F7 NORMALFÖRDELNINGEN, LINJÄRKOMBINATIONER, M.M.

Kap. 6.1-6.5 (t.s. 153), 6.7

Linnéa Gustafsson
linneag2@kth.se

- Normalfördelningen
- Linjärkombinationer av oberoende normalfördelade s.v.
- Centrala gränsvärdessatsen

KAP. 5

Några olikheter

$$X = \theta + \delta + \varepsilon$$

uppmätt värde, korrekt värde, systematiskt fel, slumpmässigt fel

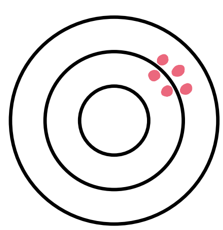
Det vi sysslar med i denna kurs

$$E(X) = E[\theta + \delta + \varepsilon] = \theta + \delta + 0 = \theta + \delta = \mu$$

$$\delta = \mu - \theta$$

$$\varepsilon = X - E(X) = X - \mu$$

Precision och noggrannhet



Detta är ett exempel på god precision och dålig noggrannhet

God precision – litet slumpmässigt fel
God noggrannhet – litet systematiskt fel

Markovs olikhet

$$P(Y \geq a) \leq \frac{E(Y)}{a}$$

Bevis: $E(Y) = \int_0^\infty y \cdot f_Y(y)\,dy = \int_0^a y \cdot f_Y(y)\,dy + \int_a^\infty y \cdot f_Y(y)\,dy \geq$

$$\geq \int_a^\infty y \cdot f_Y(y)\,dy \overset{y \geq a}{\geq} a \underbrace{\int_a^\infty f_Y(y)\,dy}_{P(Y \geq a)} = a \cdot P(Y \geq a)$$

D.v.s. $E(Y) \geq a \cdot P(Y \geq a) \Rightarrow P(Y \geq a) \leq \frac{E(Y)}{a}$ V. S. V.

Stora talens lag

$$\lim_{n\to\infty} P(|\bar{X}_n-\mu|<\varepsilon) \to 1$$

Gör man bara tillräckligt många observationer så kommer medelvärdet att komma hur nära väntevärdet som vi önskar.

Kan även skrivas:

$$\lim_{n\to\infty} P(|\bar{X}_n-\mu|>\varepsilon) \to 0$$

Bevis: $\overbrace{P(|\bar{X}_n-\mu|^2>\varepsilon^2)}^{\to 0} = \left[\text{Markovs olikhet}\right] \le \dfrac{E[|\bar{X}_n-\mu|^2]}{\varepsilon^2} = \dfrac{E[(\bar{X}_n-\mu)^2]}{\varepsilon^2} =$

$$= \frac{V(\bar{X}_n)}{\varepsilon^2} = [V(X_i)=\sigma^2] = \left[V(\bar{X}_n) = \frac{V(X_i)}{n}\right] = \frac{\sigma^2}{n\cdot\varepsilon^2} =$$

$$= [\text{om } n\to\infty] \to 0$$

D.v.s. $\lim_{n\to\infty} P(|\bar{X}_n-\mu|<\varepsilon) \to 1$

$\Rightarrow$ Ju fler mätningar n vi gör desto närmre kommer vi sanningen

"Det mest negativa som kan hända"

Tjebysjovs olikhet

se §7 i F.S.

$$P(|X-\mu|>k\cdot\sigma) \le \frac{1}{k^2}$$

Bevis: $P\left(\overbrace{\dfrac{|X-\mu|^2}{\sigma^2}}^{Y} > \overbrace{k^2}^{a}\right) = [\text{Markov}] \le \dfrac{E[(X-\mu)^2]}{\sigma^2\, k^2} = \dfrac{\sigma^2}{\sigma^2\cdot k^2} = \dfrac{1}{k^2}$ V.S.V.

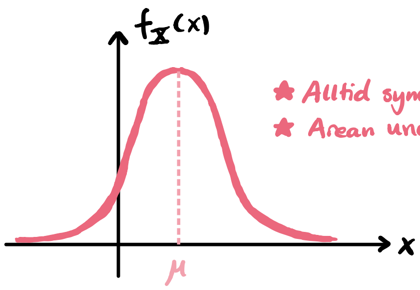
NORMALFÖRDELNINGEN

Världens vanligaste fördelning

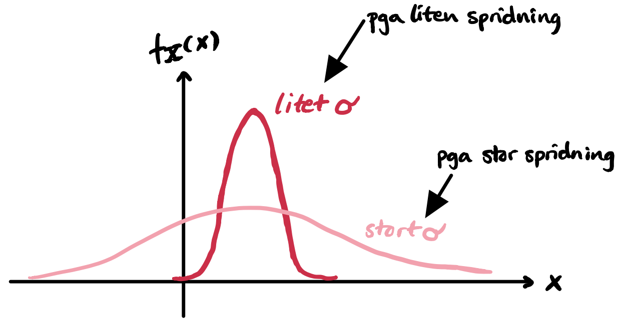
F.S. §4
Tabeller 1 & 2

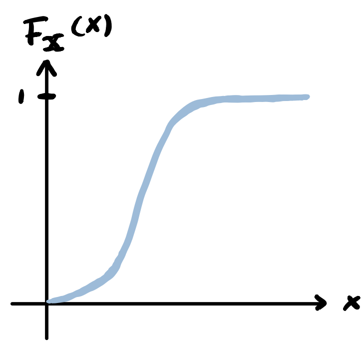
Blir krångligt att integrera
$\Rightarrow$ använder tabeller

