H6: schattingsmethodes

6.1 methode van de momenten

6.1.1 principe van de methode	
def: methode van de momenten	Een parameter a wordt afgeschat door te stellen dat het verwachte gemiddelde gelijk moet zijn aan het effectief geobserveerde gemiddelde uit de steekproef $$E_X(\hat{a}) = < X >= \int xP(x;\hat{a})dx = \bar{x}$$
stelling: fout op â 6.1.1	fout op â is gegeven door: $$V(\hat{a}) = \left(\frac{d\hat{a}}{d\bar{x}}\right)^2 \frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i-\bar{x})^2$$
6.1.2 generalisatie	
def: algemene MM	er is een set van N metingen x_i en een te fitten fistributie P(x;θ) > bevat k onbekende parameters **θ** = $\theta_1, \ldots, \theta_k$ > construeer k lineair onafhankelijke functies $g_j(x)$ > deze $g_j(x)$ zijn functies van meetgegevens met een verwachtingswaarde: $$< g_j(x) >= \int g_j(x)\; P(x;\vec{\theta})\; dx \equiv e_j(\vec{\theta})$$ de methode van de momenten voor **θ** wordt dan gegeven door: $$e_1(\widehat{\vec{\theta}}) = \frac{1}{N}\sum_{i=1}^{N} g_1(x_i)$$ $$\vdots$$ $$e_k(\widehat{\vec{\theta}}) = \frac{1}{N}\sum_{i=1}^{N} g_k(x_i)$$
functies van momenten	Een mogelijke keuze voor $g_j(x)$ zijn de veeltemen $x^1, x^2, \ldots, x^k$ > verwachtingswaarde is het j-de moment van x: $$< g_j(x) >=< x^j >$$ we kunnen echter ook orthonormale functies kiezen > bvb machten van sin,cos of polynomen
stelling: covariantie van schatters	schatters worden uit eenzelfde set van N waarnemingen gehaald > er zal een zekere correlaties tss de schatters zijn > bepaal via de covariantiematrix: $$\mathrm{cov}(\hat{\theta}_i, \hat{\theta}_j) = \sum_{l,m} \frac{\partial \hat{\theta}_i}{\partial \hat{e}_l}\frac{\partial \hat{\theta}_j}{\partial \hat{e}_m}\mathrm{cov}(\hat{e}_l, \hat{e}_m)$$ de schatter van de covariantie tss de verwachtingswaardefuncties is: $$\widehat{\mathrm{cov}}(\hat{e}_l, \hat{e}_m) = \frac{1}{N(N-1)}\sum_{i=1}^{N}(g_l(x_i)-\bar{g}_l)(g_m(x_i)-\bar{g}_m)$$

6.2 maximum (log)Likelihood	
6.2.1 principe van de methode	
principe vd maximum likelihood MLLH	voor een steekproef **x** = $\{x_1,\dots,x_N\}$ is de maximum likelihood schatter â die waarde van a waarbij de likelihood maximaal word > deze was gedefinieerd is als: $$\mathcal{L}(a:\vec{x}) = \prod_{i=1}^{N} P(x_i;a)$$ in praktijk is het makkelijker om de logaritmes de maximaliseren: $$\ln\mathcal{L}(a:\vec{x}) = \sum_{i=1}^{N} \ln P(x_i;a)$$ Het maximum vinden we dan: $$\left.\frac{\mathrm{d}\ln\mathcal{L}}{\mathrm{d}a}\right\|_{a=\hat{a}} = 0$$
6.2.2 Gauss paramater schatten	
MLLH op Gauss	naam aan dat er $\{x_i\}$ metingen zijn van dezelfde grootheid met versch. precisie > ie: x_i is getrokken uit een Gauss met gekende standaardafwijking σ_i > waarvoor we nu een schatting voor µ willen maken De waarschijnlijkheidsfunctie: $$P(x_i;\mu,\sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}}e^{-\frac{(x_i-\mu)^2}{2\sigma_i^2}}$$ zodat de logaritme van de likelihood gelijk wordt aan $$\ln\mathcal{L} = \sum_i\left(-\ln(\sqrt{2\pi\sigma_i^2})\right) - \sum_i\frac{(x_i-\mu)^2}{2\sigma_i^2}$$ en afleiden naar μ om het maximum te vinden geeft: $$\sum_i\frac{(x_i-\hat{\mu})}{\sigma_i^2} = 0$$ $$\hat{\mu} = \frac{\sum(x_i/\sigma_i^2)}{\sum(1/\sigma_i^2)}$$ > x_i waarden uitgemiddeld met elk een gewicht σ^{-2} en correct genormeerd > dit hebben we al gevonden voor het gewogen gemiddelde
