$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} \qquad \lim_{n \to \infty} \underbrace{P(|\overline{X}_n - \mu| > \varepsilon)}_{\le \frac{\sigma^2}{n \cdot \varepsilon^2}} = 0$$

Př: Měření hmotnosti má chybu $\sim N(0, \sigma^2)$, $\sigma = 1g$ (předpokládáme z. měření)

Chceme $\overline{X}_n$ s chybou $< 0{,}1g$ s pstí 95% → Moc polní věc

Postup 1: Použití slabého zákona velkých čísel:

Chceme $\varepsilon = 0{,}1$, $\sigma = 1$, n t.ž. $\frac{\sigma^2}{n \cdot \varepsilon^2} < 0{,}05 \rightarrow n > 2000$ ✓

Postup 2: Vlastnosti normálního rozdělení:

$$X_1 + \dots + X_n \sim N(0, \underset{\text{mám n hodnot}}{n} \cdot \sigma^2), \text{ tedy } \overline{X}_n \sim N(0, \frac{\sigma^2}{\underset{\text{dělím n-kem}}{n}})$$

$$P\left(|X_n - 0| < \underbrace{2\sqrt{\frac{\sigma^2}{n}}}_{< 0{,}1}\right) = 0{,}95 \qquad n > 20^2 = 400$$

↳ Tohle díky normálnímu rozdělení

Proč je 1. větší? Protože první postup je obecný a platí pro všechna rozdělení.

Thm: $X_1, X_2 \dots$ n.n.v., všechny stejně rozdělené, s $EX_i = \mu$, rozptylem σ^2

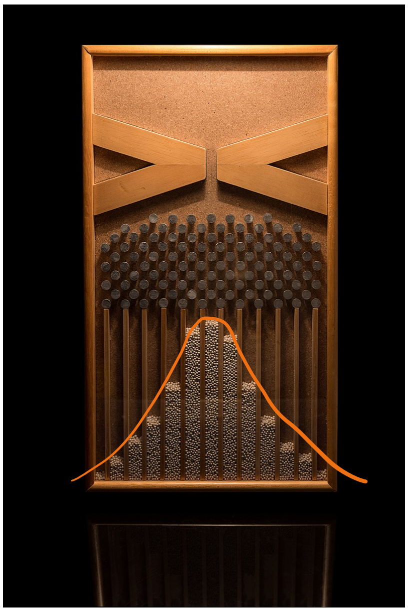
$$Y_n = \frac{(X_1 + \dots + X_n) - n \cdot \mu}{\sigma \sqrt{n}} \quad \text{pak} \quad Y_n \xrightarrow{d} N(0,1),$$

což znamená: $\lim_{n \to \infty} F_{Y_n}(t) = \Phi(t) = \lim_{n \to \infty} P(Y_n \le t) = \Phi(t)$

$Y_n = \text{norm}(X_1 \dots X_n)$, kde normalizace je $\text{norm}(S) = \frac{S - ES}{\sqrt{\text{var}(S)}}$

$\text{norm}(S)$ má $E = 0$

$\text{var} = 1$



Y_{10} popisuje polohu zrnek na konci

$$P(Y_{10} < t) \doteq \Phi(t)$$

$$P(a < Y_{10} < b) = \Phi(b) - \Phi(a)$$

$$\text{Model náhody} \underset{\text{statistika}}{\overset{\text{pravděpodobnost}}{\rightleftarrows}} \text{data}$$

To jsme doposud dělali

Tohle budeme dělat

Populace ... obyvatelé ČR (Ω)

Vzorek ... podmnožina populace

Zajímá nás nějaká vlastnost té populace a její výskyt.

Zejména nás zajímají ty vlastnosti, které se nedají všechny projít výčtem

popisná statistika ... explorativní analýza

mat. statistika .... konfirmační analýza

inferenční statistika

Problémy:

- nejasná otázka
- volba vzorku
- neúplná data
- outliers

chceme reprezentující vzorek, tedy aby číslo vzorku bylo = číslu populace. Tedy podle zákonu o velkých číslech stačí volit náhodně.

Uniformě

IQR - inter quantil range: $Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)$

Jaké úlohy statistika řeší?

- bodové úlohy ... odhad nějakého čísla (průměr něčeho v populaci etc.)
- intervalové odhady ... nalezení (a,b), který obsahuje průměr s velkou pstí
- testování hypotéz ...

  kauzalita vs. korelace

- přirozený postup ... náh. výběr bez opak.
- mat. hezký postup ... —//— s opakováním