

Populärvetenskaplig sammanfattning

Avhandlingen fokuserar på att introducera nya modeller för framförallt spatial statistik. Dessa modeller är icke-Gaussiska och utvecklade särskilt för att vara praktiskt användbara. Modellerna är dessutom beräkningseffektiva, för att kunna användas för analys av stora datamängder, något som blivit mycket viktigt i modern spatial statistik. I spatial statistik används i dagsläget nästan utan undantag Gaussiska modeller, dvs. varje observation antas ha normalfördelning. Om man ansätter en sådan modell gör man en rad antagande som man bör vara medveten om, t.ex:

- observationerna kommer alltid antas kunna ha både negativa och positiva värden,
- fördelningen för observationerna antas vara symmetriska kring medelvärdet, dvs. det är lika troligt att en observation är under som över medelvärdet,
- observationerna kommer ha lätta svansar, det betyder att mycket stora värden på observationerna kommer vara extremt ovanligt. Detta antagandet är främsta anledningen till varför man har lämnat de Gaussiska modellerna inom finansiell statistik.

Det är väl känt att mycket av spatialt observerad data inte kan vara Gaussisk, men det har saknats alternativa modeller som är praktiskt användbara. I avhandlingen konstrueras just sådana icke-Gaussiska modeller. De skapas med hjälp av den Generaliserade Hyperboliska (GH) fördelningen, vilket gör observationerna från modellen mycket mer flexibla än om de kom från en normalfördelning. GH modeller har framgångsrikt använts i tidsserieanalys för att prediktera aktiekurser.

I avhandlingen visas GH modellernas flexibilitet på data som består av nederbördmättningar från mätstationer i Brasilien. Eftersom nederbörd aldrig kan vara negativ är det onaturligt att anta den är Gaussisk, dessutom har man empiriskt visat att svansarna av fördelningen för nederbörden är tyngre än de Gaussiska svansarna, dvs. observerade extremvärden är större än vad modellen predikterar. I avhandlingen visas att man kan justera GH modellerna så att de bara tar icke negativa värden och att modellens svansar kan efterlikna svansarna man ser i data.

Modellerna i denna avhandling är parametriska, vilket innebär att ett antal parametrar helt bestämmer beteendet av modellerna. Ofta vet man inte på förhand vilka parametrar som bör användas till ett specifikt problem, och man vill sätta parametrarna så att modellen efterliknar den data man studerar. Därför är det viktigt att ha metoder som med hänsyn till datan kalibrerar parametrarna. I avhandlingen introduceras metoder som väljer den parameteruppsättning som ger störst sannolikhet till observationerna.

För spatial statistik är det mycket viktigt att kunna prediktera. Det vill säga givet att man har observationer på ett antal punkter kunna säga hur man förväntar sig att en observation skall bete sig vid en punkt där man saknar observationer. Alla metoder för estimation och prediktion som är utvecklade i avhandlingen är skapade så att de skall vara beräkningseffektiva. Särskilt i (modern) spatial statistik är det viktigt att ha beräkningseffektiva metoder eftersom man ofta gör prediktion på/med stora datamängder, t.ex. satellitdata och komplexa miljömodeller.