$X \in N(\mu, \sigma)$, $\quad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



★ Alltid symmetrisk
★ Arean under alltid = 1





Standardnormalfördelning, N(0,1)

Varje fördelning kan lätt transformeras till standardnormalfördelningen:

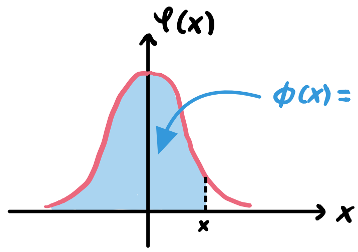
Om $X \in N(\mu, \sigma)$ gäller att $Y = \frac{X-\mu}{\sigma} \in N(0,1)$

Om $Y \in N(0,1)$ så tillhör Y standardnormalfördelningen.

Då kallas $f_Y(y)$ för $\varphi(y)$ och $F_Y(y)$ kallas $\Phi(y)$

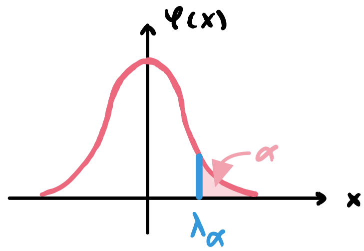
Standardnormalfördelningen finns i tabell 1 och 2.

★ <u>Tabell 1</u> används när gränsen är känd och sannolikheten söks



$\underbrace{\Phi(x) = P(X \le x)}$ finns i tabellen

★ <u>Tabell 2</u> används när sannolikheten är känd och gränsen söks



$P(X > \underbrace{\lambda_\alpha) = \alpha}$

visas parvis i tabell 2

λ_α kallas α-kvantilen

Tabell 1

<u>Ex</u> <u>Sökt</u>: $P(\mu - 2\sigma < X < \mu + 2\sigma)$

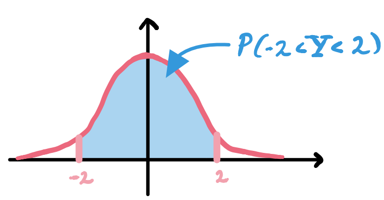
Sannolikheten att vi hamnar högst två standardavvikelser från väntevärdet.

om vi antar att $X \in N(\mu, \sigma)$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = [\text{gör om till } N(0,1)] = P\left(\frac{\mu - 2\sigma - \mu}{\sigma} < \underbrace{\frac{X - \mu}{\sigma}}_{Y \in N(0,1)} < \frac{\mu + 2\sigma - \mu}{\sigma}\right) =$$

$$= \underbrace{P(-2 < Y < 2)} =$$

Har gränser, söker sannolikhet
⇒ tabell 1



$$= P(Y < 2) - P(Y < -2) =$$

$$= \Phi(2.00) - \underbrace{\Phi(-2.00)} = \Phi(2.00) - P(Y > 2) = \Phi(2.00) - [1 - P(Y < 2)] = \Phi(2.00) - [1 - \Phi(2.00)] =$$

För negativa x, utnyttja att $\Phi(-x) = 1 - \Phi(x)$ ← se tabell 1

$$= 2\Phi(2.00) - 1 = 2 \cdot 0.97725 - 1 = \underline{\underline{0.954}}$$

Se sid. 149 (Blom):

- $P(\mu-\sigma < X < \mu+\sigma) = 0.682$
- $P(\mu-2\sigma < X < \mu+2\sigma) = 0.954$
- $P(\mu-3\sigma < X < \mu+3\sigma) = 0.997$

Ex: Upptäckten av Higgspartikeln

$1 - P(\mu-6\sigma < X < \mu+6\sigma) = 2\cdot 10^{-9}$ risk att Higgspartikeln inte fanns

Tabell 2

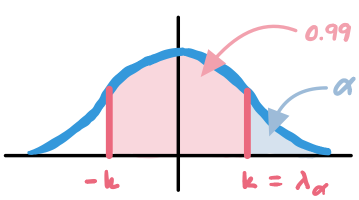
Ex Bestäm k så att $P(\mu-k\cdot\sigma < X < \mu+k\sigma) = 0.99$ då $X \in N(\mu,\sigma)$

Gör om till $N(0,1)$: $P\left(\underbrace{\frac{\mu-k\cdot\sigma-\mu}{\sigma}}_{-k} < \underbrace{\frac{X-\mu}{\sigma}}_{Y} < \underbrace{\frac{\mu+k\cdot\sigma-\mu}{\sigma}}_{k}\right) = 0.99$

$\underbrace{P(-k < Y < k) = 0.99}_{\text{Sannolikheten känd, gränsen söks}}$

$\Rightarrow$ Anpassa till tabell 2

Behöver ha på formen $P(Y > \lambda_\alpha) = \alpha$



$\Rightarrow P(Y > k) = 0.005 \Rightarrow \lambda_\alpha = \lambda_{0.005} = [\text{tabell 2}] = \underline{\underline{2.5758}}$

