SF1915 Sannorlikhet & Statistik

Grundlägande Begrepp

Slumpförsök: Experiment där resultatet inte är på förhand givet.

Utfall: Resultatet av ett slumpförsök

Utfallsrum: Mängden av möjliga utfall. (Ω)

De Morgans lag: Det gäller:

$$(A\cap B)^* = A^* \cup B^*$$
$$(A\cup B)^* = A^* \cap B^*$$

Snitt ($A\cap B$): De delade elementen av två mängder. $A\cap B = \{1,2\}\cap\{2,3\} = \{2\}$

Union ($A\cup B$): De element som fins i en elle båda mängder. $A\cup B = \{1,2\}\cup\{2,3\}=\{1,2,$

Kompliment (A^*): Motsatsen till A alltså 

Distributivitet: Det gäller:

$$A\cap(B\cup C) = (A\cap B)\cup(A\cap C)$$
$$A\cup(B\cap C) = (A\cup B)\cap(A\cup C)$$

Kolmogorovs Axioms

Om:

1) $0 \leq P(A) \leq 1$
2) $P(\Omega) = 1$
3) $A_1, A_2, A_3 \ldots$ är disjunkta

Så gäller:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

$$\Rightarrow P(A^*) = 1 - P(A)$$

Ickedisjunkta Händelser

Disjunkt $\Rightarrow A\cap B = \emptyset$

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

Multiplikationsprincipen:

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

Kombinationsteori

Välja k element från n

	Med återläggning	Utan återläggning
Med ordning	n^k	$\frac{n!}{(n-k)!}$
Utan ordning	$\binom{n+k-1}{k}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

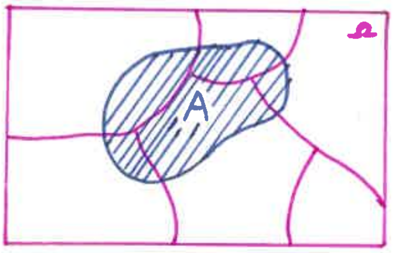
Lagen om total sannolikhet

Låt: $\Omega = H_1 \cup H_2 \cup H_3 \cup \dots \cup H_n$

$H_x \cap H_y = \emptyset$

Då gäller:

$$P(A) = \sum_{i=1}^{n} P(A|H_i) \cdot P(H_i)$$



Oberoende Händelser

A och B är oberoende om:

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

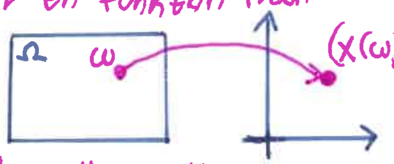
Det gäller också

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

två händelser som är disjunkta kan inte vara oberoende.

Flerdimintionella Variabler

En n-dimintionell stokastisk variabel X är en funktion från Ω till $\mathbb{R}^n$



Det gäller att:

$$P((X,Y) \in A) = \begin{cases} \sum_{(J,K) \in A} P_{XY}(J,K) = \sum_{(J,K) \in A} P(X=J, Y=K) \\ \iint_A f_{XY}(x,y)\, dx\, dy \end{cases}$$

Oberoende Flerdim. Variabler

X och Y är oberoende om

$$P(X \in C, Y \in D) = P(X \in C) P(Y \in D)$$

$\forall C, D$

Bayes Sats

Under samma sammanfattning som till vänster gäller också:

$$P(H_i|A) = \frac{P(A|H_i) \cdot P(H_i)}{P(A)}$$

$$\Leftrightarrow$$

$$P(H_i|A) = \frac{P(A|H_i) \cdot P(H_i)}{\sum_{j=1}^{n} P(A|H_j) \cdot P(H_j)}$$

Stokastiska Variabler

En stokastisk variabel X är en funktion från utfallsrum Ω till $\mathbb{R}$

$$X : \Omega \to \mathbb{R}$$

En diskret stokastisk variabel är när Ω är ändlig eller uppräkneligt oändlig.

$$P_X(k) = P(X=k) \quad k \in \Omega$$

En kontinuerlig stokastisk variabel är när Ω är överuppräknelig. Då gäller:

$$P(x \in \underset{\text{interval}}{A}) = \int_A f_X(t)\, dt$$

Det gäller:

$$\int_{-\infty}^{\infty} f_X(t)\, dt = 1 \qquad f_X(t) := P(X \leq t)$$

Täthetsfunktion

$f_X(x)$ is a contineous function of the probability of x for any interval.

