Partiell integration

$$\int u(x)\cdot v'(x)\,dx = u(x)\cdot v(x) - \int u'(x)\cdot v(x)\,dx$$

Funktion	Derivata
x^n	nx^{n-1}
a^x $(a>0)$	$a^x \ln a$
$\ln x$ $(x>0)$	$\frac{1}{x}$
e^x	e^x
e^{kx}	ke^{kx}
$\frac{1}{x}$	$-\frac{1}{x^2}$
$k\cdot f(x)$	$k\cdot f'(x)$
$f(x)+g(x)$	$f'(x)+g'(x)$
$f(x)\cdot g(x)$	$f(x)\cdot g'(x)+f'(x)\cdot g(x)$ (Produktregel)
$\frac{f(x)}{g(x)}$ $(g(x)\neq 0)$	$\frac{f'(x)\cdot g(x)-f(x)\cdot g'(x)}{(g(x))^2}$ (Kvotregel)

Om $y=f(z)$ och $z=g(x)$, så är derivatan av $y=f(g(x))$:

$$y'=f'(g(x))\cdot g'(x)$$

Fördelning av en slumpvariabel (SV): KAPITEL 3.1 SIDA 91

- En kontinuerlig sv kan anta alla värden inom ett intervall (t.ex. tid, längd eller vikt).

En diskret sv kan endast ta ett ändligt eller uppräkneligt(countable) oändligt antal värden (t.ex. x1, x2, x3,...). **Kontinuerlig**

Kontinuerlig fördelning och PDF: KAPITEL 3.2 SIDA 100

För kont.sv gäller täthetsfunktion (Probability Density Function), tätheten ≠ sannolikhet och har följande egenskaper:

$Pr(a \le X \le b) = \int_a^b f(x)\,dx$

$f(x)\ge 0 \;\; för\ alla\ x$

$\int_{\{-\infty\}}^{\infty} f(x)dx = 1$

KUMULATIV FÖRdelningsfunktion (CDF) KAPITEL 3.3 SIDA 107

Defineras enligt (samma för kont/disk): $F(x)=\Pr(X\le x)\ for\ -\infty<x<\infty$

CDF är alltid kontinuerlig från höger: $F(x)=F(x^+)=\Pr(X\le x)$

Sannolikheter för intervall: $P(x_1<X\le x_2)=F(x_2)-F(x_1)$

SAMBANDET med KONT SV (PDF OCH CDF):

$F(x)=\int_{-\infty}^{x} f(t)\,dt$ och konsekvent är $f(x)=dF(x)/dx$

PDF OM VI HAR EN MARGINAL FÖRDELNING

Om f(x,y) är kontinuerlig bivariat fördelning (joint p.d.f) med stöd ex x>0 och y>0, och 0<2x+y<4. Om vi då vill få t.ex f(x) måste vi integrera med avseende på y (dy) där vi hade integrerat på stödet [0;4-2x].

Föreläsning 3:

Kvantiler: det borde stå minus ett dvs invers

p-kvantil = $F^{-1}(p)$ är det minsta värde av x där $\boldsymbol{F(x)\ge p}$

Median: $F^{-1}(0.5)$, Nedre kvartil: $F^{-1}(0.25)$, Övre kvartil: $F^{-1}(0.75)$.

Funktion av en SV KAPITEL 3.8 SIDA 167

Fördelningen för X är känd men behöver hitta fördelning av $Y=r(X)$

DISKRET: $Y=r(X)$, f(x) är PF för X, g(y) är PF för Y: $\boldsymbol{g(y)=P(Y=y)=P(r(X)=y)=\sum_{x:r(x)=y} f(x)}$. Summan av alla f(x) för alla x-värden som uppfyller r(x) =y.	KONTINUERLIG Y och X (Y=r(X) är en 1:1 funktion och deriverbar på X:s stöd (a,b)) ➔Inversen X= $r^{-1}(Y)=s(Y)$. CDF för Y: $G(y)=\Pr(Y\le y)=\Pr[r(X)\le y]=Pr[X\le s(y)]$=$\int_a^{s(y)} f(x)dx = F[s(y)]-F[a]$.	$g(y)=f(s(y))*\left\lvert\frac{d(s(y))}{dy}\right\rvert$ $för\ \alpha<y<\beta$. Alternativt (Ex): Derivera G(y) = F(lny) ➔g(y)=f(lny)*1/y. **Formel om linjärkomb:** $g(y)=\frac{1}{abs(b)}*f\left(\frac{y-a}{b}\right)$

Föreläsning 7: MARKOV/CHEBY ÄR KAP 6 SIDA 248

Markovs olikhet är ett verktyg inom sannolikhetsteori som ger en övre gräns för sannolikheten att en icke-negativ slumpvariabel X antar ett värde större än eller lika med ett givet tal t. För alla t > 0: $\boldsymbol{\Pr(X\ge t)\le \frac{E(X)}{t}}$.

Chebychevs olikhet är en sannolikhetsgräns som mäter hur stor sannolikheten att en slumpvariabel X avviker från sitt väntevärde E(X) med minst ett visst värde t. Den använder variansen Var(X) som mått på spridningen och gäller för alla fördelningar med ändlig varians. För alla t > 0:

$P(\lvert X-E[X]\rvert\ge t)\le \dfrac{Var[X]}{t^2}$	**ALT:** $\Pr(\lvert X-E(X)\rvert\ge k\sigma)\le\frac{1}{k^2}$

Konvergens i sannolikhet: En följd av slumpvariabler $Z_1, Z_2, \ldots$ konvergerar i sannolikhet mot ett värde b om för varje $\varepsilon>0$: $\boldsymbol{\lim_{n\to\infty}\Pr(\lvert Z_n-b\rvert<\varepsilon)=1}$.Tolkning: Ju större n blir, desto större sannolikhet att Z_n ligger inom ett godtyckligt litet intervall $\boldsymbol{(b-\varepsilon, b+\varepsilon)}$. Notation: $Z_n \xrightarrow{p} b$. Om $Z_n\xrightarrow{p} b$ och *g(z)* är en **kontinuerlig funktion** i punkten z = b, då gäller:$\boldsymbol{g(Z_n)\xrightarrow{p} g(b)}$.
Konvergens i sannolikhet är svagare: Den säger inget om *hur snabbt* Z_n närmar sig *b*, bara att sannolikheten ökar med n.

Stora talens lag **Slumpmässigt stickprov:** $\boldsymbol{X_1,\ldots X_n}$ är oberoende och identiskt fördelade (i.i.d) från en fördelning med **ändligt medelvärde** ($\boldsymbol{\mu=E(X_i)}$) och **ändlig varians** ($\boldsymbol{Var(X_i)<\infty}$). $\boldsymbol{\bar{X}_n=\frac{X_1+\cdots+X_n}{n}\xrightarrow{p}\mu}$ **när** $\boldsymbol{n}\to\infty$. Ju fler observationer (*n*) vi har, desto närmare kommer stickprovsmedelvärdet $\bar{X}_n$ att ligga det sanna populationsmedelvärdet μ. **Konvergens i sannolikhet ($\xrightarrow{p}$):** Sannolikheten för att $\bar{X}_n$ avviker från μ med mer än en liten toleransnivå ε går mot 0.

