

Statistik och Dataanalys I

Föreläsning 16 - Kontinuerliga sannolikhetsmodeller

Mattias Villani



Statistiska institutionen
Stockholms universitet



mattiasvillani.com



@matvil






mattiasvillani

- Likformig fördelning
- Exponentialfördelning
- Student- t
- Sannolikhetsmodeller och verkligheten

Likformig fördelning

Likformig fördelning

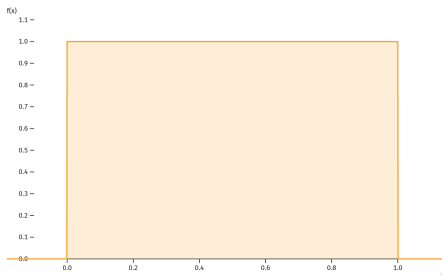
a : 
 b : 
Kvantil: 

Om $X \sim \text{Uniform}(0, 1)$ så gäller att

$$E(X) = \frac{a+b}{2} = 0.500$$

$$\text{Var}(X) = \frac{(b-a)^2}{12} = 0.0833$$

$$P(X \leq 1) = 1.000$$



Exponentialfördelning

- Om $X \sim \text{Expon}(\lambda)$ så är sannolikhetsfunktionen

$$f(x) = \lambda e^{-\lambda x}, \text{ för } x > 0$$

- $e \approx 2.71$ är Eulers tal.
- Väntevärdet är

$$E(X) = \frac{1}{\lambda}$$

- **Exponentialfördelning** vanlig modell för **väntetider**.

- ▶ Tid mellan samtal till stödlinje.
- ▶ Tid mellan mjukvarureleaser.

- Exponential och Poisson-fördelningen hänger ihop:

- ▶ Om **antalet samtal** till stödlinje per timme är **Poisson($\lambda = 3$)** så förväntar vi oss $\lambda = 3$ st samtal i timmen.
- ▶ Då är **tiden mellan samtal** **Expon($\lambda = 3$)** och vi förväntar oss $1/\lambda = 1/3$ timmar (20 minuter) mellan samtal.

Exponentialfördelning

Exponentialfördelningen

λ : 

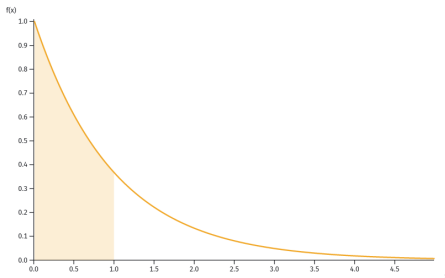
Kvantil: 

Om $X \sim \text{Expon}(1.01)$ så gäller att

$$E(X) = \frac{1}{\lambda} = 0.990$$

$$\text{Var}(X) = \frac{1}{\lambda^2} = 0.980$$

$$P(X \leq 1) = 0.6358$$



Exponentialfördelning i R

- $X \sim \text{Expon}(\lambda = 3)$. Parametern λ kallas `rate` i R.

Beräkning	R kommando	Kommentar
$f(0.5)$	<code>dexp(x = 0.5, rate = 3)</code>	$f(x)$ vid $x = 2$
$P(X \leq 0.5)$	<code>pexp(q = 0.5, rate = 3)</code>	
Kvantil	<code>qexp(p = 0.5, rate = 3)</code>	Medianen
10 slumpstal	<code>rexp(n = 10, rate = 3)</code>	

- Se programkoden [exponential.R](#) på kurssidan.

