

Kurssammanfattning

MVE055

Obs: Detta är enbart tänkt som en översikt och innehåller långt ifrån allt som ingår i kursen (vilket anges exakt på hemsidan). Fullständiga antaganden i satser kan saknas och fel kan förekomma så kontrollera allt som är osäkert med boken.

15 oktober 2007

Översikt

- 1 Sannolikhets teori
- 2 Statistik
- 3 Diskret matematik

Grundläggande definitioner

- Experiment: Dra ett spelkort
- Utfallsrum: $S = \{2,3,4,5,6,7,8,9,10,Kn,D,K,A\}$
- Händelser: Delmängder av S . Ex: $B = \{Kn,D,K,A\}$
- Sannolikhetsfunktion: $P(A)$
- Modell: S och P

Räknerregler

Axiom

- $P(S) = 1$
- $P(A) \geq 0, \forall A$
- A_1, \dots, A_n parvis disjunkta
 $\Rightarrow P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$

Räknerregler

- $P(A^C) = 1 - P(A)$
- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Betingade sannolikheter

Betingad sannolikhet

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ om } P(A) \neq 0$$

Oberoende

A, B ober. omm $P(A \cap B) = P(A)P(B)$, dvs om $P(B|A) = P(B)$

Bayes sats

Totala sannolikhetslagen

Om A_1, \dots, A_n är en partition av S så är

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

Bayes sats

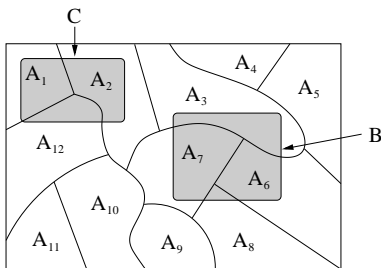
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \text{ om } P(B) \neq 0$$

Ex: Slumpmässigt utvald bil

$A_i = \{\text{bilen tillv. på fabrik } i\}$

$B = \{\text{bilen har ett fel}\}$

$A = A_1$



Diskreta stokastiska variabler

Definition

- X är en s.v. om X bestäms av utfallet
- X diskret om X antar uppräkneligt antal värden

Frekvens- och fördelningsfunktion

- $f_X(x) = \mathbf{P}(X = x)$
- $F_X(x) = \mathbf{P}(X \leq x)$

Krav och samband

- $f_X(x) \geq 0$
- $\sum f_X(x) = 1$
- $F_X(x) = \sum_{i \leq x} f_X(i)$
- $f_X(x) = F_X(x) - F_X(x - 1)$,
om X är heltalsvärd

Mått på s.v.

Väntevärde

- $E[X] = \sum_x x \cdot f_X(x) = \sum_x x P(X = x)$
- $E[h(X)] = \sum_x h(x) f_X(x) = \sum_x h(x) P(X = x)$

Spridningsmått

- $\text{Var } X = E[(X - E[X])^2] = E[X^2] - (E[X])^2$
- $\sigma_X = \sqrt{\text{Var } X}$

Räknerregler

för väntevärden

- $E[a] = a$
- $E[aX] = aE[X]$
- $E[X + Y] = E[X] + E[Y]$
- $E[XY] = E[X]E[Y]$, om X, Y är oberoende

för varianser

- $\text{Var}[a] = 0$
- $\text{Var}[aX] = a^2 \text{Var}[X]$
- $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$, om X, Y är oberoende

Standardfördelningar

Bernoulli(p)-försök

Ett försök som lyckas med sannoliken p .

Geometrisk fördelning

X : Antal (ober.) försök tills lyckat

$$X \sim \text{Geom}(p)$$

Binomial-fördelning

Y : Antal lyckade av n oberoende försök.

$$Y \sim \text{Bin}(n, p)$$

Kontinuerliga s.v.

Definition

Y kont. s.v. med täthetsfunktion $f_Y(y)$ om

- $P(a \leq Y \leq b) = \int_a^b f_Y(y) dy$
- $f_Y(y) \geq 0, \forall y$
- $\int_{-\infty}^{\infty} f_Y(y) = 1$

Samband

- $F_Y(y) = P(Y \leq y) = \int_{-\infty}^y f_Y(t) dt$
- $f_Y(y) = F'_Y(y)$, om F_Y är deriverbar i y

Väntevärde

$$E[Y] = \int_{-\infty}^{\infty} y \cdot f_Y(y) dy$$

Normalfördelningen

Definition

$Y \sim \text{Norm}(\mu, \sigma)$, om:

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

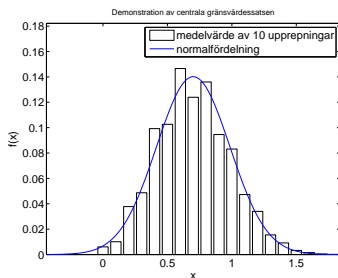
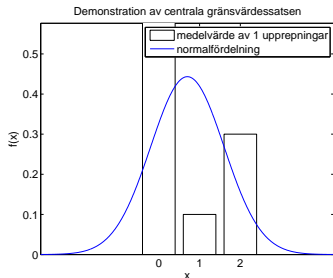
Som approximation

$\text{Bin}(n, p) \approx \text{Norm}(np, \sqrt{np(1-p)})$ om $n \cdot \min(p, 1-p) > 5$

Transformationer

- X normalfördelad $\Rightarrow aX + b$ normalfördelad
- X, Y normalfördelade $\Rightarrow X + Y$ normalfördelade

Centrala gränsvärdessatsen



CGS

Om X_1, \dots, X_n är ober. och likafördelade så är $\bar{X} = \frac{1}{n} \sum X_i$ approximativt normalfördelad.

Tumregel

$n \geq 30$ ger oftast bra noggrannhet.