$$\int_a^b f_X(x)\, dx = \underbrace{P(a \leq X \leq b)}_{\text{Fördelningsfunktion}}$$

Sannolikhetsfunktion

Det gäller:

$$X = P_X(x) = P(X = x) \qquad \sum_{x \in \Omega} P(x) = 1$$

Fördelningsfunktion

Det gäller

$$F_X(x) = P(X \leq x) = \sum_{\substack{\forall \text{ utfall} \\ \leq x}} P(x) \quad \in [0,1]$$

är alltid mellan 0 och 1

Väntevärden

Väntevärdet för den s.v. X definieras som:

$$\mu = E[X] := \begin{cases} \sum_k k \cdot p_X(k) & \text{diskret} \\ \int_{-\infty}^{\infty} x \cdot f_X(x) & \text{Kontinuerlig} \end{cases}$$

Varians

Variansen för den s.v. X definieras som:

$$V(X) := E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Standardavvikelsen

för en s.v. X så är och betecknas:

$$D(X) := \sqrt{V(X)} = \sigma$$

Räkneregler

om (a,b) är konstanter då gäller:

$$E[aX+b] = aE[X] + b$$

och:

$$V(aX+b) = V(aX) = a^2 V(X)$$

om X, Y är stokastiska variabler då gäller:

$$E[X+Y] = E[X] + E[Y]$$

$$V(X+Y) = V(X) + V(Y) + \underbrace{2C(X,Y)}_{\text{Covarians}}$$

och om de är oberoende:

$$E[XY] = E[X] \cdot E[Y]$$

Covarians

Covarians betecknar variansen mellan två s.v. X, Y

$$C(X,Y) = E[XY] - E[X]E[Y]$$

(mäter det linjära beroendet mellan X och Y.

$$V(X) = C(X,X)$$

Korrelationskoefficient

$$\rho(X,Y) := \frac{C(X,Y)}{D(X)D(Y)} \in [-1,1]$$

$$C(aX+bY,\ cZ+dW) = ac \cdot C[X,Z] + ad \cdot C[X,W] + bc \cdot C[Y,Z] + bd \cdot C[Y,W]$$

Stora Talens Lag

För alla $\varepsilon > 0$ gäller att

$$P(\mu - \varepsilon < \bar{X}_n < \mu + \varepsilon) \to 1$$

när $n \to \infty$

Medelvärdet tenderar mot μ för fler och fler försök.

Fel

För alla mätningar gäller:

$$\bar{X} = \Theta + \delta + \varepsilon$$

Uppmätt värde, Korrekt värde, Systematiskt fel, Slumpmässigt fel

$$E[\bar{X}] = E[\Theta + \delta + \varepsilon] = \Theta + \delta = \mu$$

$$\delta = \mu - \Theta \qquad \varepsilon = \bar{X} - E[\bar{X}] = \bar{X} - \mu$$

Precision: ε är låg

noggranhet: δ är låg

Tvåpunktsfördelning/Bernoullifördelning

Om $\Omega=\{0,1\}$ då gäller $p_X(1)=a$

och $p_X(0)=1-p$

För första gången fördelning

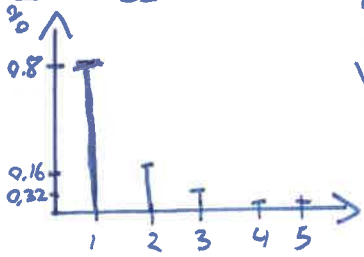
Kontinuerliga försök av ett slumpvart oberoende event tills ett gynnfall händelse.

$p_X(1)=a$ $\quad p_X(2)=(1-a)\cdot a$ $\quad p_X(3)=(1-a)^2\cdot a$ $\quad \dots$

$$\Rightarrow p_X(n)=p_X(1)\cdot\left(1-p_X(1)\right)^{n-1}$$

$$E(X)=\frac{1}{p}$$

$$V(X)=\frac{1-p}{p^2}$$



Hypergeometrisk Fördelning

Dragning utan återläggning av n från N enheter där p är andelen enheter med egenskap A. Då är X andelen enheter med egenskapen a.