Student- t fördelning (standard)

Student- t distribution

$$X \sim t_{\nu}(0, 1)$$

$$E(X) = 0 \text{ om } \nu > 1$$

$$Var(X) = \frac{\nu}{\nu - 2} \text{ om } \nu > 2$$

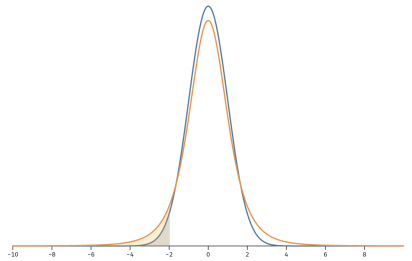
Frihetsgrader ν :

Kvantil:

Normalfördelning: $P(X \leq -1.96) = 0.02500$

Student- t fördelning: $P(X \leq -1.96) = 0.06078$

☒ normal ☐ student-t



Varför Student- t är viktig för inferens

- Låt X_1, X_2, \dots, X_n var oberoende data from $N(\mu, \sigma^2)$.
- Stickprovmedelvärdet


$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

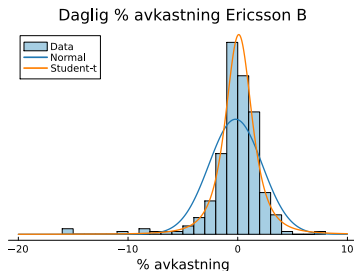
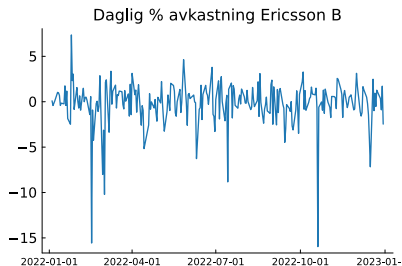
- Inferens: intresserad av fördelningen för det **standardiserad medelvärde**

$$\frac{\bar{X} - \mu}{SD(\bar{X})}$$

- Om variansen i populationen σ^2 **är känd** så är det **standardiserade medelvärde normalfördelat**.
- Om variansen i populationen σ^2 **är okänd**, och måste skattas med s^2 , så är det **standardiserade medelvärde student-t fördelat** med $\nu = n - 1$ frihetsgrader.

Aktieavkastning

- Daglig avkastning Ericsson B aktie under hela år 2022. 
- Finansiella data har ofta extremvärden. **Tunga svansar.**
- Maximum likelihood: $\mu = 0.094$, $\phi = 1.279$ och $\nu = 2.706$.



Allmän Student- t fördelning för datamodellering

Allmän Student- t distribution

$$X \sim t_{\nu}(\mu, \phi^2)$$

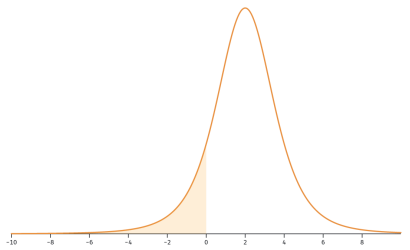
$$E(X) = \mu \text{ om } \nu > 1$$

$$\text{Var}(X) = \frac{\nu}{\nu - 2} \phi^2 \text{ om } \nu > 2$$

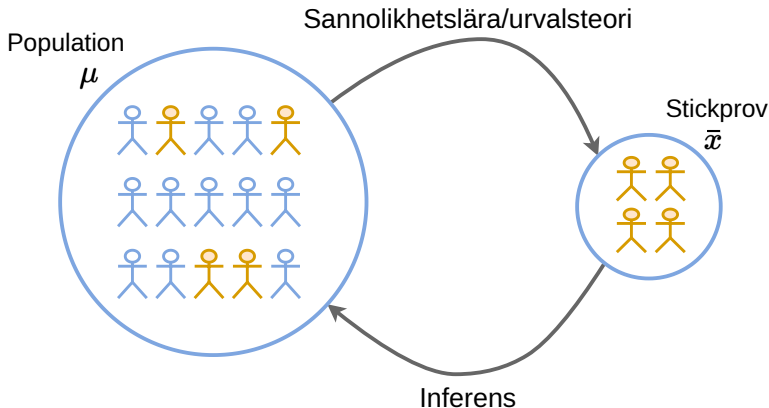
Läge μ :	<input type="text" value="2"/>	
Skala ϕ :	<input type="text" value="1.5"/>	
Frihetsgrader ν :	<input type="text" value="4"/>	
Kvantil:	<input type="text" value="0"/>	