Centrala gränsvärdessatsen CLT För stora stickprov (n ➔∞) är stickprovsmedelvärdet $\bar{X}_n$ approximativt normalfördelat, även om populationen inte är normalfördelad. CLT gäller även för diskreta variabler (t.ex. binomiala, Poisson, eller enkla kategorier som myntkast). CLT säger att för alla x:$\boldsymbol{\lim_{n\to\infty} Pr\left(\frac{\bar{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}}\le x\right)=\Phi(x)}$, **d**är $\boldsymbol{\Phi(x)}$ är fördelningsfunktionen för en standardnormalfördelning N(0,1).Approximativ normalfördelning: $\boldsymbol{\bar{X}_n\approx N(\mu,\frac{\sigma^2}{n})}$. OBS! CLT gäller inte bara för stickprovsmedelvärden. T.ex. för summor av oberoende slumpvariabler $S_n=\sum_{i=1}^n X_i\approx N(n\mu, n\sigma^2)$).
OBS! CLT förutsätter (i.i.d) data, s**torlek på n (**$n\ge 30$**) och väntevärde och varians existeras men behövs ej vara kända.**

Föreläsning 4:

Bivarat fördelning (joint dist) KAPITEL 3.4 SIDA 118

Är en samling av alla sannolikheter för händelser där både X och Y är inblandade.

$$\Pr[(X,Y)\in C]\ där\ C\ är\ en\ mängd\ av\ möjliga\ utfall\ (t.ex. X=0\ och\ Y\ge 1).$$

Diskret joint PF: $f(x,y)=P(X=x,Y=y)$ och $\sum_{All\ x,y} f(x,y)=\sum_x\sum_y f(x,y)=1$.

Kont joint PDF: $P[(X,Y)\in C]=\iint_c f(x,y)dxdy$ där $f(x,y)\ge 0$. Sannolikheten av volymen under täthetsytan över regionen C.

Bivariat normal fördelning ***KOLLA FORMEL I BOKEN (s. 337)***

JOINT CDF - Den kumulativa fördelningsfunktionen är: $F(x,y)=P(X\le x, Y\le y)$

Sannolikheter för rektangulära områden: $P(a<X\le b, c<Y\le d)=F(b,d)-F(a,d)-F(b,c)+F(a,c)$

$F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(r,s)drds$. Täthetsfunktionen f(x,y) är den andra partiella derivatan av F(x,y): $f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}$

Marginalfördelning KAPITEL 3.5 SIDA 131

beskriver sannolikhetsfördelningen för en enskild variabel (t.ex. X eller Y) i en bivariat (tvådimensionell) fördelning. Den erhålls genom att ”ignorera” eller sammanfatta den andra variabeln.

$$f(x)=\sum_{All\ y} f(x,y)\ \ [DISKRET]$$

$f(x)=\int_{-\infty}^{\infty} f(x,y)dy$, $F_1(x)=\lim_{y\to\infty} F(x,y)$ eller $F_1(x)=\int_{-\infty}^{x}\int_{-\infty}^{\infty} f(r,s)drds$ [*KONTINUERLIG*]

OBEROENDE VARIABLER (kunskap om X ger ingen information om Y, vice verca). Två SV är oberoende om $P(X\in A\ och\ Y\in B)=P(X\in A)*P(Y\in B)$. För kontinuerliga variabler kan oberoende även definieras via den gemensamma täthetsfunktionen om: $f(x,y)=h_1(x)*h_2(y)$, där $h_1(x)$ och $h_2(y)$ är täthetsfunktioner som enbart beror på x respektive y. **OBS! $h_1(x)$ och $h_2(y)$ behöver inte nödvändigtvis vara marginalfördelningarna, men vid oberoende gäller $h_1(x)=f_1(x)$ och $h_2(y)=f_2(y)$.** Faktoriseringen $f(x,y)=h_1(x)*h_2(y)$ kräver att **stödet (området där f(x,y) > 0) är rektangulärt (dvs. x och y varierar oberoende).**

BETINGADE FÖRDELNINGAR(VAR1 givet VAR2) KAPITEL 3.6 SIDA 141

Den betingade PDF av X givet Y=y_0: $\boldsymbol{g_1(x|y)=\frac{f(x,y)}{f_2(y)}}$ där $\boldsymbol{\int g_1(x|y)dx=1}$

Multiplikationsregeln $\boldsymbol{f(x,y)=f_2(y)*g_1(x|y)}$

X och Y är oberoende om och endast om den betingade fördelningen för X sammanfaller med dess marginella fördelning:

$$g_1(x|y)=f_1(x)\ för\ alla\ y\ där\ f_2(y)>0.$$

FINN F(x) genom f(x) för kontinuerlig SV - Om X är en SV med PDF f(x) och det frågas efter CDF. Då tar man intergralen ex:

$$F(x)=\int_0^x f(x)dx=\int_0^x \frac{t}{2}dt=\left[\frac{t^2}{4}\right]_0^x=\frac{x^2}{4}\ och\ om\ F(X)=0, x\le 0, Om\ F(X)=1, x\ge 2$$

Beteckningar **Fördelningsfunktion – CDF gäller för både diskret och kontinuerlig.**

- Sannolikhetsfunktion (PF eller PMF) = Diskret
- Täthetsfunktion (PDF) = Kontinuerlig

Föreläsning 5: VÄNTEVÄRDE: KAPITEL 4.1 SIDA 207

Diskret SV: $E[X]=\sum_{All\ x} xf(x)$

Väntevärdet existerar endast om minst en av följande summor är ändlig: $\boldsymbol{\sum_{Positiva\ x} x*f(x)\ eller\ \sum_{Negativa\ x} x*f(x)}$.

Kont SV med pdf f(x)**:** $E[X]=\int_{-\infty}^{\infty} xf(x)\,dx$. Väntevärdet existerar endast om integralen konvergerar (dvs. inte går mot $\pm\infty$).

Väntevärde av en funktion (där Y = r(X)) $E[r(X)]=\int_{-\infty}^{\infty} r(x)f(x)\,dx$

Väntevärde av en linjär funktion (Y=r(x)=aX+b) $E[Y]=aE[X]+b$

GAUSS APPROXMIATION av väntevärde av en funktion KAPITEL 4.1 SIDA 207

I allmänhet gäller inte E[r(X)] = r(E[X]). Gauss metod bygger på att **linjärisera r(X)** kring väntevärdet E[X]=μ: $\boldsymbol{r(X)\approx r(\mu)+r'(\mu)(X-\mu)}$, där $r'(\mu)$ är derivatan av r(x) i punkten x= μ. $E[r(X)]\approx E[r(\mu)+r'(\mu)(X-\mu)]=r(\mu)+r'(\mu)*E[X-\mu]$. Termen $r'(\mu)*E[X-\mu]$ försvinner eftersom $E[X-\mu]$=0 ➔ $\boldsymbol{E[r(X)]\approx r(\mu)=r(E[X])}$. Den fungerar bra vid låg icke-linjäritet i r(X) och låg spridning i X:s fördelning.

Väntevärde av en produkt av oberoende variabler

Om $X_1,\ldots,X_n$ är oberoende SV med ändliga väntevärden gäller: $E(\prod_{i=1}^n X_i)=\prod_{i=1}^n E(X_i)$.

VARIANS

$$VAR[X]=E[(X-\mu)^2]=E[X^2]-E[X]^2$$

$Var[X]\ge 0\ och\ Var[X]=0$ om och endast om $P(X=c)=1$ för en konstant c.

Varians av en linjär funktion (Y=aX+b) $Var[Y]=a^2*Var[X]$

Om $X_1,\ldots,X_n$ är **oberoende** slumpvariabler, och $Y=a_1X_1+\cdots+a_nX_n+b$, så gäller: $Var(Y)=a_1^2*Var(X_1)+\cdots+a_n^2*Var(X_n)$

GAUS APPROXIMATION FOR A VARIANCE $Var[r(X)]\approx r'(\mu)^2*Var(X)\ (Derivera\ med\ avseeende\ på\ X\ och\ sätter\ in\ \mu)$

Detta underlättar beräkningar men kräver att **funktionen är ”tillräckligt linjär” nära μ och Var(X) är liten.**

Higher order moments, Skewness, Kurtosis KAPITEL 4.4 SIDA 234,

K*th* order moment: $E[X^k]$	Kth order central moment $E[(X-\mu)^k]$ OBS k = 2 blir => Var[X]	Skevhet av en fördelning $E[(X-\mu)^2]/\sigma^2$ + **Positiv skevhet:** Höger svans är längre (fördelningen drar åt höger). + **Negativ skevhet:** Vänter svans är längre (fördelningen drar åt vänster).	Kurotsis (tugna svansar) $E[(X-\mu)^3]/\sigma^3$ **+ Hög kurtos (>3):** Tunga svansar och spetsig topp (t.ex. t-fördelning med få frihetsgrader). **+ Låg kurtos (<3):** Lättare svansar och plattare topp (t.ex. likformig fördelning).

