Lognormalfördelningen – några enkla iakttagelser

Vid något tillfälle kom jag i kontakt med ett antal forskare inom nanovetenskapen och jag fick ta del av deras data men endast i mycket reducerad form. Detta innebar medelvärdet, standardavvikelsen och det faktum att man mätte yta (tillväxt) av något vilket då innebär att all data var positiv.

Det numeriska värdet på medelvärdet och standardavvikelsen visade att den statistiska fördelningen av mätresultatet var skev. Det fanns alltså anledning att begrunda lognormalfördelningen. Det visade sig också behövas en illustration av lognormalfördelningen och detta ville jag göra med parametrar som åtminstone approximativt gav de värden forskarna hade på sina data.

Antag att väntevärdet på lognormalfördelningen skall vara 40 och standardavvikelsen 140. Hur får man tag på lognormalfördelningens parametrar som ger dessa värden? Nedan visas tre inte helt oberoende sätt att försöka hitta rätt värden på de två parametrarna (pkt 2-4).

1. Lognormalfördelningens väntevärde och standardavvikelse

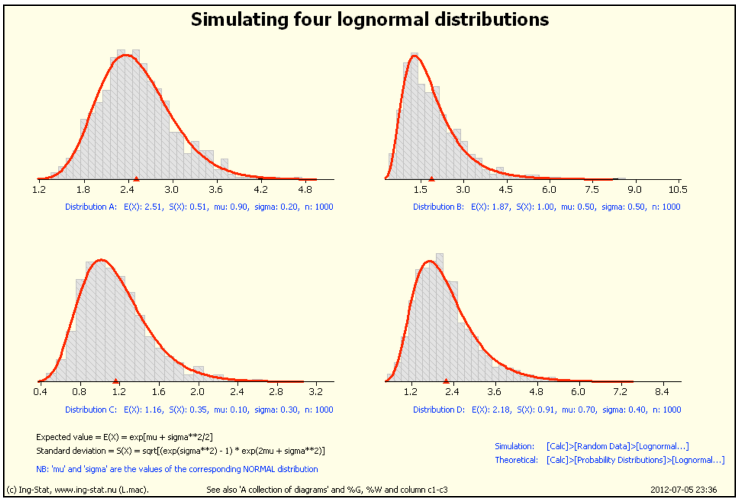
Lognormalfördelningens väntevärde och standardavvikelse – och mycket, mycket mer – kan lätt hittas i litteraturen eller på nätet:

$$\mu L = e^{(\mu+\sigma^2/2)}$$

$$\sigma L = \sqrt{\left(e^{\sigma^2}-1\right)\cdot e^{\left(2\cdot\mu+\sigma^2\right)}}$$

- μ är väntevärdet av den normalfördelning som erhålles då lognormalfördelningen logaritmeras.
- σ är standardavvikelsen hos den normalfördelning som erhålles då lognormalfördelningen logaritmeras.
- μL är väntevärdet av lognormalfördelningen.
- σL är standardavvikelsen hos lognormalfördelningen.

Om de lognormalfördelade värdena logaritmeras erhålls en vanlig normalfördelning. Observera dock att detta inte gäller väntevärdet.

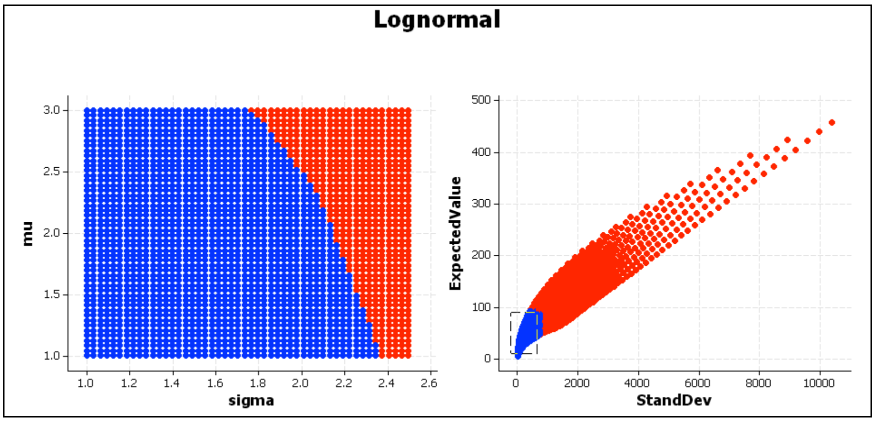


Några olika lognormalfördelningar.

2. Ett rent grafiskt försök

Ett första försök var att skapa en grid över en stor mängd värden på de två parametrarna och sedan med 'brush'-teknik försöka hitta de parameteruppsättningarna som åtminstone approximativt ger det önskade väntevärdet och standardavvikelse.

Nedanstående diagram är en skärmdump av ett resultat. Det visade sig dock vara inte helt lätt att lirka och ändra parameteruppsättningarna ty varje förändring av den ena parametern påverkade både väntevärdet och standardavvikelsen (diagrammet nedan är ett av flera försök att finna rätt parametervärden):



Det vänstra diagrammet visar en mängd kombinationer av 'my' och 'sigma'.

Det högra diagrammet visar väntevärde och standardavvikelse i motsvarande lognormalfördelning.

De blå punkterna i respektive diagram korresponderar.