MLLH op Gauss met onbekende σ	Stel nu dat ook σ onbekend is > we moeten afleiden naar beide grootheden: $$\begin{cases} \mu & \rightarrow & \sum_i \frac{(x_i-\hat{\mu})}{\hat{\sigma}^2} = 0 \\ \sigma & \rightarrow & -\frac{N}{\hat{\sigma}} + \frac{\sum(x_i-\hat{\mu})^2}{\hat{\sigma}^3} = 0 \end{cases}$$ we vinden voor µ hetzelfde > we vinden $\hat{\mu} = \bar{x}$, zoals tevoren. > dan hebben we de vgl: $\hat{\sigma}^2 = \frac{1}{N}\sum(x_i-\bar{x})^2$
6.2.3 eigenschappen van de MLLH methode	
stelling: consistentie van MLLH	MLLH schatters zijn meestal consistent, maar niet altijd
stelling: parametertransformatie en MLLH	MLLH schatters zijn invariant onder parameter transformaties: $$\widehat{f(a)} = f(\hat{a})$$

6.2.4 asymptotische limiet en MBV	
bias in asymptotische limiet	Voor N→∞ wordt elke consistente schatter onbevooroordeeld > bias van consistente MLLH verdwijnt in asymptotische limiet
efficiëntie in asymptotische limiet	Veronderstel dat de echte waarde voor a a_0 is > de schatter â voldoet aan: $$\left.\frac{d\ln\mathcal{L}}{da}\right\vert_{a=\hat{a}}=0$$ in een Taylorreeks rond a_0: $$\left.\frac{d\ln\mathcal{L}}{da}\right\vert_{a_0}+(\hat{a}-a_0)\left.\frac{d^2\ln\mathcal{L}}{da^2}\right\vert_{a_0}+\mathcal{O}(\hat{a}^2)=0$$ $(\hat{a}-a_0)$ wordt klein voor grote N > enkel eerste twee termen bekijken Echter nu, â is niet identiek aan a_0 > want afgeleide van ln$\mathcal{L}$ bij a_0 wijkt af van 0 door statistische fluctuatie van gegevens: $$\left.\frac{d\ln\mathcal{L}}{da}\right\vert_{a_0}\neq 0$$ we weten ook dat: $$\left\langle\frac{d\ln\mathcal{L}}{da}\right\rangle=0$$ Vermits de afgeleide bekomen wordt als som van N onafh. termen zegt het Centraal Limiet theorema ons dat de verdeling Gaussisch is
variantie van dln$\mathcal{L}$	We vinden: $$V\left(\left.\frac{d\ln\mathcal{L}}{da}\right\vert_{a_0}\right)=\left\langle\left(\frac{d\ln\mathcal{L}}{da}\right)^2\right\rangle-\underbrace{\left\langle\frac{d\ln\mathcal{L}}{da}\right\rangle^2}_{0}=-\left\langle\frac{d^2\ln\mathcal{L}}{da^2}\right\rangle \qquad (6.10)$$ De vewachtingswaarden worden uitgewerkt voor $a=a_0$ en is dus de tweede term van de variantie gelijk aan nul.
variantie van $(\hat{a}-a_0)$	De Taylorreeks geeft ons dat $(\hat{a}-a_0)$ evenredig is met dln$\mathcal{L}$/da > wordt ook beschreven door Gauss met gemiddelde: $$<\hat{a}-a_0>=\left\langle\frac{d\ln\mathcal{L}}{da}\right\rangle=0$$ nu dus: $$V(\hat{a}-a_0)=\left(\frac{d(\hat{a}-a_0)}{d(d_a\ln\mathcal{L})}\right)^2 V\left(\frac{d\ln\mathcal{L}}{da}\right)=-\left\langle\frac{d^2\ln\mathcal{L}}{da^2}\right\rangle\Big/\left(\left.\frac{d^2\ln\mathcal{L}}{da^2}\right\vert_{a_0}\right)^2$$
Variantie bij grote N	In de limiet van grote N convergeren verwachtingswaarden naar de echte waarde, $$\left\langle\frac{d^2\ln\mathcal{L}}{da^2}\right\rangle=\left.\frac{d^2\ln\mathcal{L}}{da^2}\right\vert_{a_0}$$ zodat we de verwachtingswaarde in de teller mogen vervangen door zijn waarde bij a_0 . Daardoor valt één van de factoren in de noemer weg, zodat we voor de variantie van $\hat{a}$ vinden: $$V(\hat{a})=\sigma_{\hat{a}}^2=-\frac{1}{\left(\left.\frac{d^2\ln\mathcal{L}}{da^2}\right\vert_{a_0}\right)}=-\frac{1}{\left\langle\frac{d^2\ln\mathcal{L}}{da^2}\right\rangle} \qquad (6.11)$$ hetgeen precies de MVB is.
