Introduktion

I denna datorövning ska vi undersöka hypotesprövning[1]. Om du har tid finns det en extrauppgift på sluten av instruktionen som behandlar testvariabelmetoden[2]. I uppgiften betraktar vi ett räkneexperiment, där slumpvariabelen N är antal händelser[3] vi observerar i vårt experiment. Vår nollhypotes H_0 är att N är distribuerad enligt en Poissonfördelning med väntvärdet b, där b är "bakgrund" som vi mäter genom ett kalibreringsexperiment. Den alternativa hypotesen H_1 är att väntevärdet är $b + s$, där s är väntevärdet för antalet "signal"-händelser.

Användbara funktioner

Notera; på engelska (och i Python) kallas fördelningsfunktionen *Cumulative Distribution Function* (CDF). I python finns också en funktion som räknar ut $1 - CDF$, som kallas SF, *survival function.*

```
>>>> import scipy.stats as sps
>>>> #For discrete distributions, Python calls the probability
>>>> #distribution the "pmf". For an expectation value of 1.5,
>>>> #the probability to observe 4 events is
>>>> sps.poisson(1.5).pmf(4)
0.047066518156309411
>>>> #You may compute the CDF for a Poisson
>>>> #with expectation value 1.5 and observed events=4:
>>>> sps.poisson(1.5).cdf(4)
0.98142406377785929
>>>> #Computing the survival function:
>>>> sps.poisson(1.5).sf(4)
0.018575936222140675
#note that CDF+SF=1
```

Uppgifter

- Kom ihåg definitionen av p-värde från boken[4]; *P-värdet är sannolikheten, förutsatt att nollhypotesen är sann, att observera ett minst lika extremt utfall som det observerade.*
  - Hur räknar du ut p-värdet för ett räkneexperiment där nollhypotesen är att värdena är fördelade enligt en Poissonfördelning med väntevärde 1.5?
  - Räkna ut p-värdet för 1,2 och 3 antal observerade händelser
  - Generera 10000 slumpmässiga tal enligt nollhypotesen. Hur många av talen är större än 0? 1? 2? Stämmer resultatet med det du beräknade i förra uppgiften?
- Nu vill vi testa $H_0; N \sim \text{Pois}(1.5)$ mot $H_1 : N \sim \text{Pois}(4)$. Generera 10000 slumpässiga tal enligt H_1. Vad är testets styrka, dvs. sannolikheten att förkasta H_0 vid signifikansnivå 0.95?

[1] kapitel 7.4 i boken
[2] sida 309 i kursboken.
[3] stjärnor, muoner, bilar etc.
[4] sida 310

- Upprepa uppgiften ovan mer generellt, testa $H_0; N \sim \text{Pois}(1.5)$ mot $H_1(s) : N \sim \text{Pois}(1.5 + s)$. Räkna ut testets styrka för ett antal[5] s mellan 0 och 10. Plotta styrkan β som funktion av s.
- (extra uppgift) Nu vill vi testa $H_0; N \sim \text{Pois}(5.5)$ mot $H_1 : N \sim \text{Pois}(10)$.
  - Som tidigare kan du räkna ut p-värdet utifrån Poisson-statistik, och styrkan för att testa H_0 mot H_1.
  - Vad händer om du i stället betraktar distributionen av F, definierad av en funktion av det slumpmässiga talet N; $F := f(n)$[6]?
  - Generera 100000 slumpässiga tal från H_0, och räkna ut $f(n) = |n - 5.5|$ för observationerna.
  - Plotta ett histogram av $f(n)$ för nollhypotesen.
  - Kan du räkna ut[7] hur stor F måste vära för att endast 5% av värdene är högre?
  - Hur stor är styrkan för testet som använder $f(n)$ istället för p-värdet av Poisson-fördelningen?

[5] 5-10 stycken

[6] Märk att i et forsök på å følge bokens notation, bruker vi N,F för slumpvariabler, och n,f för de numeriske värdene (som kan brukes i funktioner).

[7] hint: kolla på dokumentationen til scipy.stats.scoreatpercentile()