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

Föreläsning 3

Johan Lindström

4 september 2017

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.

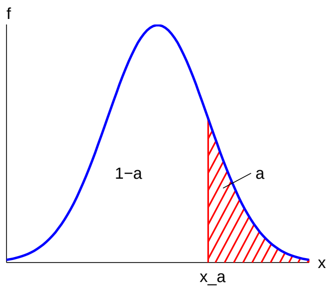
$$V(X) = E\left\{\left[X - E(X)\right]^2\right\} = E(X^2) - E(X)^2$$

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

α-kvantil, x_α (Kap. 3.5 & 3.6.3)

En **kvantil**, x_α, till en s.v. X är en gräns som överskrids med slh α.

$$F_X(x_\alpha) = 1 - \alpha \iff \int_{-\infty}^{x_\alpha} f_X(x)\,dx = 1 - \alpha \iff \int_{x_\alpha}^{\infty} f_X(x)\,dx = \alpha$$

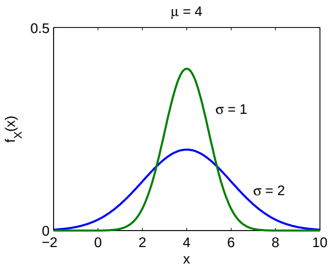


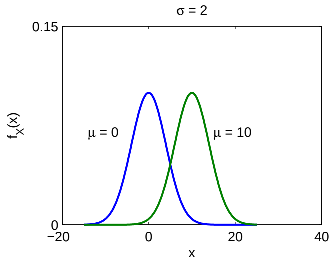
Normalfördelning (Kap. 3.6)

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

Täthetsfunktion:

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





Normalfördelning (Kap. 3.6)

Standardiserad Normalfördelning (Kap. 3.6.2)

Om $X \in \mathrm{N}\left(\mu, \sigma^2\right)$, $E(X) = \mu$, $V(X) = \sigma^2$ så är

$$\frac{X-\mu}{\sigma} \in N(0,1)$$

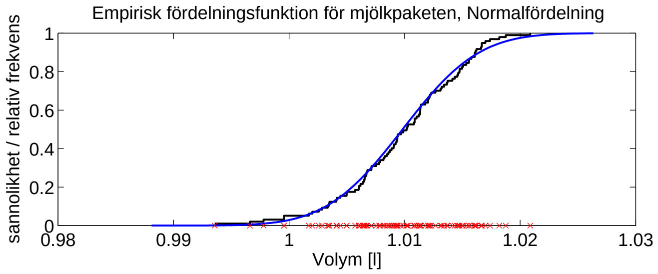
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \; x \in \mathbb{R}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \Phi(\frac{x-\mu}{\sigma}), \; x \in \mathbb{R}$$

$$x_\alpha = \mu + \sigma\lambda_\alpha$$

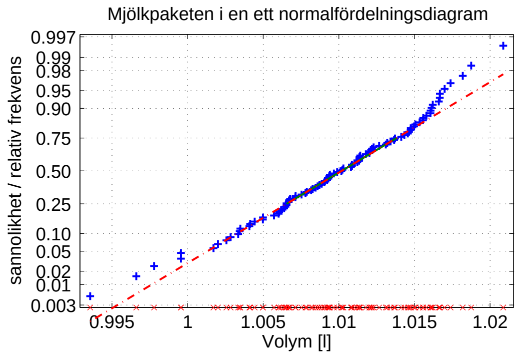
Empirisk fördelningsfunktion

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**.



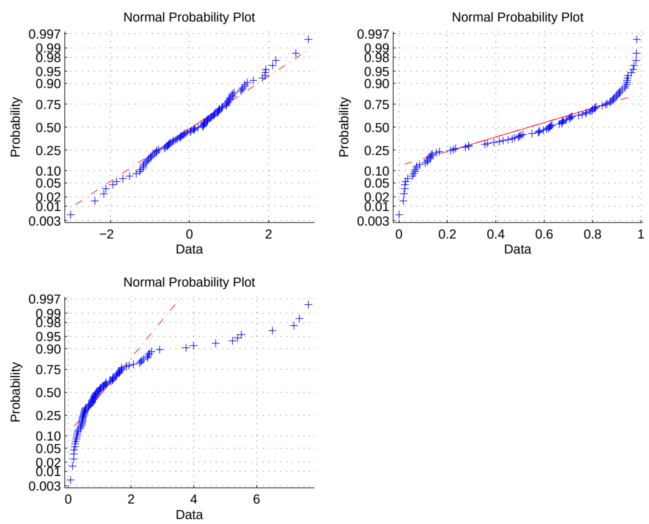
Matlab: `stairs(sort(x), (1:length(x))/length(x))`