Moment genererande funktion: KAPITEL 4.4 SIDA 234,

MGF är: $\boldsymbol{\psi(t)=E(e^{tX})}$

$\boldsymbol{E(X^n)=\psi^{(n)}(0)}$ Användbar då moment är derivator av $\psi(t)$ evaluted at t=0.

MGF för linjär komb (Y=aX+b): $\psi_Y(t)=e^{bt}\cdot\psi_X(at)$

MGF för en summa: $\psi_Y(t)=\psi_{X_1}(t)*\psi_{X_2}(t)*\ldots*\psi_{X_n}(t)$, där $X_1,\ldots,X_n$ är oberoende variabler och $Y=X_1+\cdots+X_n$.

MEAN, MEDIAN, MODE (MÅTT FÖR CENTRAL TENDENS (LÄGESMÅTT) I EN FÖRDELNING S241 KAPITEL 4.5

Median: $P(X\le m)=1/2$ (Delar fördelningen i 2 lika delar). Minimerar **absolut förlust:** $\boldsymbol{E[\lvert X-d\rvert]}$.	Mode x_{mode}:Det vanligaste värdet (maximerar sannolikhetstätheten f(x)). Används för att identifiera ”toppen” av fördelningen.	Medelvärde (Mean): Genomsnittet av alla värden.Minimera **kvadratisk förlust:** $\boldsymbol{E[(X-d)^2]}$.Känsligt för extremvärden (stora avvikelser ”straffas hårt”).

KOVARIANS $\boldsymbol{Cov(X,Y)=E[(X-\mu_X)*(Y-\mu_Y)]}=E(XY)-\mu_X\mu_Y$. Definieras som medelvärdet av produkten av avvikelserna från variablernas medelvärden som mäter **linjärt samband** mellan två variabler. **Positiv kovarians:** När X är över sitt medelvärde, tenderar Y också att vara över sitt. **Negativ kovarians:** När X är över sitt medelvärde, tenderar Y att vara under sitt.

Linjärkombination: $Var(aX+bY+c)=a^2*Var(X)+b^2*Var(Y)+2ab*Cov(X,Y)$

Korrelation **(Korrelation mäter ENDAST linjära samband!).** Variabler kan fortfarande vara beroende på ett **icke-linjärt sätt** (t.ex. Y = X²). **Undantag: För normalfördelade variabler: Nollkorrelation = Oberoende.**

$$\rho[X,Y]=\frac{Cov[X,Y]}{\sigma_X\sigma_Y}$$ **FÖR BETINGAT VÄNTEVÄRDE KAPITEL 4.7 SIDA 256,**

Betingat väntevärde $\boldsymbol{E(Y|x)=\int_{-\infty}^{\infty} y*g_2(y|x)dy}$. $\boldsymbol{g_2(y|x)}$ är den betingade täthetsfunktionen för Y givet X = x. *E(Y|X)* är en **funktion av X**, inte av Y (Y har ”integrerats bort” i *E(Y|X)*). Eftersom *X* är en slumpvariabel ➔*E(Y|X)* blir också en **slumpvariabel.**

Law of iterated expectations: $\boldsymbol{E[E(Y|X)]=E(Y)}$. **Tolkning:** Genom att först ta det betingade väntevärdet av Y givet X, och sedan ta väntevärdet över alla X ➔Vi får det obetingade väntevärdet av Y.

Betingad varians. $\boldsymbol{Var(Y\mid X)=\int_{-\infty}^{\infty}(y-\mu_{Y|X})^2 * g_2(y\mid x)dy}$ **(**Mäter spridningen av Y **givet ett specifikt värde på X).**

Total varians och dess uppdelning: $\boldsymbol{Var(Y)=E[Var(Y\mid X)]+Var[E(Y\mid X)]}$. $\boldsymbol{E[Var(Y\mid X)]}$: Medelvärdet av de betingade varianserna (spridningen *inom* varje grupp av X). $\boldsymbol{Var[E(Y\mid X)]}$: Variansen av de betingade medelvärdena (spridningen *mellan* grupper av X).

KOMMENTAR ANGÅENDE MOMENT GENEREANDE FUNKTION: Om $X_1,\ldots,X_n$ är **oberoende och identiskt fördelade (samma fördelning),** och $Y=X_1+\cdots+X_n$, så är: $\psi_Y(t)=\psi(t)*\ldots*\psi(t)\ (n\ gånger)=[\psi(t)]^n$. **OBS! iid medför att** $\boldsymbol{\psi_{X_1}(t)=\psi_{X_2}(t)=\cdots=\psi_{X_n}(t)=\psi(t)}$.

Föreläsning 9:

Sannolikhetsteori beskriver hur data $X_1,\ldots,X_n$ genereras från en fördelning $f(x|\theta)$ **med kända parametrar** θ (t.ex. μ och σ för normalfördelning). **Statistisk inferens** drar slutsatser om **okända parametrar** θ utifrån observerade data $x_1,\ldots,x_n$ från $f(x|\theta)$.

Frekventist: Parametrar (t.ex. μ i en normalfördelning) ses som fasta konstanter. Sannolikhet definieras som den relativa frekvensen av en händelse vid många upprepningar. **Bayesian**: Parametrar kan tilldelas en sannolikhetsfördelning om deras värde är okänt. Sannolikhet är subjektiv och representerar en grad av tro (baserat på data och tidigare kunskap).

Likelihoodfunktionen: Likelihoodfunktionen $L(\theta|data)$ mäter hur troligt det är att observerad data uppstår för olika värden på en parameter θ. **Likelihoodfunktionen är *inte* en pdf för $\boldsymbol{\theta}$.** Likelihoodfunktionen $L(\theta|x)=f(x|\theta)$ är samma matematiska uttryck, men den tolkas som en funktion av θ (inte som en funktion av x). Integralen över θ behöver inte vara 1: $\int_{-\infty}^{\infty} f(x|\theta)d\theta\neq 1\ (i\ allmänhet)$.

INTRODUKTION TILL BAYES Jag är osäker om θ. Denna osäkerhet beskrivs av min subjektiva prior probablitiy distribution för θ, p(θ). Priorn beskriver min grad av tro givet olika värden för theta INNAN jag hämtar data. Jag samlar in min data X1 = x1, Xn=xn från $f(x\_1,\ldots,x\_n\,|\,\theta)$. Jag har då lärt mig om min **posterior** distribution $f(\theta|X\_1=x\_1,\ldots,X\_n=x\_n)$ med avseende på θ

Bayes teori för dis.var och kont.var: $p(A_i|B)=\frac{p(B|A_i)*p(A_i)}{P(B)}=\frac{p(B|A_i)*p(A_i)}{P(B)}, p(\theta|x_1,\ldots,x_n)=\frac{f(x_1,\ldots,x_n|\theta)p(\theta)}{\int f(x_1,\ldots,x_n|\theta)p(\theta)d\theta}$

En uppdaterad sannolikhetsfördelning för ***θ efter*** att ha sett datan. Den kombinerar prior och data via **Bayes sats:** $\boldsymbol{p(\theta|data)=\frac{f(data|\theta)*p(\theta)}{p(data)}}$. **Likelihood:** $f(data|\theta)$ (hur troligt är data givet θ). **Prior:** $p(\theta)$ (min initiala tro eller tidigare kunskap). **Marginal likelihood:** $p(data)=\int f(data|\theta)*p(\theta)d\theta$ (En normaliseringskonstant). **Tolkning:** Posteriorn $p(\theta|data)$ är en **sannolikhetstäthetsfunktion (pdf) för $\boldsymbol{\theta}$**, vilket innebär att $\int p(\theta|data)d\theta=1$.

