MTH3028 STATISTICAL INFERENCE: THEOREY AND PRACTICE

scroll down to bottom to tick for all chapters or 2 week

can post questions at bottom of each weeks tile

each week read some of lecture notes & associated videos

recordings of sessions go on recording sessions tile

full lecture notes available

each week study given parts of each section

prerecorded videos for exercises

very bottom page before each week requests & feedback

overview tuesdays   feedback session thursday

coursework due first day of term?

office hours tues 1.30pm fri 3.30pm

<u>distri</u> <u>func</u> $F(x;\theta) = Pr(X_1 \le x) = \ldots = Pr(X_n \le x)$

$F(x;\theta) = \int_{-\infty}^{x} f(y;\theta)\,dy$ & $f(x;\theta) = \frac{d}{dx} F(x;\theta)$

when r.v. discrete $f(x;\theta)$ denotes Z <u>mass func</u> $Pr(X_1 = x)$

\Ex 1.1/ $F(x;\theta) = 1 - \exp(-\theta x)$

density func $f(x;\theta) = \theta \exp(-\theta x)$ for $x > 0$

$\hat{\theta}$ $\underline{x} = (x_1, \ldots, x_n)$ into esti $\hat{\theta}(\underline{x})$ of θ

$m_k(\theta) = E(X_1^k)$ be k^{th} moment of F ∴ $\hat{\theta}$ for θ:

$m_k(\hat{\theta}) = \frac{1}{n} \sum_{i}^{n} x_i^k$ for $k = 1, 2, \ldots$

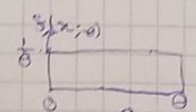
\Ex 1.2/ $X_1, \ldots X_n$ independent $Uni(0, \theta)$ density

$f(x;\theta) = 1/\theta$ for $0 \le x \le \theta$ ∴

$E(X_1) = \int_0^\theta x f(x;\theta)\,dx = \int_0^\theta \frac{x}{\theta}\,dx = \frac{1}{\theta}\left[\frac{x^2}{2}\right]_0^\theta = \frac{\theta}{2}$ ∴ $\hat{\theta}/_2 = \bar{x}$ ∴

Method of moments esti $\hat{\theta} = 2\bar{x}$

$f(x;\theta)$



∴ $f(x;\theta) = \frac{1}{\theta}$ for $x \in (0, \theta)$

$m_1(\theta) = E(X) = \int_0^\theta x f(x;\theta)\,dx = \int_0^\theta x \frac{1}{\theta}\,dx = \frac{1}{\theta}\left[\frac{1}{2}x^2\right]_0^\theta = \frac{\theta}{2}$

$m_1(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$ ∴ $\bar{x} = \frac{\hat{\theta}}{2}$ ∴ $\hat{\theta} = 2\bar{x}$

\Ex 1.3/ $X_1, \dots, X_n$ indep Exp(θ) density $f(x;\theta)=\theta\exp(-\theta x)$

for $x>0$ ∴

$E(X_1)=\int_0^1 x f(x;\theta)\,dx = \int_0^\infty x\theta e^{-\theta x}\,dx = \left[-xe^{-\theta x}\right]_0^\infty - \left[\theta^{-1}e^{-\theta x}\right]_0^\infty = \frac{1}{\theta}$

∴ $1/\hat\theta = \bar{x}$ ∴ $\hat\theta = 1/\bar{x}$ $\quad \bar{x} = \sum_{i=1}^{n} x_i/n$ is Z sample mean

independent joint density or mass func $\underline{X}=(X_1, \dots, X_n)$ is:

$f_n(\underline{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta)$ when $f_n(\underline{x};\theta)$ is a func of θ

\Def 1.1/ Z likelihood based on data $\underline{x}$ is $L(\theta;\underline{x}) = f_n(\underline{x};\theta)$