6.2.5 fouten op de MLLH schatters	
stelling: fout op MLLH schatter	Voor elke onbevooroordeelde efficiënte MLLH schatter kan de fout geëvalueerd worden: $$V(\hat{a})^{-1}=-\left.\frac{d^2\ln\mathcal{L}}{da^2}\right\vert_{a=\hat{a}}$$
stelling: waarden van log-likelihood	$$\ln\mathcal{L}(a)=\ln\mathcal{L}_{\max}-\frac{(a-\hat{a}(\vec{x}))^2}{2\sigma^2}$$ Bij het punt 1σ weg van de piek is ln$\mathcal{L}$ met 0.5 verminderd 2σ 2 3σ 4.5 … …

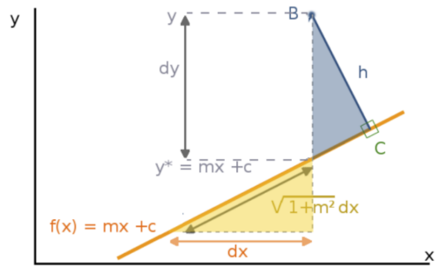
vor van de log-likelihood	Een beetje herwerken van de formule vinden we: $$\ln \mathcal{L}_{\max} - \ln \mathcal{L} = \frac{A}{2}[\hat{a}(\vec{x}) - a]^2$$ De log-likelihood is dus een parabool (Fig.6.3). Exponentiëren leidt tot $$\mathcal{L}(\vec{x}; a) = \mathcal{L}_{\max} e^{\frac{A}{2}[a - \hat{a}(\vec{x})]^2}$$
asymmetrische fouten	Als N niet groot genoeg is, is de Likelihood ftie geen Gauss en log-likelihood geen parab. > bekijk waar de functie 0.5 gezakt is > dit zal niet meer perfect symmetrisch liggen > asymmetrische fouten, notatie vb: $$a = 0.21^{+0.29}_{-0.27}$$
6.2.6 meerdere parameters	
MLLH voor meerdere parameters	Veralgemeen alles tot nu toe gewoon voor een meerdimensionaal geval > we zoeken nu dus: $$\frac{\partial \ln \mathcal{L}(x_1 \ldots, x_N; a_1, \ldots, a_k)}{\partial a_j} = 0 \qquad \forall j = 1, \ldots, k$$ > zelfde voor invariantie > likelihood ftie is een meerdimensionale Gauss
stelling: covariantiematrix	We vinden: $$\sigma_{\hat{a}_j}^{-2} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial a_j^2} \right\rangle \qquad \mathrm{cov}^{-1}(a_i, a_j) = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial a_i \partial a_j} \right\rangle = -\left. \frac{\partial^2 \ln \mathcal{L}}{\partial a_i \partial a_j} \right\|_{a=\hat{a}}$$
vorm van likelihood functie	voor grote N in 2D vormt ze een 2D Gauss: $$\ln \mathcal{L}(a_1, a_2) = (\ln \mathcal{L})_{\max} - \frac{1}{2(1-\rho^2)} \left[\left(\frac{a_1 - \hat{a}_1}{\sigma_{\hat{a}_1}} \right)^2 + \left(\frac{a_2 - \hat{a}_2}{\sigma_{\hat{a}_2}} \right)^2 - 2\rho \left(\frac{a_1 - \hat{a}_1}{\sigma_{\hat{a}_1}} \right) \left(\frac{a_2 - \hat{a}_2}{\sigma_{\hat{a}_2}} \right) \right]$$ waarin $\rho = \mathrm{cov}(\hat{a}_1, \hat{a}_2)/(\sigma_{\hat{a}_1} \sigma_{\hat{a}_2})$ de correlatiecoëfficiënt voor a_1 en a_2 is. Het contour van $\ln \mathcal{L}(a_1, a_2) = \ln \mathcal{L}_{\max} - 0.5$ is dus gegeven door: $$\frac{1}{1-\rho^2} \left[\left(\frac{a_1 - \hat{a}_1}{\sigma_{\hat{a}_1}} \right)^2 + \left(\frac{a_2 - \hat{a}_2}{\sigma_{\hat{a}_2}} \right)^2 - 2\rho \left(\frac{a_1 - \hat{a}_1}{\sigma_{\hat{a}_1}} \right) \left(\frac{a_2 - \hat{a}_2}{\sigma_{\hat{a}_2}} \right) \right] = 1$$ Dit is een ellips met als centrum de MLLH schatters $(\hat{a}_1, \hat{a}_2)$ die een hoek ϕ maakt met de a_1-as: $$\tan 2\phi = \frac{2\rho \sigma_{\hat{a}_1} \sigma_{\hat{a}_2}}{\sigma_{\hat{a}_1}^2 - \sigma_{\hat{a}_2}^2}$$ Merk in het bijzonder op dat de raaklijnen aan de ellips parallel met de assen de ellips snijden bij $a_1 = \hat{a}_1 \pm \sigma_{\hat{a}_1}^2$ en $a_2 = \hat{a}_2 \pm \sigma_{\hat{a}_2}^2$.
