Matematisk statistik för B, K, N, BME och Kemister

Föreläsning 2

Johan Lindström

30 augusti 2017

Numerisk beskrivning av data

Medelvärde (Kap. 2.2)

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

Medelvärdet anger **tyngdpunkten** för observationerna.

Varians (Kap. 2.2)

$$s^2 = \frac{1}{n-1}\left((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2) + \ldots + (x_n - \bar{x})^2\right)$$

Grundläggande begrepp (Kap. 3.1)

- **Utfall** – resultatet av ett slumpmässigt försök.
  Bet. $\omega_1, \omega_2, \ldots$
- **Händelse** – en samling av ett eller flera utfall.
  Bet. $A, B, \ldots$
- **Utfallsrum** – mängden av möjliga utfall.
  Bet Ω

Kolmogorovs axiomsystem (Kap. 3.2)

$0 \leq P(A) \leq 1$
En sannolikhet är ett tal mellan 0 och 1

$P(\Omega) = 1$
Sannolikheten att **något** skall hända är 1

$P(A \cup B) = P(A) + P(B)$
Om och endast om A och B är **oförenliga**

Oberoende händelser (Kap. 3.2.4)

Händelserna A och B är **oberoende** av varandra

$$\Longleftrightarrow$$

$$P(A \cap B) = P(A)P(B)$$

Obs. Skilj mellan **oberoende** och **oförenliga**.

Kan två oberoende händelser vara oförenliga?

Läs själva 3.2.3 **Bayes Sats**.

Exempel II

Kasta 3 tärningar vad är sannolikheten att få:

1. Alla (3 stycken) 3:or?
2. Inga 5:or?
3. Minst ett udda (1:a, 3:a, 5:a) nummer?

Stokastisk variabel (Kap. 3.3.1)

En **stokastisk variabel** eller **slumpvariabel** är ett **tal** vars värde styrs av slumpen.
Bet. $X, Y, \ldots$.

Sannolikhetsfunktion (Kap. 3.3.2)

För en **diskret** s.v. X definieras **sannolikhetsfunktionen** som

$$p_X(k) = \mathrm{P}(X = k)$$

Några egenskaper:

- $0 \leq p_X(k) \leq 1$, eftersom det är sannolikheter
- $\mathrm{P}(a \leq X \leq b) = \sum_{k=a}^{b} p_X(k)$
- $\sum_{\text{alla } k} p_X(k) = 1$. Slh att X skall anta **något** värde är 1.

Täthetsfunktion (Kap. 3.3.3)

En **kontinuerlig** s.v X har i stället en **täthetsfunktion** $f_X(x)$.

$$\mathrm{P}(X \in A) = \int_A f_X(x)\,dx$$

Några egenskaper:

- $f_X(x) \geq 0$
- $\mathrm{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx$
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$. Slh att X skall anta **något** värde är 1.

Fördelningsfunktion (Kap. 3.4)

För att räkna ut sannolikheter behöver man summera $p_X(k)$ eller integrera $f_X(x)$. Det kan därför vara användbart att ha en **fördelningsfunktion** (borde heta kumulativ förd.funk.)

$$F_X(x) = \mathrm{P}(X \leq x)$$

Några egenskaper:

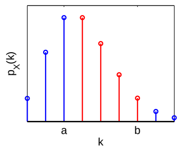
- $0 \leq F_X(x) \leq 1$, eftersom det är en sannolikhet
- $F_X(x)$ är växande.

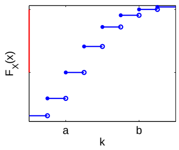
Diskret	Kontinuerlig
$F_X(x) = \sum_{k \leq x} p_X(k)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$
$p_X(k) = F_X(k) - F_X(k-1)$	$f_X(x) = \frac{d}{dx} F_X(x)$

Fördelningsfunktion

Diskret

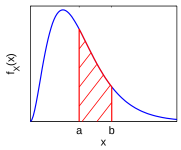
$$P(a < X \leq b) = \sum_{k=a+1}^{b} p_X(k) \qquad P(a < X \leq b) = F_X(b) - F_X(a)$$

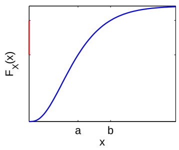




Kontinuerligt

$$P(a < X \leq b) = \int_a^b f_X(x)\,dx \qquad P(a < X \leq b) = F_X(b) - F_X(a)$$





Väntevärde, $E(X)$, μ, μ_X, m, ... (Kap. 3.5)

Väntevärdet anger **tyngdpunkten** för fördelningen och kan tolkas som det värde man får i "**medeltal** i långa loppet".

$$E(X) = \begin{cases} \int_{-\infty}^{\infty} x f_X(x)\,dx & \text{Kont.} \\ \sum_k k p_X(k) & \text{Diskr.} \end{cases}$$