Z log-likelihood is $l(\theta;\underline{x}) = \log L(\theta;\underline{x})$

Z maximum likelihood estimate is Z val of θ that maximises Z likelihood

\Ex 1.4/ $X_1, \dots, X_n$ indep Exp(θ) ∴ $f(x;\theta)=\theta e^{-\theta x}$

$L(\theta;\underline{x}) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} (\theta e^{-\theta x_i}) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right)$ ∴

$f(x;\theta) = \theta e^{-\theta x}$ ∴

$f_n(\vec{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$ ∴

$L(\theta;\underline{x}) = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$ ∴

$l(\theta;\underline{x}) = n\log\theta - \theta \sum_{i=1}^{n} x_i$ ∴

$l'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$ ∴ $l'(\hat\theta)=0$ ∴ $\frac{n}{\hat\theta} = \sum_{i=1}^{n} x_i$ ∴

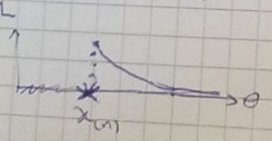
$\hat\theta = \frac{1}{\bar{x}}$ ∴

$l''(\hat\theta) = -\frac{n}{\hat\theta^2} < 0$ ∴ $\hat\theta = 1/\bar{x}$ is mle ∴

$l'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$ ∴ $l'(\hat\theta)=0$ ∴ $\hat\theta = \frac{1}{\bar{x}}$ ∴ $l''(\hat\theta) = -\frac{n}{\hat\theta^2} < 0$

\Ex 1.5/ $X_1, \dots, X_n$ Uni$(0,\theta)$

$L(\theta;\underline{x}) = \prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \theta^{-n} & \text{if } \theta \geq x_{(n)} \\ 0 & \text{if } \theta < x_{(n)} \end{cases}$ $\quad x_{(n)} = \max\{x_1, \dots, x_n\}$



∴ mle is $\hat\theta = x_{(n)}$ likelihood is not diffable at $\hat\theta$ in this case

estimator is a func $\hat{\theta}(\underline{X})$ of r.v. $\underline{X}$ & ∴ itself is a r.v.
an estimate is a num, $\hat{\theta}(x)$ by subing x for $\underline{X}$
quality of estimation by its sampling distri
$l(\theta)$ for $L(\theta;\underline{X})$

\Def 1.2/ Z score func is Z first deriv of Z log-likelihood:
$U(\theta;\underline{X}) = \frac{\partial l}{\partial \theta}$ if θ is a vec with ith elem θ_i then $U(\theta;\underline{X})$ is a vec with ith elem $\partial l/\partial\theta_i$. U, $U(\theta)$, $U(\theta;\underline{X})$, a random quantity

\Def 1.3/ Z observed information is $J(\theta;\underline{X}) = -\frac{\partial^2 l}{\partial \theta^2}$
if θ is a vec then $J(\theta,\underline{X})$ is a matrix with (i,j)th elem $-\partial^2 l/\partial\theta_i\partial\theta_j$; $J(\theta)$, $J(\theta;\underline{X})$ a random quantity
Z expected information is $I(\theta) = E[J(\theta)]$
if θ is a vec then $I(\theta)$ is Mat (i,j)th elem $-E(\partial^2 l/\partial\theta_i\partial\theta_j)$
since $I(\theta)$ is an expectation, its const rather than a random quantity

\Ex 1.6/ score func for Exp(θ) in (Ex 1.4): $U(\theta) = \frac{\partial l}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} X_i$
observed info: $J(\theta) = -\frac{\partial^2 l}{\partial\theta^2} = \frac{n}{\theta^2}$
& since it doesnt involve $\underline{X}$ have:
$I(\theta) = E[J(\theta)] = n/\theta^2$

