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 2008

Översikt

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: $\mathrm{P}(A)$
- Modell: S och P

Räkneregler

Axiom

- $\mathrm{P}(S) = 1$
- $\mathrm{P}(A) \geq 0$, $\forall A$
- $A_1, \ldots, A_n$ parvis disjunkta
  $\Rightarrow \mathrm{P}(A_1 \cup \cdots \cup A_n) = \mathrm{P}(A_1) + \cdots + \mathrm{P}(A_n)$

Räkneregler

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

Betingade sannolikheter

Betingad sannolikhet

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

Oberoende

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

Bayes sats

Totala sannolikhetslagen

Om $A_1, \ldots, A_n$ är en partition av S så är

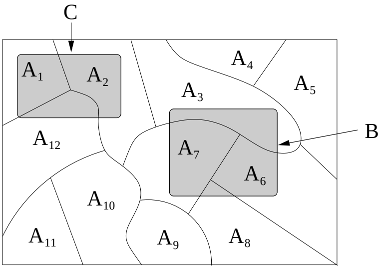
$$\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(B \cap A_i) = \sum_{i=1}^{n} \mathrm{P}(A_i)\,\mathrm{P}(B|A_i)$$

Bayes sats

$\mathrm{P}(A|B) = \dfrac{\mathrm{P}(B|A)\,\mathrm{P}(A)}{\mathrm{P}(B)}$ om $\mathrm{P}(B) \neq 0$

Ex: Slumpmässigt utvald bil

$A_i = \{$bilen tillv. på fabrik $i\}$
$B = \{$bilen har ett fel$\}$
Ta $A = A_1$ i Bayes sats



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

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

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

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

Spridningsmått

- $\operatorname{Var} X = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$
- $\sigma_X = \sqrt{\operatorname{Var} X}$

Räkneregler

för väntevärden

- $\mathrm{E}[a] = a$
- $\mathrm{E}[aX] = a\,\mathrm{E}[X]$
- $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$
- $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$, om X, Y är oberoende

för varianser

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

Standardfördelningar

Bernoulli(p)-försök

Ett försök som lyckas med sannoliket 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

- $F_Y(y) = \mathrm{P}(Y \leq y) = \int_{-\infty}^{y} f_Y(y)dt$
- $f_Y(y) = F'_Y(y)$, om F_Y är deriverbar i y

Väntevärde

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

Normalfördelningen

Definition

$Y \sim$ Norm(μ, σ), om:

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

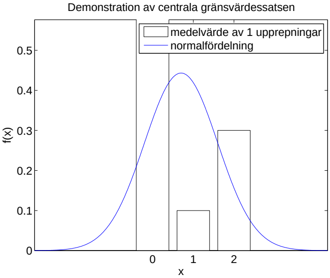
Som approximation

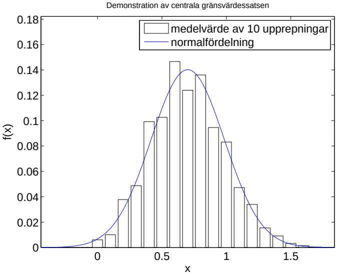
Bin$(n, p) \approx$ Norm$(np, \sqrt{np(1-p)})$ om $n \cdot \min(p, 1-p) > 5$

Transformationer

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

Centrala gränsvärdessatsen





CGS

Om $X_1, \ldots, X_n$ 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

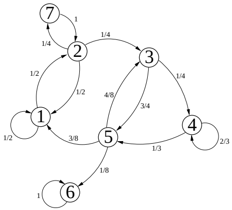
- $\mathrm{E}[h(X,Y)] = \sum_x \sum_y h(x,y) f_{XY}(x,y)$
- $\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$
- $\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$

Markovkedjor

Definition

$X_0, X_1, X_2, \ldots$ är en Markovkedja med övergångsmatris P om

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



Sannolikhetsvektorer

Om X antar värden $1, \ldots, k$ kan fördelningen skrivas som $\mathbf{u} = [u_1 u_2 \ldots u_k]$, 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

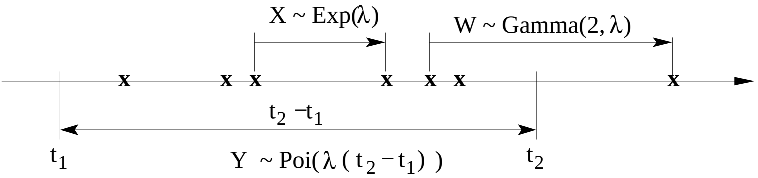
Uppkomst

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 $\mathrm{E}[X] = \lambda L$
- $\mathrm{P}[2 \text{ händelser på } \textit{kort} \text{ tid}] = \textit{liten}$

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

$X_1, \ldots, X_n$ är slumpmässigt stickprov av X om

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

Skattningar

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

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

Allmänna v.v.r. skattningar

- $\mathrm{E}[X]$ skattas med $\bar{X} = \frac{1}{n} \sum X_i$
- $\operatorname{Var} X$ 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 $\forall \theta$:

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

Standardkonfidensintervall

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

Skillnad mellan undersökningar?

Konf.int. för skillnader

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

Genererande funktioner

Definition

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

$$A(x) = a_0 + a_1 x + a_2 x^2 \dots$$

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

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

Inversion

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