Några vanliga fördelningar – från ett GUM-perspektiv

I denna PM redovisas några av de vanligaste statistiska fördelningarna och deras hantering inom ramen för GUM: Guide to the Expression of Uncertainty in Measurement.

Den 1-dimensionella normalfördelningen

$N(\mu,\sigma)$ betecknar en 1-dimensionell normalfördelning med väntevärdet μ och medelfelet σ [1]. I HMK finns ett koncept för felgränser som baseras på 1σ, 2σ, 3σ och som har sitt ursprung i denna fördelning:

- Fel större än 3σ betraktas som grova fel. Därför är 3σ-gränsen att betrakta som en ren *kassationsgräns* och ommätning krävs.
- 2σ används som "*varningsgräns*". Om felen överskrider denna gräns bör den bakomliggande orsaken analyseras. Denna gräns överensstämmer med den konfidensnivå som är dominerande inom statistiken, nämligen 95 %.
- 1σ-gränser används för ett test av att antagandet om normalfördelade fel är korrekt – ett *fördelningstest*. 2/3 av mätmaterialet bör ha fel som är mindre än 1σ.

Konfidensnivåerna för normalfördelningen är

- $3\sigma \rightarrow$ 99,7 %, mycket osannolikt
- $2\sigma \rightarrow$ 95,5 % ($1{,}96\sigma \rightarrow$ 95 %)
- $1\sigma \rightarrow$ 68,3 % (≈ 2/3)

varur relationen till ovanstående felgränser tydligt framgår.

Exempel: Bestäm ett konfidensintervall på nivån 95 % för väntevärdet μ av en normalfördelad stokastisk variabel $X \in N(\mu,\sigma)$ med det kända medelfelet $\sigma = \sqrt{2}$. Ur 8 st. oberoende mätningar har medeltalet beräknats till $\bar{x} = 4{,}00$.

•

Standardosäkerheten för medeltalet ges av uttrycket

$$u(\bar{x}) = \sigma / \sqrt{n} = \sqrt{2} / \sqrt{8} = 0{,}50$$

Normalfördelningens täckningsfaktor på konfidensnivån 95% $k_{95} = \lambda_{95} = 1{,}96$ ger sedan

$$P\{\mu \in \bar{x} \pm 1{,}96 * u(\bar{x})\} = P\{\mu \in 4{,}00 \pm 0{,}98\} = P\{\mu \in [3{,}02; 4{,}98]\} = 95\%$$

t-fördelningen

Den vanligaste tillämpningen av t-fördelningen är vid konstruktion av konfidensintervall för väntevärdet av normalfördelade stokastiska variabler där även σ är okänt (jfr. föregående exempel). Standardosäkerheten skattas då med standardavvikelsen

$$s = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)}$$

[1] Det teoretiska spridningsmåttet benämner vi *medelfel,* med beteckningen σ, till skillnad mot *standardosäkerheten* som är kopplad till den reella mätningen. Ibland beräknas dock standardosäkerheten ur σ.

där $n-1$ är antalet frihetsgrader (överbestämningar). I analogi med normalfördelnings-exemplet ovan får vi vidare täckningsfaktorn

$$k_\alpha = t_\alpha(n-1)$$

där α är konfidensnivån, och medeltalets standardosäkerhet skattas av

$$u(\bar{x}) = s / \sqrt{n}$$

I tabellen nedan ser vi att t-fördelningen konvergerar mot normalfördelningen när antalet frihetsgrader ökar.

Frihetsgrader = f	3	7	15	30	50	80	120	∞
$t_{95}(f)$	3,182	2,365	2,131	2,042	2,009	1,990	1,980	1,960

<u>Exempel:</u> Bestäm ett konfidensintervall på nivån 95 % för väntevärdet μ av en normalfördelad stokastisk variabel $X \in N(\mu, \sigma)$ där medelfelet σ är okänt. Ur 8 st. oberoende mätningar har medeltalet beräknats till $\bar{x} = 4,00$ och standardavvikelsen till $s = 1,4$.

•

Standardosäkerheten för medeltalet ges av uttrycket

$$u\left(\bar{x}\right) = s / \sqrt{n} = 0,495$$

och t-fördelningens täckningsfaktor för $n-1=7$ frihetsgrader på 95 % nivå blir

$$k_{95} = t_{95}(7) = 2,365$$

Sammantaget får vi

$$P\left\{\mu \in \bar{x} \pm 2,365 * u(\bar{x})\right\} = P\left\{\mu \in 4,00 \pm 1,17\right\} = P\left\{\mu \in \left[2,83; 5,17\right]\right\} = 95\%$$

