Introduction to Probability
Tail Bounds and Limit Theorems

Dominance of Random Variables

Theorem

Let X and Y be two random variables with $P(X \geq Y) = 1$.

Then $E(X) \geq E(Y)$.

Markov's inequality

Markov's inequality gives an upper bound on the probability of a right tail event for a nonnegative random variable.

For a random variable $X \geq 0$ with $E(X) = \mu$,

$$\textbf{Markov's inequality: } P(X \geq a) \leq \frac{\mu}{a}.$$

Note that Markov's inequality is only useful if $\mu < a$.

Markov's inequality yields very weak bounds, since no distribution information is given about X besides $X \geq 0$.

Markov's inequality

Example

For $X \sim Exp(5)$, Markov's inequality yields a simple upper bound on $P(X \geq 10)$:

$$P(X \geq 10) \leq \frac{E(X)}{10} = \frac{1}{50} = 0.02.$$

Note that the actual probability is

$$P(X \geq 10) = \int_{10}^{\infty} 5e^{-5x} dx = e^{-5(10)} = e^{-50} \approx 1.929 \times 10^{-22},$$

so Markov's inequality does not always yield such good bounds.

Chebyshev's inequality

If X has mean $E(X) = \mu$, then $Y = X - \mu$ has mean $E(Y) = 0$.

If $Var(X) = \sigma^2$, then $Var(Y) = \sigma^2$ as well.

We can use Markov's inequality to get bounds on the probabilty of X being a certain distance away from its mean μ.

Chebyshev's inequality: $P(|X - \mu| \geq a) \leq \dfrac{\sigma^2}{a^2}$.

Again, note that Chebyshev's inequality is only useful if $\sigma < a$.

Chebyshev's inequality

To get a quick upper bound on the probability of being a certain number of *standard deviations* away from the mean of X, use $a = k\sigma$ for some number k:

$$\textbf{Chebyshev's inequality: } P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Chebyshev's inequality example: $Exp(5)$

For $X \sim Exp(5)$, we know that $E(X) = \frac{1}{5}$ and $Var(X) = \sigma^2 = \frac{1}{25}$.

Thus, $SD(X) = \sigma = \frac{1}{5}$.

Hence, Chebyshev's inequality gives a bound on X being too far from $\frac{1}{5}$: letting $k = 2$, we get

$$P\left(\left|X - \frac{1}{5}\right| \geq \frac{2}{5}\right) \leq \frac{1}{4}.$$

The true probability is, since $X \geq 0$,

$$P\left(\left|X - \frac{1}{5}\right| \geq \frac{2}{5}\right) = P\left(X \geq \frac{3}{5}\right) = \int_{3/5}^{\infty} 5e^{-5x}\,dx = e^{-3} \approx 0.0498.$$

Law of Large Numbers

Various **Laws of Large Numbers** states that a sum of IID random variables with finite mean have their **sample average converge to the mean** in various ways (depending on the method of proof).

Weak Law of Large Numbers

Theorem

Weak Law of Large Numbers (WLLN): *If $X_1, X_2, X_3, \ldots$ are IID with*

$$\textit{mean } E(X_1) = \mu \textit{ and variance } Var(X_1) = \sigma^2,$$

then, for any $a > 0$,

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| > a\right) = 0.$$

That is, the probability that the sample average $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ differs from its mean μ, goes to 0 as n increases.

Proof of the Weak Law of Large Numbers

We prove WLLN via Chebyshev's inequality.

