

# Stokastiska processer och tidsserier

Med tanke på att nästan allt som händer omkring oss påverkas av tillfälligheternas spel, kan man hävda att hela vår civilisation är en produkt av oändligt många slumpmässiga processer.

Slumpen har påverkat vad som varit och den fortsätter att styra vår framtid. Vi inser ju att historiska fakta inte kan ge helt säkra förutsägelser om vår framtid. Ändå är det många gånger precis vad vi försöker åstadkomma.

Från vad som varit tycker vi oss se trender och lagbundenheter och med hjälp av dem försöker vi spå om framtiden. Men våra förutsägelser visar alltid mer eller mindre fel.

**DEN UTVECKLING VI SÖKER** förklara och den framtid vi söker förutspå är underkastad slumpens oförutsägbarhet. I den här artikeln ska vi betrakta statistiska tidsserier för vad de faktiskt är – resultatet av ett stort antal stokastiska, eller slumpmässiga, processer.

Ett exempel på en tidsserie är den summerade fruktsamheten, TFR, under perioden 1900–2005. Varje års värde är resultat av ett mycket stort antal stokastiska processer under just det året. Tidsserien kan skrivas som  $TFR(t)$  där indexet  $t$  anger kalenderåret.

Det finns ingen modell som med säkerhet fastställer ett framtida  $TFR(t)$  utifrån tidigare observationer  $TFR(t-1)$ ,  $TFR(t-2)$ , ...,  $TFR(t-k)$ . Även om vi i stora drag anser oss kunna förklara fruktsamhetsutvecklingen utifrån sociala, teknologiska, ekonomiska och andra kunskaper så kvarstår det rörelser och tendenser i tidsseriekurvan för TFR som inte kan förklaras med hjälp av de kunskaper vi har. Sammantaget innebär detta att TFR-processen är en stokastisk process.

**STATIONÄRA TIDSSERIER** saknar trend. Avsaknaden av trend innebär inte att värdena är desamma år från år. De varierar mer eller mindre. Men variationerna betraktas som slumpmässiga kring en konstant linje, det vill säga kring ett tänkt konstant medelvärde.

En stationär tidsserie karakteriseras inte bara av att den har ett konstant medelvärde. Den har också konstant varians. Det innebär att variationerna kring medelvärdet är av likartad storlek.

Det är svårt att avgöra om en observerad tidsserie är stationär eller inte. Om serien uppvisar en tydlig systematisk förändring över tid är den inte stationär. Vi säger att den har en trend. Men det vi i dagligt tal kallar för trend kan mycket väl vara resultatet av slumpmässigheter. Sådana trender är ett vanligt inslag i stationära tidsserier.

Typiskt för tidsserier är att observationerna är korrelerade med

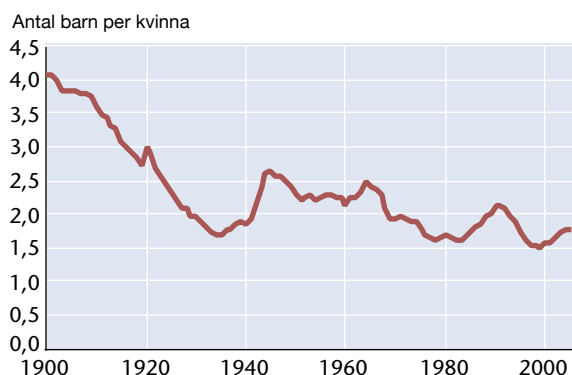
varandra. Det vill säga att det finns ett samband mellan intilliggande värden i tidsserien, ett högt värde följs av ett annat högt värde och ett lågt värde av ett annat lågt. Korrelationen mellan observationer som tillhör samma tidsserie kallas för autokorrelation.

