



NTNU

Det skapende universitet

**Normalfordelingen =
Gaussfordelingen =**
Den viktigste fordelingen

TMA4245 Statistikk

Gunnar Taraldsen
SINTEF ICT

Kilder

- ▶ **Walpole, Myers, Myers and Ye: Probability and Statistics**

6.2	Normal Distribution	172
6.3	Areas under the Normal Curve	176
6.4	Applications of the Normal Distribution	182

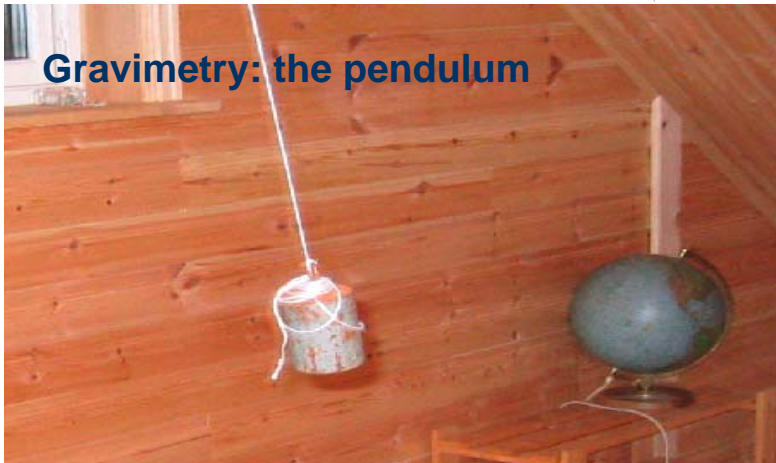
- ▶ Gonick and Smith: THE CARTOON GUIDE TO STATISTICS
- ▶ <http://no.wikipedia.org/wiki/Normalfordeling> og http://en.wikipedia.org/wiki/Normal_distribution
- ▶ Løvås: Statistikk for universiteter og høyskoler
- ▶ Taraldsen: Grunnkurs i statistikk og sannsynlighetsteori

Litt historie

- ▶ Abraham de Moivre (1667-1754) utviklet teorien for normalfordelingen omkring 1773 som en grense av binomialfordelingen.
- ▶ Store deler av teorien for statistisk inferens slik den brukes i dag bygger på eller kan knyttes til normalfordelingen.
- ▶ Normalfordelingen kalles også **Gaussfordelingen** etter Karl Friedrich Gauss (1777-1855) som også studerte normalfordelingen. Han utledet Gaussfordelingen i forbindelse med feilen ved mange repeterte målinger av samme størrelse.

Et pendelforsøk

Gravimetry: the pendulum



Pendelen og tyngdens akselerasjon

- ▶ Tyngdens akselerasjon g er gitt ved

$$g = \left(\frac{2\pi}{T} \right)^2 l$$

hvor T er perioden og l er pendelens lengde.

- ▶ Dermed kan g måles ved å måle T og l .

Anvendelser av normalfordelingen

- ▶ Førstevalg ved modellering av usikkerhet i et eksperiment.
- ▶ En sum av mange uavhengige stokastiske variable er tilnærmet normalfordelt. Dette lærer du mer om i kapittel 6.4 og Sentralgrenseteoremet mer generelt i kapittel 8.
- ▶ Målefeil til et instrument.
- ▶ Statistisk kvalitetskontroll i industrien (kap.17).
- ▶ Statistisk inferens gitt ved estimering (k9), hypotesetesting (k10), lineær regresjon (k11-12, pensum er ca k1-11), ANOVA (Analysis Of Variance, k13-15), og mer generelle lineære SEM (Structural Equations Modeling) og ikke-lineære modeller bygger oftest på normalfordelingen.
- ▶ ... og mye mye mer ...

En viktig anvendelse



GUIDE 98-3

Uncertainty of measurement —

Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)

Incertitude de mesure —

*Partie 3: Guide pour l'expression de l'incertitude de
mesure (GUM:1995)*

BIPM: Bureau International des Poids et Mesures

IEC: International Electrotechnical Commission

IFCC: International Federation of Clinical Chemistry**

ISO: International Organization for Standardization

IUPAC: International Union of Pure and Applied Chemistry**

IUPAP: International Union of Pure and Applied Physics**

OIML: International Organization of Legal Metrology

Standard normalfordeling

- ▶ En stokastisk variabel Z er standard normalfordelt dersom den har en sannsynlighetstetthet

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

- ▶ Det kan vises at

$$E(Z) = 0, \quad \text{Var}(Z) = 1$$

Kumulativ fordeling

- ▶ Den kumulative fordelingen

$$\Phi(z) = P(Z \leq z)$$

er tabulert for $z = -3.79, -3.78, \dots, 3.78, 3.79$ i formelhefte.