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)



Linjärkombination (Kap. 3.5.4)

- $E(aX + b) = aE(X) + b$
- $V(aX + b) = a^2 V(X)$
- $D(aX + b) = |a| D(X)$
- $E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$
- $V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i)$ om oberoende

Specialfall av oberoende och likafördelade s.v. (Kap. 4.4)

Låt $E(X_i) = \mu, \quad V(X_i) = \sigma^2$

Summa: $Y = \sum_{i=1}^{n} X_i$

- $E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu = n\mu$
- $V(Y) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} 1^2 V(X_i) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2$

Medelvärde: $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

- $E(\overline{X}_n) = \sum_{i=1}^{n} \frac{1}{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu$
- $V(\overline{X}_n) = \sum_{i=1}^{n} \frac{1}{n^2} V(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

Exempel: Brädor

Kapa brädor med oberoende längder X_i. $E(X_i) = 1$ m och $V(X_i) = 0.1$ m^2. Bestäm $E(Y)$ och $V(Y)$ om Y ges av

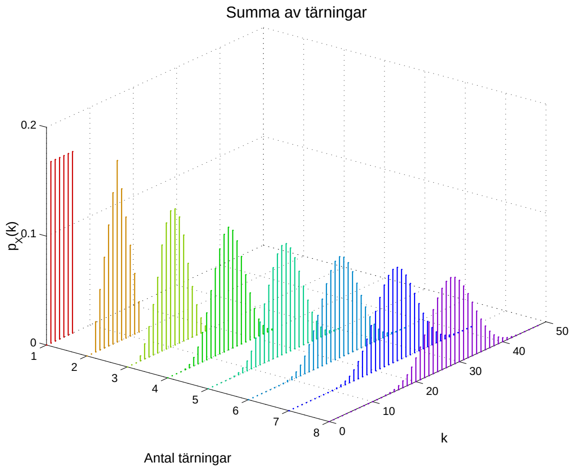
a) Sammanlagda längden av 10 stycken.

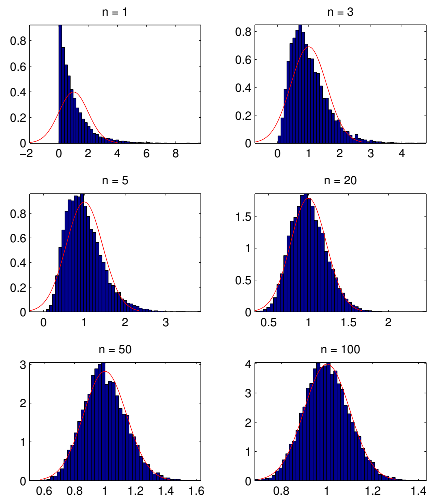
b) Tag en bräda, kapa nio till exakt lika långa.

Centrala gränsvärdessatsen CGS (Kap. 4.5)

Om $X_1, X_2, \ldots, X_n$ är **oberoende likafördelade** stokastiska variabler med $E(X_i) = \mu$, $V(X_i) = \sigma^2$ så är

$$\sum_{i=1}^{n} X_i \underset{\sim}{\in} \mathrm{N}\left(n\mu, n\sigma^2\right) \qquad \text{då } n \text{ stort} \quad (n \to \infty)$$





Medelvärde av $X_i \in Exp(1)$

Äpplen

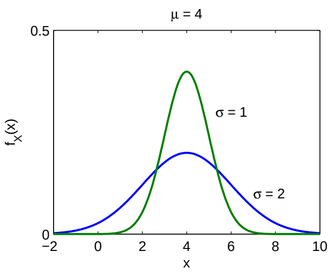
Tag 25 äpplen från ett äppelträd. Låt X_i vara vikten av äpple nr i. Vad är slh att den sammanlagda vikten överstiger $3\,150\ g$ om $E(X_i) = 120\ g$ och $V(X_i) = 400\ g^2$?