6.2.7 opmerkingen bij Maximum Likelihood methode	
opmerkingen	- 'maximum' betekend niet 'meest waarschijnlijke' waarde positief: - voor grote N heeft â een waarsch.verdeling die onbevooroordeeld is en Gaussisch rond de echte waarde van a met variantie gegeven door de MVB > heel goede schatter - er gaat geen info verloren door gegevens samen te nemen in bins negatief: - voor kleine N vertoont de MLLH een bias - je moet de vorm van de ouderdistributie kennen - de vgl van dln$\mathcal{L}$/da is niet altijd op te lossen
6.2.8 maximum likelihood en Monte Carlo	
max likelihood en monte carlo	in sommige gevallen is likelihood niet analytisch op te lossen > bereken numeriek > veronderstel dat de schatter efficiënt en onbevooroordeeld is > gebruik Monte Carlo om MLLH uit te breiden

6.2.9 uitgebreide maximum likelihood	
uitgebreide maximum likelihood EML	Standaard is de waarschijnlijkheidsdichtheid genormeerd: $$\int P(x;a)dx = 1$$ laat deze voorwaarde echter los > neem een functie Q, zonder een vaststaande normering: $$\int Q(x;a)dx = \nu$$ > handig voor experimenten waarbij aantal gebeurtenissen a priori bekend is Een bepaalde Q(x;a) voorspeld dus v elementen in een gebied > verm. de likelihood met de Poisson dat N evenementen observeert als gemiddelde v: $$e^{-\nu}\frac{\nu^N}{N!}$$ dan is de log-likelihood: $\begin{aligned}\ln\mathcal{L} &= \sum \ln P(x_i;a) - \nu + N\ln\nu \\ &= \sum \ln(\nu P(x_i;a)) - \nu \\ &= \sum \ln Q(x_i;a) - \nu\end{aligned}$ een toename in normering van Q zal de likelihood doen stijgen > waargenomen evenement wordt meer waarschijnlijk
eigenschappen van EML	Zelfde eigenschappen als MLLH
covariantiematrix EML	$$\text{cov}(a_i,a_j)^{-1} = -\left\langle \frac{\partial^2 \ln\mathcal{L}}{\partial a_i \partial a_j}\right\rangle = -\int \frac{\partial \ln Q}{\partial a_i}\frac{\partial \ln Q}{\partial a_j} Q(x;a)dx$$ > zelfde als MLLH op een factor N na

6.3 kleinste kwadraten	
gebruik kleinste kwadraten	Je hebt twee variabelen x en y en: 1. een aantal exact gekende x waarden, 2. een corresponderende verzameling y waarden, gemeten met een nauwkeurigheid σ_y, 3. een functie $f(x;\vec{\theta})$ die de waarde van y voorspelt voor om het even welke x; de vorm van deze functie is bekend maar bevat onbekende parameters $\vec{\theta}$ die je probeert te bepalen. >> om onbekende parameters te bepalen uit een aantal gegevens
6.3.1 principe van de methode	