\Ex 1.7/ $X_1, \dots, X_n$ indep r.v. each $N(\mu,\sigma^2)$ distri & density
$f(x;\mu,\sigma) = (2\pi)^{-1/2}\sigma^{-1}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$ for $-\infty < x < \infty$
Z likelihood of Z vec param (μ,σ) is
$L(\mu,\sigma;x) = \prod_{i=1}^{n} f(x_i;\mu,\sigma) = (2\pi)^{-n/2}\sigma^{-n}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$ ∴
$l(\mu,\sigma;x) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$ ∴
score func: $U = (\partial l/\partial\mu, \partial l/\partial\sigma)$ where
$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)$ $\quad \frac{\partial l}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2$
∴ $U=0$ ∴ mle: $\hat{\mu} = \bar{X}$ ∴ $\hat{\sigma} = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]^{1/2}$
∴ observed info:

$$J(\mu,\sigma) = \begin{bmatrix} -\partial^2 l/\partial\mu^2 & -\partial^2 l/\partial\mu\partial\sigma \\ -\partial^2 l/\partial\mu\partial\sigma & -\partial^2 l/\partial\sigma^2 \end{bmatrix} \quad \therefore$$

$\partial^2 l/\partial\mu^2 = -\frac{n}{\sigma^2}$ $\quad \frac{\partial^2 l}{\partial\mu\partial\sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)$ &

$\frac{\partial^2 l}{\partial\sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$

Z expected info: $I(\mu,\sigma) = \begin{bmatrix} -E\left(\frac{\partial^2 l}{\partial\mu^2}\right) & -E\left(\frac{\partial^2 l}{\partial\mu\partial\sigma}\right) \\ -E\left(\frac{\partial^2 l}{\partial\mu\partial\sigma}\right) & -E\left(\frac{\partial^2 l}{\partial\sigma^2}\right) \end{bmatrix} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$

since $E(X_i)=\mu$ & $E[(X_i-\mu)^2] = var(X_i) = \sigma^2$

\1.2 properties of pt estis/

\Ex 1.8/ $X_1,\dots,X_n$ indep $Uni(0,\theta)$ r.v estimator $\hat\theta = \max\{X_1,\dots,X_n\}$ for θ $\therefore$ sampling distri of $\hat\theta$:

$Pr(\hat\theta \le x) = Pr(X_1 \le x,\dots,X_n \le x) = \prod_{i=1}^{n} Pr(X_i \le x) = \prod_{i=1}^{n}\frac{x}{\theta} = \frac{x^n}{\theta^n}$ for $0 \le x \le \theta$

\Ex 1.9/ $X_1,\dots,X_n$ indep $N(\mu,\sigma^2)$ estimator $\hat\mu = \bar X$ for μ $\therefore$

$\hat\mu$ is normal with expectation $E(\bar X) = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n}\sum_{i=1}^{n}E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$

& variance $var(\bar X) = var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$

\Def 1.4/ Z <u>bias</u> of an estimator $\hat\theta$ is $E(\hat\theta)-\theta$ if $E(\hat\theta)=\theta$ then $\hat\theta$ is unbiased for θ. Z <u>standard error</u> of an estimator is Z square root of its variance. Z <u>mean squared error</u> (mse) of an esti $\hat\theta$ is $E[(\hat\theta-\theta)^2]$ mse in terms of bias & var

$mse(\hat\theta) = var(\hat\theta) + bias(\hat\theta)^2$

\Ex 1.10/ let $X_1,\dots,X_n$ indep $Uni(0,\theta)$ r.v estimator $\hat\theta = \max\{X_1,\dots,X_n\}$

$Pr(\hat\theta \le x) = (x/\theta)^n$ for $0 \le x \le \theta$ & $\therefore$ Z density of $\hat\theta$:

$\frac{d}{dx}Pr(\hat\theta \le x) = \frac{nx^{n-1}}{\theta^n}$ for $0 \le x \le \theta$ $\therefore$ Z expectation of $\hat\theta$:

$$E(\hat\theta) = \int_0^\theta x\frac{nx^{n-1}}{\theta^n}dx = \int_0^\theta \frac{nx^n}{\theta^n}dx = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n\theta}{n+1}$$

