

# Hvad kan rekorder benyttes til?

**Kort svar:** Ikke ret meget

**Langt svar:** Se vedlagt pdf-dokument.

## 1. Introduktion

I forbindelse med den globale opvarmning eller afkøling har der været en heftig diskussion om ekstreme observationer mht. temperatur, isdække, hastighed, nedbør og ødelæggelser fra naturkræfterne.

Der er ofte blevet henvist til, at der nu er observeret en rekord i forhold til en tidsperiode på f.eks. 100 år. Dette betyder at observationen er den største eller mindste over de sidste 100 år. Det er nogle gange tolket som et bevis på den globale opvarmning eller den globale afkøling.

Det er oplagt, at nyhedsmedierne har stor fokus på antal rekorder. Man kan næste altid høre om et udsædvanligt naturfænomen, at der er det værste i over 100 år. Der er efter min opfattelse tale om systematisk datafiske, når medierne tale om rekorder.

Spørgsmålet er derfor. Hvad kan vi sige om forekomsten af rekorder og antal rekorder over en bestemt tidsperiode?

De følgende bemærkninger om ekstreme observationer vedr. klima er givet ud fra et statistik- og sandsynlighedsteoretisk synspunkt uden et stort kendskab klimatiske og meteorologiske forhold.

## 2. Definition af en rekord

En rekord ud af en serie af observationer er en observation, der er større/mindre end alle foregående observationer. Det kan være tale om temperaturen, vandstanden, vindhastighed eller barometertrykket målt på et givet sted på en given dag. Selv om der er tale vidt forskellige typer af målinger vil fordelingen af antal rekorder være den samme blot, der er tale om målingerne er identisk fordelte og har en kontinuert fordelingsfunktion. Et ganske overraskende resultat. Dette følger af at hvis en stokastisk variabel  $X$  har fordelingsfunktion  $F$ , så vil den stokastisk variabel  $Y = F^{-1}(X)$  være uniformfordelt på intervallet  $[0;1]$  samtidig med at rækkefølgen af  $X$ 'erne og  $Y$ 'er er bibeholdt.

## 3. Hvad er sandsynligheden for at observere en rekord?

Under antagelse af, at jeg har en serie af identisk fordelte målinger af længden  $n$ , vil sandsynligheden for at observere en rekord i periode  $n+1$  netop være  $1/(n+1)$  dvs.

$$P(\text{en rekord i periode } n + 1) = 1/(n + 1)$$

Hvis man har 1000 måleserier af temperaturen over en periode på 50 år og der ikke er nogen systematisk udvikling, vil man forvente i alt  $1000/(50+1) \approx 19.6$  rekorder dvs. ca. 20 rekorder.

Dette betyder, at hvis der ikke er nogen globale opvarmning eller global afkøling, vil man med stor sandsynlighed kunne finde en rekorder blot ved at gøre antal måleserier stort nok.

Hvis der er tale om en voksende/faldende tendens i en måleserie vil man forvente flere/færre rekorder end  $1/(n+1)$ . Man kan derfor benytte et binomialtest til at test hypotesen om  $p=1/(n+1)$ .

Derfor er det vanskelig uden en nærmere analyse, at benytte en rekord som et argument for en global opvarmning eller afkøling.

#### 4. Antal rekorder over en tidsperiode

Hvis  $N_n$  er lige med antal rekorder gælder og det antages, at der er tal om uafhængige måleresultater og der ikke er nogen systematisk udvikling gælder, at

$$E(N_n) = \sum_{k=1}^n \frac{1}{k}$$

$$Var(N_n) = \sum_{k=1}^n \frac{1}{k} - \frac{1}{k^2}$$

og at

$$E(N_n) - \ln(n) \rightarrow \gamma \text{ for } n \rightarrow \infty$$

$$Var(N_n) - \ln(n) \rightarrow \frac{\pi^2}{6} \text{ for } n \rightarrow \infty$$

og hvor  $\gamma = 0.5772$  angiver Eulers konstat.

Den følgende tabel viser værdier af middelværdien, spredning og 99%-fraktilen for  $N_n$  for udvalgte værdier af  $n$ .

N	middelværdi	Sprednings	99% -fraktilen
10	2.9	1.2	5.7
100	5.2	1.9	9.6
150	5.6	2.0	10.2
1000	7.5	2.4	13.1
10000	9.8	2.9	16.4

Fra denne tabel følger, at hvis man har en serie af 150 uafhængige identisk fordelte målinger forvente 5.6 rekorder og 99%-fraktilen er på 10.2. Selv om jeg har en måleserie på 10000 observation vil det forventede antal rekorder kun være på 9.8 ! Ganske overraskende.

Nu kan der selvfølgelig indvendes: Hvad skal dette nytte? Mange klimamålinger ikke er identisk fordelte målinger.

Nytten af denne viden ligger i, at der hermed er angivet et Benchmark for antal rekorder. Hvis vi f.eks. har 150 målinger, at der var tale om identisk fordelte målinger forventes 5.6 rekorder og ikke noget i

retning af 140 eller 150. Hvis vi observeret 11 eller flere rekorder i en måleserie på 150 må vi forkaste hypotesen om identisk fordelte målinger.

Derfor kan den ovenstående tabel være nyttig i klimadiskussionerne.

### **5. Eksempel:**

I perioden fra til 1880 til 2009 år har vi observeret 15 globale temperatur rekorder. Bemærk at første observation altid tælles med som en rekord..

Jeg har foretaget optælling på grafen for den globale temperatur på Wikipedia med forbehold for aflæsningsfejl.

Hvis vi antager, at antal observationerne alle stammer fra en måleserie med identisk fordelte målinger, kan vi beregne sandsynligheden for, at få flere end 15 maksimum rekorder ved en normalfordelings-tilnærmelse.

Nu fremgår perioden 130 år ikke af tabellen, men ved direkte beregning få middelværdien til 5.4486 og spredningen til 1.9522 , således at

$P(15 \text{ eller flere maksimumsrekorder}) = 0.0000005$  dvs. 1 ud af 2 mio. år.

Det er altså helt usandsynligt, at få 15 eller flere rekorder. Derfor forkastes antagelse om identisk fordelte temperaturmålinger. Næste problem er så at forklare hvorfor vi få denne signifikans.