Föreläsning 6, Matematisk statistik 7.5hp för E
Normalfördelning

Stas Volkov

Kovarians, $C(X, Y)$

$$C(X,Y) = E\{[X - E(X)][Y - E(Y)]\} = E(XY) - E(X)E(Y)$$

Korrellationskoefficient, ρ, $\rho_{X,Y}$

$$\rho_{X,Y} = \frac{C(X,Y)}{D(X)D(Y)} \in [-1, 1]$$

Räkneregler

$$E\left(\sum a_i X_i + b\right) = \sum a_i E(X_i) + b$$

$$V(aX + b) = a^2 V(X)$$

$$C\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) = \sum_i \sum_j a_i b_j C(X_i, Y_j)$$

$$V\left(\sum_i a_i X_i\right) = \sum_i a_i^2 V(X_i) + 2\underbrace{\sum_{i<j} a_i a_j C(X_i, X_j)}_{=0 \text{ om okorrelerade}}$$

Stora talens lag

Om $X_1, X_2, \ldots, X_n$ är oberoende och likafördelade med $E(X_i) = \mu$ så gäller

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

för alla $\varepsilon > 0$.

Det vill säga medelvärdet konvergerar i sannolikhet mot väntevärdet då n växer mot oändligheten!

Gauss approximationsformler i en variabel

$Y = g(X)$. Taylorutveckla funktionen g kring $\mu = E(X)$

$$g(X) \approx g(\mu) + (X - \mu)g'(\mu) \quad \Longrightarrow$$

- $E(Y) \approx g(E(X))$
- $V(Y) \approx g'[E(X)]^2 V(X)$

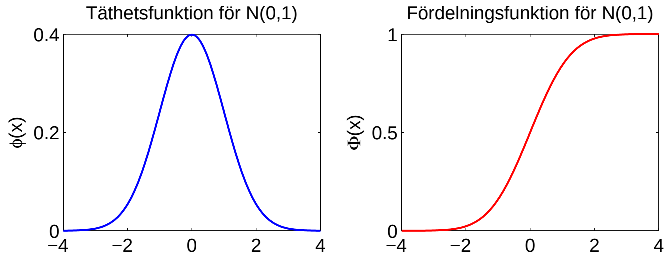
Standardiserad normalfördelning

$$X \in N(0,1),\ E(X)=0,\ V(X)=1,\ x_\alpha \equiv \lambda_\alpha$$

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

$$F_X(x) = \int_{-\infty}^{x} \varphi(t)\,dt \equiv \Phi(x),\ x \in \mathbb{R}$$

$\Phi(x)$ räknas ut numeriskt eller m.h.a. tabell (1).



Allmän normalfördelning

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

Sats 6.1

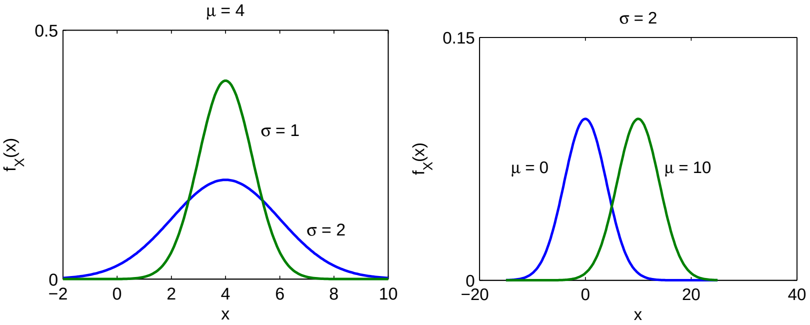
Om $X \in N(\mu,\sigma)$, $E(X) = \mu$, $V(X) = \sigma^2$ så är

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

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

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

Täthetsfunktioner för några normalfördelningar



Exempel

Om $X \in N(0,1)$, beräkna:

1. $P(X \leq 1)$
2. $P(X > -1)$
3. $P(X > 1.96)$
4. $P(-1 \leq X \leq 1.96)$
5. $x_{0.025}$

Om $Y \in N(2,5)$, beräkna:

1. $P(Y \leq 7)$
2. $P(Y > -3)$
3. $P(Y > 11.8)$
4. $P(-3 \leq X \leq 11.8)$
5. $y_{0.025}$

Linjärkombinationer av normalfördelningar

Linjärkombinationer av simultant normalfördelade s.v. är normalfördelade.