**KORRELATIONEN MELLAN** variablerna  $x$  och  $y$  anger hur mycket  $x$  påverkar  $y$ . En positiv korrelation mellan  $x$  och  $y$  innebär att när  $x$  ökar, så ökar även  $y$ . En negativ korrelation innebär att när  $x$  ökar, så minskar  $y$ . Hur mycket den ena variabeln påverkar den andra beror på storleken av korrelationen. Korrelationer kan anta värden mellan  $-1$  och  $1$ .

Tidsserier framställs ofta som hjälpmedel för att förutsäga framtiden. Det som karakteriserar en positivt korrelerad tidsserie som TFR är ju att observationer som ligger nära varandra i tiden skiljer sig mycket litet åt. Obser-

## Barnafödandets upp- och nedgångar

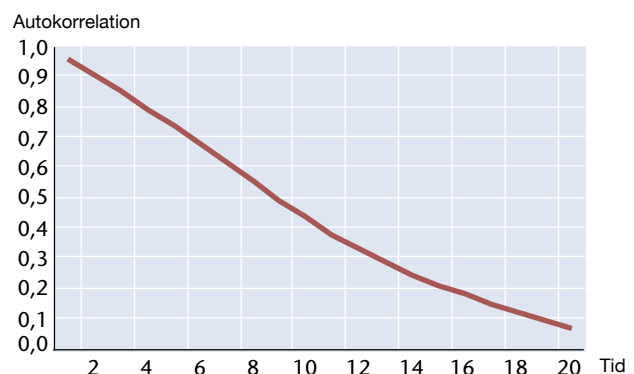
Den summerade fruktsamheten i Sverige 1900–2005



Den summerade fruktsamheten betecknas TFR, total fertility rate. Figuren visar TFR för varje kalenderår mellan 1900 och 2005.

## Korrelationen minskar med tiden

Autokorrelationen för TFR i Sverige 1900–2005



Autokorrelationen mellan ett års TFR och de följande årens värden är hög och positiv. Hög för att processens värde ett år fram i tiden i stor utsträckning styrs av observationen året innan. Positiv för att hög fruktsamhet ett år oftast följs av hög fruktsamhet nästföljande år, och omvänt. Korrelationen avtar dock med tiden.

vationerna är mer eller mindre bundna till varandra.

Den egenskapen utnyttjas när man gör prognoser på kort sikt. Det finns dock en stor principiell skillnad mellan prognoser på kort och lång sikt.

**MEDAN KORTA PROGNOSE** i hög grad kan grundas på de senast uppmätta observationerna så är inte det nödvändigtvis fallet för långa prognoser.

När man gör prognoser på lång sikt försöker man därför skatta processens framtida medelvärde på annat sätt. Ofta anlitas så kallade expertpaneler, vilket innebär att man låter ett antal experter på området diskutera en sannolik framtida utveckling. Deras åsikter vägs sedan samman till ett antagande om framtiden.

Det ligger i sakens natur att prognosen nästan aldrig stämmer överens med verkligheten.

Detta misstolkas ibland som att prognosen blev fel. Det beror på en missuppfattning. Man har inte klart för sig att den långa prognosen uttalar sig om en framtida nivå eller medelvärde och inte om ett enskilt framtida värde. Individuella observationer skiljer sig ju i allmänhet från medelvärdet.

Den verklighet som visar sig vartefter brukar i tidsseriesammanhang kallas för den stokastiska processens fortsatta realisation. I förväg kan man ju inte veta exakt hur den tar sig ut. Man gissar sig till medelvärdena men precis hur de verkliga värdena kommer att variera kring dessa medelvärden kan man inte förutspå. Det man vet är att värdena kommer att variera.

Genom att studera hur variationen hittills sett ut kan man bilda sig en uppfattning om variansen kring dessa medelvärden. Utifrån dessa antaganden skulle framtiden i princip kunna anta

ett oändligt antal olika utseenden – ett oändligt antal alternativa värden kring de skattade medelvärdena.