Prior: Det är priorn p(θ) som hälper konvertera liklihoodfunktionen till en posterior densitet för θ.

Formel för posterior

$$\boldsymbol{Posterior\ \propto Likelihood*Prior}$$

$p(\theta|x)\propto p(x|\theta)*p(\theta)$ där x är en vektor **glöm ej det**

EXEMPEL PÅ MOMENTMETODEN FÖR ATT SKATTA LAMBDA

EXEMPLE

(X,Y) är SV, PDF: $f(x,y)=\frac{3}{16}(4-2x-y)\ då\ x>0, y>0, 0<2x+y<4$. **A) Beräkna f(x):** För att göra detta finn integreringsområdet för y => 0<y<4-2x. Beräkna sedan följande integral $\int_0^{4-2x}\frac{3}{16}(4-2x-y)dy$. Detta ger då f(x) = 3/8(x-2)^2. **B) Bestäm f(Y|X):** Formeln för betingad p(x) säger $f(X|Y)=\frac{f(x,y)}{f(x)}$ det vill säga $\frac{3}{16}(4-2x-y)/\frac{3}{8}(x-2)^2$ detta ger f(Y|X) = $\frac{4-2x-y}{2(x-2)^2}$. **C) Beräkna E[Y|X=1/2]:** Använda formel $E[Y]=\int y(f(y)$. Ger $\int y*\frac{1}{2}*\frac{(4-2*0.5-y))}{2(0.5-2)^2}=\int\frac{2}{9}*(3-y)\,dy, 0<y<3\ldots=1$

UPPGIFT 2: Låt X ha PDF $f(x)=2x$, $0<x<1$ Bestäm PDF för $Y=1-X^2$. Så använda formel $P(Y\le y)=P(X\le\sqrt{y-1})=F_X(\sqrt{y-1})$. Derivera nu mha kedjeregeln. Så först deriverar vi Fx som yttre derivatan och sedan y-1 som inre. Det ger $\frac{dF_X(x)}{d}=f_x(\sqrt{y-1})=yttre\ derivatan$ och $\frac{d\sqrt{y-1}}{d}=\frac{1}{2\sqrt{y-1}}*(1)=inre\ derivatan$. Sedan skriv ut hela uttrycket med f(x) = $2(\sqrt{y-1})*\frac{1}{2\sqrt{y-1}}=\frac{2\sqrt{y-1}}{2\sqrt{y-1}}=1$ Detta är Y’s PDF.

EXEMPEL UPPGIFT: **Låt SV X ha PDF $\boldsymbol{f(x)=\theta x^{\theta-1}}$ där 0<x<1 annars 0.**

Beräkna E[X]=$\boldsymbol{\int_0^1 x\ \theta x^{\theta-1}dx=\frac{\theta}{\theta 01}}$

Medianen för X: $P(X<m)=0.5\ där\ m=medianen\ P(X<m)=\boldsymbol{\int_1^m \theta x^{\theta-1}dx=m^{\theta})=0.5=>m=0.5^{\frac{1}{\theta}}}$

EXEMPEL GAUS APPROXIMATION: Låt $X_1 och\ X_2$ vara oberoende och exponentialfördelade med lambda = 3.

Visa att 6(X1+X2 är chifördelat med två frihetsgrader: Använd mfg för exp $\psi Xi(t)=\frac{\lambda}{\lambda-t}$. Där Z = 6(X1+X2). $\psi_z(t)=E[e^{t6(X1+X2)}]=\cdots=E[e^{t6X1}]E[e^{t6X2}]=\psi_{X1}(t6)\psi_{X2}(t6)=\left(\frac{\lambda}{\lambda-t}\right)^2=\cdots(\lambda=3)=\left(\frac{3}{3-6t}\right)^{\frac{4}{2}}=(1-2t)^{-2}\ vilket\ är chi2(4)$

Beräkna $\boldsymbol{E\left[\frac{2}{X_1+X_2}\right]}$ ***mha gauss approximation***: **X1+X2 är gamma(2,3) så E[X1+X2] = 2/3 det vill säga E[2/X1+X2] = 2/2/3=3**

Formel för m.g.f.: Se "Lagen om väntevärdet för en funktion av X"

$$\psi(t)=E(e^{tX})=\begin{cases}\int_{-\infty}^{\infty} e^{tx}\cdot f(x)\,dx & \text{(kontinuerlig } X\text{)},\\ \sum_x e^{tx}\cdot p(x) & \text{(diskret } X\text{)}.\end{cases}$$

Inverse CDF-metoden: Om $Y\sim U(0,1)$ och *F* är en kontinuerlig CDF av *X*, så kan vi generera *X* med fördelning *F* genom att sätta $X=F^{-1}(Y)$. Generera först $u_1,\ldots,u_n \overset{iid}{\sim} U(0,1)$. Beräkna sedan $x_i=F^{-1}(u_i)$. Då blir $x_1,\ldots,x_n \overset{iid}{\sim} F$. (Det innebär att om F är normalfördelningens CDF ➔ $x_1,\ldots,x_n$ är normalfördelad). Metoden kan anpassas för **diskreta slumpvariabler,** men kräver extra hantering eftersom CDF:en är styckvis konstant. Ex: För att simulera $X\sim Binomial(n,p)$, leta upp det minsta x där $F(x)\ge u$ för ett givet $u\sim U(0,1)$.

Föreläsning 13 – Pivotal kvantitet: En statistisk variabel/funktion av stickprovsdata och parametern θ vars känd fördelning **inte beror på okända parametrar** (här μ eller σ^2).

FÖRELÄSNING 11 KAPITEL 7 I BOKEN

Bayes estimators: Bayesiansk metodik ger en fullständig beskrivning av osäkerheten kring parametern θ genom posteriorfördelningen *p(θ|x)*. **Förlustfunktionen $L(a,\theta)$** mäter kostnaden av att välja skattningen a när det sanna värdet är $\boldsymbol{\theta}$. Det Bayesianska svaret är att välja $\boldsymbol{\alpha}$ som minimerar den ***förväntade förlusten*** över posteriorn: $E[L(a,\theta)|x] = \int L(a,\theta) * p(\theta|x)d\theta$. **Kvadratisk förlust** $L(a,\theta) = (\theta - a)^2$: Minimera den förväntade kvadratiska felkostnaden ➔ **Posteriormedelvärdet** är optimalt. **Absolut förlust** $L(a,\theta) = |\theta - a|$: Minimera den förväntade absoluta felkostnaden ➔ **Posteriormedianen** är optimalt.**Asymmetriska förlust:** T.ex. högre kostnad för överskattning än underskattning ➔ Anpassa skattning.

Maximum Likelihood Estimation: Likelihood-funktionen $f(x_1,\dots,x_n|\theta)$ beskriver sannolikheten (eller tätheten) för de observerade datan $x_1,\dots,x_n$ som en funktion av den okända parametern θ. **MLE:** Maximum likelihood estimator är Hitta det värde på θ som ***maximerar likelihood-funktionen.*** Detta är det θ-värde som gör datan mest sannolik.

Skillnad mellan estimator och estimate: Estimator ($\hat{\theta}$): En regel eller funktion som används för att beräkna skattning, t.ex. $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (Medelvärdesestimatorn). **Estimate:** Det specifika numeriska värdet som fås när estimatorn tillämpas på en datasets, t.ex. $\hat{\theta} = 6.9$ för en given datamängd.

