جلسه‌ی شانزدهم آمار و احتمال مهندسی «دکتر مرتضی زارعی - دانشگاه شریف»

اگر تعداد زیادی X_i رو با هم جمع بزنیم $X_1+X_2+\dots+X_n$ یعنی $\sum_{i=1}^{n}$ و تمام X_i ها لزوما مستقل باشند و همه از یک توزیع یکسان پیروی کنند. آن گاه حاصل جمع آن‌ها از یک توزیع نرمال پیروی خواهد کرد. به شرطی که $n \to \infty$

سوال اینه که چرا برای هر توزیعی این اتفاق میوفته و تبدیل میشه به یک توزیع نرمال؟ و σ و μ اون توزیع نرمال حاصل شده، چه چیزی خواهند بود؟

قضیه‌ی حد مرکزی

$X_1 \dots X_n$

$n \to \infty$

$\forall i \leq n : X_i \sim$ توزیع یکسان

$$\Rightarrow \quad E\left(\sum_{i=1}^{n} X_i\right) = n\mu, \qquad Var\left(\sum X_i\right) = n\sigma^2$$

متغیر تصادفی خودمون رو نرمال سازی می‌کنیم.

$$Z = \frac{X-\mu}{\sigma} \Rightarrow \frac{\sum X_i - n\mu}{\sqrt{n}\,\sigma}$$

و قضیه‌ی حد مرکزی بیان میکنه که μ ، σ در این حالت ۰ و ۱ خواهند بود.

$$\forall a \in \mathbb{R} : \mathbb{P}\left(\frac{\sum X_i - n\mu}{\sqrt{n}\,\sigma} \leq a\right) \longrightarrow \Phi(a)$$

این قضیه را اثبات نمی‌کنیم چون خیلی تکنیکه (ریاضیاتی) رو به عهده‌ی خواننده میذاریم.

سوال: اگر که توزیع‌های X_i ها یکسان نباشند چی؟ اون موقع میتونه به هر چیزی تبدیل بشه و لزوماً نرمال نباشه.
اگر که هم وابسته باشن چی؟ در این حالت هم به توزیع نرمال نخواهیم رسید.

مثال: فاصله‌ی یک ستاره تا ما توسط یک ستاره‌شناس به دفعات اندازه گیری میشه. هر بار اندازه گیری یک خطا داره و ستاره‌شناس به دنبال بدست آوردن یک تخمین خوب از فاصله هست. میانگین فاصله‌های اندازه گیری شده μ و واریانس 9 هست. ستاره‌شناس باید چند بار اندازه گیری انجام بده تا با احتمال 95 درصد مطمئن بشه که فاصله تخمین زده شده (میانگین اندازه گیری‌ها) برابر 1 سال نوری هست؟

هر کدوم از X_iها از یک توزیع با میانگین μ، و $\sigma^2 = 9$ پیروی می‌کنند.

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} x_i - n\mu}{3\sqrt{n}} \sim N(0,1)$$

طبق قضیه‌ی حد مرکزی استفاده میکنیم.

میانگین تخمین زده شده $= \mu_e =$ معدل گرفتن از اندازه‌ها $= \dfrac{\sum x_i}{n}$

Error: ماکزیمم برابر ۱ $\rightarrow -1 \le \mu_e - \mu \le 1$

$$P(-1 \le \mu_e - \mu \le 1) = P\left(-1 \le \frac{\sum x_i}{n} - \mu \le 1\right) = P\left(-\frac{\sqrt{n}}{3} \le Z \le \frac{\sqrt{n}}{3}\right)$$

$$\approx \phi\left(\frac{\sqrt{n}}{3}\right) - \phi\left(-\frac{\sqrt{n}}{3}\right) = 2\phi\left(\frac{\sqrt{n}}{3}\right) - 1 = \underbrace{0.95}_{\text{خواسته‌ی سوال}} \Rightarrow \phi\left(\frac{\sqrt{n}}{3}\right) = 0.975$$

$$\Rightarrow \frac{\sqrt{n}}{3} = 1.96 \Rightarrow n = 34.57$$

یعنی اگر 34.57 بار آزمایش را تکرار کنیم، با احتمال 95 درصد خطای محاسبه‌ی ما کمتر از یک سال نوری خواهد بود.

نکته: اصولاً طبق قضیه‌ی حد مرکزی باید تعداد n ها بیشتر از $\underline{30}$ باشد.

در عمل خیلی از کارخونه‌ها میان مثلاً میانگین عمر یه تعداد محصول رو اندازه گیری می‌کنن و با استفاده از روش‌های مختلف واریانس رو هم بدست میارن. می‌خوان بدونن چندتا لامپ رو میانگین بگیرن با دقت خوبی به μ طول عمر کل کالاها می‌رسن. و این رو هزینه‌ی بدست آوردن μ طول عمر تأثیر داره. اینکه چندتا کالا رو خرابی کنیم مهم و واقعاً ...