Proof Note that, for any n, since the X_j are IID,

$$E\left(\overline{X}_n\right) = \frac{1}{n}\sum_{j=1}^{n} E\left(X_j\right) = \frac{1}{n}(nE(X_1)) = E(X_1) = \mu$$

and

$$Var\left(\overline{X}_n\right) = \frac{1}{n^2}\sum_{j=1}^{n} Var\left(X_j\right) = \frac{1}{n^2}(nVar(X_1)) = \frac{\sigma^2}{n}.$$

Thus, by Chebyshev's inequality, for any $a > 0$,

$$P\left(\left|\overline{X}_n - \mu\right| \geq a\right) \leq \frac{\sigma^2}{na^2} \to 0 \text{ as } n \to \infty. \blacksquare$$

Central Limit Theorem

One of the fundamental theorems of probability theory is the

Central Limit Theorem,

which allows approximation of a sum of IID random variables (with finite mean and variance) by the normal distribution.

Central Limit Theorem

Theorem

Central Limit Theorem (CLT): *If $X_1, X_2, X_3, \ldots$ are IID with*

$$E(X_1) = \mu \text{ and } Var(X_1) = \sigma^2,$$

then the standardized version of their sum,

$$Z_n = \frac{\sum_{j=1}^{n} X_j - n\mu}{\sigma\sqrt{n}},$$

converges in distribution to a standard normal random variable $Z \sim N(0, 1)$ as $n \to \infty$; for any $a \in \mathbb{R}$,

$$\lim_{n\to\infty} F_{Z_n}(a) = \lim_{n\to\infty} P\left(Z_n \leq a\right) = P(Z \leq a) = N(a).$$

Proof of CLT via MGF

Idea of the Proof We can prove that a sequence of random variables W_n converges in distribution to another random variable W's distribution.

We do this by showing that the limit of the MGFs of the sequence, $M_{W_n}(t)$, converges to the limit of the MGF of W, $M_W(t)$.

This works because the MGF, like the CDF, is a way to uniquely identify a distribution.

This is precisely what we'll do to prove the Central Limit Theorem: show that

$$lim_{n\to\infty} M_{Z_n}(t) = M_Z(t).$$

Proof of CLT via MGF

Proof First, the MGF of $Z \sim N(0,1)$ is

$$M_Z(t) = e^{t^2/2}.$$

Next, with

$$Y_j = \frac{X_j - \mu}{\sigma}: \;\; E(Y_j) = 0, \;\; Var(Y_j) = E(Y^2) = 1,$$

the MGF of the standardized sum

$$Z_n = \frac{\sum_{j=1}^n X_j - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{X_j - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{j=1}^n Y_j,$$

of IID standardized random variables Y_j is

$$M_{Z_n}(t) = E(e^{tZ_n}) = E\left(e^{\frac{t}{\sqrt{n}}\sum_{j=1}^n Y_j}\right) = E\left(e^{\frac{t}{\sqrt{n}}Y_1}\right)^n = M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)^n.$$

Proof of CLT via MGF

Writing out the **Taylor expansion** of $M_{Z_n}(t)$ around $t = 0$,

$$M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)^n = \left[1 + E(Y)\frac{t}{\sqrt{n}} + \frac{E(Y^2)}{2}\frac{t^2}{n} + \frac{E(Y^3)}{6}\frac{t^3}{n^{3/2}} + \cdots\right]^n.$$

We know $E(Y) = 0$, $E(Y^2) = 1$, and get the approximation

$$M_{Z_n}(t) = M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)^n = \left[1 + \frac{t^2}{2n} + o(n^{-1}) + \cdots\right]^n,$$

which, by some calculus, limits as $n \to \infty$, to

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + o(n^{-1}) + \cdots\right]^n = e^{t^2/2},$$

which is exactly the MGF of $Z \sim N(0,1)$. $\blacksquare$

Applications of the CLT

We can use the CLT to approximate probabilities for any sum of several IID random variables (assuming finite mean and variance).

We will see a couple examples with continuous and discrete random variables.

Applications of the CLT: Aggregated Waiting Times

Say your daily wait for public transit (train, bus, etc.)
is an exponential random variable with mean 5 minutes.

What is the probability that you spend the equivalent of an entire day each year standing on the platform while commuting?

Applications of the CLT: Aggregated Waiting Times

Let $X_j \sim Exp(\frac{1}{5})$ for day j's wait, and assume each day's wait is independent from the others.

