

Allmän info om litteraturen, undervisningen (föreläsningar och övningar), examinationen (tentamen, kontrollskrivning och laborationer), kurshemsidan, registrering samt kursens upplägg.

Introduktion till sannolikhets teori och statistikt teori

Ämnet matematisk statistik som denna kurs ger introduktion till har många användningsområden. Matematisk statistik behövs för att göra orimons undersökningar och väderprognoser, för att analysera resultatet i läkemedelstester, för att bestämma priset på optioner och för att göra kvalitetskontroll av en tillverkningsprocess mm. Den matematiska statistiken har två delar: sannolikhets teori och statistikt teori. Sannolikhets teori handlar om att formulera och analysera slumpmodeller, dvs matematiska modeller där utfallet inte kan specificeras exakt (det beror på slumpen) utan det varierar slumpmässigt från gång till gång.

Ex: K kronor placeras på ett konto med fast ränta r . Efter n år är beloppet på kontot $K(1+r)^n$ (deterministisk modell). Om de K kronorna placeras i en aktie vet vi inte på förväg hur stort värdet är efter n år. Vi kan dock sätta upp en slumpmodell för aktieprisets utveckling och utifrån modellen besvara frågeställningar såsom "Hur stor är sannolikheten att beloppet efter n år är större än $2K$?"

Vi kan från en slumpmodell bara dra slutsatser om hur stor sannolikhet det är att olika händelser inträffar, exvis att en tärning visar sex vid tio tärningskast i rad eller att det snöar i Stockholm i morgon. Första halvan av kursen ägnas åt sannolikhets teori.

Statistikt teori handlar om att analysera slumpmässiga datamängder med matematiska metoder. Säg att vi gjort ett antal mätningar av en fysikalisk konstant. Naturliga frågeställningar är då exvis:

- Vilket värde på konstanten stämmer bäst överens med datat? (punktskattning)
- Ange ett intervall som med 95% sannolikhet innehåller det korrekta värdet på konstanten (konfidensintervall)
- Kan vi med 99% sannolikhet utesluta att konstanten är noll? (hypotesprövning)

Ett viktigt moment i analysen av datamängder är att visualisera datat (deskriptiv statistik). Vi demonstrerar detta mha datafilen bmrh.dat på kurshemsidan som innehåller födelsevikten (avrundad till närmaste total gram) hos $n=747$ nyfödda spädbarn. Datat är ogrupperat. Om vi bestämmer de absoluta frekvenserna f_i hos de olika förekommande vikterna y_i eller de relativa frekvenserna $p_i = f_i/n$, så får vi

grupperade data som kan åskådliggöras i stolpsdiagram. Eftersom y_i kan anta många olika värden är det praktiskt att dela in datat i klasser och göra ett histogram där höjden av varje stapel är proportionell mot totala frekvensen för vikterna i klassen. Ofta är det intressant att bestämma ett "genomsnittligt värde" eller lägesmått för en datamängd. Ofta används (aritmetiska) medelvärde.

Om $x_1, x_2, \dots, x_n$ betecknar observationerna så ges medelvärdet av $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ (I exemplet är $\bar{x} = 3,40$ kg). Vi är även intresserade av hur mycket datat varierar, dvs ett spridningsmått för datat. De vanligaste spridningsmåten är variansen och standardavvikelsen som ges av

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ resp. } s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{I exemplet är } s = 0,57 \text{ kg})$$

Läs resten av kapitel 10 på egen hand.

Sannolikhetssteorins grunder

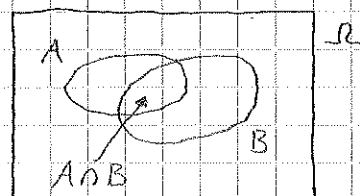
Vi inleder med ett par grundläggande definitioner.

Def: Ett slumpförsök är ett försök där resultatet inte kan avgöras på förhand (ex.vis ett tärningskast med en sexsidig tärning.)