visa
normalfördelning ☐

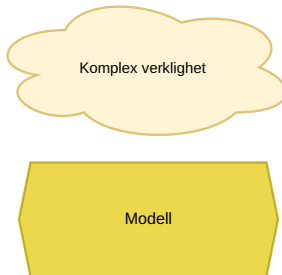
Student- t fördelning: $P(X \leq 0) = 0.1266$



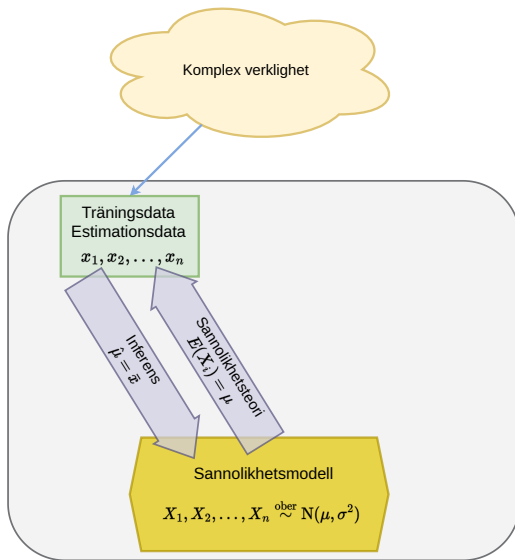
Population och stickprov - ändliga populationer



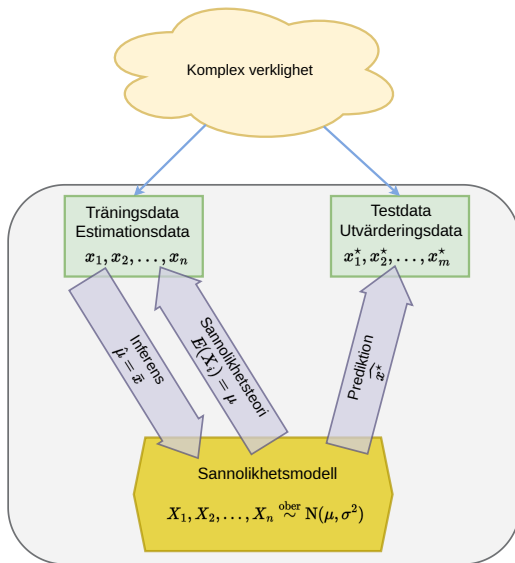
Modeller som en förenkling av verkligheten



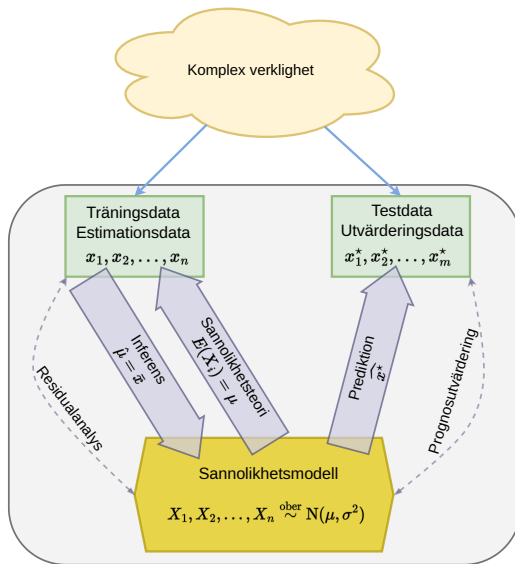
Sannolikhetsmodeller och inferens



Sannolikhetsmodeller möter verkligheten - prediktion



Modellering är en iterativ process



Slutmålet är ofta beslutsfattande i en osäker värld