& Z bias of $\hat\theta$ is $E(\hat\theta)-\theta = -\theta/(n+1)$ $\quad$ similar shows:

$E(\hat{\theta}^2) = n\theta^2/(n+2)$ & $\therefore$ $var(\hat{\theta}) = E(\hat{\theta}^2) - E(\hat{\theta})^2 = n\theta^2/[(n+1)^2(n+2)]$

\Ex 1.11/ $N(\mu, \sigma^2)$ estimation $\hat{\mu} = \bar{X}$ for μ $\therefore$
$E(\bar{X}) = \mu$ $\therefore$ $\hat{\mu}$ is unbiased

\Def 1.5/ an estimator $\hat{\theta}$ is <u>consistent</u> for θ is $\forall \varepsilon > 0$,
$\lim_{n\to\infty} Pr(|\hat{\theta} - \theta| > \varepsilon) = 0$

consistency usually required to be accepted
sufficient (but not necessary) for $\hat{\theta}$ consistent is:
$mse(\hat{\theta}) \to 0$ as $n \to \infty$

\Ex 1.12/ $Uni(0, \theta)$ estimator $\hat{\theta} = \max\{X_1, ..., X_n\}$ for θ
in (Ex 1.10) $bias(\hat{\theta}) = -\theta/(n+1)$ & $var(\hat{\theta}) = n\theta^2/[(n+1)^2(n+2)]$ $\therefore$
both these converge to 0 as $n \to \infty$ & $\therefore$ $\hat{\theta}$ is consistent

\Ex 1.13/ expectation μ var σ^2 $\therefore$ $\bar{X}$ is consistent for μ since
$\bar{X}$ is unbiased for μ & $var(\bar{X}) = \sigma^2/n \to 0$ as $n \to \infty$
implies $\hat{\mu}$ is consistent for μ in (Ex 1.9)

report <u>standard error</u> of an estimator
report Z estimated standard error by replacing Z param
with its pt esti

\Ex 1.14/ in (Ex 1.12) Z standard error of Z estimator $\hat{\theta} = \max\{X_1, ..., X_n\}$ is $\sqrt{n}\,\theta/\sqrt{(n+1)^2(n+2)}$ & $\therefore$ Z estimated standard
error is $\sqrt{n}\,\hat{\theta}/\sqrt{(n+1)^2(n+2)}$

eg: for data θ: $\hat{\theta} = 0.94$ with estimated standard error 0.08

if have estimator $\hat{\theta}$ for param θ but want an esti func,
$\phi = h(\theta)$ of θ a possible estimator for ϕ is $\hat{\phi} = h(\hat{\theta})$

\Ex 1.15/ $Poi(\theta)$ estimate $\phi = Pr(X_1 = 0) = \exp(-\theta)$
estimator $\hat{\theta} = \bar{X}$ is unbiased for θ with var: $var(\hat{\theta}) = \theta/n$
a possible estimator for ϕ is $\hat{\phi} = \exp(-\hat{\theta})$

Sampling properties of ϕ approx: approx the var of any func of an estimator: delta Method

let $\hat{\theta}$ be consistent estimator for θ & let $\phi = h(\theta)$

$\hat{\theta} - \theta$ is small   taylor expansion yields $h(\hat{\theta}) \approx h(\theta) + (\hat{\theta} - \theta) h'(\theta)$

Z only random quantity in this approx is $\hat{\theta}$ & ∴

$\text{var}[h(\hat{\theta})] \approx \text{var}[h(\theta) + (\hat{\theta} - \theta) h'(\theta)] = \text{var}[\hat{\theta} h'(\theta)] = [h'(\theta)]^2 \text{var}(\hat{\theta})$

\Ex1.16/ $h(\theta) = \exp(-\theta)$ delta method: approx of var of $\hat{\phi}$:

$\text{var}(\hat{\phi}) \approx (-e^{-\theta})^2 \text{var}(\hat{\theta}) = \frac{\theta \exp(-2\theta)}{n}$

usually $\hat{\phi} = h(\hat{\theta})$ $\hat{\theta}$ consistent & unbiased for θ

usually $\hat{\phi} = h(\hat{\theta})$ biased for $\phi = h(\theta)$ reduce bias.