def: kleinste kwadraten	Je minimaliseert het (gewogen) kwadratisch verschil tss een groep van N metingen (x_i,y_i) en de voorspelde waarden f(x_i;**θ**): $$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x_i;\vec{\theta})}{\sigma_i} \right]^2$$ > laat de parameters die je wilt bepalen variëren > voorspelde waarden worden zodanig aangepast dat je dicht bij waarneming uitkomt > kwadrateren zorgt voor een meer uitgesproken verschil
werking van KK	Uit een set van N precieze waarden x={x_1,…,x_N} met hiervoor y={y_1,…,y_N} met σ_i nauwkeurigheid op y_i > methode om een parameter a van een functie f(x;a) te schatten die de echte waarde van y voorspelt voor om het even welke x > kies dan die waarde van a die de kleinste $\mathcal{X}^2$ geeft > als f afgeleid kan worden, kan deze kleinste $\mathcal{X}^2$ gevonden worden via: $$\frac{d\chi^2}{da} = \sum_i \frac{2}{\sigma_i^2} \frac{df(x_i;a)}{da} [y_i - f(x_i;a)] = 0$$ de bekomen schatting â op a heeft een fout de we kunnen uitrekenen > nl: â wordt gegeven als functie van y_i deze een fout σ_i hebben > formule voor propagatie van fouten geeft de fout op â
KK in meerdere dimensies	de functie hangt nu af van een **θ**={θ_1,…,θ_k} ipv a > er zijn k gekoppelde vgln in k onbekenden: $$\frac{\partial\chi^2}{\partial\theta_i} = 0 \qquad \forall i = 1, \dots k$$
KK vanuit maximum likelihood	als de metingen y Gaussische verdeeld zijn > neem aan dat de y gegeven worden door een ftie f van x die afh. is ve parameter a > functie geldt evenwel enkel voor de ideale y > onze metingen wijken daarvan af door de resolutie Centraal Limiet Theorema: verdeling vd gemeten y rond de ideale waarde is Gaussisch > waarschijnlijkheid om bepaalde y_i te hebben voor gegeven x_i: $$P(y_i;a) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(y_i - f(x_i;a))^2}{2\sigma_i^2}}$$ de log-likelihood vd volledig steekproef is dus: $$\ln \mathcal{L} = -\frac{1}{2} \sum_{i=1}^{N} \left[\frac{y_i - f(x_i;a)}{\sigma_i} \right]^2 - \sum_{i=1}^{N} \ln \sigma_i \sqrt{2\pi}$$ om de likelihood te maximaliseren moet de volgende geminimaliseerd worden: $$\sum_{i=1}^{N} \left[\frac{y_i - f(x_i;a)}{\sigma_i} \right]^2$$ wat niet anders dan de KK is

6.3.2 een rechte fitten	
toepassing KK	aanpassen van een rechte is meest voorkomende toepassing > metingen y_i hebben allemaal een fout σ en liggen op een lijn y=mx+c
y=mx fitten	we moeten $\mathcal{X}^2$ minimaliseren door m aan te passen: $\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - m x_i}{\sigma_i}\right]^2$ (6.18) Dit afleiden naar m geeft: $\frac{d\chi^2}{dm} = \sum -2x_i \frac{y_i - m x_i}{\sigma_i^2}$ Als alle σ_i gelijk zijn wordt dit een gemeenschappelijke factor die buiten de som kan gebracht worden, hetgeen leidt tot: $\frac{-2}{\sigma^2} \sum (x_i y_i - m x_i^2)$ Voor de kleinste kwadraten schatting van de helling, $\hat{m}$ is deze som gelijk aan 0 en dus: $\sum x_i y_i = \hat{m} \sum x_i^2$
