

Grundläggande regel för sannolikhetsberäkningar:

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

Definition av betingad sannolikhet:

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

Bayes formel:

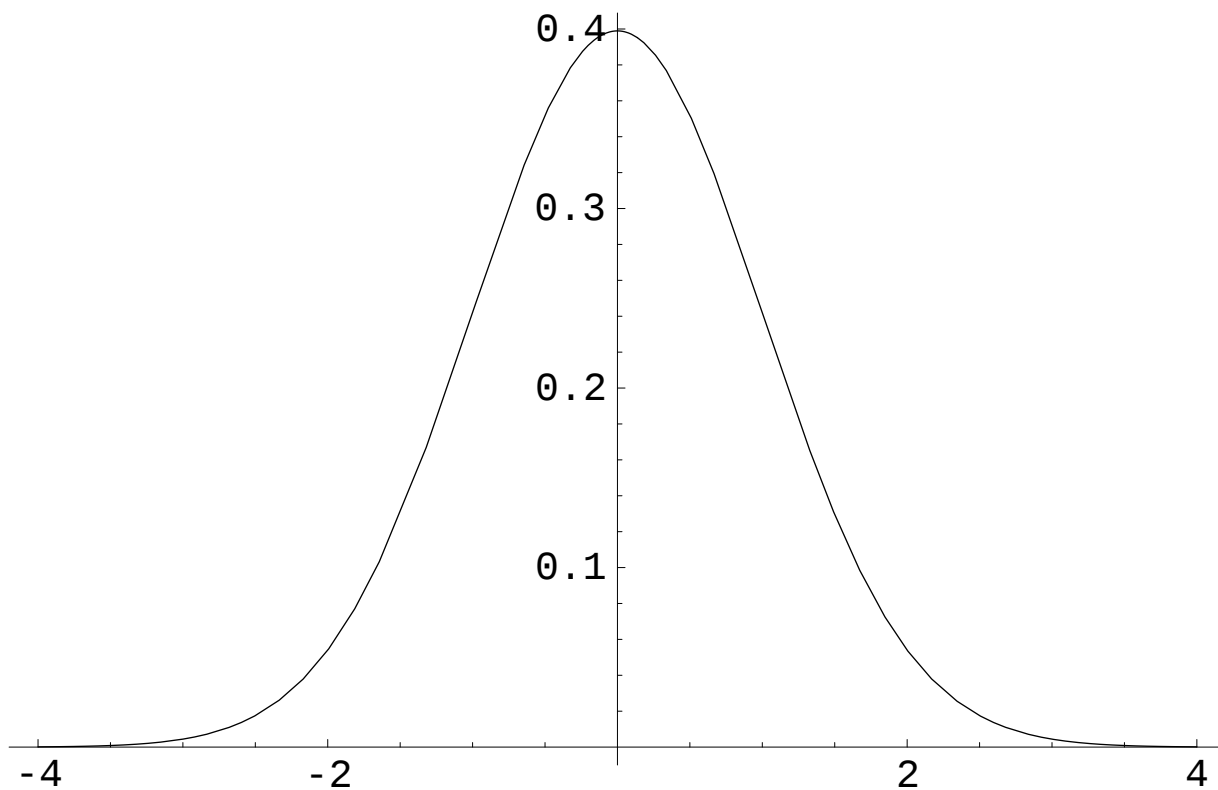
$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\text{icke-}A)P(B|\text{icke-}A)}$$

Den allmänna normalfördelningskurvan:

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

Den standardiserade normalfördelningskurvan:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$



Verklighet:

12.34 7.76 9.06 9.97 11.65
10.29 11.01 6.25 11.07

Vi har 9 mätningar av μ .

Vi tror att mätfelen följer normalfördelningskurvan med okänt σ .

Tänkbara problem:

- punkt- och intervallskatta μ
- visa att $\mu > \mu_0$

Verklighet: Av 100 slumpvis utvalda och kontrollerade bilar visade det sig att 27 st hade början till sprickbildning i en av bromsslangarna fram.

Låt p vara proportionen bilar av den aktuella modellen som har detta fel.

Tänkbara problem:

- punkt- och intervallskatta p
- hur stor kan p som mest vara?

Händelser och deras sannolikheter

Vi ska utföra ett slumpmässigt försök.

Utfallsrummet S är mängden av alla möjliga försöksresultat.

Elementen ω i S kallas för utfall.

Delmängderna $A, B, \dots$ till S kallas för händelser.

Sannolikheten att A inträffar skrives $P(A)$ eller $P(\omega \in A)$.

Låt A, B vara händelser. Då betecknar

$A \cap B$ händelsen A och B

$A \cup B$ händelsen A eller B

$A^c = A'$ händelsen icke- A

Dessutom definierar vi

$$A \setminus B = A \cap B^c$$

$$A \Delta B = A \setminus B \cup B \setminus A$$

Händelserna A och B sägs vara ömsesidigt uteslutande eller disjunkta, om $A \cap B = \emptyset$.

Händelserna $A_1, \dots, A_n$ partitionerar (delar in) en händelse B , om $B \subseteq \bigcup_{i=1}^n A_i$ och $A_i \cap A_j = \emptyset$ då $i \neq j$.