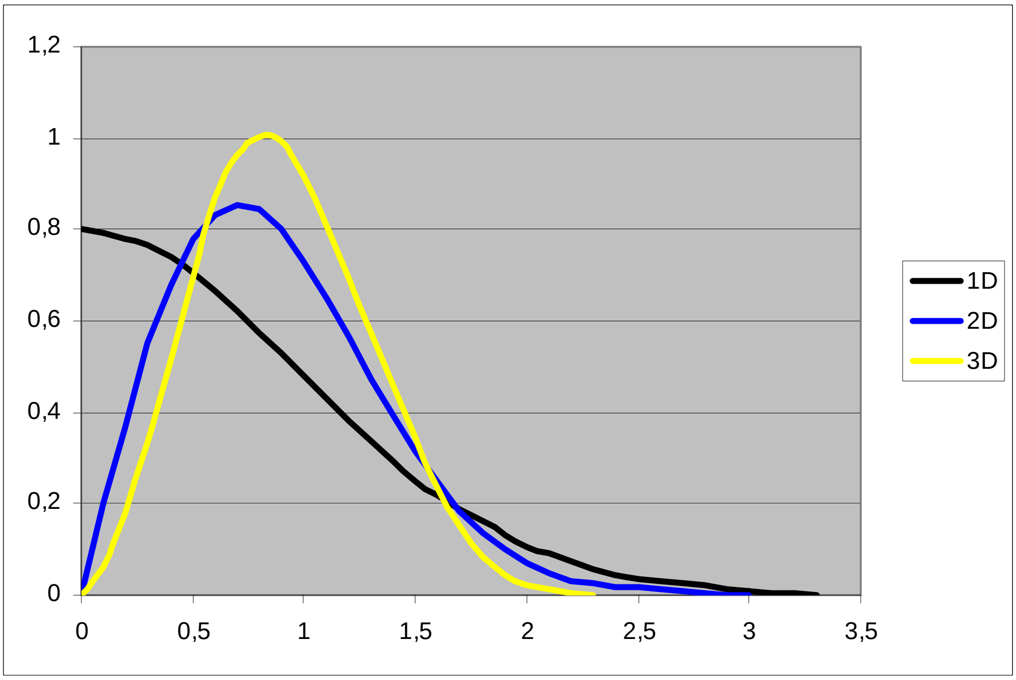
Normalfördelningen i två och tre dimensioner

Eftersom positioner redovisas i 2D eller 3 D så är även normalfördelningen i två och tre dimensioner av intresse. Dessa fördelningar ser ganska annorlunda ut jämfört med motsvarande 1-dimensionella fördelning, se figur på nästa sida.

Eftersom de 2- och 3-dimensionella fördelningarna endast antar positiva värden så representeras den 1-dimensionella fördelningen i figuren av <u>absolutbeloppet</u> av en normalfördelning; man så att säga ”viker ihop” den karaktäristiska Gauss-klockan.

Så här distinkta är dock fördelningarna endast i teorin – om samtliga komponenter är lika stora och om inga korrelationer föreligger mellan komponenterna. Så är det inte i verkligheten, och då blir gränserna mellan olika dimensioner mer diffusa.

Om osäkerheten är större i <u>en</u> av komponenterna – t.ex. höjdkomponenten i en 3-dimensionell position – så beter sig fördelningen nästan som om den vore 1-dimensionell. Om osäkerheten i <u>två</u> av komponenterna i en 3-dimensionell fördelning är större än den tredje så får fördelningen egenskaper som liknar den 2-dimensionella.



Teoretiskt gäller för 1σ, 2σ, 3σ i $1D$, $2D$ och $3D$:

	1σ	2σ	3σ
$1D$	$68,27\%$	$95,45\%$	$99,73\%$
$2D$	$63,21\%$	$98,17\%$	$99,99\%$
$3D$	$60,80\%$	$99,30\%$	$100-\varepsilon$ %

Men mot bakgrund av ovanstående resonemang så vet man inte i det verkliga fallet var i respektive kolumn man ska läsa av %-värdet. Man får vanligen nöja sig med att konstatera att:

- 3σ $\rightarrow$ nära 100 %, mycket osannolikt
- 2σ $\rightarrow$ minst 95 % (95-99 %)
- 1σ $\rightarrow$ ca. 2/3 (61-68 %)

Dvs. HMK-konceptet fungerar rätt bra i samtliga dimensioner, och t.ex. täckningsfaktorn 2 (2σ) ger minst 95 % konfidensnivå i såväl $1D$ som i $2D$ och $3D$.

