CS70: Lecture 32.

Inequalities: Markov and Chebyshev

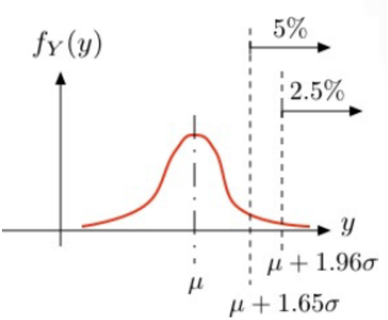
1. Review: Gaussian RV, CLT
2. Inequalities: Markov, Chebyshev
3. Examples
4. Confidence Intervals: Cheybshev Bound

Normal (Gaussian) Distribution.

For any μ and σ, a **normal** (aka **Gaussian**) random variable Y, which we write as $Y = \mathcal{N}(\mu, \sigma^2)$, has pdf

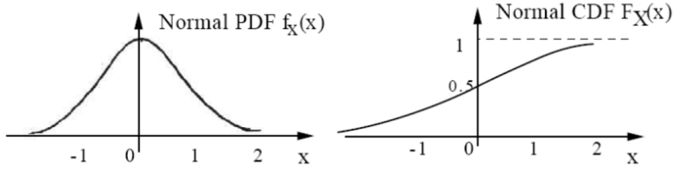
$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}.$$

Standard normal has $\mu = 0$ and $\sigma = 1$.



Note: $Pr[|Y - \mu| > 1.65\sigma] = 10\%$; $Pr[|Y - \mu| > 2\sigma] = 5\%$.

Standard Normal Variable



X is a standard normal variable if

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \quad x \in \Re$$

Since $f(x) = f(-x)$ and $\int_0^\infty x f_X(x) dx$ is finite, $E[X] = 0$.

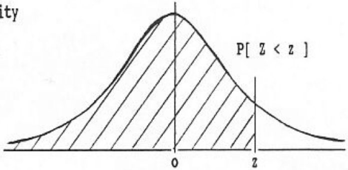
Table

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z i.e.

$$P[Z < z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830

Recap: Crown Jewel of Normal Distribution

Central Limit Theorem

For any set of independent identically distributed (i.i.d.) random variables X_i, define $T_n = \sum X_i$ to be the "total sum" as a function of n.

(and we can define $A_n = \frac{1}{n} \sum X_i$ to be the "running average.")

Suppose the X_i's have expectation $\mu = E(X_i)$ and variance σ^2.

Then the Expectation of T_n is $n\mu$, and its variance is $n\sigma^2$.

Interesting question: What happens to the **distribution** of T_n as n gets large?

Note: We are asking this for **any arbitrary original distribution X_i!**

Review: Central Limit Theorem

Central Limit Theorem

Let $X_1, X_2, \ldots$ be i.i.d. with $E[X_1] = \mu$ and $var(X_1) = \sigma^2$. Define

$$S_n := \frac{T_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

$$E(S_n) = \frac{1}{\sigma\sqrt{n}}(E(T_n) - n\mu) = 0$$

$$Var(S_n) = \frac{1}{\sigma^2 n} Var(T_n) = 1.$$

Then,

$$S_n \rightarrow \mathcal{N}(0,1), \text{as } n \rightarrow \infty.$$

That is,

$$Pr[S_n \leq \alpha] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2} dx.$$

Implications of CLT

1. The Distributions of S_n and M_n wipe out all the information in the original information except for μ and σ^2.
2. If there are large number of small and independent factors, the aggregate of these factors will be normally distributed. E.g. Noise.
3. The Gaussian Distribution is very important – many problems involve sums of iid random variables and the only thing one needs to know is the mean and variance.

Example: Guessing numbers

Alice picks 100 numbers uniformly from $[0,1]$. Bob must guess the sum within 2. What is the probability he is right if he guesses 55?
If X_i is Alice's i^{th} guess, it is difficult to calculate the CDF of $S_{100} = \sum_i X_i$.
$E[S_n] = 50, \mathrm{var}(S_n) = 100/12$.
So by the CLT, $\frac{S_{100}-50}{\sqrt{100/12}}$ is approx $N(0,1)$. Thus

$$P(53 \leq S_{100} \leq 57) \approx \phi(\frac{7}{2.887}) - \phi(\frac{3}{2.887}) = \phi(2.4247) - \phi(1.039)$$