N = Totala enheter $\quad$ n = antal dragningar

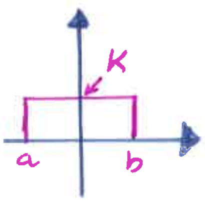
a = en egenskap $\quad$ p = andelen från N med egenskap a.

$$p_X(k)=\frac{\text{gynsamma}}{\text{möjliga}}=\frac{\binom{Np}{k}\cdot\binom{N(1-p)}{n-k}}{\binom{N}{n}} \qquad E(X)=np$$

Likformig fördelning (Kontinuelig)

Täthetsfunktionen är konstant under hela intervallen. Eftersom $\int_{-\infty}^{\infty} f_X(x)\,dx=1$ så är höjden:

$$K=\frac{1}{b-a}$$



$$E(X)=\frac{a+b}{2}$$

$$V(X)=\frac{(b-a)^2}{12}$$

$X\in U(a,b)$

Likformig fördelning (Diskret)

Alla fördelningar är lika stor sannolikhet.

$|\Omega|=n \Rightarrow p_X(x)=\frac{1}{n} \quad x\in\Omega$

Binomialfördelning

Oberoende försök av N händelser med samma sannolikhet. och X är antalet lyckade försök.

$X\in Bin(n,p)$ $\quad$ n - antalet försök, p - gynsamhet sannolikhet.

ordning spelar ingen roll:

●●○ = ●○● = ○●●

3 sätt

$$p_X(k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$$

$$E(X)=np \qquad V(X)=\frac{1-p}{p^2}$$

Poissonfördelning

Antar att vi har en intervall $[0,t]$ och att ett event har uniform sannolikhet att hända under hela intervallen. X är då antalet händelser under $[0,t]$ om händelserna har en intensitet λ som är $\lambda=\frac{\text{händelser}}{\text{tidsenhet}}$

$$p_X(k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t} \qquad \mu=\lambda\cdot t$$

$$\Leftrightarrow$$

$$p_X(k)=\frac{\mu^k}{k!}e^{-\mu}$$

$$E(X)=\mu \qquad V(X)=\mu$$

Exponentialfördelning

antal att en händelse har lika stor förhandelse över en intervall för varje enhet oberoende om den har inträffat innan. Då har vi Poisson fördelning men om den s.v. T är tiden till nästa händelse så gäller det att T är en exponentiell fördelning av X

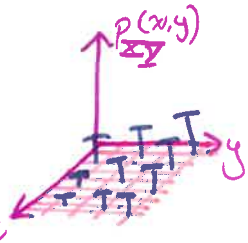
$$f_T(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & \text{annars.}\end{cases}$$

Flerdimensionella S.V. (Diskret)

givet de S.V. X och Y då är deras diskreta sannolikhets funktion

$p_{XY}(x,y)$ $\qquad \sum_{\forall(x,y)} p_{XY}(x,y) = 1$

Det gäller också:



$$p_X(x) = \sum_{\forall y} p_{XY}(x,y)$$

Samma för $p_Y(y)$

om oberoende:

$$p_{XY}(x,y) = p_X(x) \cdot p_Y(y)$$

Allmänt:

$$F_{XY}(x,y) = P(X \leq x \cap Y \leq y)$$

Flerdiminsionella S.V. (kontinuerlig)

givet X Y SV då är deras kontinuerliga sannolikhetsfunktion:

$$P_{X,Y}(x,y) = \iint_{(x,y)\in D} f_{XY}(x,y)\,dx\,dy \qquad \iint_{-\infty}^{\infty} f(x,y) = 1$$

det gäller också:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy \qquad \text{Samma för } f_Y(y)$$

Och om oberoende:

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

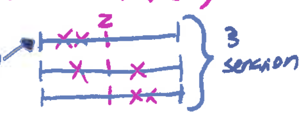
allmänt

$$F_{XY}(x,y) = P(X \leq x \cap Y \leq y)$$

Max och Min

givet X Y som oberoende S.V. och $Z = \max(X,Y)$ då är

$$F_Z(z) = P(Z \leq z) = P(\max(X,Y) \leq z)$$



$$P(Z \leq z) = p(X \leq z) \cdot p(Y \leq z)$$

$$= F_X(z) \cdot F_Y(z)$$

givet X Y som oberoende S.V. och $Z = \min(X,Y)$

$$F_Z(z) = P(\min(X,Y) \leq z) = P(\text{minsta värdet} \leq z)$$

$$= 1 - P(\text{alla värden} > z) = 1 - P(X > z) \cdot P(Y > z)$$

$$= 1 - (1 - F_X(z))(1 - F_Y(z))$$

Summa (Diskret)