stelling: ^m	De KK schatting van m is: $\hat{m} = \frac{\overline{xy}}{\overline{x^2}}$ met nauwkeurigheid $V(\hat{m}) = \sum_{i=1}^{N} \left(\frac{d\hat{m}}{dy_i}\right)^2 \sigma_i^2 = \sum_{i=1}^{N} \left(\frac{x_i}{N\overline{x^2}}\right)^2 \sigma^2$ > dus: $V(\hat{m}) = \frac{\sigma^2}{N\overline{x^2}}$
stelling: fitten van y=mx+c	als alle σ_i gelijk zijn, dan: $\hat{m} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\text{cov}(x,y)}{V(x)}$ en $\hat{c} = \bar{y} - \hat{m}\bar{x} = \frac{\overline{x^2}\bar{y} - \bar{x}\overline{xy}}{\overline{x^2} - \bar{x}^2}$ de eerste uitdrukking voor ^c geeft aan dat de lijn door het zwaartepunt $(\bar{x}, \bar{y})$ gaat
stelling: fouten op schatters	fouten op schatter worden gegeven door $V(\hat{m}) = \frac{1}{N(\overline{x^2} - \bar{x}^2)}\sigma^2$ en $V(\hat{c}) = \frac{\overline{x^2}}{N(\overline{x^2} - \bar{x}^2)}\sigma^2$ De covariantie en correlatiecoëfficiënt tussen de schatters $\hat{m}$ en $\hat{c}$ wordt gegeven door : $\text{cov}(\hat{m}, \hat{c}) = \frac{-\bar{x}}{N(\overline{x^2} - \bar{x}^2)}\sigma^2$ $\rho_{\hat{m},\hat{c}} = \frac{\text{cov}(\hat{m}, \hat{c})}{\sqrt{V(\hat{m})V(\hat{c})}} = -\frac{\bar{x}}{\sqrt{\overline{x^2}}}$
stelling: $\mathcal{X}^2$ zonder ^m en ^c	De $\mathcal{X}^2$ kan berekend worden zonder de voorafgaande kennis van ^m en ^c > zonder de som van de afwijkingen te nemen: $\chi^2 = N\frac{V_y}{\sigma^2}(1 - \rho_{x,y}^2)$
stelling: systematische fout	Een systematische fout y heeft geen invloed op de helling van ^m maar wordt (kwadratisch) opgeteld bij de fout op het intercept ^c
$\mathcal{X}^2$ bij ongelijke σ_i	als de σ_i niet allen gelijk zijn dan: $\chi^2 = \sum_i \frac{1}{\sigma_i^2}(y_i - m x_i - c)^2$ > geeft dezelfde vgln voor ^m en ^c maar: $\bar{x}$ *en* $\bar{y}$ *zijn de gewogen gemiddelden met gewichtsfunctie* $w_i = \sigma_i^{-2}$ de normering wordt dan door het totale gewicht W gegeven, niet door N: $W = \sum_{i=1}^{N} \sigma_i^{-2}$ op precies dezelfde manier moet de grootheid σ^2 vervangen worden door: $\overline{\sigma^2} = \frac{1}{W}\sum_{i=1}^{N} w_i \sigma_i^2 = \frac{1}{W}\sum_{i=1}^{N} \sigma_i^{-2}\sigma_i^2 = \frac{N}{W}$

Extrapolatie van y	Voor een gegeven X > de voorspelling Y is gelijk aan: $$\hat{m}X + \hat{c}.$$ De fout hierop: $$V(Y) = V(\hat{c}) + X^2 V(\hat{m}) + 2X \operatorname{cov}(\hat{m}, \hat{c})$$ > we willen echter deze laatste term vermijden > als $\bar{x} = 0$ valt deze term weg > je moet je later niet herinneren om hem mee te nemen > $Y = \hat{m}(X - \bar{X}) + \hat{c}'$ >> geeft de ongecorreleerde schattingen voor $\hat{m}$ en $\hat{c}'$.