$\hat{\theta} - \theta$ is small & a taylor expansion yields:

$h(\hat{\theta}) \approx h(\theta) + (\hat{\theta} - \theta) h'(\theta) + \frac{1}{2}(\hat{\theta} - \theta)^2 h''(\theta)$

expectation: $E(\hat{\phi}) \approx h(\theta) + [E(\hat{\theta}) - \theta] h'(\theta) + \frac{1}{2} E[(\hat{\theta} - \theta)^2] h''(\theta) =$

$\phi + \frac{1}{2} \text{var}(\hat{\theta}) h''(\theta)$

if have estimator for Z var of $\hat{\theta}$: esti bias of $\hat{\phi}$:

$E(\hat{\phi}) - \phi \approx \frac{1}{2} \widehat{\text{var}}(\hat{\theta}) h''(\hat{\theta})$

bias corrected estimator: $\tilde{\phi} = \hat{\phi} - \frac{1}{2} \widehat{\text{var}}(\hat{\theta}) h''(\hat{\theta})$

\Ex1.17/ estimator $\hat{\phi} = \exp(-\hat{\theta})$ is biased for ϕ

reduce bias: $h(\theta) = \exp(-\theta)$ ∴ $h''(\theta) = \exp(-\theta)$ & esti $\text{var}(\theta)$ with

$\widehat{\text{var}}(\hat{\theta}) = \hat{\theta}/n$ obtain Z bias-corrected estimator

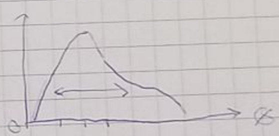
$\tilde{\phi} = \hat{\phi} - \frac{\hat{\theta}}{2n} \exp(-\hat{\theta}) = \left(1 - \frac{\bar{X}}{2n}\right) \hat{\phi}$

Monte Carlo simulation to ~~esti~~ approx an estimators properties by generating sample's & calc estimate from each & they give frequency distri of an approx to Z sampling distri of Z estimator ∴ can approx properties of Z estimator

bias is mean of estis subtract true param value

standard error approx from standard deviation

```
>> theta = 4
>> p1 = p2 = numeric(1000)          ## vectors to store ests #
>> p1bias = p2bias = numeric (20)    # vecs to store biases #
p1serr = p2serr = numeric(20)        # vecs to store standard errors #
>> for (n in 1:20) {          # for each n=1,...,20 #
>> for (i in 1:1000) {        # repeat Z following 1000 times
>> x = rpois(n, theta)        << simulate a sample from poi(theta) >>
>> p1[i] = exp(-mean(x))      # calc & store \hat{phi} #
>> p2[i] = (1-mean(x)/(2*n))*p1[i]   # calc & store \tilde{phi} #
>> }
>> p1bias[n] = mean(p1) - exp(-theta)    # approx Z biases #
>> p2bias[n] = mean(p2) - exp(-theta)    ##
>> p1serr[n] = sd(p1)        # approx Z standard errors #
p2serr[n] = sd(p2)           ##
>> }
>> par(mfrow=c(1,2))
>> plot(1:20, p1bias, type="l", xlab="n", ylab="Bias")
>> lines (1:20, p2bias, lty=2)
>> plot (1:20, p1serr, type="l", xlab="n", ylab="standard error")
>> lines (1:20, p2serr, lty=2)
```



Model eg $X_1, \dots, X_n$ iid with cdf $F(x;\theta)$

inference: estimate θ: Method of moments, maximum likelihood

properties of estimators: sampling distrib

- bias, standard error, mse, consistency

Delta Method: $\phi = h(\theta)$, $\hat{\theta}$

$\hat{\phi} = h(\hat{\theta})$ $var(\hat{\phi}) \approx [h'(\theta)]^2 var(\hat{\theta})$

bias$(\hat{\phi}) \approx \frac{1}{2} h''(\theta) var(\theta)$

Monte Carlo Simulation:

\Ex1.17/ $X_1, \ldots, X_n$ iid poi(θ) $\phi = e^{-\theta}$

$\hat{\theta} = \bar{x}$ is unbiased for θ with $var(\hat{\theta}) = \frac{\theta}{n}$

$\hat{\phi} = e^{-\hat{\theta}}$

we have $h(\theta) = e^{-\theta}$ so $h'(\theta) = -e^{-\theta}$

& $h''(\theta) = e^{-\theta}$

approx to bias is $\frac{1}{2} h''(\theta) var(\hat{\theta}) = e^{-\theta} \frac{\theta}{2n}$

remove bias from $\hat{\phi}$ to get

$\tilde{\phi} = \hat{\phi} - \frac{\hat{\theta} e^{-\hat{\theta}}}{2n} \overset{?}{=} (1 - \frac{\bar{x}}{2n}) \hat{\phi}$

$= \hat{\phi} \left[1 - \frac{\hat{\theta} e^{\hat{\theta}}}{2n\hat{\phi}} \right] = \hat{\phi} \left[1 - \cancel{\hat{\theta} e^{-\hat{\theta}}} \frac{\hat{\theta} \hat{\phi}}{2n \hat{\phi}} \right] = \hat{\phi} \left(1 - \frac{\hat{\theta}}{2n}\right)$

\1.3/ \efficiency & Z Cramér-Rao lower bound/

Z variance of two estimators $\hat{\theta}$ & $\tilde{\theta}$ are usually compared by their relative efficiency or asymptotic relative frequency

\Def 1.6/ let $\hat{\theta}$ & $\tilde{\theta}$ be two estimators based on samples of size n. The relative efficiency of $\hat{\theta}$ to $\tilde{\theta}$ is $var(\tilde{\theta}) / var(\hat{\theta})$ & Z asymptotic relative efficiency is Z limit of this ratio as $n \to \infty$

\Ex1.18/ $N(\mu, \sigma^2)$ two possible estimators for μ are Z sample mean & median   Z var of Z mean is σ^2/n & its poss to show Z var of Z median is $\pi \sigma^2 / (2n)$. The relative efficiency of Z median to Z mean is $2/\pi \approx 64\%$.

\ex1.19/ poi(θ)   Z expectation & var of this distri are both equal to θ. Z sample mean $\bar{X}$ & Z sample var $S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1)$ are both unbiased for θ now $var(\bar{X}) = \theta/n$ & its possible to show $var(S^2) = \frac{2n\theta^2 + (n-1)\theta}{n(n-1)}$

Z relative efficiency of $\bar{X}$ to S^2 is : $\frac{var(S^2)}{var(\bar{X})} = \frac{2n\theta + n - 1}{n-1} = 1 + \frac{2n\theta}{n-1}$

this depends on θ & n but always exceeds 1 so we would prefer $\bar{X}$ to S^2 regardless of n or Z true val of θ.

require estimators be unbiased then seek lowest variance, if exists its called Z <u>minimum</u> <u>variance</u> <u>unbiased</u> <u>estimator</u>

let $\hat{\theta} = \hat{\theta}(\underline{X})$ be unbiased estimator for θ where $\underline{X}$ has joint density $f_n(\underline{x};\theta)$ seek a lower bound for Z variance of $\hat{\theta}$, let U be arbit r.v