Flerdimensionella s.v.

Definition

X, Y s.v. på samma utfallsrum (S) har 2-dim. frekvensfunktion
 $f_{XY}(x, y) = P(X = x, Y = y)$

Marginalfördelningar

$$f_X(x) = \sum_y f_{XY}(x, y)$$

Oberoende

X, Y oberoende om $f_{XY}(x, y) = f_X(x)f_Y(y), \forall x, y$

Väntevärde och korrelation

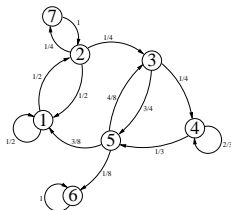
- $E[h(X, Y)] = \sum_x \sum_y h(x, y)f_{XY}(x, y)$
- $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]$
- $\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$

Markovkedjor

Definition

X_0, X_1, X_2, \dots är en Markovkedja med övergångsmatrix P om

$$\begin{aligned} P(X_{n+1} = j | X_n = i, X_{n-1} = a_{n-1}, \dots, X_0 = a_0) = \\ = P(X_{n+1} = j | X_n = i) = P_{ij} \end{aligned}$$



Sannolikhetsvektorer

Om X antar värden $1, \dots, k$ kan fördelningen skrivas som

$\mathbf{u} = [u_1 u_2 \dots u_k]$, där $u_i = P[X = i]$

Markovkedjor II

Steg

s.v.	sannolikhetsvektor
X_0	\mathbf{u}
X_1	$\mathbf{u}P$
X_k	$\mathbf{u}P^k$

Poissonprocesser

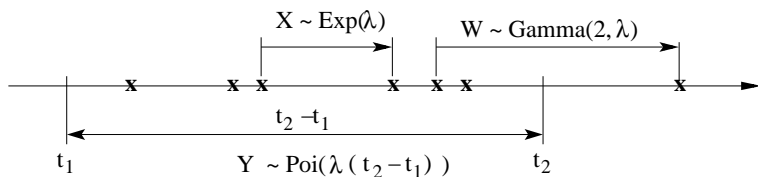
Uppkomst

Punkter utspridda slumpmässigt i tiden/rummet s.a.

- Antal punkter i två ej överlappande intervall är oberoende.
- Om X är antal punkter i ett intervall av längd L så är $E[X] = \lambda L$
- $P[2 \text{ händelser på } \textit{kort tid}] = \textit{liten}$

Poissonfördelning II

Poissonprocess med intensitet λ



Poissonfördelning

Y : antal punkter i intervall av längd L

$$Y \sim \text{Poi}(\lambda L)$$

Exponentialfördelning

X : tid(avstånd) till nästa punkt

$$X \sim \text{Exp}(\lambda)$$

Statistik

Stickprov

X_1, \dots, X_n är slumpmässigt stickprov av X om

- X_1, \dots, X_n är oberoende
- X_1, \dots, X_n har samma fördelning som X

Skattningar

En bra skattare $\hat{\theta}$ för en parameter θ bör vara:

- Väntevärdesriktig: $E[\hat{\theta}] = \theta$
- Lågvariant: $\text{Var } \hat{\theta}$ liten

Allmänna v.v.r. skattningar

- $E[X]$ skattas med $\bar{X} = \frac{1}{n} \sum X_i$
- $\text{Var } X$ skattas med $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$

Konfidensintervall

Definition

Ett stokastiskt intervall (L, U) är ett konfidensintervall för θ med konfidensgrad $1 - \alpha$ om $\forall \theta$:

$$P(L \leq \theta \leq U) = 1 - \alpha$$

Standardkonfidensintervall

$X \sim \text{Norm}(\mu, \sigma)$	μ	$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	Approx. via CGS
$X \sim \text{Norm}(\mu, \sigma)$	μ	$\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$	
$X \sim \text{Norm}(\mu, \sigma)$	σ^2	$\left[\frac{(n-1)S^2}{\chi_{\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2} \right]$	
$X \sim \text{Bin}(n, p)$	p	$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	

Skillnad mellan undersökningar?

Konf.int. för skillnader

$X \sim \text{Norm}(\mu_X, \sigma_X)$ $Y \sim \text{Norm}(\mu_Y, \sigma_Y)$	$\mu_X - \mu_Y$	$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$
$X \sim \text{Norm}(\mu_X, \sigma)$ $Y \sim \text{Norm}(\mu_Y, \sigma)$	$\mu_X - \mu_Y$	$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{S_p^2 \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)},$ $S_p^2 = \frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X + n_Y - 2}$
$X \sim \text{Bin}(n_X, p_X)$ $Y \sim \text{Bin}(n_Y, p_Y)$	$p_X - p_Y$	$\hat{p}_X - \hat{p}_Y \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}$ (Approx. via CGS)

Genererande funktioner

Definition

En serie a_0, a_1, a_2, \dots har gen. fun.

$$A(x) = a_0 + a_1x + a_2x^2 \dots$$

Inversionsformel

$$a_n = \frac{A^{(n)}(0)}{n!}$$

Standardserier

$1, c, c^2, \dots$	$\frac{1}{1-cx}$
$\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \dots$	$(1+x)^n$
$\binom{k}{k}, \binom{k+1}{k}, \binom{k+2}{k}, \dots$	$\frac{1}{(1-x)^{k+1}}$

Momentgenererande funktioner

Definition

$$m_X(t) = \mathbb{E}[e^{tX}] = \frac{\mathbb{E}[X^0]}{0!} + \frac{\mathbb{E}[X^1]}{1!}t + \frac{\mathbb{E}[X^2]}{2!}t^2 + \dots,$$
 om väntevärdet existerar i intervall kring 0.

Inversion

$$\mathbb{E}[X^n] = m_x^{(n)}(0)$$