**ALLA TIDSSERIER ÄR DOCK INTE** att betrakta som autokorrelerade. Tidsserier vars observationer inte är korrelerade kallas ofta för ”vitt brus”. Den ryske matematikern Eugene Slutsky visade på 1920-talet att man i sådana tidsserier ofta kan skapa en ”falsk korrelation” genom att göra om dem till en serie av löpande medelvärden. Löpande medelvärden beräknas genom att successivt slå samman i tiden intilliggande observationer och dividera summan med antalet sammanslagna observationer.

Denna insikt att man kunde skapa *falsk korrelation* hade två konsekvenser. Å ena sidan antydde detta att byggstenarna för vissa tidsserier skulle kunna utgöras av okorrelerade slumpstal. Å andra sidan att man genom att medvetet utjämna observerade tidsserier skulle kunna införa falska korrelationer i den ursprungliga datastrukturen. Men det är en annan historia.

## FAKTA

### Historia

William Playfair publicerade år 1821 den första tidsserien i modern tid. Det var dock först på 1920-talet som teorin för tidsserier på allvar började utvecklas av vetenskapsmän som Eugene Slutski från Ryssland och George Udny Yule från England. I dag spelar teorin för tidsserier en viktig roll inom en lång rad vetenskaper exempelvis inom ekonometrin.

Källa: Kendall, M. 1976. *Time-Series Analysis*, Charles Griffin and Company: London.

Marquis Pierre-Simon de Laplace (1749–1827) som var en ledande vetenskapsman och matematiker hävdade att allt kring oss är deterministiskt, även mänskligt beteende. Detta var den allmänna åsikten till dess att den moderna kvantfysiken i början på det 20:e århundradet avvisade denna teori. Förkastandet av en deterministisk värld till förmån för en indeterministisk utgör en dramatisk vetenskaplig och filosofisk brytning med tidigare uppfattningar.

Källa: Hawking, S. 1996. *A Brief History of Time*. Bantam Dell Pub Group. New York.

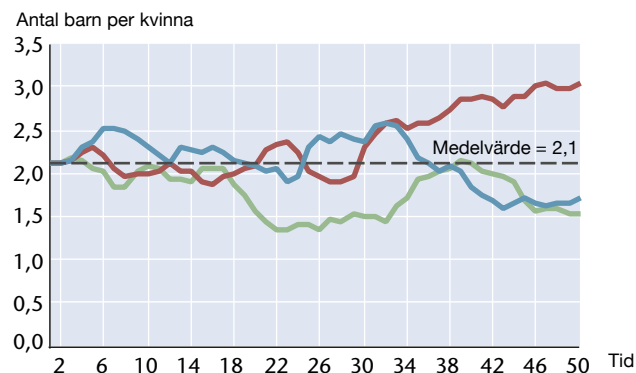
### Terminologi

Ordet *stokastisk* betyder slumpmässig och kommer av grekiskans ord för ”skicklig på att gissa”.

*Den summerade fruktsamheten* anger det antal barn som kvinnor och män skulle få i genomsnitt om det enskilda årets fruktsamhet i varje ålder skulle räda.

## Samma antaganden ger olika utfall

### Tre stokastiska realisationer av en TFR-process



En stokastisk process har oändligt många utfall. Realisationerna i figuren visar tre av de utfall som en TFR-process med medelvärdet 2,1 och variansen 0,01 barn per kvinna ger. Ju längre fram i tiden som realisationerna sträcker sig desto mer skiljer de sig från det sist uppmätta observationen eftersom korrelationen avtar med tiden.



### Författare

#### Michael Hartmann

arbetar med stokastiska prognoser på Prognosinstitutet vid SCB  
tfn: 08-506 943 74  
e-post: [michael.hartmann@scb.se](mailto:michael.hartmann@scb.se)

### Lästips

Gottman, J.M (1981). *Time-series analysis*. Cambridge University Press: London.