MLE för POISSON	MLE FÖR BERNOULLI	MLE FÖR NORMAL(var[x] känd)
$X_1,\dots,X_n \mid \theta \sim Pois(\theta)$med $f(x) = \frac{\theta^x * \exp(-\theta)}{x!}$ **Likelihoodfunktionen:** $f(x_1,\dots,x_n \mid \theta) = \prod_{\{i=1\}}^{n} f(x_i \mid \theta) = \prod_{\{i=1\}}^{n} \frac{\theta^{xi}\exp(-\theta)}{xi!} = \frac{\theta^{\sum xi} * \exp(-n\theta)}{\prod_{\{i=1\}}^{n} xi!}$ MLE: $\frac{d\ln f(x_1,\dots,x_n\|\theta)}{d\theta}$=0 vilket är $\sum \frac{x_i}{\theta} - n = 0$ och **skattning** $\hat{\theta} = \sum x_i/n = \bar{x}$ **vi kan kontrollera att detta är MLE genom att ta andra derivatan av** $\frac{d2\ln f(x_1,\dots,x_n\|\theta)}{d2\theta} = -\sum x_i/\theta^2 < 0$	$X_1,\dots,X_n \mid \theta \sim Bern(\theta)$ med $f(x) = \theta^x(1-\theta)^{1-x}$ för x [0,1] **Likelihoodfunktion:** $f(x_1,\dots,x_n \mid \theta) = \prod_{\{i=1\}}^{n} f(x_i \mid \theta) = \prod_{\{i=1\}}^{n} \theta^{xi}(1-\theta)^{1-xi} = \theta^{\sum xi}(1-\theta)^{n-\sum xi}$ Log-**likelihoodfunktion** MLE: $\frac{d\ln f(x_1,\dots,x_n\|\theta)}{d\theta}$= 0 vilket ger $\frac{\sum xi}{\theta} - \frac{(n-\sum xi)}{1-\theta} = 0$ så $\hat{\theta} = \sum x_i/n = \bar{x}$ och verifera: $\frac{d2\ln f(x_1,\dots,x_n\|\theta)}{d2\theta} = -$ ger $\frac{\sum xi}{\theta^2} - \frac{(n-\sum xi)}{1-\theta^2} < 0$	$X_1,\dots,X_n \mid \theta \sim N(\theta,\sigma^2)$ med $f(x\|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} * \exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2\right)$ **Likelihoodfunktion**: $f(x_1,\dots,x_n \mid \theta)=\prod_{\{i=1\}}^{n} f(x_i \mid \theta) = \prod_{\{i=1\}}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} * \exp\left(-\frac{1}{2\sigma^2}(xi-\theta)^2\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} * \exp\left(-\frac{1}{2\sigma^2}\sum(x-\theta)^2\right)$ Log-**likelihoodfunktion**: $=-\frac{n}{2} * \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(x-\theta)^2$ MLE: : $\frac{d\ln f(x_1,\dots,x_n\|\theta)}{d\theta} = 0$ ger oss $\frac{1}{\sigma^2} * \sum(xi-\theta) = 0$ blir $\sum(xi-\theta) = n\bar{x} - n\theta = 0$ där **skattningen** $\hat{\theta} = \frac{\sum xi}{n} = \bar{x}$ och verifera att detta är MLE $\frac{d2\ln f(x_1,\dots,x_n\|\theta)}{d2\theta} = -\frac{n}{\sigma^2} < 0$

Properties of the MLE: Invariansegenskapen för MLE: Om $\hat{\theta}$ är MLE för θ, så är $g(\hat{\theta})$ MLE för $g(\theta)$, förutsatt att g är en-till-en funktion (dvs bijektiv).Konsistens: $\hat{\theta} \xrightarrow{p} \theta$ $när\ n \to \infty$. För stora stickprov är MLE:s samplingfördelning ungefär **normalfördelad** med: **Medelvärde:** Det sanna värdet θ. **Varians:** $\frac{1}{I(\theta)}$, där $I(\theta)$ är Fisher-informationen.

Method of moments: Jämföra teoretiska moment (förväntade värden) med empiriska moment (beräknade från data).

Teoretiskt moment $\boldsymbol{\mu_j(\theta)}$: Väntevärde av X^j givet parametern θ för j 1 till k. Empiriskt moment ($\boldsymbol{m_j}$): Genomsnittet av datan upphöjt till j: $m_j = \frac{\sum_{i=1}^{n} X_i^j}{n}$ för j 1 till k. The method of moments estimator kan då fås genom att lösa θ från systemet av ekvationer: $\mu_1(\theta) = m_1$ ... $(tänk\ er\ att\ den\ går\ vertikalt)$ .... $\mu_k(\theta) = m_k$ (**Välj antal moment (*k*):** Använd lika många moment som antal parametrar jag vill skatta.) Första momentet (j=1): $\mu_1(\theta) = E(X|\theta)$ (Medelvärde). Andra momentet (j=2): $\mu_2(\theta) = E(X^2|\theta)$ (Relaterat till variansen).

Föreläsning 12 KAPITEL 8.1 8.2 8.4

SAMPLING FÖRDELNINGAR Låt $T = r(X_1,\dots,X_n,\theta)$ är en funktion som beror på både datan $(X_1,\dots,X_n)$ och parametern θ. Ex: $T = \sum_{i=1}^{n}(X_i - \mu)^2$, där μ är en parameter (t.ex. populationsmedelvärdet). Om *T* inte beror på θ → T är **statistika.** Samplingfördelningen är ***fördelningen av en statistika från populationen.***

Användning av en sampling fördelning

Sampling fördelning från en Normal modell: MLE för μ (känd varians) är $\hat{\mu} = \bar{X}$. Samplingsfördelningen för $\hat{\mu}$ är $\hat{\mu}|\mu \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. MLE för σ^2med känd μ $blir$: $\sigma_0^2 = \frac{1}{n}\sum(X_i - \mu)^2$. Samplingfördelning för $\boldsymbol{\sigma_0^2|\mu,\sigma^2 \sim \frac{\sigma^2}{n} * \chi2(n)}$.

CHIFÖRDELNINGEN Chi2fördelningen med m frihetsgrader är gammafördelningen $Gamma(\frac{m}{2},\frac{1}{2})$ med notation $X \sim X_m^2$. Moments of chisquare fördelningen är $E[X] = m, Var[X] = 2m$.

Den additiva egenskapen för chi2fördelningen(om en Standardnormal kvadreras)

Om $X \sim N(0,1)$ då är $X^2 \sim X_1^2$(det vill säga om man kvadrerar en std normal 0,1 så blir den chi fördelad med 1 frihetsgrad). **Den additiva egenskapen blir då $\boldsymbol{Om\ X_1,\dots,X_n}$ är $\boldsymbol{oberoende\ X_i \sim X_m^2\ d}$å är $\boldsymbol{X_1 + \dots + X_n \sim X_m^2\ med\ m = m_1 + \dots + m_k}$.** Om $X_1,\dots,X_m$ är **oberoende och identiskt fördelade (iid)** standardnormala variabler (*N(0,1)*), så följer summan av deras kvadrater en Chi2-fördelning med *m* frihetsgrader: $X_1^2 + \dots + X_m^2 \sim \chi_m^2$.

Samplingsfördelning av stickprovsvarians i en Normal modell

MLE för varians i en normalfördelning med känt μ är $\widehat{\sigma_0^2} = \frac{1}{n}\sum(X_i - \mu)^2$ och då är $\frac{\widehat{\sigma_0^2}n}{\sigma^2} = \sum((X_i - \mu)/\sigma)^2 = \sum Z_i^2$ Där $Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$ Eftersom $Z_i^2 \sim \chi_1^2$, följer summan av *n* oberoende Z_i^2-termer en Chi2-fördelning med *n* frihetsgrader:$\frac{n\sigma_0^2}{\sigma^2} \sim \chi_n^2$.