$$= .9925 - .9251 = 0.0649$$

Example: Counting Customers

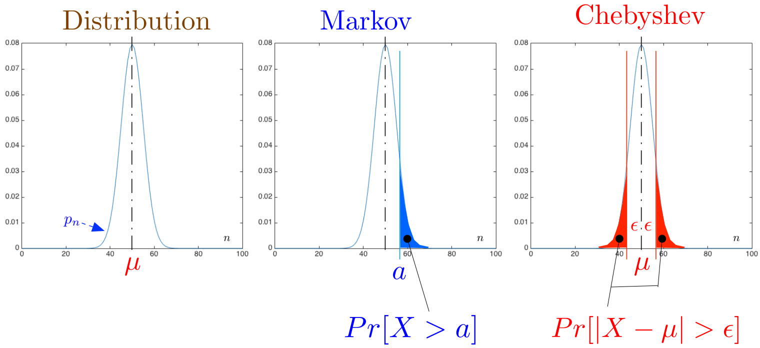
A store has one cashier who can serve one customer at a time.
The time taken to serve a customer $X \sim$ uniform on $[1, 5]$ minutes.
What is the probability that at least 200 customers can be served in 8 hours?
We want $P(\sum_{i=1}^{200} X_i \leq 480)$.
$E[S_{200}] = 600$, $var(S_{200}) = 200\ 16/12 = 800/3$.

$$P(200 \leq \sum_{i=1}^{200} X_i \leq 480) \approx \phi(\frac{480-600}{20\sqrt{2/3}}) - \phi(\frac{-400}{20\sqrt{2/3}})$$

$$= \phi(-6\sqrt{1.5}) - \phi(-20\sqrt{1.5}) = \phi(24.4949) - \phi(7.348).$$

$$P(\sum_{i=1}^{200} X_i \leq 480) \approx 0$$

Inequalities: An Overview



Andrey Markov

Andrey (Andrei) Andreyevich Markov

Born	14 June 1856 N.S. Ryazan, Russian Empire
Died	20 July 1922 (aged 66) Petrograd, Russian SFSR

Andrey Markov is best known for his work on stochastic processes. A primary subject of his research later became known as Markov chains and Markov processes.

Pafnuty Chebyshev was one of his teachers.

Markov was an atheist. In 1912 he protested Leo Tolstoy's excommunication from the Russian Orthodox Church by requesting his own excommunication. The Church complied with his request.

Markov Inequality

If X can only take non-negative values then

$$P(X \geq a) \leq \frac{E[X]}{a}$$

for all $a > 0$.

This inequality makes no assumptions on the existence of variance and so it can't be very strong for typical distributions. In fact, it is quite weak.

Markov Inequality

Example: X is the height of a random adult in Berkeley. If $E[X] = 68$ inches, the Markov Inequality says that

$$P(X > 144) \leq \frac{68}{144} = 0.47$$

On the other hand since it is general, we can try to see what happens to it as we add more assumptions on the distribution. Think of this inequality as being the building block for others...

Markov's inequality (General Form)

The inequality is named after Andrey Markov, although it appeared earlier in the work of Pafnuty Chebyshev. It should be (and is sometimes) called Chebyshev's first inequality.

Theorem Markov's Inequality

Assume $f : \Re \to [0, \infty)$ is nondecreasing. Then,

$$Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}, \text{ for all } a \text{ such that } f(a) > 0.$$

Proof:

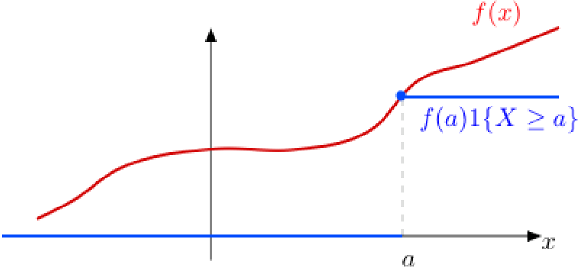
Observe that

$$1\{X \geq a\} \leq \frac{f(X)}{f(a)}.$$

Indeed, if $X < a$, the inequality reads $0 \leq f(X)/f(a)$, which holds since $f(\cdot) \geq 0$. Also, if $X \geq a$, it reads $1 \leq f(X)/f(a)$, which holds since $f(\cdot)$ is nondecreasing.

