En egenskap hos normalfördelningen

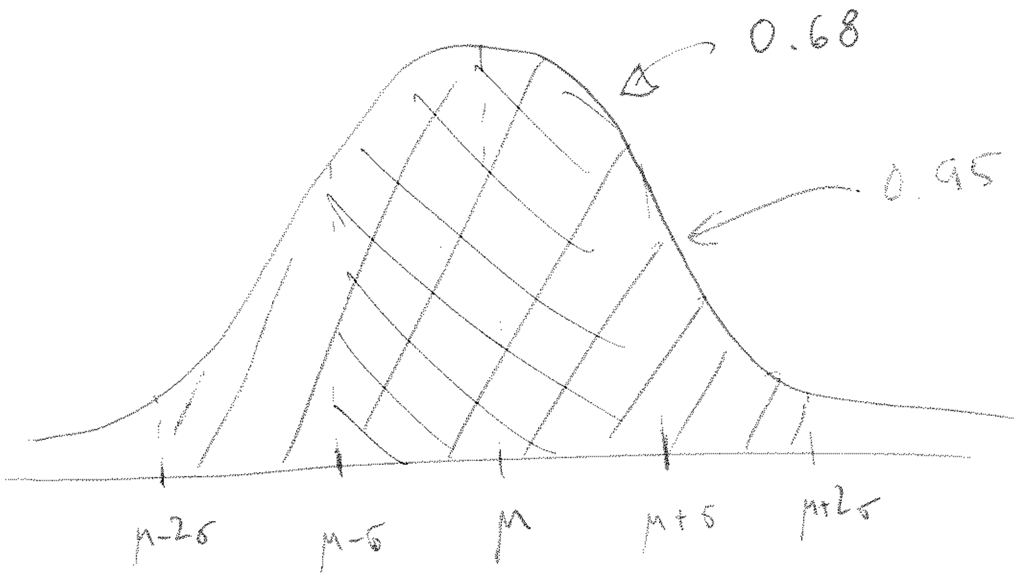
För en $N(\mu, \sigma^2)$-fördelad variabel X gäller att

1) $P(\mu - \sigma < X < \mu + \sigma) \approx 0.68$

2) $P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95$

3) $P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997$

Det innebär att om vi har ett stickprov från en $N(\mu, \sigma^2)$-fördelning så kommer ca 68% av observationerna att hamna i intervallet $[\mu-\sigma, \mu+\sigma]$, ungefär 95% i intervallet $[\mu-2\sigma, \mu+2\sigma]$ och ungefär 99.7% i intervallet $[\mu-3\sigma, \mu+3\sigma]$.



Centrala gränsvärdessatsen (Central limit theorem)

I vissa fall vet vi mer om $\bar{X}$ än bara dess väntevärde och varians.

Sats: Om $X_1, X_2, \ldots, X_n$ är ett stickprov från en ~~normalfördelning~~ $N(\mu, \sigma^2)$-fördelning så gäller att $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Denna sats kommer till användning snart.

En ännu mer användbar sats är följande

Sats (Centrala gränsvärdessatsen) (CGS)

Om $X_1, X_2, \ldots, X_n$ är ett stickprov från någon fördelning (vilken som helst!) med väntevärde μ och varians σ^2 så gäller att

$$\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$$

approximativt

Ju större n desto bättre approximation

CGS är det absolut viktigaste resultatet inom statistik. Om n är stort så är datans ursprung (nästan) irrelevant!

PLOT! (Tärning)

② Oftast vill man också analysera hur stor osäkerhet det finns i våra skattare. PLOT! Vi tar då hjälp av intervallskattningar.

Intervallskattningar (Konfidensintervall)

När man gör konfidensintervall vill man ta fram två ~~punkter~~ tal: en övre gräns U (upper) och en undre gräns L (lower) för intervallet. Vi skriver då [L, U].

Idé: Vi bestämmer oss i förväg för ett procenttal, t ex 95%. Vi vill sen skapa intervallet [L, U] på ett sådant sätt att vi är 95% säkera på att den riktiga parametern täcks av intervallet.

Om det t. ex. är ett väntevärde μ vi vill göra ett konfidensintervall för så vill vi att

$$0.95 = \mathbb{P}(\text{konfidensintervallet täcker } \mu) =$$
$$= \mathbb{P}(L \leq \mu \leq U).$$

Observera att μ är fix (dvs icke slumpmässig) men okänd. Det är L och U som är slumpmässiga. PLOT!

Procenttalet vi väljer (95% ovan) kallas för konfidensgrad och 0.05 ~~är~~ kallas då för signifikansnivå, och betecknas med α.