Triangelfördelningen

$T(b,d)$, en (symmetrisk) triangelfördelning i intervallet $[b,d]$, har följande egenskaper:

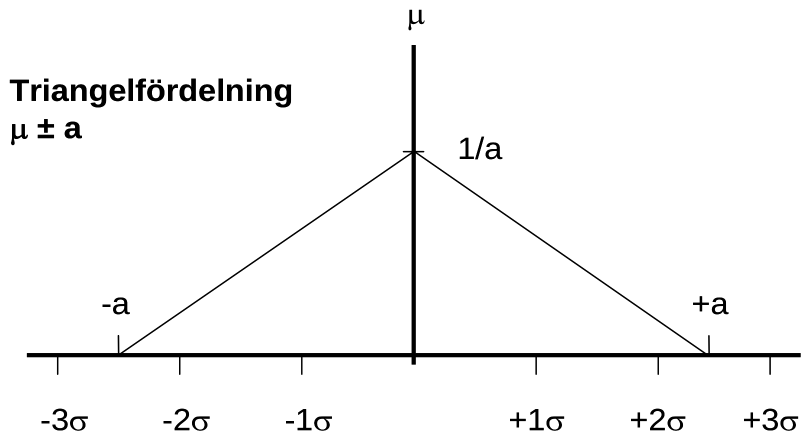
Väntevärde: $\mu = (b+d)/2$

Varians: $\sigma^2 = (d-b)^2/24$

Fördelningen kan även uttryckas $\mu \pm \mathrm{a}$ där $\mathrm{a} = (d-b)/2$. Det ger oss medelfelet:

$$\sigma = (d-b)/\sqrt{24} = \mathrm{a}/\sqrt{6}$$

1σ har konfidensnivån 64,98 %, 2σ har konfidensnivån 96,63 % och 3σ ligger utanför fördelningens gränser (se figuren). $k_{95} = 1,90$ är triangelfördelningens täckningsfaktor på 95 % konfidensnivå.



Exempel: Bestäm medelfelet för $T(8,10)$, dvs. en triangelfördelning i intervallet $[8,10]$.

•

Vi har $b = 8$, $d = 10$ och därmed $\mathrm{a} = (10-8)/2 = 1$. Det ger oss

$$\sigma = (10-8)/\sqrt{24} = 1/\sqrt{6} \approx 0,41$$

Rektangelfördelningen

$R(b,d)$, en rektangelfördelning i intervallet $[b,d]$, har följande egenskaper:

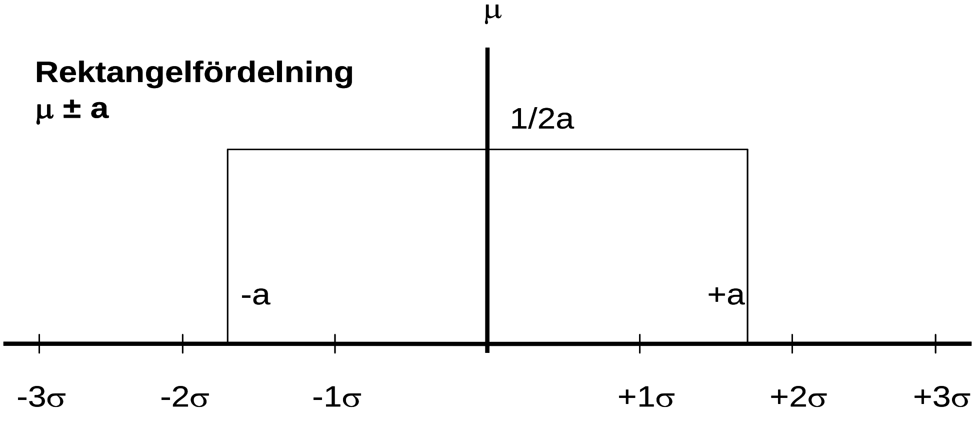
$$\text{Väntevärde:} \quad \mu = (b+d)/2$$

$$\text{Varians:} \quad \sigma^2 = (d-b)^2/12$$

Fördelningen kan även uttryckas $\mu \pm \mathrm{a}$ där $\mathrm{a} = (d-b)/2$. Det ger oss medelfelet:

$$\sigma = (d-b)/\sqrt{12} = \mathrm{a}/\sqrt{3}$$

1σ har konfidensnivån 57,7 %. Såväl 2σ som 3σ ligger utanför fördelningens gränser (se figuren). $k_{95} = 1,65$ är rektangelfördelningens täckningsfaktor på 95 % konfidensnivå.



Exempel: Bestäm medelfelet för en rektangelfördelning i intervallet -3 till +5, $R(-3,5)$.

•

Vi har $b=-3$, $d=5$ och därmed $a=\left(5+3\right)/2=4$. Det ger oss

$$\sigma=\left(5+3\right)/\sqrt{12}=4/\sqrt{3}\approx 2,309$$

Sammanställning

Slutligen görs följande jämförelse mellan den 1-dimensionella normalfördelningen, triangel-fördelningen och rektangelfördelningen:

Fördelning	**1σ**	**2σ**	**3σ**	**k_{95}**
Normalfördelning (1D)	68,3 %	95,5 %	99,7 %	1,96
Triangelfördelning	65,0 %	96,6 %	100 %	1,90
Rektangelfördelning	57,7 %	100 %	100 %	1,65

/Clas-Göran Persson