3. Ett bättre semi-grafiskt försök

Genom att manipulera det matematiska uttrycket för väntevärdet och lösa ut σ-parametern blev uppgiften lite lättare och gav dessutom en smula insikt:

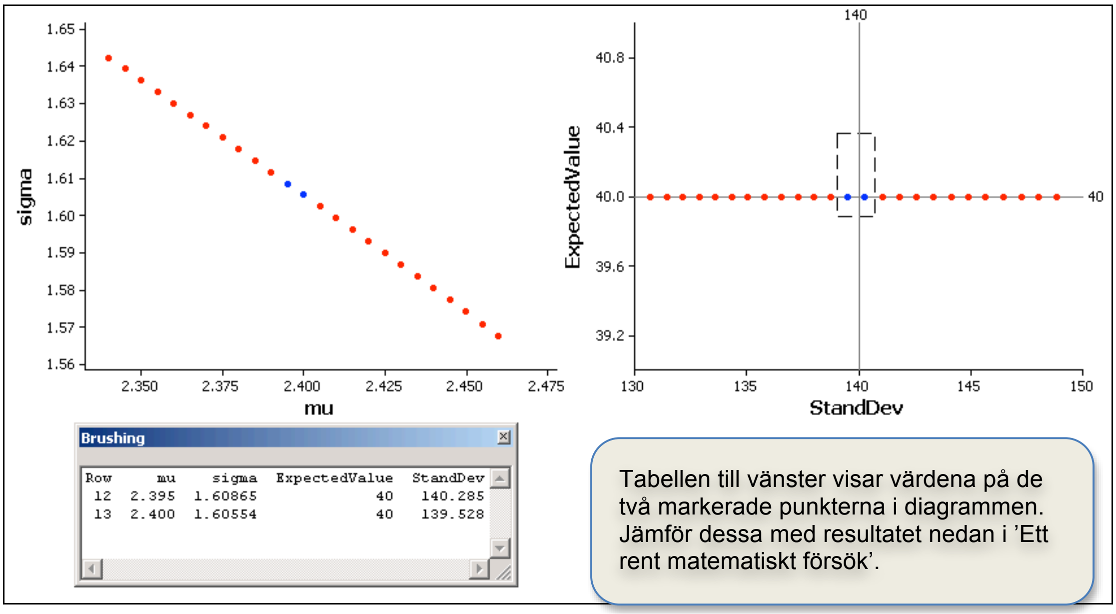
$$\sigma = \sqrt{2 \cdot (\ln \mu L - \mu)}$$

- det framgår att μ måste vara mindre än logaritmen av väntevärdet av lognormalfördelningen (lnμL), (ty annars blir uttrycket under rottecknet negativt!).
- man ser också att om μ = lnμL blir σ = 0 och μL = *exp*(μ). (OBS att datorprogrammen inte accepterar σ = 0 men man kan ju pröva med värden såsom 0.00001 som är nästan noll.)

Nu har vi en funktion som vi kan illustrera i en graf och dessutom ger funktionens alla punkter ett konstant värde på μL och olika värden på σL. Det blir alltså lättare att plocka fram den kombination av μ och σ som ger de önskade värdena på μL = 40 och σL= 140.

Det vänstra diagrammet nedan visar funktionen ovan då μL = 40. Y-axeln visar μ -parametern och X-axeln visar σ-parametern. Den streckade rektangeln är s.k. 'brushing' och innehåller två punkter på diagrammet. Dessa två punkter visas också i det högra diagrammet. Alla punkter i det högra diagrammet ligger vågrätt dvs på samma μL-värde (40) och det framgår också att de två blåmarkerade punkterna ligger nära det önskade värdet på σL-värdet (140).

Genom att rita om diagrammen med ändrade skalor kan man komma de önskade värdena med önskad noggrannhet.



Den vänstra grafen ovan visar sambandet mellan parametrarna 'my' och 'sigma' för ett givet väntevärde (40) på lognormalfördelningen. Den högra grafen visar väntevärde och standardavvikelsen i en lognormalfördelningen där väntevärdet är 40. De två blå punkterna i respektive diagram ligger närmast det önskade resultatet (40 och 140). Önskas fler decimaler (dock osannolikt) på 'my' och 'sigma' kan diagrammet lätt ritas om med annan skala.

4. Ett rent matematiskt försök

Om man söker på internet kan man hitta mycket. Nedan finns de sökta uppgifterna i rent matematisk form. Genom att lägga in värden på μL och σL kan parametrarna μ och σ beräknas:

$$\mu = 2 \cdot \ln \mu L - \frac{\ln\left(\mu L^2 + \sigma L^2\right)}{2} \qquad \sigma = \sqrt{\ln\left(\mu L^2 + \sigma L^2\right) - 2 \cdot \ln \mu L}$$

Genom att sätta in värdena $\mu L = 40$ och $\sigma L = 140$ i de två uttrycken erhålles det önskade resultatet:

$$\mu = 2 \cdot \ln \mu L - \frac{\ln\left(\mu L^2 + \sigma L^2\right)}{2} = 2 \cdot \ln 40 - \frac{\ln\left(40^2 + 140^2\right)}{2} \approx 2.39688$$

$$\sigma = \sqrt{\ln\left(\mu L^2 + \sigma L^2\right) - 2 \cdot \ln \mu L} = \sqrt{\ln\left(40^2 + 140^2\right) - 2 \cdot \ln 40} \approx 1.60748$$

4. Sammanfattning

Detta dokument visar några olika sätt att hitta de parametervärden i en lognormalfördelning som ger den ett visst givet väntevärde och given standardavvikelse. Det snabbaste sättet är naturligtvis genom två matematiska formler men kanske ger grafiska eller semi-grafiska metoder ytterligare insikter. ■

Ingemar Sjöström