anta att vi har $p_X(x)$ och $p_Y(y)$ och vi har $Z = X + Y$ och $p_Z(z)$

$$p_Z(k) = P(X+Y = k)$$

$$= \sum_{j=0}^{k} p_X(j) \cdot p_Y(k-j)$$

Summa (kontinuerligt)

anta att X, Y S.V. och att $Z = X + Y$, $f_X(x)$ och $f_Y(y)$ och vi söker $F_Z(z)$

$$F_Z(z) = P(x + z \leq z)$$

$$= \iint_{x+y \leq z} f_{XY}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_X(x) \cdot f_Y(y)\,dy\,dx$$

Normalfördelning

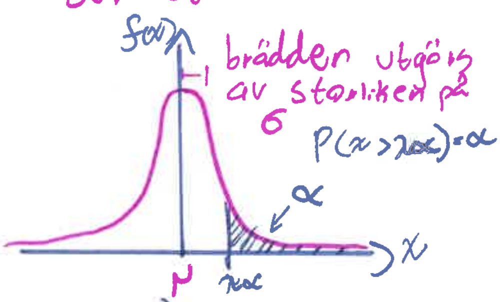
betecknas som:

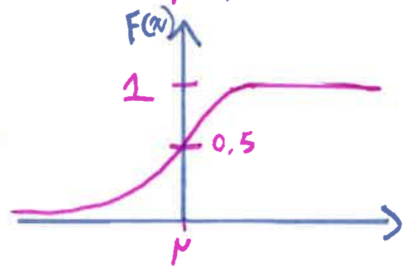
$X \in N(\mu, \sigma)$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

Ser ut som





Standardnormalfördelning

$X \in N(0,1)$

finns som tabell

$\varphi(x) \quad \Phi(x)$

generellt så gäller för alla $X \in N$

$E(X) = \mu$

$V(X) = \sigma^2$

$D(X) = \sigma$

$E(X) = \mu + \sigma E(Y)$

Allmänn Normalfördelning

Det gäller:

$$f_X(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

$N(0,1)$

och

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Det är alltså möjligt att moddellera varje godtycklig normalfördelning med standard normalfördelningen.

Medelvärde av S.V.

Om en stokastisk variabel (inte nödvändigtvis normalfördelad) mäts många gånger så kommer medelvärdet av mätningarna ta form utav en normalfördelning

Linjärkombinationer

för normalfördelning av oberoende variabler är kombinationer ganska lätta.

Multiplikation och addering av konstanter:

$X \in N(\mu, \sigma)$

$Y = aX + b \in N(a\mu + b, |a|\sigma)$

Addition och subtraktion av oberoende s.v.

$$X + Y \in N\left(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$

$$X - Y \in N\left(\mu_X - \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$

Generellt:

$$\sum_{i=1}^{n} a_i X_i + b_i \in N\left(\sum_{i=1}^{n} a_i X_i + \sum_{i=1}^{n} b_i, \sqrt{\sum_{i=1}^{n} a_i^2 \sigma_i^2}\right)$$

Medelvärden för $\{x_1, x^2 \ldots x_n\}$, $x_i \in N(\mu, \sigma)$

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Följdsats

$\{X_1, X_2 \ldots X_{n_1}\} \in N(\mu_x, \sigma_x)$ $\{Y_1, Y_2 \cdots Y_{n_2}\} \in N(\mu_y, \sigma_y)$

$$\bar{X} - \bar{Y} \in N\left(\mu_x - \mu_Y, \sqrt{\sigma_x^2/n_1 + \sigma_y^2/n_2}\right)$$

Centrala Gränsvärdesatsen

Normalfördelning uppträder ofta i statistik på oväntade ställen. Nämligen är summan av oberoende s.v. ungefär normalfördelad. så länge som antalet komponenter är stort nog.

Sats:

Om $X_1, X_2 \dots$ är en oändlig följd av oberoende likfördelade s.v. med Väntevärdet μ och standardavikelsen σ så gäller:

$$Y_n = X_1 + \dots X_n$$

$$P\left(a < \frac{Y_n - n\cdot\mu}{\sigma\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a)$$

när $n \to \infty$

Halvkorrektion

När man approximerar en diskret fördelning till kontinuerlig uppstår ett problem angående intervaler.