Fördelning av stickprovsmedelvärde och varians i Normal data

MLE av variansen då $\boldsymbol{\mu}$ är okänd: $\widehat{\sigma^2} = \frac{1}{n}\sum(X_i - \bar{X})^2$. Om $X_1,\dots,X_n|\mu,\sigma^2 \sim (iid)N(\mu,\sigma^2)$ då är $\bar{X}$ oberoende och $\bar{X} \sim N(\mu,\frac{\sigma^2}{n})$ och $\frac{\widehat{\sigma_0^2}n}{\sigma^2} \sim X_{n-1}^2$ **NOTE:** Oberoendet av Xbar och sigma2 hålls endast för normaldata. **Förlust av en frihetsgrad (*n-1*)** beror på att vi uppskattat $\mu\ med\ \bar{\boldsymbol{X}}$.

THE MLE OF THE VARIANS IS NOT THE SAMPLE VARIANCE MLE för $\boldsymbol{\sigma^2}$ (när μ är okänt) är partisk.

Notera att MLE för sigma2 är inte den normala stickprovvariansen $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ som oftast är sedd som en bättre skattning. Men Samplingsfördelningen av en funktion av variansen i stickprovet blir $\frac{(n-1)S^2}{\sigma^2} \sim X_{n-1}^2$

T FÖRDELNINGEN då variansen är okänd i populationen

Hur nära är stickprovsmedelvärdet till populationsmedelvärdet? Om variansen är känd kan vi kolla på fördelningen av $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$. Men om sigma2 är okänd: Standardiserat medelvärde, $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$. Skalad stickprovsvarians, $Y = \frac{(n-1)\sigma_0^2}{\sigma^2} \sim \chi_{n-1}^2$. Dessa är oberoende. Genom att kombinera dem får vi: $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{Z}{\sqrt{Y/(n-1)}} \sim t_{n-1}$.

SATSEN SÄGER: Om $X_1,\dots,X_n \sim (iid)N(\mu,\sigma^2)$ då är $\frac{(\bar{X}-\mu)}{S/\sqrt{n}} \sim t(n-1)$

HUR MAN BERÄKNAR STYRKA MHA EXEMPELUPPGIFT

A) Pröva med 5% signifikans H0=$\boldsymbol{\mu = 8}$ mot Ha $\boldsymbol{\mu < 8}$. *Xbar är 7.30, s=0.942.* . Detta blir t-test då **sigma är okänd i uppgiften** $T = {\bar{X} - \mu_0}\Big/{s/\sqrt{n}} \sim t(n-1) om\ H0\ är\ sann = {7.30 - 8}\Big/{0.942/\sqrt{10}} = -2.396 \sim t(9) = -1.833^{FÖRKASTA!!!}$

B) Anta nu att sigma är känd och är 0.924(Normal). Beärkna styrkan givet att $\boldsymbol{\mu = 7}$
$\pi(7) = P(Förkasta\ H0|\boldsymbol{\mu = 7})$ vi kommer ihåg sedan tidigare uppgift att förkasta H0 är vänstersidigt och vi gör det med 5% signfikans så $P(Z \le -1.645|\boldsymbol{\mu = 8}) = \boldsymbol{0.05}$) $[I\ Z\ tabellen\ ger\ 95\% => 1{,}645]$ men vi har dock X så vi behöver gå från X till standard normal: $P\left(\frac{X-8}{\frac{0.924}{\sqrt{10}}} \le -1.645|\boldsymbol{\mu = 7}\right) = [sedan\ gå\ X\ endast) = P\left(X \le -1645 * \frac{0.924}{\sqrt{10}} + 8\middle|\boldsymbol{\mu = 7}\right) =$

$$\boldsymbol{P(X \le 7.5193|\mu = 7) = [Gör\ till\ STD\ normal] = P\left(\frac{X-7}{\frac{0.924}{\sqrt{10}}} - \frac{7.5193-7}{\frac{0.924}{\sqrt{10}}}\middle|\mu = 7\right) = P(Z \le 1.78) = 0.9625}$$

EXEMPEL MED ML SKATT/ASYM EGENSKAP;ANTA ATT DU HAR N OBER OBS FRÅN X ~POI(lambda)

A) Maximum likelihood skattningen av λ är $\bar{X}$. Visa att denna skattning är väntevärdesriktig och konsistent skattning av lamba. Väntevärdeskattning E[X]$= \lambda = \mu\ och\ Var[X] = \frac{\lambda}{n} som\ går\ mot\ 0\ om\ n\ går\ mot\ oändligheten$

B) Beräkna 95% KI för λ MHA ML SKATTNINGARS ASYMPTOTISKA EGENSKAPER. DET ÄR OK OM DU SKATTAR PARAMETRAR I MEDELFEL. SÄTT ÄVEN IN DET OBSERVERADE STICKPROVET SOM GER $\bar{X} = 1.6$, n =100.
Vi börjar med att ta fram $f(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$, $\ln f(x|\lambda) = xln\lambda - lnx! - \lambda$. Första derivatan blir $\frac{x}{\lambda} - 1$ och andra blir $-\frac{x}{\lambda^2}$. Då sätter vi in detta i formel för asymptotisk egenskap $-E\left[-\frac{x}{\lambda^2}\right] = \frac{E[X]}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$. Då är X asymp N$(\lambda, \frac{1}{\frac{1}{\lambda}*n}) = (\lambda, \frac{\lambda}{n})$. Lambda/n får skattas med Xbar. **Nu kan vi beräkna KI.** $\frac{\bar{X}-\lambda}{\sqrt{\frac{\bar{X}}{n}}}$ asym N(0,1) med ett 95% KI fås $(\bar{X} - 1.96\sqrt{\frac{\bar{x}}{n}}; \bar{X} + 1.96\sqrt{\frac{\bar{x}}{n}})$

EXEMPEL HYPOTESPRÖVNING: 10 ober obs har samlats med resultat: 0,0,1,0,2,0,2,0,1,2. $\boldsymbol{H_0}$: $\lambda = 1, H_1: \lambda < 0$

A) Använda $\sum_{i=1}^{10} X_i$ som testvariabel. Som förkastningsområde (kritiskt område) väljer du att förkasta H_0 om summan av Xi är liten. Bestäm Förkastelse område för signifikansnivå 5%. Vad blir storleken på testet? $n = 10, \sum X_i = 8, \bar{X} = 0.8$, $\sum X_i \le c\ är\ förkastelse\ område$. P$(\sum X_i \le c) < 0.05$ och $\sum X_i \sim Poi(10 = 10 * 1)\ Då\ h0\ är\ sann$. Kolla tabell för poisson gå dit vi har lambda 10. Och sedan är c de siffror 0,1. Summera de sannolikheter under 10 0.005+0.0023 osv. När summan går över 0.05 det är det c som vi kommer förkasta. I detta fall, om c = 4 har vi 0.05 och c = 5 blir 0.06. Så vi förkastar H0 om $\sum X_i \le 4$. P$(\sum X_i \le 4) = (0.005 + 0.0023 + 0.0076 + 0.0189)$ k=1+k=2+k=3+k=4 i tabellen. Storlken blir då 0.029.

För att skatta andelen defekta enheter p i en produktion vill man först utvärdera två olika förfaringsätt. Metod 1 $\boldsymbol{X_1 \sim Bin(50, p)}$. Andra metoden $\boldsymbol{X_2 \sim Geo(p)}$. ML skatta p för båda metoderna. Använd data från simueringstudien ovan sätt in skattningarna.

Försök nr	1	2	3	4	5
Metod 1	1	2	0	0	1
Metod 2	197	19	83	11	30

Metod 1: $L(p) = p^{xi}(1-p)^{50-xi} = p^{\sum xi}(1-p)^{50*5-\sum xi}$ där $lnL(p) = \sum xi * lp + (250 - \sum xi) * \ln(1-p)$ derivera sedan för ML skatta: $\frac{\sum xi}{p} - \frac{250-\sum xi}{1-p} = 0 => \hat{p} = \frac{\sum xi}{250} = \frac{4}{250} = 0.016$.

