Brief introduction to probability
part 2

Last time we defined EX - the expected value of a RV X.

Think of EX as follows: if you "measured" X a lot of times, it would be EX on average.
The precise statement behind that is

Thm: (The weak law of large numbers)
let $X_1, X_2, \ldots$ be independent, identically distributed (i.i.d.) random variables with finite expectation.
let $S_n = X_1 + \ldots + X_n$ & $\mu = EX_1$. Then

$$\frac{S_n}{n} \xrightarrow[\text{in probability}]{P} \mu$$

$$\text{i.e. } \forall \varepsilon > 0,\ P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0.$$

Will prove a weaker version.

Preliminaries

While EX - measures the expected value (mean) of X the variance $Var(X)$ measures how much it varies around its mean $\mu = EX$.

$$Var(X) := E(X-\mu)^2 \underset{\text{check}}{\overset{\uparrow}{=}} E(X^2) - E(X)^2.$$

Exercise: Expectation is linear: $\forall \alpha_1, \ldots, \alpha_n \in \mathbb{R}$

$$E(\alpha_1 X_1 + \ldots + \alpha_n X_n) = \alpha_1 EX_1 + \ldots + \alpha_n EX_n.$$

If $X_1, X_2, \ldots, X_n$ indep, then also have

$$Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n).$$

Hint: 1st show that if X, Y indep, then $E(XY) = EX\,EY$.

Thm: (Markov's inequality)

If $X \geq 0$ has finite expectation, then $\forall a > 0$

$$P(X \geq a) \leq \frac{EX}{a}.$$

Pf: $EX = E(X 1_{X<a} + X 1_{X \geq a}) = E(X 1_{X<a}) + E(X 1_{X \geq a})$

$$\geq E(X 1_{X \geq a}) \geq E(a 1_{X \geq a}) = a E(1_{X \geq a}) = a P(X \geq a)$$

$\triangle$

Cor: (Chebyshev's inequality)

If X has finite variance, then

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2} \quad \forall a > 0.$$

Pf: $P(|X - E(X)| \geq a) = P((X - E(X))^2 \geq a^2) \leq \frac{E((X-EX)^2)}{a^2} = \frac{Var(X)}{a^2}$

$\triangle$

Pf: (of the weak LLN, assuming finite variance $Var(X_i) = C < \infty$).

Note that

$$E\left(\frac{S_n}{n}\right) = \frac{1}{n} E(X_1 + \dots + X_n) = \frac{1}{n}(EX_1 + \dots + EX_n) = \frac{1}{n}(n\mu) = \mu.$$

By Chebyshev's ineq., $\forall \varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{Var\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{\frac{1}{n^2} Var(X_1 + \dots + X_n)}{\varepsilon^2}$$

$$= \frac{\frac{1}{n^2}(Var X_1 + \dots + Var X_n)}{\varepsilon^2} = \frac{\frac{1}{n^2}(n\, Var X_1)}{\varepsilon^2} = \frac{Var(X_1)}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

$\triangle$

A stronger result actually holds

Thm (Strong LLN)

If $X_1, X_2, \dots$ are pairwise indep., identically distributed RVs w/ $\mu = \mathbb{E}X_i$, & $S_n = X_1 + \dots + X_n$, then $\frac{S_n}{n} \xrightarrow[n\to\infty]{} \mu$ almost surely, i.e. $P(\lim_{n\to\infty} \frac{S_n}{n} = \mu) = 1$.

What is the difference between the two LLN?
The key difference is the type of convergence.

Recall that $ES_n = n\mu$, so

$$\frac{S_n}{n} - \mu = \frac{S_n - n\mu}{n} = \frac{S_n - E(S_n)}{n}$$

So LLN says if we center S_n, (i.e. $S_n - \mathbb{E}S_n$) & scale it by n, then the randomness disappears & it goes to 0, so the fluctuations of S_n around its mean $\mathbb{E}S_n$ are of smaller order than n. In fact they are of order $\sqrt{n}$. The Central Limit Theorem makes this precise.

Thm (The central limit theorem CLT)

Suppose $X_1, X_2, \dots$ are iid RVs with finite variances $Var(X_i) = \sigma^2 \in (0,\infty)$. If $S_n = X_1 + \dots + X_n$, & $\mu = \mathbb{E}X_i$, then $\forall\, a < b$

$$P\left(a < \frac{S_n - n\mu}{\sqrt{n}} < b\right) \xrightarrow[n\to\infty]{} P(a \le N(0,\sigma^2) \le b)$$

Say $\frac{S_n - n\mu}{\sqrt{n}}$ cv. to $N(0,\sigma^2)$ in distribution.