För diskret gäller.

$$P_X(x < 120) \neq P_X(x \leq 120)$$

Medans för kontinuerlig är de lika Alltså använder man medelvärdet.

$$x = \frac{120 + 119}{2} = 119.5$$

Aproximera Bin(n,p) till Po(np)

Dela upp intervallen av bin tills varje delintervall är så litet att sannolikheten är p för en händelse.

Detta gäller om:

$$p \leq 0.1$$

Approximera Hyp till Bin

När man drar ett från ett mycket stort N så ändras sannolikheten inte mycket. Till exempel att man drar en person från sveriges befolkning så är sannolikheten att nästa person har blåa ögon nästan helt orörd

Det går att approximera Hyp(N,n,p) till Bin(n,p)

$$\frac{n}{N} \leq 0.1$$

Approximera Bin(n,p) till $N(n\cdot p, \sqrt{n\cdot p(1-p)})$

Om $X \in Bin(n,p)$ där X är antalet lyckade försök $X = I_1 + I_2 \dots I_n$ $\quad I_i = 1$ om lyckat, $I_i = 0$ om ej lyckat

$P(I_k = 0) = 1 - p \quad P(I_k = 1) = p$

$$\Rightarrow E[X] = p\cdot k \qquad E[I_k^2] = p$$

$$V[X] = V[I_k]\cdot n \qquad V[I_k] = p - p^2$$

Detta gäller om:

$np(1-p) \geq 10$ $\qquad$ Ju mer likfördelat desto mindre n.

Approximera Po(μ) till $N(\mu, \sqrt{\mu})$

Man kan dela upp Possion intervallen i flera intervaller och tillslut får man en Normalfördelning. Detta gäller om.

$$\mu \geq 15$$

Sammanfattning

$$\begin{tikzcd}
Hyp \arrow[r, "\frac{n}{N} \leq 0.1"] & Bin \arrow[r, "np(1-p) \geq 10"] \arrow[d, "p \leq 0.1"] & N \\
 & Po \arrow[r, "\mu \geq 15"'] & N
\end{tikzcd}$$

Ogrupperad Data (Tabulering)

(Raw data) är en samling av mätvärden i en lista. Tex antalet tändstickor i en tändsticksask.

51 52 53 51 50 49 50 51 49 52

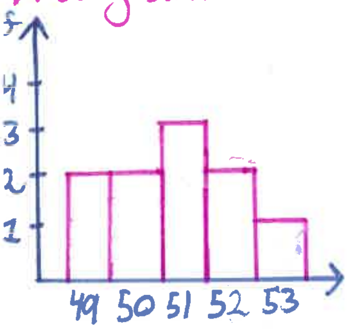
Grupperad data är däremot är samma data fast grupperad i frekvens
Till exempel en Frekvenstabell

Klass	Frekvens	Relativfrekvens (%)
49	2	20
50	2	20
51	3	30
52	2	20
53	1	10

Vid indelning av klasser gäller:

1) Klassbredd bör vara konstant
2) öppna klasser bör undvikas. eg 50+
3) Varje mätvärde ska tydligt tillhöra en grupp eller klass.

Histogram:



Kovarians

Ogrupperad Data (Läge/spridningsmått)

Om $x_1, x_2 \dots x_n$ är data så definieras medelvärde:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1+x_2+\dots+x_n}{n}$$

Spridningsmått/stickprovsvariansen är då

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i-\bar{x})^2$$

Då är stickprovsavvikelsen.

$$s=\sqrt{s^2}$$

och kvadratsumman:

$$Q=\sum_{i=1}^{n} (x_i-\bar{x})^2 = (n-1)s^2$$

Undre/övre kvartilen är mittenvärdet av den undre/övre halvan av de sorterade mätvärden.