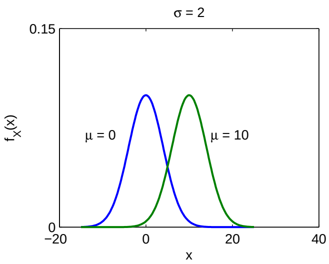
Normalfördelning (Kap. 3.6)

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

Täthetsfunktion:

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



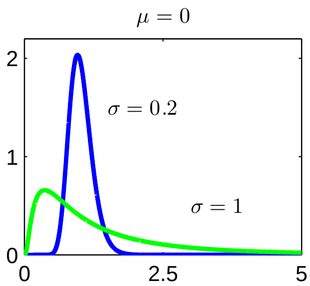


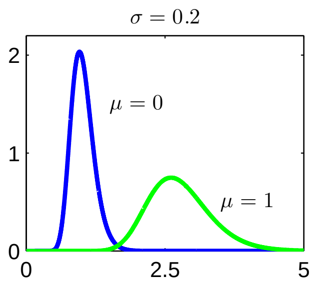
log-Normalfördelning (Kap. 3.6.5)

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

Täthetsfunktion:

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



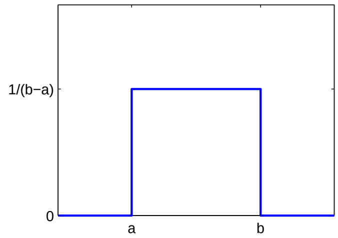


Rektangel- eller likformig fördelning (Kap. 6.2.1)

Beteckning: $X \in R(a, b)$ eller $X \in U(a, b)$ (eng. uniform)
$E(X) = (a + b)/2 \quad V(X) = (b - a)^2/12$

Täthetsfunktion:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & \text{f.ö.} \end{cases}$$

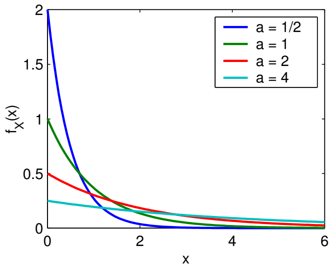


Exponentialfördelning (Kap. 6.2.2)

Beteckning: $X \in Exp(a) \qquad E(X) = a \quad V(X) = a^2$

Täthetsfunktion:

$$f_X(x) = \begin{cases} \frac{1}{a}e^{-x/a}, & x \geq 0 \\ 0 & x < 0 \end{cases}$$



Binomialfördelning (Kap. 6.1.1)

Beteckning: $X \in \text{Bin}(n, p)$

Förekomst: En händelse A med $P(A) = p$ upprepas n oberoende gånger. $X =$ Antalet gånger A inträffar.

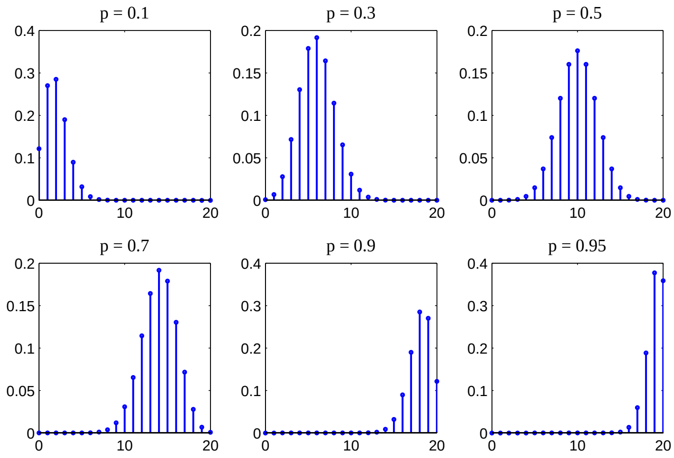
Egenskaper:

$$p_X(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n, \quad q = 1 - p$$

$$E(X) = np, \quad V(X) = npq$$

- Om $X \in Bin(n_1, p)$ och $Y \in Bin(n_2, p)$, ober. så är $X + Y \in Bin(n_1 + n_2, p)$
- Om $npq \geq 10$ är X ungefär normalfördelad.
- Om $n \geq 10$ och $p \leq 0.1$ är X ungefär Poissonfördelad, $X \underset{\sim}{\in} Po(E(X))$.

Binomialfördelning, $X \in Bin(20, p)$



Poissonfördelning (Kap. 6.1.3)

Beteckning: $X \in \mathrm{Po}(\mu)$

Egenskaper:

$$p_X(k) = e^{-\mu} \cdot \frac{\mu^k}{k!} \quad k = 0, 1, \ldots$$
$$E(X) = \mu, \quad V(X) = \mu$$

- Om $X \in Po(\mu_1)$ och $Y \in Po(\mu_2)$, ober. så är $X + Y \in Po(\mu_1 + \mu_2)$
- Om $\mu \geq 15$ är X ungefär normalfördelad.

Poissonfördelning, $X \in Po(\mu)$