Will sketch a proof under stronger assumptions

Preliminaries

Moments

Given a RV X we defined 2 numbers associated with it: EX, $\operatorname{Var}X = EX^2 - (EX)^2$.

The quantity EX^2 is called the second moment of X. More generally EX^k, where $k \in \mathbb{N}$, is called the k'th moment of X.

The expectation & variance alone don't contain enough information to identify the distribution of X, but generally all moments together do. I.e. the list of numbers $EX, EX^2, \ldots$ identify the distribution of X uniquely.

This is the case for example for all the distributions we have looked at.

Given a sequence of constants $c_0, c_1, c_2, \ldots$ a useful way to pack the information contained in them into one object is the generating function of them

$$f(z) := c_0 + c_1 z + c_2 z^2 + \cdots$$

Sometimes it is more useful to use the exponential generating function

$$g(z) := c_0 + c_1 \frac{z}{1!} + c_2 \frac{z^2}{2!} + \cdots + c_n \frac{z^n}{n!} + \cdots$$

The exponential generating fcn has better convergence properties.

In the case of moments we will work with the exponential generating function of the moments

$$M_X(z) := \underbrace{EX^0}_{=1} + (EX) z + (EX^2) \frac{z^2}{2!} + \cdots$$

$$M_X(z) = \sum_{k \geq 0}^{\infty} (\mathbb{E}X^k) \frac{z^k}{k!}.$$

This is called the moment generating function of X. If it exists, it will completely determine the distribution of X. Note that $M_X(z)$ might not exist: for example moments could be infinite or the series might not converge for any non-zero z.

We can rewrite $M_X(z)$ as follows:

$$M_X(z) = \sum_{k \geq 0}^{\infty} (EX^k) \frac{z^k}{k!} = \sum_{k \geq 0}^{\infty} E\left(\frac{X^k z^k}{k!}\right) = E\left(\sum_{k=0}^{\infty} \frac{X^k z^k}{k!}\right) = \mathbb{E}(e^{zX}).$$

linearity of expectation

this is not simply due to linearity since we have an infinite sum

Rmk: You can get the moments of X from its MGF:

$$E(X^n) = \left.\frac{d^n M_X(z)}{dz^n}\right|_{z=0}.$$

Rmk: Often $M_X(z) = \mathbb{E}(e^{zX})$ is used as the defn of MGF.

The MGF can be very useful when showing convergence in distribution.

Thm: (Convergence thm)

Suppose X has a continuous cdf & $M_X(t)$ is finite in $(-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$.

As mentioned before, the MGF determines the distr of X.

Thm: (Uniqueness theorem). Suppose X, Y have cts MGFs which are finite in some interval $(-\varepsilon, \varepsilon)$. If

$$M_X(z) = M_Y(z) \quad \forall z \in (-\varepsilon, \varepsilon),$$ then

X & Y have the same distribution.

If the MGF's of $Y_1, Y_2, \ldots$ satisfy

$$\lim_{n\to\infty} M_{Y_n}(t) = M_X(t) \quad \forall\, t \in (-\varepsilon, \varepsilon), \text{ then}$$

$$Y_n \xrightarrow[n\to\infty]{d} X \quad (Y_n \text{ converges to } X \text{ in distribution as } n \to \infty)$$

I.e $\forall\, a \in \mathbb{R}$

$$\lim_{n\to\infty} P(Y_n \le a) = P(X \le a).$$

We will use this in the sketch of the proof of the CLT. Instead of showing $\frac{S_n - n\mu}{\sqrt{n}}$ cv to $N(0,\sigma^2)$ in distribution

i.e. instead of $P\left(a \le \frac{S_n - n\mu}{\sqrt{n}} \le b\right) \longrightarrow P(a \le N(0,\sigma^2) \le b)$

we will show that $M_{\frac{S_n - n\mu}{\sqrt{n}}}(t) \longrightarrow M_{N(0,\sigma^2)}(t)$.

Rmk: A related fcn, called the characteristic fcn of X is defined by $ch_X(t) := E(e^{itX})$

Unlike the MGF it always exists & the actual pf of the CLT goes through the characteristic fcn.

Rmk: If X has density, then $ch_X(t)$ is the Fourier transform of the density fcn.

Since $S_n = X_1 + \cdots + X_n$, we will need to know how the MGF behaves under sums & also what $M_{N(0,\sigma^2)}(t)$ is.