49 49 50 50 51 | 51 51 52 52 53

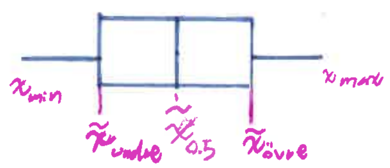
$\tilde{x}_{undre}$ Median $\tilde{x}_{övre}$

Kvartilavstånd: $\tilde{x}_ö - \tilde{x}_u = 52-50=2$

Kvartilintervall: $(\tilde{x}_u, \tilde{x}_ö) = (50, 52)$

Variansbredd: $x_{max}-x_{min} = 53-49=4$

Boxplot



För grupperad data kan medel och andra värden lätt hittas.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i y_i$$

k- antalet grupper
f- frekvensen
y- mätvärde

$$Q=\sum_{i=1}^{k} f_i(y_i-\bar{x})^2 = \sum_{i=1}^{k} f_i y_i^2 - n\cdot\bar{x}^2$$

Om grupperna är klassindelade så blir y_i klassens mittelvärde

Korrelation

Korrelation mäter hur mycket två variabler följer varandra. Den mäts med formeln:

$$C_{x,y} = \frac{1}{n+1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{n+1}\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

och korrelationskoefficienten mäts med formeln:

$$r = \frac{C_{xy}}{S_x \cdot S_y} \qquad r \in (-1,1)$$

Konsistens

Om för alla $\theta \in \Omega$ och för varje givet $\varepsilon > 0$

$$P(|\theta_n^* - \theta| > \varepsilon) \to 0$$

Då $n \to \infty$, säges θ^*_{obs} vara konsistent.

Maximum Likelihood-Metoden

Objektivt sätt att matcha en funktion med mätdata.
Antagande: De mätvärden man får är det mest troligaste mätvärdet.
Maximera sedan θ^* så att antagandet stämmer

Diskret:
$$L(\theta) = P(X_1 = x_1, X_2 = x_2 \dots X_n = x_n)$$

Kontinuerlig:
$$L(\theta) = f_{X_1 X_2 \dots X_n}(x_1, x_2 \dots x_n)$$

Värdet av θ^*_{obs} som $L(\theta)$ har det högsta värdet kallas ML-skattningen

Minsta kvadratmetoden

Låt $X_i = \mu_i(\theta) + \varepsilon_i$ Alltså X är en känd funktion av θ plus ett försöksfel, som ofta kan antas vara noll. Låt

$$Q(\theta) = \sum_{i=1}^{n}[x_i - \mu_i(\theta)]^2$$

Värdet θ^*_{obs} för vilket $Q(\theta)$ tar det minsta värdet kallas för MK skattningen av θ. Detta görs med derivatan.

Puktskattning θ^*_{obs}

En punktskattning av en parameter är en funktion av mätdatan $x_1, x_2 \dots x_n$ som är utfall av de stokastiska variablerna $X_1, X_2 \dots X_n$

$$\theta^*_{obs} = \theta^*(x_1, x_2, \dots x_n)$$

Stickprovsvariabel θ^*

En stickprovsvariabel är en funktion av en samling stokastiska variabler. Den kan realiseras av en skattning.

$$\theta^* = \theta^*(X_1, X_2 \dots X_n)$$

Givet en punktskattning θ^*_{obs} så sägs det vara så att θ^*_{obs} är väntevärdesriktig om.

$$E(\theta^*) = \theta \qquad \forall \theta \in \Omega_\theta$$

"skattningen har rätt värde"
Detta kan vara ett nödvändigt antagande

Medelkvadratfelet

För en punktskattning θ^*_{obs} är:

$$MSE = E((\theta^* - \theta)^2)$$

Det gäller att

$$MSE = E((\theta^* - \theta)^2) = E((\theta^* - E(\theta^*) + E(\theta^*) - \theta^2))$$
$$= V(\theta^*) + \underbrace{(E(\theta^*) - \theta)^2}_{\text{systematiskt fel}}$$

om skattningen är väntevärdesriktig dvs om $E(\theta^*) = \theta$ då är $MSE = V(\theta^*)$

Väntevärdesriktighet

en skattning är v.v.r av θ (θ^*_{obs} av θ) om:

$$E(\theta^*) = \theta$$

Alltså θ^* beskriver den stora θ och kan därför användas för att estimera θ.

Effektivitet

Om flera väntevärdesriktiga skattningar $(\Theta_1^*, \Theta_2^*, \ldots \Theta_n^*)$ av Θ då är den skattningen Θ_k^* med lägst varians den mest effektiva skattningen av Θ.