stelling: fout op Y	de fout op de geëxtrapoleerde waarde Y is: $$V(Y) = \frac{\sigma^2}{N(\overline{x^2} - \bar{x}^2)}(X - \bar{x})^2 + \frac{\sigma^2}{N}$$
6.3.3 gebinde gegevens	
principe van bins	Bij grote steekproeven zijn er veel computaties nodig > oplossing: neem gegevens samen in bins veronderstel er zijn N evenementen met waarschijnlijkheidsfunctie P(x;a) > deze zijn gesorteerd in intervallen van 1 tot N_b > interval j is gecentreerd op punt x_j, heeft breedte B_j en bevat n_j elementen ideaal verwachte aantal elementen in een bin j is $f_j = NB_jP(x;a)$ > eigenlijke aantal wordt beschreven door Poisson statistiek > variantie is gelijk aan gemiddelde > de totale $\mathcal{X}^2$ gesommeerd over alle intervallen is: $$\chi^2 = \sum_j^{N_b} \frac{(n_j - f_j)^2}{f_j}$$ of vereenvoudigd: $$\chi^2 \approx \sum_j^{N_b} \frac{(n_j - f_j)^2}{n_j}$$
6.3.4 de $\mathcal{X}^2$-verdeling	
def: $\mathcal{X}^2$	= kwadraat van verschil in waargenomen y_i^w en theoretische y_i^t gewogen volgens de fout op de waarnemingen σ_i $$\chi^2 = \sum_i \frac{(y_i - f(x_i))^2}{\sigma_i^2} = \sum_{i=1}^{N} \left(\frac{y_i^w - y_i^t}{\text{verwachte fout}} \right)^2$$
def: aantal vrijheidsgraden	= aantal punten N in de som min het aantal variabelen k dat aangepast is om $\mathcal{X}^2$ te minimaliseren
betekenis van $\mathcal{X}^2$	- als de functie goed overeenkomt zal $\mathcal{X}^2$ klein zijn - als $\mathcal{X}^2$ te klein is, betekend dit dat de fout σ op de metingen fout ingeschat is
st: waarschijnlijkheidsdichtheid	de waarsch.dichtheid van $\mathcal{X}^2$ met n vrijheidsgraden is: $$P(\chi^2; n) = \frac{2^{-n/2}}{\Gamma(n/2)} \chi^{n-2} e^{-\chi^2/2}$$ waarbij $\Gamma(x)$ de standaard gamma functie is.
stelling: som van $\mathcal{X}^2$	Als 2 variabelen een $\mathcal{X}^2$ verdeling volgen met N_1 en N_2 metingen > dan wordt de som van de 2 beschreven door een $\mathcal{X}^2$ met N_1+N_2 metingen

6.3.5 fouten op x en y	
casestudy: fouten op x en y	Veronderstel dat er fouten zijn op zowel x als y: 1: fout op x = fout op y = σ 2: x en y verschillende nauwkeurigheid
geval 1: gelijke fouten	fit een rechte lijn en pas de maximum likelihood toe > een gemeten punt B kan afkomstig zijn van eender welk punt A op de ideale rechte > waarschijnlijkheidsdichtheid hiervoor is: $$\begin{aligned} P(A \to B) &= \frac{1}{2\pi\sigma^2} e^{-\frac{(x_A - x_B)^2}{2\sigma^2}} e^{-\frac{(y_A - y_B)^2}{2\sigma^2}} \\ &= \frac{1}{2\pi\sigma^2} e^{\frac{-r^2}{2\sigma^2}} \\ &= \frac{1}{2\pi\sigma^2} e^{-\frac{u^2}{2\sigma^2}} e^{-\frac{h^2}{2\sigma^2}} \end{aligned}$$ om de totale waarschijnlijkheid te vinden dat B afkomstig is van A, integreren over alle u: $$P(B) = \frac{1}{2\pi\sigma^2} \int e^{-\frac{u^2}{2\sigma^2}} e^{-\frac{h^2}{2\sigma^2}} du = \frac{1}{2\pi\sigma^2} \sqrt{2\pi\sigma^2} e^{-\frac{h^2}{2\sigma^2}}$$ h is de loodrechte afstand van het punt tot de lijn ie: van het gemeten punt tot het punt C > C is het meest waarschijnlijke punt op de rechte om $\mathcal{X}^2$ te vormen: sommeer alle gekwadrateerde afstanden tss B en overeenkomstige C > uit gelijkvormige driehoeken uit de figuur vinden we h: $$h_i = \frac{y_i - (mx_i + c)}{\sqrt{1 + m^2}}$$ dus we vinden voor $\mathcal{X}^2$ $$\chi^2 \propto \sum_i \frac{(y_i - mx_i - c)^2}{1 + m^2}$$ - afleiden naar c levert de formule $\bar{y} = \hat{m}\bar{x} + \hat{c}$ - afleiden naar m levert: $$\hat{m} = A \pm \sqrt{A^2 + 1} \qquad \text{met} \qquad A = \frac{V(y) - V(x)}{2\text{cov}(x, y)}$$ > twee oplossingen voor ^m staan loodrecht op elkaar > één geeft de beste oplossing, de ander de slechtste 