Varians, $V(X)$, σ^2, σ_X^2 (Kap. 3.5)

Variansen anger hur utspridd X är kring sitt väntevärde.

$$V(X) = E\left\{\left[X - E(X)\right]^2\right\} = E(X^2) - E(X)^2$$

Standardavvikelse:, $D(X)$, σ, σ_X $\qquad D(X) = \sqrt{V(X)}$

Exempel

Antag att X har täthetsfunktionen $f_X(x) = \frac{1}{2}e^{-x/2}$, $x \geq 0$.
Bestäm:

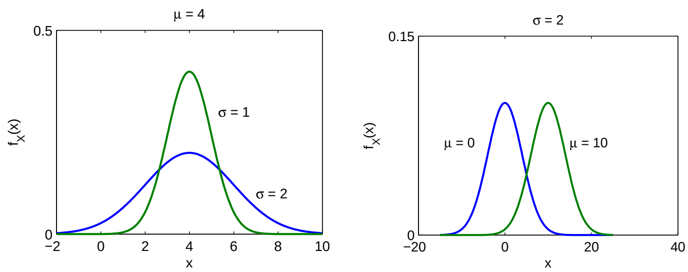
1. Fördelningsfunktionen $F_X(x)$.
2. Sannolikheten $P(2 \leq X \leq 4)$.
3. Väntevärdet $E(X)$.
4. Medianen $x_{0.5}$.

Normalfördelning (Kap. 3.6)

Beteckning: $X \in \mathrm{N}(\mu, \sigma^2)$

Täthetsfunktion:

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

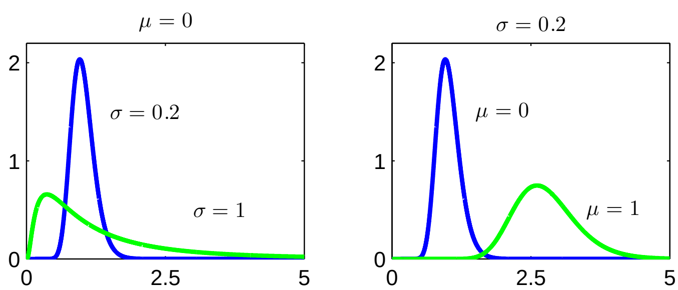


log-Normalfördelning (Kap. 3.6.5)

Beteckning: $X \in \log \mathrm{N}(\mu, \sigma^2)$ $\quad \ln(X) \in \mathrm{N}(\mu, \sigma^2)$

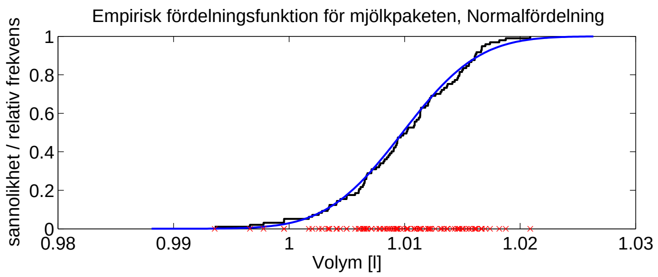
Täthetsfunktion:

$$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}, \quad 0 < x < \infty$$



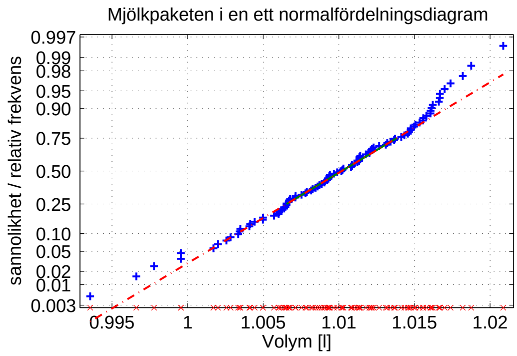
Empirisk fördelningsfunktion

En **empirisk fördelningsfunktion** konstrueras genom att sortera de n mätvärdena och plotta mätvärde i mot i/n. Vid ett givet x-värde kan man då avläsa andelen mätvärden som är mindre än detta x. Denna kan jämföras med en **fördelningsfunktion**.



Matlab: `stairs(sort(x), (1:length(x))/length(x))`

Vanligt är att man skalar om axlarna i den empiriska fördelningsfunktionen så att en given fördelnings fördelningsfunktion blir en rät linje. T.ex en **normalfördelningsplot**. Denna är användbar för att se om datamaterialet passar den givna fördelningen.



Matlab: `normplot(x)`
`qqplot(exprnd(1,1000,1), expinv((1:1e3)/1e3))`

Exempel normplot (Kap. 3.6.4)