Matematisk statistik V
Föreläsningsanteckningar
Tommy Norberg
16 mars 2005

Introduktion

Det som man normalt brukar ägna sig åt i en grundkurs i matematisk statistik är

1. sannolikhetsräkningar
2. slumpmodeller
3. statistisk slutledning

Den förstnämnda delen handlar om sannolikhetsbegreppet. Man får lära sig ordentligt vad som menas med sannolikheten för en händelse A, vilket skrives $P(A)$ eller $P[A]$. En händelse är något som antingen inträffar eller inte inträffar då man gör ett (slump-) försök.

Det är väldigt praktiskt att tänka sig att man gör ett försök i vilket man kan observera huruvida ett antal händelser inträffar eller ej, så fort man studerar något fenomen i vilket slumpen spelar en roll. Det spelar ingen roll om man studerar något så simpelt som ett tärningskast eller något så avancerat som hur en viss akties pris flukturerar på börsen.

Man går igenom hur man med logiska regler kombinerar ihop händelser och hur man räknar fram sannolikheten för den kombinerade händelsen i termer av händelserna man startade med. Om t ex A och B är två händelser så kan man definiera händelserna

- icke-A (A inträffar inte; betecknas ofta A^c eller A')
- A och B (båda händelserna inträffar; betecknas ofta $A \cap B$)
- A eller B (minst en av händelserna inträffar; betecknas ofta $A \cup B$)

Känner man tre av sannolikheterna

$$P(A), \quad P(B), \quad P(A \text{ och } B), \quad P(A \text{ eller } B)$$

så kan man räkna ut den fjärde m h a den mest grundläggande formeln för addition av sannolikheter

$$P(A) + P(B) = P(A \text{ eller } B) + P(A \text{ och } B) \tag{1}$$

Att ordentligt förstå denna formel är ett av delmålen med kursen.

Ibland vet man att en händelse B inträffat, men saknar information om huruvida en annan händelse A inträffat eller ej. I sådana situationer kan det vara praktiskt att jobba med betingade sannolikheter. Man definierar den betingades sannolikheten för A givet B, enl

$$P(A|B) = \frac{P(A \text{ och } B)}{P(B)} \tag{2}$$

Om $P(A|B) \neq P(A)$, så finns ett beroende mellan A och B, och man säger att A och B är *oberoende* om

$$P(A \text{ och } B) = P(A)P(B) \tag{3}$$

Kombinera (3) med (2) och konkludera att oberoende innebär

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

Formel (1) för addition av sannolikheter kombinerad med definitionen (2) av betingning leder till en väldig rik, djup och elegant matematisk teori, med vilken man kan göra verifierbara förutsägelser om slumpfenomen. Att ordentligt förstå och kunna använda teorin för betingning så som den t ex uttrycks i Bayes formel

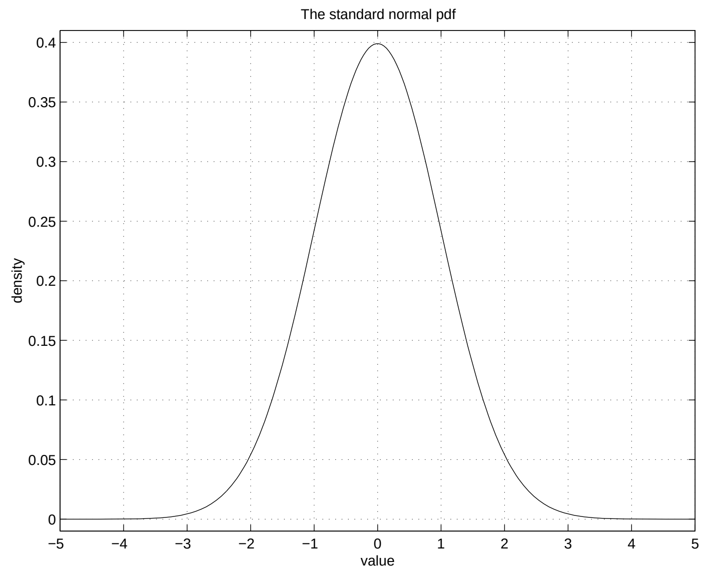
$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\text{icke-}A)P(B|\text{icke-}A)} \tag{4}$$

är därför ett av kursens delmål.

Notera att ofta skriver man $A \cap B$ istället för A och B, $A \cup B$ istället för A eller B och A' eller A^c istället för icke-A.

Den mest kända slumpmodellen är förmodligen normalfördelningen. Den s k normalfördelningskurvan, given av

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty \tag{5}$$



Figur 1: Den standardiserade normalfördelningskurvan har $\mu = 0$ och $\sigma = 1$

talar om för oss hurpass troliga olika utfall av t ex ett mätfel är. Parametrar är $\mu \in R$, $\sigma > 0$. Utfallsrum är hela R. Den standardiserade normalfördelningskurvan har $\mu = 0$ och $\sigma = 1$. Se figur 1.