Simplifying the number of days to $n = 300$, what is the (approximate) probability the sum of the X_j is more than 24 hours?

Applications of the CLT: Aggregated Waiting Times

Probabilistically: for IID $X_j \sim Exp(\frac{1}{5})$, what is the probability that

$$\sum_{j=1}^{300} X_j > 24(60) = 1440?$$

For IID $X_j \sim Exp(\frac{1}{5})$, $j = 1, 2, ..., 300$, we can approximate this probability via the CLT with

$$\mu = E(X_1) = 5, \; \sigma^2 = Var(X_1) = 25.$$

Thus, $\sigma = SD(X_1) = 5$ and the standardized version of the sum is approximately a standard normal $Z \sim N(0, 1)$.

Applications of the CLT: Aggregated Waiting Times

Continuity correction is not needed since exponential RVs are continuous.

$$P\left(\sum_{j=1}^{300} X_j > 1440\right) = P\left(\frac{\sum_{j=1}^{300} X_j - 300(5)}{5\sqrt{300}} > \frac{1440 - 300(5)}{5\sqrt{300}}\right)$$

$$\overset{(CLT)}{\approx} P\left(Z > \frac{1440 - 300(5)}{5\sqrt{300}}\right)$$

$$\approx P\left(Z > -\frac{2\sqrt{3}}{5}\right)$$

$$\overset{(Z-table)}{\approx} 1 - N\left(-0.69282\right) \approx 0.7558.$$

Interpretation: 75% chance that, under these assumptions, you spend a full day standing on the platform each year.

Applications of the CLT: Uniform Approximation

Example

A fair 6-sided die is rolled 100 times. Approximate the probability that the sum of the rolls is between 200 and 250.

Die roll $j = 1, 2, 3, ..., 100$ is $X_j \sim Unif(\{1, 2, 3, 4, 5, 6\})$, IID.

Hence,

$$E(X_1) = \frac{7}{2} = 3.5, \;\; Var(X_1) = \frac{35}{12} = 2.91\overline{6},$$

and so $SD(X_1) = \sqrt{\frac{35}{12}} \approx 1.707825$.

Applications of the CLT: Uniform Approximation

By CLT approximation, with $Z \sim N(0,1)$, $Z \approx \frac{\sum_{j=1}^{100} X_j - 100(3.5)}{10(1.707825)}$,

$$P\left(200 \leq \sum_{j=1}^{100} X_j \leq 250\right) \approx P\left(199.5 < \sum_{j=1}^{100} X_j < 250.5\right)$$

$$\approx P\left(-8.8124 \leq Z \leq -5.8261\right) \approx 0,$$

which means that there is practically no chance that the sum with mean of 350 and SD 17 is more than 5 SDs from its mean.

(The sum of this many rolls will, by the Law of Large Numbers, be very close to the mean $n\mu = 100(3.5) = 350$.)

Monte Carlo method

The so-called **Monte Carlo method** of simulation is a cornerstone of probability-based financial, economic, social, physical, chemical, and biological modeling.

It uses the Law of Large Numbers to approximate values based on the relative frequency approach just displayed to approximate more complicated calculations.

Monte Carlo method: approximate π

Example

Approximate π via random sampling.

This can be done by simulating several thousand bivariate uniform random variables in the unit square

$$(X_k, Y_k) \in [0,1] \times [0,1],$$

and then counting the ones below the curve $y = \sqrt{1 - x^2}$.

Monte Carlo method: approximate π

The relative frequency of those points

$$f(n) = \#\left\{k \in \{1, 2, ..., n\} : \;\; Y_k \leq \sqrt{1 - X_k^2}\right\}$$

approximates $\frac{\pi}{4}$ as n grows:

$$\lim_{n\to\infty} \frac{f(n)}{n} = \frac{\pi}{4} \approx 0.78539816339745.$$