Konfidensintervall

En intervall I_Θ som med sannolikhet $(1-\alpha)$ täcker över Θ kallas för en konfidensintervall med konfidensgrad $(1-\alpha)$.

$$P(\Theta \in I_\Theta) = 1-\alpha$$

$$P(a_1(\bar{X}) < \Theta < a_2(\bar{X})) = 1-\alpha \Rightarrow (a_1(\bar{X}), a_2(\bar{X}))$$

För normalfördelning gäller:

$$P\left(\frac{-a}{\sigma/\sqrt{n}} < Y < \frac{a}{\sigma/\sqrt{n}}\right) = 1-\alpha$$

Konfidensintervallen beror på 3 saker

1) stor spridning på mätdatan stort (σ) ger bredare intervall.
2) Flera observationer ger smalare konfidensintervall.
3) Större säkerhet att μ är i intervallen betyder att intervallen blir bredare.

Låt $x_1 \ldots x_n$ vara ett slumpmässigt stickprov från $N(\mu, \sigma)$ där μ är okännt. Då är:

$$I_\mu = (\bar{x} - \lambda_{\alpha/2} D, \bar{x} + \lambda_{\alpha/2} D) \qquad D = \frac{\sigma}{\sqrt{n}}$$

om σ är känt.

$$I_\mu = (\bar{x} - t_{\alpha/2}(f) d, \bar{x} + t_{\alpha/2}(f) d) \qquad d = \frac{s}{\sqrt{n}} \quad f = n-1$$

om σ är okännt

en tvåsidig konfidensintervall för μ med konfidensgrad $1-\alpha$

Medelfelet

En skattning av $D(\Theta^*)$ kallas medelfelet för Θ^* och betecknas $d(\Theta^*)$. Medelfelet för skattningen är alltså

$$D^*(\Theta^*)_{obs}$$

Approximerad konfidensintervall

Även när fördelningen för Θ inte är normalfördelad går det att approximera en konfidensintervall för väntevärdet förutsatt att stickprovet är stort nog. Man kan anta att Θ^* är en normalfördelning med väntevärde Θ och standardavvikelse $D = D(\Theta^*)$ då gäller.

$$\frac{\Theta^* - \Theta}{D} \in N(0,1)$$

Då gäller också:

$$I_\Theta = (\Theta^* - \lambda_{\alpha/2} D, \Theta^* + \lambda_{\alpha/2} D) \quad \text{om } D \text{ ej beroende av } \Theta$$

$$I_\Theta = (\Theta^*_{obs} - \lambda_{\alpha/2} d, \Theta^*_{obs} + \lambda_{\alpha/2} d) \quad \text{om } D \text{ är beroende av } \Theta$$

för konfidensintervall av Θ med approximationen konfidensgraden $1-\alpha$

Chi kvadrat

om $(X_1, X_2 \ldots X_n) \in N(0,1)$ och de är oberoende

$$\sum_{i=n}^{n} X_i^2 \in \chi^2(f)$$

$$\Rightarrow \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \in \chi^2(f)$$

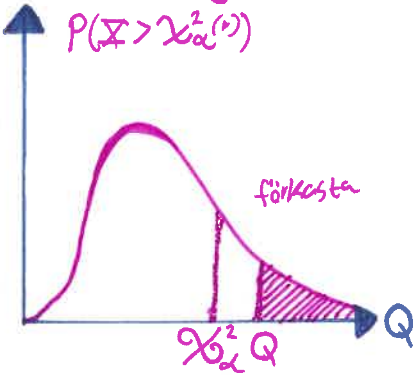
för $X_i \in N(\bar{x}, \sigma)$

χ^2 Test

Test av en given fördelning. Används när ett H_0 har en fördelning efter en sannolikhetsfunktion.

$$Q=\sum_{i=1}^{r}\frac{(x_i-np_i)^2}{np_i}$$

Om Q är stort förkastas H_0. $\Rightarrow Q > \chi^2_\alpha(f)$ (tabell 4)



Homogenitetstest

Samma sannolikhetsfunktion i olika grupper. (H_0)

$$Q=\sum_{i=1}^{s}\sum_{j=1}^{r}\frac{\left(x_{ij}-\frac{n_i m_j}{N}\right)^2}{\left(\frac{n_i m_j}{N}\right)}$$

n_i = försök i serie i
m_j = antal resultat j
N = totala mätningar

Alla $\frac{n_i m_j}{N} \geq 5$

Oberoendetest

Om man vill undersöka om egenskaper A och B är oberoende där A har utfallen $A_1, A_2 \dots A_r$ och B har utfallen $B_1, B_2 \dots B_s$ och H_0 är att A och B är oberoende.