Används i kap. 12 (konfidensintervall)

Jämför med sid. 149 (Blom):

- $P(\mu-1.96\sigma < X < \mu+1.96\sigma) = 0.95$
- $P(\mu-2.58\sigma < X < \mu+2.58\sigma) = 0.99$
- $P(\mu-3.29\sigma < X < \mu+3.29\sigma) = 0.999$

Ex 6.2 a) $X \in N(28.0,\ 0.25)$. Sökt: $P(27.5 < X < 28.5)$

$$P(27.5 < X < 28.5) = P\left(\frac{27.5-28}{0.25} < \underbrace{\frac{X-28}{0.25}}_{Y \in N(0,1)} < \frac{28.5-28}{0.25}\right) = P(-2 < Y < 2) =$$

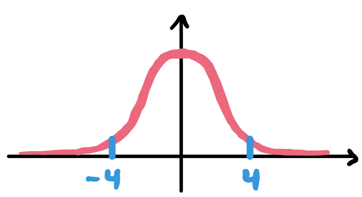
$= [\text{se ovan}] = \underline{\underline{0.954}}$

Se §4: Varje linjärkombination av oberoende N-fördelade stok. var. är N-fördelad.

Ex 6.2 b) Nu görs 4 mätningar och vi bildar $\bar{X} = \frac{X_1 + \ldots + X_4}{4}$ där $X_i \in N(28.0,\ 0.25)$

Sökt: $P(27.5 < \bar{X} < 28.5)$

$$P(27.5 < \bar{X} < 28.5) = \left[\bar{X} \in N\left(28.0,\ \frac{0.25}{\sqrt{4}}\right)\right] = P\left(\frac{27.5-28.0}{0.25/\sqrt{4}} < \underbrace{\frac{\bar{X}-28.0}{0.25/\sqrt{4}}}_{Z \in N(0,1)} < \frac{28.5-28.0}{0.25/\sqrt{4}}\right) =$$

$= P(-4 < Z < 4) =$ [] $= P(Z < 4) - P(Z < -4) =$

$= \Phi(4) - \Phi(-4) = \Phi(4) - (1 - \Phi(4)) = 2 \cdot \Phi(4) - 1 = [\text{tabell 1}] =$

$= 2 \cdot 0.99997 - 1 = 0.99994$

Kärnan i hela den här kursen: Att man ska göra fler mätningar så att det blir bättre

CENTRALA GRÄNSVÄRDESSATSEN ← CGS

F.S. § 5

Antag att n är stort och X_i:na är likafördelade och oberoende och $E(X_i) = \mu$, $D(X_i) = \sigma$

$$\Rightarrow \sum_{i=1}^{n} X_i \underset{\text{approximativt}}{\sim} N(n\cdot\mu, \sigma\cdot\sqrt{n})$$

Hur vet man om n är stort?

* Det cyniska svaret

Fråga examinatorn på tentan om n är stort (men ska vara tydligt).

* Det mer intressanta svaret (hur man ska tänka i verkligheten)

Har med symmetri att göra.

Om man har en ursprungsfunktion som är väldigt symmetrisk behövs det inte alls många s.v. i summan för att summan approximativt ska bli normalfördelad.
(För tärning med 6 sidor, $\frac{1}{6}$ sannolikhet kanske det räcker med 10 kast.)

Har man en sned fördelning (t.ex. lotto P(vinner) = liten stolpe, P(inte vinner) = "jättestor stolpe som står långt ifrån"): kanske inte ens 1 miljon s.v. räcker.

P.g.a. linjärkomb. etc. $\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ (om CGS gäller)

Ex 6.6 Avrundningsfel

X_i = avrundningsfel i kr

$X_i \in U[-0.5, 0.5]$ eg. diskr. men approx. som kont. rektangelfördelad

$Y = X_1 + X_2 + \ldots + X_{48}$

Sökt: $P(|Y| < 6) = P(-6 < Y < 6)$

X_i:na är många, ober., likafördelade $\Rightarrow Y \sim N(n\cdot\mu, \sigma\cdot\sqrt{n})$ enl. CGS

$n = 48$ se §4, $U[a,b]$: $\mu = \frac{a+b}{2} = \frac{-0.5+0.5}{2} = 0$, $\sigma^2 = \frac{(b-a)^2}{12} = \frac{[0.5-(-0.5)]^2}{12} = \frac{1}{12}$

$$\Rightarrow Y \sim N\left(48\cdot 0, \sqrt{\tfrac{1}{12}}\cdot\sqrt{48}\right) = N(0, 2)$$

$$P(-6 < Y < 6) = P\left(\frac{-6-0}{2} < \underbrace{\frac{Y-0}{2}}_{Z \sim N(0,1)} < \frac{6-0}{2}\right) = P(-3 < Z < 3) = \Phi(3) - \Phi(-3) = 2\cdot\Phi(3) - 1 = \underline{\underline{0.997}}$$