Taking the expectation yields the inequality, because expectation is monotone. □

A picture



$$f(a)1\{X \geq a\} \leq f(x) \Rightarrow 1\{X \geq a\} \leq \frac{f(X)}{f(a)}$$

$$\Rightarrow Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}$$

Chebyshev's Inequality

This is Pafnuty's inequality:
Theorem:

$$Pr[|X - E[X]| > a] \le \frac{var[X]}{a^2}, \text{ for all } a > 0.$$

Proof: Let $Y = |X - E[X]|$ and $f(y) = y^2$. Then,

$$Pr[Y \ge a] \le \frac{E[f(Y)]}{f(a)} = \frac{var[X]}{a^2}.$$

□

This result confirms that the variance measures the "deviations from the mean."

Chebyshev Inequality

If X is a random variable with finite mean and variance σ^2, then

$$P(|X - E[X]| \geq c) \leq \frac{\sigma^2}{c^2}$$

for all $c > 0$.
Also, letting $c = k\sigma$:

$$P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

Example

Example: X is the height of a random adult in Berkeley. If $E[X] = 68$ inches, and $\sigma_X^2 = 49$, the Chebyshev Inequality says that

$$P(X > 144) \leq P(|X - 68| > 76) \leq \frac{49}{(76)^2} = 0.0084$$

Example: Random Walk

Starting at the origin, I flip a coin 10,000 times: on each flip, if it is heads I step forward; if it is tails I step backward. Prob of heads is 0.5. Estimate the prob that I will be greater than 400 steps away from the origin.

For now let's say the coin is tossed n times. Let $X_i = 1$ if toss i is heads and -1 otherwise.

$E[X_i] = 0$ and $E[X_i^2] = 1$ so $var(X_i) = 1$

Let $X = \sum_i X_i$. $E[X] = 0$ and $var(X) = n$. We want $|X|$ since that's the magnitude of the distance.

Chebyshev says:

$$P(|X| \geq k\sqrt{n}) \leq \frac{1}{k^2}$$

$$P(|X| \geq 100k) \leq \frac{1}{k^2}$$

So, the prob I am more than 400 steps away is less than $\frac{1}{16}$.

Chebyshev Inequality

Example: We want to estimate $E[X]$ by measuring random adults and computing the average height (M_n). How many should we measure to ensure that the estimate is within 1 inch of $E[X]$ with probability 0.99? Assume that $\sigma_X^2 = 36$.
$var(M_n) = \frac{36}{n}$. We want

$$P(|M_n - E[M_n]| \leq 1) = 1 - P(|M_n - E[M_n]| > 1) \geq 0.99$$

$$P(|M_n - E[M_n]| > 1) \leq 0.01$$

$E[M_n] = E[X]$, so Chebyshev tells us that

$$P(|M_n - E[X]| \geq 1) \leq \frac{36}{n}$$

$$\frac{36}{n} \leq .01 \Rightarrow n \geq 3600.$$

In general to have 0.99 confidence that one is within c of the mean: $n \geq \frac{100\sigma^2}{c^2}$.

Fraction of H's

Here is a classical application of Chebyshev's inequality.

How likely is it that the fraction of H's differs from 50%?

Let $X_m = 1$ if the m-th flip of a fair coin is H and $X_m = 0$ otherwise.

Define

$$M_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$

We want to estimate

$$Pr[|M_n - 0.5| \geq 0.1] = Pr[M_n \leq 0.4 \text{ or } M_n \geq 0.6].$$

By Chebyshev,

$$Pr[|M_n - 0.5| \geq 0.1] \leq \frac{var[M_n]}{(0.1)^2} = 100 var[M_n].$$

Now,

$$var[M_n] = \frac{1}{n^2}(var[X_1] + \cdots + var[X_n]) = \frac{1}{n} var[X_1] \leq \frac{1}{4n}.$$

$Var(X_i) = p(1 - lp) \leq (.5)(.5) = \frac{1}{4}$

Fraction of H's

$$M_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$

$$Pr[|M_n - 0.5| \geq 0.1] \leq \frac{25}{n}.$$

For $n = 1,000$, we find that this probability is less than 2.5%.

As $n \to \infty$, this probability goes to zero.

In fact, for any $\varepsilon > 0$, as $n \to \infty$, the probability that the fraction of Hs is within $\varepsilon > 0$ of 50% approaches 1:

$$Pr[|M_n - 0.5| \leq \varepsilon] \to 1.$$

This is an example of the (Weak) Law of Large Numbers.

We will address WLLN next time.

Summary

Inequalities: Markov and Chebyshev

1. Inequalities: Markov and Chebyshev Tail Bounds
2. Confidence Intervals: Chebyshev Bounds