- ▶ Dette kan brukes til å beregne sannsynligheter

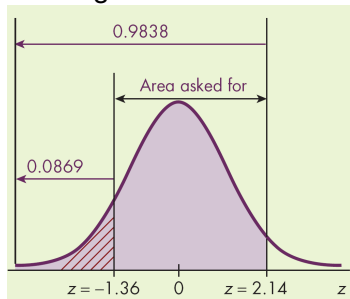
$$\Pr(z_1 < Z < z_2) = \Pr(Z < z_2) - \Pr(Z < z_1) = \Phi(z_2) - \Phi(z_1)$$

- ▶ Det kan og brukes motsatt vei til å beregne z verdier for intervall som skal ha en gitt sannsynlighet.

Areal mellom to grenser I

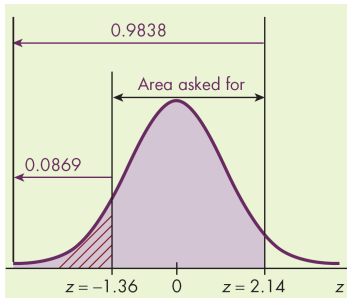
Finn arealet under normalkurven mellom $z = -1.36$ og $z = 2.14$, dvs finn $\Pr(-1.36 < Z < 2.14)$.

Løsning:



- ▶ Kumulativ sannsynlighet = Areal under graf til venstre for en gitt verdi.
- ▶ Dette kan brukes til løsning av alle oppgaver av denne typen.

Areal mellom to grenser II



Arealet til venstre for den øvre grensen, $z = 2.14$, inkluderer både arealet vi søker og arealet til venstre for den nedre grensen $z = -1.36$. Derfor må vi trekke fra det siste arealet.

$$\begin{aligned}\Pr(-1.36 < Z < 2.14) &= \Pr(Z < 2.14) - \Pr(Z < -1.36) \\ &= \Phi(2.14) - \Phi(-1.36) \\ &= 0.9838 - 0.0869 = 0.8969\end{aligned}$$

Normalfordeling

- ▶ La Z være standard normalfordelt. Den stokastiske variabelen

$$X = \mu + \sigma Z$$

er da normalfordelt $N(\mu, \sigma^2)$.

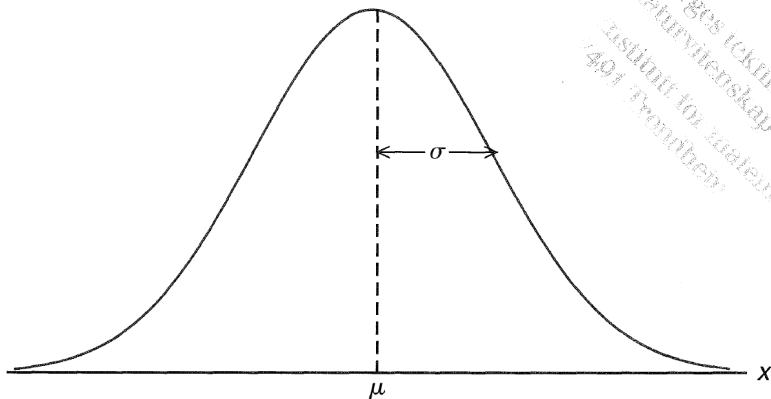
- ▶ Det kan vises at sannsynlighetstettheten er

$$f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- ▶ Det kan vises at

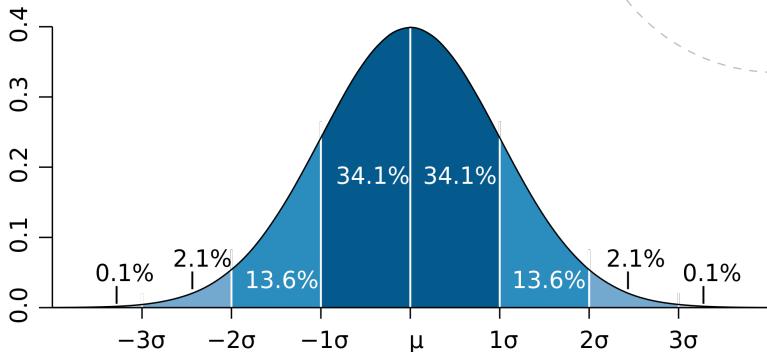
$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2$$