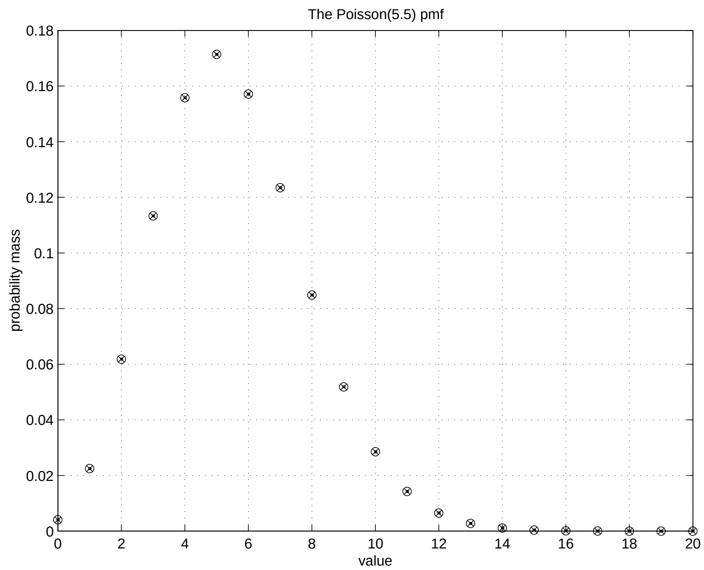
Normalfördelningen är en s k kontinuerlig modell. Kännetecknande för kontinuerliga fördelningar är att man beräknar sannolikheter medelst integration.

Vi ska också studera s k diskreta modeller. Kännetecknande för dessa är att sannolikheter beräknas medelst summation. Poissonfördelningen, se figur 2, är ett bra och viktigt exempel på en diskret modell. Den beskrivs av funktionen

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \tag{6}$$

Parameter är $\lambda > 0$. Utfallsrum är $Z_+ = \{0, 1, 2, \ldots\}$.

I kursen ska vi gå igenom ett antal slumpmodeller. Ett av kursens delmål är att vi ska lära oss känna igen situationer när våra slumpmodeller är relevanta modeller av verkliga skeenden där slumpen har en icke försumbar inverkan.



Figur 2: Ett exempel på en diskret fördelning

Antag att vi har n st mätningar $x_1, \ldots, x_n$ av en obekant storhet μ, där vi vet att mätfelen hyfsat bra följer normalfördelningskurvan (5) med okänt σ. I den statistiska slutledningen studeras hur man genom att ”punktskatta” μ skaffar sig en bra uppfattning om μ:s verkliga värde och hur man genom att ”intervallskatta” μ skaffar sig en bra uppfattning om felet i punktskattningen. Ett av delmålen med kursen är att vi ska kunna tillämpa denna teori i situationer då mätvärdena kan antas vara oberoende.

Ett annat av kursens delmål är att vi ska skaffa oss en grundläggande insikt om hur statistiska hypotestest göres och analyseras. Ofta gör man nämligen jämförande mätningar. Antag att de n mätningarna ovan gjordes i syfte att visa att μ är större än ett referensvärde μ_0. Då sätter man upp en s k nollhypotes

$$H_0 : \mu \leq \mu_0$$

(som är komplementet till det man vill visa) och en alternativhypotes

$$H_1 : \mu > \mu_0$$

(som är det man vill påvisa). I den statistiska analysen av mätvärdena bildar man sig en uppfattning om hurpass otrolig nollhypotesen H_0 är. Ju otroligare H_0 är, desto troligare är det att alternativet H_1 (det man vill påvisa) är sant.

Traditionella kurser i matematisk statistik börjar med en rejäl portion sannolikhetsteori, sedan följer ett avsnitt där man lär sig ett antal standardmodeller för olika sorters slumpfenomen och så avslutar man med att studera statistisk slutledning.

Våra läroboksförfattarer Devore & Farnum har ett modernare upplägg. De utgår ifrån data, diskuterar hur man kan studera olika datamängder och inför olika slumpmodeller innan de går igenom de viktigaste sannolikhetsteoretiska bitarna. Sist kommer statistikteorin. En orsak till att bygga upp kursen kring data är att det kan vara lita svårt att förstå annars varför man måste lära sig att räkna med sannolikheter och studera ett antal abstrakta slumpmodeller innan man kan börja med statistikteorin, som för många civilingenjörer är den viktigaste biten av kursen.

Jag har kompletterat boken med ett häfte om stokastisk simulering samt de tre häftena

- Bayesiansk uppdatering av sannolikhetsskattningar (betastatistik)
- Poissonprocessen och extrema laster
- Något om risk

Dessa tre häften innehåller material som en V-ingenjör behöver känna till, men som inte återfinns i läroböcker. Jag kommer att använda ca ett föreläsningspass (en dubbeltimma) till varje.

Det första handlar om hur man på ett hyfsat objektivt sätt kan föra in subjektiva kunskaper som experterfarenhet när man ska skatta en sannolikhet. Det andra beskriver en modell för när katastrofer (ovanliga händelser) inträffar och en teknik för skattning av sannolikheter för händelser som är så ovanliga att vi ännu inte sett någon inträffa (extrapolation). I det tredje ska vi introducera risk-begreppet såsom det ofta används i V-sammanhang.