Metod 2: $\boldsymbol{L}(p) = p^5(1-p)^{\sum xi}$ där $lnL(p) = 5lnp + \sum xi \ln(1-p)$ derivering ger $\frac{5}{p} - \frac{\sum xi}{1-p} = 0$. Där skattningen är $\hat{p} = \frac{5}{\sum xi + 5} = \frac{5}{345} = 0.0145$.

B) För att bedöma v ilken av metoderna som är bäst ska du beräkna Fisherinformationen för de metoderna. De räcker med att du tar ett försök. Vilken metod ger störst information? Vi använder formel I(p). För metod är E[X] = 50*p. Första derivatan och andra derivatan har vi redan fått så vi sätter in i formel: $-E[-\frac{X}{p^2} - \frac{50-x}{(1-p)^2} => \frac{E[X]}{p^2} + \frac{50-E[X]}{(1-p)^2} = \frac{50}{P(1-p)^2} = I(p)$

För Metod två blir det $E[X] = \frac{1-p}{p}$ där ... tillsut blir I(p) $= \frac{1}{p^2(1-p)}$

Föreläsning 13 KAPITEL 8.5 8.7 8.8

Konfidensintervall

Ett konfidensintervall använder sig av sampling fördelningen to derive probablity statements of the form: $P(A < \theta < B) = 0.95$ där A och B är SV bestämda från samplingfördelningen.
Notera att det är intervallet [A,B] som är slumpmässigt inte parametern θ.
Konfidensintervall för normal medelvärde: $X_1,\dots,X_n|\mu,\sigma^2 \sim N(\mu,\sigma^2)$.
Vi känner till sampling fördelningen: $U = \frac{(\bar{X}-\mu)}{S/\sqrt{n}} \sim t_{n-1}$
Från t tabellen kan vi finna konstant C så att $P(-c < U < c) = 0.95$
Olikheten -c<U<c är densamma som: $\bar{X} - c * \frac{S}{\sqrt{n}} < \mu < \bar{X} + c * \frac{S}{\sqrt{n}}$
Det vill säga att P$(\bar{X} - c * \frac{S}{\sqrt{n}} < \mu < \bar{X} + c * \frac{S}{\sqrt{n}})$ = 0.95 där A B är respektive sidor.
Notera: $0.95 = P(-c < U < c) = F_{n-1}(c) - F_{n-1}(-c) = F_{n-1}(c) - [1 - F_{n-1}(c)] = 2F_{n-1}(c) - 1$. Detta i sin tur ger att $c = F_{n-1}^{-1}[\frac{1+0.95}{2} = 0.975]$
$F_v^{-1}(.)$ är kvantil funktionen av t fördelningen med v frihetsgrader.
DENNA metode fungerar eftersom fördelningen av $U = \frac{(\bar{X}-\mu)}{S/\sqrt{n}}$ beror inte på μ vilket betyder att U är en **Pivotal kvantitet.**

Konfidensintervall för normal varians (ensidigt)

$P(\sigma^2 < A) = 0.95$ där A är en SV som vi behöver finna.
$Pivotal\ kvantitet$: $\boldsymbol{U = \frac{(n-1)S^2}{\sigma^2} \sim \chi_n^2 - 1}$
Från **Chi2-fördelningstabellen** hittar vi c så att: $\Pr(U > c) = 0.95$.
$U > c$ är samma olikhet som $\sigma^2 < \frac{(n-1)S^2}{c}$**dvs** $P\left(\sigma^2 < \frac{(n-1)S^2}{c}\right) = 0.95$

Unbiased estimators

Teori: Låt $\hat{\theta}(X_1,\dots,X_n)$ vara en skattning av θ. $\hat{\theta}$ kommer givetvis att variera från stick till stickprov, men vi hade velat att det skulle vara correct on average: $\boldsymbol{E_\theta[\hat{\theta}] = \theta}$.
An estimator $\hat{\theta}(X_1,\dots,X_n)$ is an unbiased estimator for θ if $E_\theta[\hat{\theta}(X_1,\dots,X_n)] = \theta$
Bias: $Bias[\hat{\theta}] = E_\theta[\hat{\theta}(X_1,\dots,X_n)] - \theta$
Example: $\bar{X}$ är en unbiased estimator of the mean in the iid normal model:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum Xi\right) = \frac{1}{n}\left(\sum E[Xi]\right) = \frac{1}{n}(\sum \mu) = \mu$$

Example S^2 is an unbiased estimator of σ^2 in the normal model:
$E\left[\frac{(n-1)S^2}{\sigma^2} = n - 1\right]$ detta ger $E[S^2] = \sigma^2$. Note that σ^2 is biased when μ is unknown

MSE

The mean squared error (MSE) of an estimator $\hat{\theta}(X_1,\dots,X_n)$ is defined as:
MSE($\hat{\theta}$) $= E_\theta\left[(\hat{\theta} - \theta)^2\right]$
Vi har relationen: MSE($\hat{\theta}$) $= Var[\hat{\theta}] + [Bias(\hat{\theta})]^2$
OBS: S^2 har ett mindre MSE än MLE för σ^2för alla värden av μ och σ^2i en normal population. Därför är det bättre att dela med n-1 än bara n.

Fisher Information

Fishers information är ett mått på mängden information som en stokatisk variabel X innehåller en okänd parameter θ i en statisisk modell. Den kvantifierar hur känslig sannolikhetsfördelningen för X är för förändringar i θ, vilket är avgörande i att bedöma precisionen i parameterestimeringar.

$I(\theta) = -E_\theta\left(\frac{d2}{d\theta^2}f(X \mid \theta)\right)$	För ett slumpmässigt stickprov blir det (iid observationer $I_n(\theta) = -E_\theta\left(\frac{d2}{d\theta^2}f(X_1,\dots,X_n \mid \theta)\right)$	$I_n(\theta) = n * I(\theta)$

ML SKATTNINGARS ASYMPTOTISKA EGENSKAPER:

$$Vi\ låter\ \hat{\theta}\ vara\ ML\ skattning\ till\ \theta. Om\ n\ är\ stort\ så\ är\ \hat{\theta}\ approx\ N(\theta, \frac{1}{n * I, (\theta)}\ SIDA\ 523$$

FÖRELÄSNING 14 KAPITEL 9.1 9.5 9.7

TEST STATISTICS AND CRITICAL REGIONS

En statistika är en regel som baserat på data avgör om man ska anta nollhypotesen eller alternativhypotesen. Vi benämnner utfallsrummet för data $\boldsymbol{X} = (X_1,\dots,X_n) = S$. Det är även så att $S = S_0 \cup S_1$. Där:

- $S_0 = The\ data\ set\ for\ which\ H0\ is\ not\ rejected.$
- $S_1 = The\ data\ sets\ from\ which\ H0\ is\ rejected.$

$S_1\ kallas\ för$ det$kritiskaområ$det $för\ testet.$ **Vi förkastar $\boldsymbol{H_0}$ då $\boldsymbol{X \in S_1}$**

Normal EXEMPEL: $X_1,\dots,X_n|\mu,\sigma^2 \sim N(\mu,\sigma^2)\ där\ sigma\ känd.$

- $H_0: \mu = \mu_= \ Ha: \mu \ne \mu_0$
- $Statistika: Z = {\bar{X} - \mu_0}\Big/{\sigma/\sqrt{n}}$ Förkasta H0 om $|Z| \ge c(detta\ är S_1$ det $kritiska\ området)$

POWER FUNCTION

Styrkefunktionen av ett test $\boldsymbol{\delta}$, $\boldsymbol{\pi(\theta|\delta) = P}(Förkasta\ H0|\theta)$ $\quad vilket\ är\ givet\ av\ \boldsymbol{\pi(\theta|\delta) = P(X} \in S_1|\theta)$