Om $X_i \in N(\mu_i, \sigma_i)$ och $Y = \sum_{i=1}^{n} a_i X_i$ gäller

$$Y \in N(E(Y), D(Y))$$

$$\iff$$

$$Y \in N\left(\sum_{i=1}^{n} a_i \mu_i, \sqrt{\sum_{i=1}^{n} a_i^2 \sigma_i^2}\right)$$

... om $X_1, \ldots, X_n$ är oberoende s.v.
(Lägg annars till kovarianserna i variansen.)

Exempel: Beräkna $P(3X_1 - X_2 > 2)$ om $X_i \in N(2,1)$ och oberoende samt om $C(X_1, X_2) = 0.5$.

Centrala gränsvärdessatsen CGS

Låt $X_1, X_2, \ldots, X_n$ vara oberoende stokastiska variabler med samma fördelning och $E(X_i) = \mu$, $V(X_i) = \sigma^2$ (ändliga).
Då gäller att:

$$P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq a\right) \to \Phi(a) \text{ då } n \to \infty \text{ för alla } a$$

Observera att bråket i sannolikheten hela tiden har väntevärde noll och varians ett.

Tillämpning av Centrala gränsvärdessatsen

Summa av oberoende likafördelade s.v. är ungefär normalfördelad om antalet termer är ”tillräckligt stort”. Med $E(X_i) = \mu$, $V(X_i) = \sigma^2$ fås

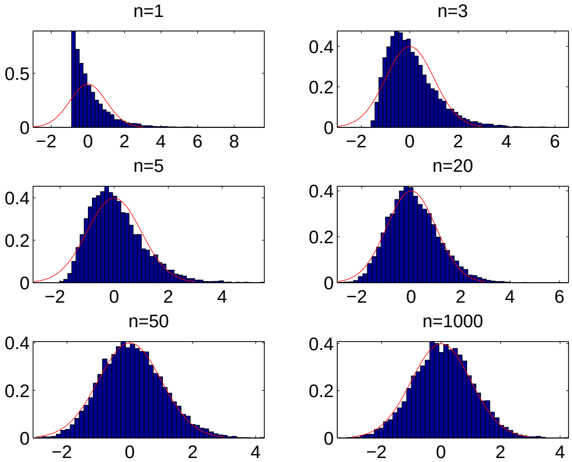
1. Om $Y = \sum_{i=1}^{n} X_i$ gäller
$$Y \underset{\sim}{\in} N(n\mu, \sigma\sqrt{n})$$
2. Om $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ gäller
$$\bar{X}_n \underset{\sim}{\in} N(\mu, \frac{\sigma}{\sqrt{n}})$$

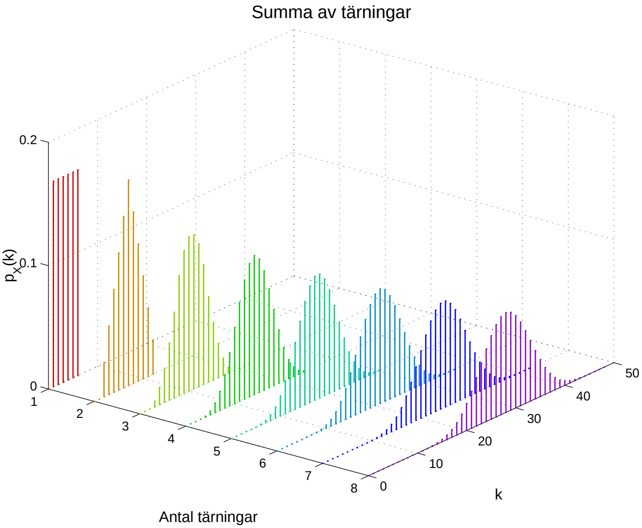
Äpplen

Tag 25 äpplen från ett äppelträd. Låt X_i vara vikten av äpple nr i. Vi vet att
$E(X_i) = 120\ g$ och $V(X_i) = 400\ g^2$.
Vad är sannolikheten att den sammanlagda vikten överstiger $3,15$ kg ?