1) Let $X \sim N(0, \sigma^2)$. What is $M_X(t)$?

$$M_X(t) = E(e^{tX}) = \int_{\mathbb{R}} e^{tx} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}} dx$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\left(\frac{x^2 - 2\sigma^2 t x}{2\sigma^2}\right)} dx$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\left(\frac{(x-\sigma^2 t)^2 - \sigma^4 t^2}{2\sigma^2}\right)} dx$$

$$= e^{\frac{\sigma^2 t^2}{2}} \int_{\mathbb{R}} \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\sigma^2 t)^2}{2\sigma^2}}}_{\text{This is the density of a } N(\sigma^2 t, \sigma^2) \text{ RV, so the integral is 1.}} dx$$

$$= e^{\frac{\sigma^2 t^2}{2}}.$$

2) If $X_1, X_2, \dots, X_n$ are indep, & $S_n = X_1 + \dots + X_n$ then

$$M_{S_n}(t) = E(e^{tS_n}) = E\left(e^{t(X_1 + \dots + X_n)}\right) = E(e^{tX_1} \cdots e^{tX_n}) = E(e^{tX_1}) \cdots E(e^{tX_n})$$

(by independence) $= M_{X_1}(t_1) \cdots M_{X_n}(t_n)$.

Sketch of CLT pf:

Let $Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$

$$M_{Y_n}(t) = E\left(e^{t\frac{S_n - n\mu}{\sigma\sqrt{n}}}\right) = E\left(e^{t\frac{X_1 - \mu}{\sigma\sqrt{n}}} \cdots e^{t\frac{X_n - \mu}{\sigma\sqrt{n}}}\right)$$

$$\underset{\text{(independence)}}{=} \prod_{i=1}^{n} E\left(e^{t\frac{X_i-\mu}{\sigma\sqrt{n}}}\right) \underset{\substack{\text{identically}\\ \text{distributed}}}{=} E\left(e^{\frac{t}{\sigma\sqrt{n}}(X_1-\mu)}\right)^n$$

$$E\left(e^{\frac{t}{\sigma\sqrt{n}}(X_1-\mu)}\right) = E\left(1 + \frac{t}{\sigma\sqrt{n}}(X_1-\mu) + \frac{t^2}{2\sigma^2 n}(X_1-\mu)^2 + \frac{1}{n^{3/2}}\infty + \frac{1}{n^2}\infty + \cdots\right)$$

$$= 1 + \frac{t}{\sigma\sqrt{n}}\underbrace{E(X_1-\mu)}_{0} + \frac{t^2}{2\sigma^2 n}\underbrace{E(X_1-\mu)^2}_{\sigma^2} + \frac{1}{n^{3/2}}\infty + \cdots$$

$$= 1 + \frac{t^2}{2n} + \frac{1}{n^{3/2}}\infty + \cdots$$

so $M_{Y_n}(t) = \left(1 + \frac{t^2}{2n} + \frac{1}{n^{3/2}}\infty + \cdots\right)^n$

Need $\lim_{n\to\infty} M_{Y_n}(t)$. Compute

$$\lim_{n\to\infty} \ln M_{Y_n}(t) = \lim_{n\to\infty} n \underbrace{\ln\left(1 + \frac{t^2}{2n} + \frac{1}{n^{3/2}}\infty + \cdots\right)}_{\approx -\frac{t^2}{2n}} = -\frac{t^2}{2}$$

so $\lim_{n\to\infty} M_{Y_n}(t) = e^{-\frac{t^2}{2}}$

$e^{-\frac{t^2}{2}}$ is the MGF of the standard normal

so by the convergence thm $Y_n \xrightarrow{d} N(0,1)$

i.e. $\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$

$\triangle$

Consider a special case of the CLT

$X_1, X_2, \ldots \sim \text{Bernoulli}(p)$. Think of independent trials where the probability of success is p & failure is $1-p$.

$EX_i = p$

Then $S_n = X_1 + \cdots + X_n$ is the number of successes in n independent trials.

LLN says $S_n \simeq np$ on the leading order &

CLT says the fluctuations are of order $\sqrt{n}$ & Gaussian.

A different regime is when the events are rare, so on average a constant number of them occur / # successes is of constant order.

What is the limit in such a regime?

<u>Thm</u> (Poisson limit theorem)

Let $X_{N,i}$, $1 \le i \le N$ be independent RVs with $X_{N,i} \sim \text{Bernoulli}(p_{N,i})$ & let $S_N = \sum_{i=1}^{N} X_{N,i}$.

Suppose that as $N \to \infty$

1) $\max_{1 \le i \le N} p_{N,i} \to 0$ (the successes are rare)

2) $ES_N = \sum_{i=1}^{N} p_{N,i} \to \lambda < \infty$ (on average have λ successes)

Then $S_N \xrightarrow{d} \text{Poisson}(\lambda)$ as $N \to \infty$.

(If have a large number of independent "rare events" & on average λ occur, then the number that occur is $\sim \text{Poisson}(\lambda)$.)

The proof structure is the same as that of the CLT.

Use characteristic fcns.