De tre sannolikhetsaxiomen

1. $P(A) \in [0, 1]$
2. $P(S) = 1$
3. $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Räkneregler för sannolikheter

Sats 1:

$$P(A \setminus B) = P(A) - P(A \cap B)$$

Följd 1:

$$P(\emptyset) = 0$$

Följd 2:

$$P(A^c) = 1 - P(A)$$

Följd 3:

$$A \subseteq B \Rightarrow P(A) \leq P(B)$$

Sats 2:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Följd 1:

$$P(A \cup B) \leq P(A) + P(B)$$

Mer om sannolikheter

Händelserna A och B sägs vara oberoende, om

$$P(A \cap B) = P(A) \cdot P(B)$$

Den betingade sannolikheten för A givet B , är

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Observera att

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

Observera att

$$A, B \text{ oberoende} \Leftrightarrow P(A|B) = P(A)$$

Sats (lagen om total sannolikhet): Låt $A_1, \dots, A_n$ partitionera B . Då gäller

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

Sats (Bayes formel): Under samma förutsättningar som i lagen om total sannolikhet, gäller

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

Stokastiska variabler

När utfallet är reellvärt brukar vi beteckna det med stora bokstäver från slutet av alfabetet $(X, Y, Z, \dots)$, och kalla det för en stokastisk variabel.

Sannolikheten att den stokastiska variabeln X antar ett värde i mängden A skrives

$$P(X \in A)$$

Ofta behöver man räkna på flera stokastiska variabler på en gång. Det är då viktigt att kunna hantera sannolikheter typ

$$P(X \in A, Y \in B, Z \in C) \text{ där } A, B, C \subseteq R$$

eller

$$P((X, Y, Z) \in A) \text{ för } A \subseteq R^3$$

Stokastiska variabler har fördelningar. Löst uttryckt kan man säga att fördelningen talar om för oss hur man beräknar sannolikheter för de olika möjliga utfallen.

Endimensionella fördelningar

Vi ska börja med att tänka oss försök där utfallet X är endimensionellt.

En stokastisk variabel X sägs vara diskret, om dess utfallsrum S_X är en diskret mängd. Funktionen

$$f_X(x) = P(X = x) \text{ för } x \in R$$

är X :s massfunktion (täthet, täthetsfunktion, frekvensfunktion). Observera att

$$P(X \in A) = \sum_{x \in A} f_X(x)$$

En stokastisk variabel X sägs vara kontinuerlig, om dess utfallsrum S_X är ett intervall eller en union av intervall, och det finns en funktion $f_X : R \rightarrow [0, \infty)$, benämnd X :s täthet eller täthetsfunktion, sådan att

$$P(X \in A) = \int_A f_X(x) dx$$

Notera att i båda fallen gäller $f_X(x) = 0$ då $x \notin S_X$.

När det är helt klart till vilken stokastisk variabel en viss täthet eller massfunktion hör, glömmer vi ofta att indexera. I sådana fall skriver vi alltså ofta $f(x)$ istället för $f_X(x)$ och S istället för S_X .

Med $\sum_{x \in A} f_X(x)$ avses summation av $f_X(x)$ över alla $x \in A$ sådana att $f_X(x) > 0$.

Fördelningsfunktion

Funktionen

$$F_X(x) = P(X \leq x)$$

kallas X :s fördelningsfunktion.

Observera att

$$F_X(x) = \sum_{y \leq x} f_X(y) \text{ då } X \text{ är diskret}$$

$$F_X(x) = \int_{-\infty}^x f_X(y) dy \text{ då } X \text{ är kontinuerlig}$$

I det kontinuerliga fallet gäller även att

$$f_X(x) = F'_X(x)$$

Och i det diskreta fallet gäller

$$f_X(x) = F_X(x) - F_X(x-)$$

där

$$F_X(x-) = \lim_{y \nearrow x} F_X(y)$$

Definition av väntevärde

För en diskret stokastisk variabel X , sådan att

$$\sum_x |x| f_X(x) < \infty$$

låter vi

$$E[X] = \sum_x x f_X(x)$$

vara X :s väntevärde.

Exempel 1: $X \sim \text{Ber}(p) \Rightarrow E[X] = p$

Exempel 2: $X \sim \text{Geo}(p) \Rightarrow E[X] = 1/p$

Det är vanligt att man kallar ett väntevärde för μ . Vissa läroboksförfattare skriver μ_X istället för $E[X]$. Det är också vanligt att man utelämnar hakparantesen, alltså att man skriver EX istället för $E[X]$, om inga missförstånd kan uppstå.

Definition av väntevärde (forts)

För en kontinuerlig stokastisk variabel X , sådan att

$$\int_{-\infty}^{\infty} |x| f_X(x) < \infty$$

låter vi

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)$$

vara X :s väntevärde.

Observera att stora talens lag säger oss att medelvärdet konvergerar mot väntevärdet då antalet observationer går mot oändligheten. Enda kravet för att stora talens lag ska gälla är att väntevärdet existerar.

Så man skulle kunna kalla väntevärdet för det teoretiska medelvärdet.

Chebyshevs olikhet

Sats: För en stokastisk variabel X med väntevärde μ och varians σ^2 gäller

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

I litteraturen stavas Chebyshev på flera sätt. Blom, t ex, väljer stavningen Tjebyshev.