Normalfordelingen $N(\mu, \sigma^2)$



8008 tekni
naturvitenskape
institutt for matem
401 Trondheim

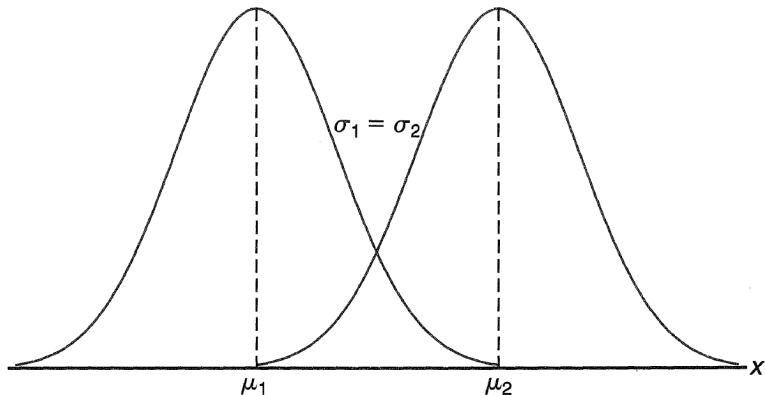
Standardavvik og areal under graf



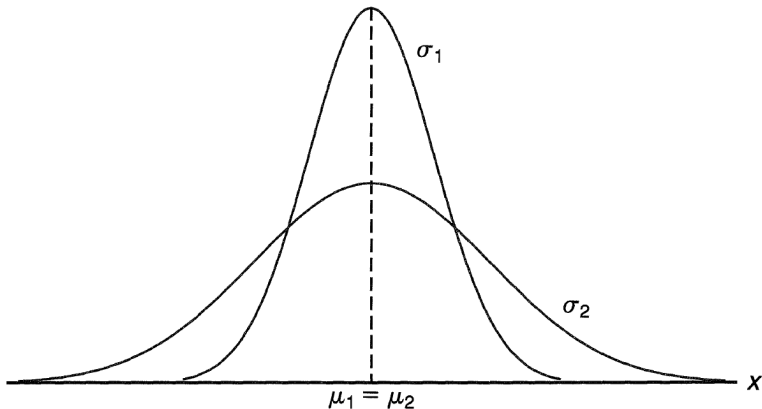
Spesielt vil området $\mu \pm 2\sigma$ inneholde litt mer enn 95% av arealet.

Normalfordelinger med ulik

μ = forventning = lokasjon = posisjon



Normalfordelinger med ulik σ^2 = varians = skala² = spredning

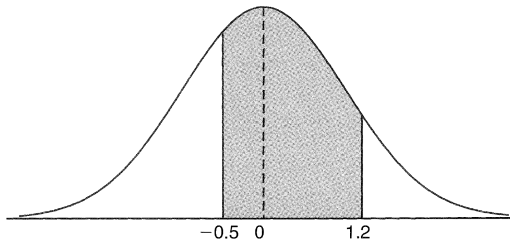


Beregning av sannsynlighet mellom to x verdier

La $X \sim N(50, 10^2)$. Finn sannsynligheten for $45 < X < 62$.

Løsning:

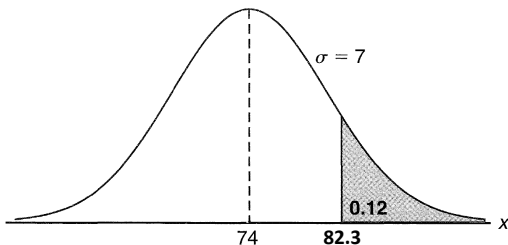
$$\begin{aligned}\Pr(45 < X < 62) &= \Pr(45 < 50 + 10Z < 62) = \Pr(-0.5 < Z < 1.2) \\ &= \Phi(1.2) - \Phi(-0.5) = 0.8849 - 0.3085 = 0.5764\end{aligned}$$



(PS: Trykkfeil i Ex 6.4 figur. Det skal stå z på aksene.)

Beregning av x verdi tilsvarende gitt sannsynlighet

Poengsummen til studenter som tok en eksamen er normalfordelt med middel 74 og standardavvik 7. De beste 12% får karakter A. Hvor mange poeng må til for å få A? Løsning: $\Pr(Z > z) = 0.12$ er det samme som at $\Phi(z) = 0.88$. Tabelloppslag gir løsningen $z = 1.175$, som gir $x = \mu + \sigma z = 74 + 7 \cdot 1.175 \approx 82.3$. De som får 83 poeng eller mer får A.



Viktige teoretiske egenskaper

- ▶ **Sentralgrenseteoremet:** En sum av mange uavhengige variable er tilnærmet normalfordelt.
- ▶ **Delta metoden:** Dersom X er normalfordelt, så er $h(X)$ tilnærmet normalfordelt når $\text{Var}(X)$ er liten.
- ▶ **Lineærkombinasjoner:** En lineærkombinasjon av uavhengige normalfordelte variable er normalfordelt. Mao er $aX + bY$ normalfordelt dersom X og Y er uavhengige og normalfordelte.

Oppsummert fra læreboka

- ▶ Den viktigste sannsynlighetstettheten er normalfordelingen = Gaussfordelingen.



- ▶ Standard normalfordelt variabel Z med tetthet $f(z) = (1/\sqrt{2\pi}) \exp(-z^2/2)$ og normalfordelt variabel $X = \mu + \sigma Z$.
- ▶ Parametrene som bestemmer normalfordelingen $N(\mu, \sigma^2)$ er forventningsverdien $\mu = E(X)$ og variansen $\sigma^2 = \text{Var}(X)$.
- ▶ Begrenset areal under graf = sannsynlighet. Det kan beregnes ved tabulerte verdier for den kumulative fordeling $\Phi(z)$.
- ▶ Den samme tabellen kan brukes til å beregne Z og X verdier tilsvarende spesifiserte sannsynligheter.