Def: Resultatet av ett slumpförsök kallas för ett utfall (betecknas ω) och mängden av alla möjliga utfall kallas för utfallsrummet (betecknas Ω) (för tärningskastet är ω antalet prickar och $\Omega = \{1, 2, 3, 4, 5, 6\}$).

Def: En händelse (betecknas $A, B, \dots$) är en mängd av utfall, $A \subset \Omega$ (för tärningskastet kan A vara ex.vis "udda antal prickar", dvs $A = \{1, 3, 5\}$).

Eftersom händelser är delmängden av ett utfallsrum kan vi använda mängdoperationerna snitt, union och komplement samt visualisera händelser med Venn-diagram. Om ex.vis $A = \{1, 3, 5\}$ och $B = \{4, 5, 6\}$, så är $A \cap B = \{5\}$. Om två händelser A och B inte kan inträffa samtidigt så sägs de vara disjunta.



Inom sannolikhetssteorin är vi intresserade av sannolikheten för att olika händelser inträffar. För födelseviktdatat från förra timmen gäller att 69% av de nyfödda hade en födelsevikt på mellan 3 och 4 kg. Vi förväntar oss därför att ett nyfött barn med ca 69% sannolikhet väger mellan 3 och 4 kg. Denna sannolikhet bygger på observationer (relativ frekvens). Om vi tar med fler observationer (ex.vis alla nyfödda i Sverige 2016), så kommer den relativa frekvensen att variera men så småningom stabilisera sig på en viss nivå. För att undvika svårigheten att bestämma sannolikheter exakt från relativa frekvenser, så postulerar vi att det till varje händelse A finns ett tal $IP(A)$ som vi kallar sannolikheten för A (ex.vis gäller $IP(\text{"tärningen visar sex prickar"}) = 1/6$) och vi kräver att den relativa frekvensen för A konvergerar mot $IP(A)$ då antalet observationer går mot oändligheten. Detta ger en entydig definition av sannolikheter. Kolmogorov tog 1933 fram följande axiomsystem för sannolikheter.

Axiom: Ett sannolikhetsmått IP är en funktion av händelser sådan att

1. $0 \leq IP(A) \leq 1$ för alla händelser $A \subset \Omega$.
2. $IP(\Omega) = 1$.
3. Om $A_1, A_2, \dots$ är parvis disjunta händelser, så gäller

$$IP\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} IP(A_i) \quad (\text{Om } A \cap B = \emptyset, \text{ så är } IP(A \cup B) = IP(A) + IP(B))$$

Vi kan härleda räkne regler för sannolikheter från axiomen, ex.vis:

$$1 = [Ax. (2)] = IP(\Omega) = [\text{def. av kompl.}] = IP(A \cup A^*) = [Ax. 3, A \cap A^* = \emptyset] = IP(A) + IP(A^*),$$

så $IP(A^*) = 1 - IP(A)$, vilket kallas komplementsatsen. En följsats till denna är att $IP(\emptyset) = IP(\Omega^*) = 0$. Från axiomen kan vi också visa att (se boken)

$$IP(A \cup B) = IP(A) + IP(B) - IP(A \cap B) \quad (\text{Insas från Venn-diagrammet eftersom } IP(A) + IP(B) \text{ "räkar" händelsen } A \cap B \text{ dubbelt.})$$

Om Ω består av m möjliga utfall $\omega_1, \dots, \omega_m$ som alla är lika sannolika, så sätter vi $IP(\omega_i) = 1/m$ och säger att sannolikhetsmålet är likförmigt (jfr tärningskast). Betrakta en händelse A som ges av g olika gynnsamma utfall. Enligt den klassiska sannolikhetsdefinitionen gäller då $IP(A) = g/m$. Kombinatorik används ofta för att bestämma g och m . Exempelvis säger multiplikationsprincipen att om det finns n_1 respektive n_2 sätt att utföra åtgärd ett och två, så kan båda åtgärderna utföras på $n_1 n_2$ sätt.