Think of Chebyshev's inequality as a tightening of Markov's inequality using the additional data of σ_X.

Derivation of Markov's inequality - Continuous RV

Under the hypothesis that $X \geq 0$ and $c > 0$, we have:

$$\mu_X = E[x] \; = \int_0^\infty x f_X(x)\, dx \quad = \quad \int_0^c x f_X(x)\, dx + \int_c^\infty x f_X(x)\, dx$$

On the range $c \leq x < \infty$ we may convert x to c, making the integrand smaller:

$$\int_c^\infty x f_X(x)\,dx \;\geq\; \int_c^\infty c f_X(x)\,dx$$

Simplify:

$$\int_c^\infty c f_X(x)\,dx \quad \ggg \quad c\int_c^\infty f_X(x)\,dx \quad \ggg \quad cP\big[\,X \geq c\,\big]$$

Also:

$$\int_0^c x f_X(x)\,dx \quad \geq \quad 0$$

Therefore:

$$\int_0^\infty x f_X(x)\,dx \quad \geq \quad cP\big[\,X \geq c\,\big]$$

$$\ggg \quad E[x] \quad \geq \quad cP\big[\,X \geq c\,\big]$$

Extra - Derivation of Chebyshev's inequality

Notice that the variable $(X-\mu_X)^2$ is always positive. Chebyshev's inequality is a simple application of Markov's inequality to this variable.

Specifically, using c^2 as the Markov constant, Markov's inequality yields:

$$P\big[\,(X-\mu_X)^2 \geq c^2\,\big] \quad \leq \quad \frac{E\big[\,(X-\mu_X)^2\,\big]}{c^2}$$

Then, by monotonicity of square roots:

$$(X-\mu_X)^2 \geq c^2 \quad \iff \quad |X-\mu_X| \geq c$$

And of course $E\big[\,(X-\mu_X)^2\,\big] = \sigma_X^2$. Chebyshev's inequality follows.

02 Illustration

Example - Markov and Chebyshev

A tire shop has 500 customers per day on average.

(a) Estimate the odds that more than 700 customers arrive today.

(b) Assume the variance in daily customers is 100. Repeat (a) with this information.

Solution

Write X for the number of daily customers.

(a) Using Markov's inequality with $c = 700$, we have:

$$P\big[\,X \geq 700\,\big] \quad \leq \quad \frac{500}{700} \approx 0.71$$

(b) Using Chebyshev's inequality with $c = 200$, we have:

$$P\big[\,|X-500| \geq 200\,\big] \quad \leq \quad \frac{100}{200^2} \approx 0.0025$$

The Chebyshev estimate is much smaller!

03 Theory - Sample mean

Sample mean and its variance

The **sample mean** of a set $X_1, X_2, \ldots, X_n$ of IID random variables is an RV giving the average value:

$$M_n(X) \;=\; \frac{1}{n}\Big(X_1 + \cdots + X_n\Big)$$

Statistics of the sample mean:

$$E\big[M_n(X)\big] \;=\; E[X_i], \qquad \mathrm{Var}\big[M_n(X)\big] \;=\; \frac{\mathrm{Var}[X_i]}{n} \qquad \text{(for any } i\text{)}$$

The sample mean is typically applied to repeated trials of an experiment. The trials are independent, and the probability distribution of outcomes should be identical from trial to trial.

Notice that the variance of the sample mean limits to 0 as $n \to \infty$. As more trials are repeated, and the average of all results is taken, the fluctuations of this average will shrink toward zero.

As $n \to \infty$ the *distribution* of $M_n(X)$ will converge to a PMF with all the probability concentrated at $E[X_i]$ and none elsewhere.

Large Numbers

04 Theory - Law of Large Numbers

Let $X_1, X_2, \ldots, X_n$ be a collection of IID random variables with $\mu = E[X_i]$ for any i and $\sigma^2 = \mathrm{Var}[X_i]$ for any i.

Recall the sample mean:

$$M_n(X) \;=\; \frac{1}{n}\Big(X_1 + \cdots + X_n\Big),$$

$$E\big[M_n(X)\big] \;=\; \mu_X, \qquad \mathrm{Var}\big[M_n(X)\big] \;=\; \frac{\sigma_X^2}{n}$$

Law of Large Numbers (weak form)

For any $c > 0$, by Chebyshev's inequality we have:

$$P\big[\,\big|M_n(X) - \mu_X\big| \geq c\,\big] \;\leq\; \frac{\sigma_X^2}{nc^2} \qquad \text{(finite LLN)}$$

Therefore:

$$\lim_{n\to\infty} P\big[\,\big|M_n(X) - \mu_X\big| < c\,\big] \;=\; 1 \qquad \text{(infinite LLN)}$$

05 Illustration

Example - LLN: Average winnings

A roulette player bets as follows: he wins \$100 with probability 0.48 and loses \$100 with probability 0.52. The expected winnings after a single round is therefore $100 \cdot 0.48 - 100 \cdot 0.52$ which equals $-\$4$.

By the LLN, if the player plays repeatedly for a long time, he expects to lose \$4 per round on average.

The 'expects' in the last sentence means: the PMF of the cumulative average winnings approaches this PMF:

$$P_{M_n(X)}(k) = \begin{cases} 1 & k = 4 \\ 0 & k \neq 4 \end{cases}$$

This is by contrast to the 'expects' of expected value: the probability of achieving the expected value (or something near) may be low or zero! For example, a single round of this game cannot result in a \$4 loss.

Exercise - Enough samples

Suppose $X_1, X_2, \ldots, X_n$ are IID samples of $X \sim \text{Ber}(0.6)$.

(a) Compute $E[X_i]$ and $\text{Var}[X_i]$ and $\text{Var}\left[M_{100}(X)\right]$.

(b) Use the finite LLN to find α such that:

$$P\Big[\ |M_{100}(X) - 0.6| \geq 0.05\ \Big] \leq \alpha$$

(c) How many samples n are needed to guarantee that:

$$P\Big[\ |M_n(X) - 0.6| \geq 0.1\ \Big] \leq 0.05$$