Hur snabbt konvergerar medelvärdet i stora talens lag?

Histogram för $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$, $X_i \in Exp(1)$ med $\mu = \sigma = 1$.




Summa av tärningar
0.2
0.1
0
$p_X(k)$
1
2
3
4
5
6
7
8
Antal tärningar
0
10
20
30
40
50
k

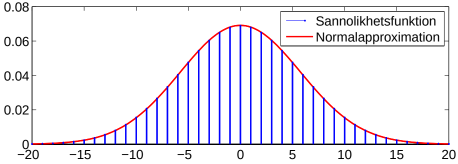
Exempel: Sten, sax och påse

Per och Lisa spelar sten, sax och påse. Vinnaren får en krona av förloraren, vid oavgjort händer inget. Antag att det är samma sannolikhet för vinst, oavgjort och förlust.

a) Bestäm fördelningen för Lisas vinst i en spelomgång.

b) Beräkna väntevärde och varians för Lisas vinst i en spelomgång.

c) Vad är sannolikheten att Lisa totalt vunnit minst **5** kr efter **50** spelomgångar?



CGS och Gaussapproximation (Delta metoden)

$E(X_i) = \mu \quad V(X_i) = \sigma^2, \quad X_1, X_2, \ldots, X_n$ oberoende likafördelade.
Vi har att

$$\sqrt{n} \cdot (g(\bar{X}_n) - g(\mu)) \approx g'(\mu)\sqrt{n}(\bar{X}_n - \mu)$$
$$= \sigma \cdot g'(\mu) \cdot \underbrace{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}_{\text{CGS ger } \underset{\sim}{\in} N(0,1)} \underset{\sim}{\in} N(0,\ \sigma \cdot |g'(\mu)|).$$

Vilket ger att

$$g(\bar{X}_n) \underset{\sim}{\in} N\left(g(\mu),\ |g'(\mu)| \cdot \frac{\sigma}{\sqrt{n}}\right).$$

Det här är användbart i statistiken sedan!

Exempel: Mätning av tyngdacceleration

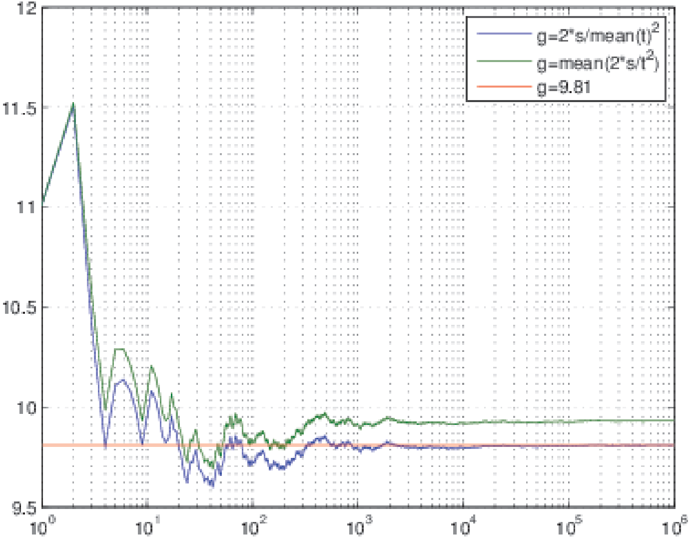
Tyngdaccelerationen, g, mäts genom att tiden, t, det tar för en kula att, från stillastående, falla $s = 1$ m mäts. Från fysiken vet vi att:

$$s = \frac{gt^2}{2} \qquad g = \frac{2s}{t^2} \qquad t = \sqrt{\frac{2s}{g}}$$

Anta att varje mätning, T_i, är oberoende och rektangelfördelade kring det rätta värdet med en osäkerhet på ± 0.05 s.

Bestäm variansen av $G = \dfrac{2s}{\bar{T}^2}$ efter n oberoende mätningar.

Exempel: Mätning av tyngdacceleration



Medelvärde av 100 mätningar