TYPE OF ERRORS

Typ 1 fel (mycket allvarligt)	Typ 2 fel(inte så allvarligt)
Vi förkastar H0 när H0 är sann Kontroll: $\boldsymbol{\pi(\theta\|\delta) \le \alpha_0}$	**Vi förkastar in H0 när H0 är falsk**

Exempel: $X_1,\dots,X_n|\mu,\sigma^2 \sim N(\mu,\sigma^2)\ där\ sigma\ känd.$

- $H_0: \mu = \mu_= \ Ha: \mu \ne \mu_0$ där $Z = {\bar{X} - \mu_0}\Big/{\sigma/\sqrt{n}} \sim N(0,1)\ om\ H0\ sann.$
- $Förkasta\ H0$ om $|Z| \ge c$ där bestäm $P(|Z| \ge c|H_0) = \alpha_0$ vilket ger $c = \Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right)$.
- $Exempel\ om\ 5\%\ signfikans$: ger $c = \Phi^{-1}\left(1 - \frac{0.05}{2}\right) = \Phi^{-1}(0.975) = 1.96\ från\ std\ Normal\ tabell$

Exempel: Testing the probablity in Bernoulli trials

$X_1,..,X_n \sim Bern(p)$	Hyp: $H_0: p \le p_o\ H_1: p > p_0$	Statistika: $Y = \sum X_i\ Y \sim bin(p)$

$X_1,..,X_n \sim Bern(p)$ $H_0: pG\ J \le p_o\ H_1: p > p_0$ Statistika: $Y = \sum X_i\ Y \sim bin(p)$ Kritiskt området $Y \ge c$. Finn c by controlling the type I error. $P(Y \ge c|p)\ is\ an\ increasing\ function\ of\ p. Choose\ c\ such\ that\ P(Y \ge c|p = p_0) \le \alpha_0$. Exempel om N = 10, p_0 = 0.5, och $\alpha_0 = 0.05$. För $c = 8, P(Y \ge 8|p = p_0) = 0.0547$, och $c = 9\ ger\ P(Y \ge 9|p = p_0) = 0.0108$. Det vill säga att $Y \ge 9\ is\ a\ level\ \alpha_0 = 0.05\ test. The\ size\ of\ the\ test\ is\ 0.0108$

THE T TEST

$X_1,..,X_n|\mu,\sigma^2 \sim N(\mu,\sigma^2), \sigma^2\ okänd$ H0: $\mu = \mu_0$ och Ha: $\mu \ne \mu_0$
Statistika: $T = {\bar{X} - \mu_0}\Big/{s/\sqrt{n}} \sim t(n-1) om\ H0\ är\ sann$ Förkasta H0 om $|T| \ge c, där\ du\ finner\ c\ från\ P(|T| \ge c|H_0) = \alpha_0$
Vilket ger $c = F_{-1n}^{-1}(1 - \frac{\alpha_0}{2})$ där $F_{n-1}(\quad)är\ CDF\ av\ t\ fördelningen\ med\ n - 1\ frihetsgrader.$

THE P VALUE

P-värdet är den minsta signifikansnivå α_0 som skulle få oss att förkasta H0. P values are tail areas. P-value is the answer to the question if H0 is true what is the probability of obtaining a test statistic that is at least as extreme as the one observed in our data?

THE F TEST

TVÅ test från olika populationer	Hypoteserna	STATISTIKA
$X_1,\dots,X_m\|\mu_x,\sigma_X^2 \sim N(\mu_x,\sigma_X^2)$ $Y_1,\dots,Y_n\|\mu_Y,\sigma_>^2 \sim N(\mu_x,\sigma_Y^2)$ Där varianserna är okända	$\boldsymbol{H_0}: \sigma_X^2 = \sigma_Y^2$ $\boldsymbol{H_1}: \sigma_X^2 \ne \sigma_Y^2$ Förkasta när $V \le c_1 och\ V \ge c_2$	$V = {S_X^2/(m-1)}\Big/{S_Y^2/(n-1)}$ $S_X^2 = \sum_{i=1}^{m}/(X_i - \bar{X})^2$ $S_Y^2 = \sum_{i=1}^{n}/(Y_i - \bar{Y})^2$

THE F DISTRIBUTION

V's fördelning kan fås genom:${\sigma_Y^2}\Big/{\sigma_X^2} * V = {S_X^2/[(m-1)\sigma_X^2]}\Big/{S_Y^2/[(n-1)\sigma_Y^2]}$ which is the distribution of the ratio of two indepedent chisquare variables with m-1 degrees of freedom in the numerator and n-1 in the denominator.

F fördelningen: Låt $Y \sim X_m^2\ vara\ oberoende\ från\ W \sim X_n^2$ då är $X = {Y/m}\Big/{W/n} \sim F_{m,n}$ Detta betyder att X följer en F fördelning med m och n frihetsgrader. **Då under H0 har vi** $\sigma_X^2 = \sigma_Y^2$ så $V \sim F_{m-1,n-1}$. F fördelningen kan användas till för att få det kritiska området $[V \le c_1 or\ V \ge c_2]$ där $c_1 = F_{m-1,n-1}^{-1}(\alpha_0/2)$ $\quad c_2 = F_{m-1,n-1}^{-1}(1 - \alpha_0/2)$

LIKELIHOOD RATIO TEST

Om vi har $X_1,\dots,X_n|\theta \sim f(X_1,\dots,X_n|\theta)$. Note that we do not assume independence and θ can be a vector (t.ex $\theta = (\mu,\sigma^2)$
Hypotes: $H_0: \theta \in \Omega_0\ H_a: \theta \in \Omega_1$
Likelihood ratio test (LRT) statistika: $\Lambda(x_1,..,x_n) = {f(x_1,\dots,x_n|\widehat{\theta_0})}\Big/{f(x_1,\dots,x_n|\hat{\theta})}$ where $\widehat{\theta_0}$ is the MLE when the parameter space $\Omega_0, and\ \hat{\theta}$ is the MLE without restriction.
Kritiskt område: Reject H0 om : $\Lambda(x_1,..,x_n) \le c$.
Stora stickprov LRT: För stora strickprov n så har vi asymptosiskt: -2ln $\Lambda(x_1,..,x_n) \sim X_k^2$ där k är antalet parameter som är restricted under h0.0

EXEMPELUPPGIFT MED PDF/INVERSE CDF METODEN: Låt Y vara likformigt fördelad på intervallet 0 till 1. Låt X= $(Y+2)^2$
A) Bestäm fördelningsfunktion(=CDF) för X: $Y \sim Lik(0,1)\ där\ X = (Y+2)^2$ om y = 1 blir x = 9 och om y = 0 då blir x = 4 det vill säga x har följande stöd 4<x<9. $F(x) = fördelningsfunk\ till\ X = P(X \le x) = P((Y+2)^2 \le x) = P(Y \le \sqrt{x} - 2) = \sqrt{x} - 2$ ty $G(y) = y, 0 < y < 1$. Där F(X) $\{1\ om\ x \ge 0,\ \sqrt{x} - 2\ om\ 4 < x < 9, och\ 0\ om\ x \le 9$
B) Anta att du har simulerat 4 slumptal från Y och dessa blev 0.521 0.075 0.384 0.977. Använd inversa CDf metoden för att generera 4 slumptal på X. $Y = F(X)\ där\ F^{-1}(y_i) = x_i = (y_i + 2)^2$. Då sätter vi in de värden som angavs: $(\boldsymbol{0.521} + 2)^2. (\boldsymbol{0.075} + 2)^2. (\boldsymbol{0.384} + 2)^2. (\boldsymbol{0.977} + 2)^2$ *om man beräknar dessa ser man att detta är rimligt då talen ligger mellan 4 och 9.*