geval 2: ongelijke fout	Herschaal de variabelen zodat ze wel dezelfde fout hebben: $$y' = \frac{y}{\sigma_y}, \; x' = \frac{x}{\sigma_x}$$ na transformatie terug naar oorspronkelijke x en y vinden we: $$\hat{m} = \frac{\sigma_y}{\sigma_x}(A \pm \sqrt{A^2 + 1}) \qquad \text{met} \qquad A = \frac{\sigma_x^2 V(y) - \sigma_y^2 V(x)}{2\sigma_x \sigma_y \text{cov}(x, y)}$$

6.3.6 lineaire kleinste kwadraten en matrices	
variabelen in vectorvorm	- **a** is een vector met k parameters $a_1,\dots,a_k$ - **y** is vector met alle y_i - **f** is vector met alle $f(x_i,\mathbf{a})$ >> allen vectoren hebben N elementen
$\mathcal{X}^2$ in matrixnotatie	we weten: $\chi^2 = \sum_i \sum_j [y_i - f(x_i;\mathbf{a})] V_{ij}^{-1} [y_j - f(x_j;\mathbf{a})]$ of dus: $\chi^2 = (\tilde{\mathbf{y}} - \tilde{\mathbf{f}}) \mathbf{V}^{-1} (\mathbf{y} - \mathbf{f})$ als er onafhankelijke metingen zijn, dan is **V** diagonaal: $V_{ij} = \sigma_i^2 \delta_{ij} \qquad V_{ij}^{-1} = \frac{\delta_{ij}}{\sigma_i^2}$ >> leidt nu af naar elke a_r en stel deze gelijk aan 0 > we krijgen n vgln waaruit we **â** kunnen halen
stelling: **â** en **V**(**â**)	$\hat{\mathbf{a}} = (\tilde{\mathbf{C}}\mathbf{V}^{-1}\mathbf{C})^{-1}\tilde{\mathbf{C}}\mathbf{V}^{-1}\mathbf{y}$ de nauwkeurigheid wordt gegeven door de covariantiematrix $\mathbf{V}(\hat{\mathbf{a}}) = (\tilde{\mathbf{C}}\mathbf{V}(\mathbf{y})^{-1}\mathbf{C})^{-1}$ > **V** is een NxN matrix **C** is een rechthoekige matrix met N rijen en n kolommen > **C** is de matrix bestaande uit alle C_{ir}: $C_{ir} = c_r(x_i)$
6.3.7 niet-lineaire kleinste kwadraten	
niet-lineaire kleinste kwadraten	indien f(x;**a**) niet lineair, gebruik een iteratieve methode voor een eerste gok $\mathbf{a}^0$ zijn de gradiënten voor onafhankelijke metingen: $\left.\frac{\partial \chi^2}{\partial a_r}\right\|_{\mathbf{a}=\mathbf{a}} = g_r(\mathbf{a^0}) = \sum_i -\frac{2}{\sigma_i^2}\left[y_i - f(x_i;\mathbf{a^0})\right]\frac{\partial f(x_i;\mathbf{a^0})}{\partial a_r}$ Je wilt en minimum raken, en dus een increment $\delta\mathbf{a}$ vinden waarvoor geldt : $g_r(\mathbf{a^0}+\delta\mathbf{a}) = \left.\frac{\partial \chi^2}{\partial a_r}\right\|_{\mathbf{a}=\mathbf{a}+\delta\mathbf{a}} = 0 \qquad \forall r$ Je doet dit door een ontwikkeling van $\frac{\partial\chi^2}{\partial a_r}$ in een Taylor reeks, waarbij je enkel de nulde en eerste orde termen behoudt : $g_r(\mathbf{a^0}+\delta\mathbf{a}) \approx g_r(\mathbf{a^0}) + \sum_s \frac{\partial g_r}{\partial a_s}\delta a_s = g_r(\mathbf{a^0}) + \sum_s \frac{\partial^2\chi^2}{\partial a_r \partial a_s}\delta a_s$ Noteren we $f(x_i;a^0)$ voor het gemak als f_i dan: $G_{rs} = \frac{\partial^2\chi^2}{\partial a_r\partial a_s} = \sum_i \frac{-2}{\sigma_i^2}\left[-\left(\frac{\partial f_i}{\partial a_r}\right)\left(\frac{\partial f_i}{\partial a_s}\right) + (y_i - f_i)\left(\frac{\partial^2 f_i}{\partial a_r \partial a_s}\right)\right]$ Eindelijk vinden we uit de matrix vergelijking : $\delta\mathbf{a} = -\mathbf{G}^{-1}\mathbf{g}$ waarbij de oplossing via iteratie bekomen wordt. >> wanneer de g_r klein genoeg worden stopt de iteratie > resulterende **a** wordt als oplossing genomen implementatie in python: standaardpakket in software bibliotheek