Z correlation between Z r.v $\hat{\theta}$ & U is defined: $\mathrm{Corr}(\hat{\theta}, U) = \dfrac{\mathrm{cov}(\hat{\theta}, U)}{\sqrt{\mathrm{var}(\hat{\theta})\,\mathrm{var}(U)}}$

where $\mathrm{cov}(\hat{\theta}, U) = E(\hat{\theta}U) - E(\hat{\theta})E(U)$ is Z covariance

know: $|\mathrm{Corr}(\hat{\theta}, U)| \leq 1$ & ∴ $\mathrm{var}(\hat{\theta})\mathrm{var}(U) \geq \mathrm{cov}(\hat{\theta}, U)^2$ ∴

can find U st $\mathrm{cov}(\hat{\theta}, U) = 1$ $\forall$ unbiased $\hat{\theta}$ ∴ have Z lower bound:
$\mathrm{var}(\hat{\theta}) \geq 1/\mathrm{var}(U)$

can find such a U? have $E(\hat{\theta}) = \theta$ since $\hat{\theta}$ is unbiased &
$\hat{\theta}$ is a func of ~~x~~ $\underline{X}$ & ∴ Z expectation of $\hat{\theta}$ can be written as

$E(\hat{\theta}) = \int \hat{\theta}(\underline{x}) f_n(\underline{x};\theta)\,d\underline{x}$ ∴

$1 = \frac{\partial}{\partial\theta}\int \hat{\theta}(\underline{x}) f_n(\underline{x};\theta)\,d\underline{x} = \int \hat{\theta}(\underline{x}) \frac{\partial f_n(\underline{x};\theta)}{\partial\theta}\,d\underline{x} = \int \hat{\theta}(\underline{x}) \frac{\partial l(\theta;\underline{x})}{\partial\theta} f_n(\underline{x};\theta)\,d\underline{x}$ where:

as $\frac{\partial}{\partial\theta} l(\theta;\underline{x}) = \frac{\partial}{\partial\theta}\log L(\theta;\underline{x}) = \frac{1}{L(\theta;\underline{x})}\frac{\partial L(\theta;\underline{x})}{\partial\theta} = \frac{1}{f_n(\underline{x};\theta)}\frac{\partial f_n(\underline{x};\theta)}{\partial\theta}$

the interchanging of order of Z differentiation & Z integration holds under Z 'regularity conditions' that Z <u>support</u> of f_n (ie set of $\underline{x}$ vals which $f_n(\underline{x};\theta) \neq 0$) is indep of θ & Z integral exists

if take U to be Z score func, $U(\theta;\underline{x}) = \partial l(\theta;\underline{x})/\partial\theta$ ∴ have shown

$1 = \int \hat{\theta}(\underline{x}) U(\theta;\underline{x}) f_n(\underline{x};\theta)\,d\underline{x} = E(\hat{\theta}U)$ &:

$E(U) = \int \frac{\partial l(\theta;\underline{x})}{\partial\theta} f_n(\underline{x};\theta)\,d\underline{x} = \int \frac{\partial f_n(\underline{x};\theta)}{\partial\theta}\,d\underline{x} = \frac{\partial}{\partial\theta}\int f_n(\underline{x};\theta)\,d\underline{x} = \frac{\partial}{\partial\theta}1 = 0$

∴ if $\hat{\theta}$ is unbiased & U is Z score then $\mathrm{cov}(\hat{\theta}, U) = E(\hat{\theta}U) - E(\hat{\theta})E(U) =$
$1 - \theta \times 0 = 1$ & ∴ $\mathrm{var}(\hat{\theta}) \geq 1/\mathrm{var}(U)$ as required ∴

$E(U) = 0$ & ∴ $\mathrm{var}(U) = E(U^2)$ but $E(U^2) = -E\left(\frac{\partial^2 l}{\partial\theta^2}\right)$ &

Z expected info, $I(\theta)$ which is of f_n for $f_n(\underline{x};\theta)$ ∴

$$\frac{\partial^2 l}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left(\frac{1}{f_n}\frac{\partial f_n}{\partial\theta}\right) = \frac{1}{f_n}\frac{\partial^2 f_n}{\partial\theta^2} - \left(\frac{1}{f_n}\frac{\partial f_n}{\partial\theta}\right)^2 = \frac{1}{f_n}\frac{\partial^2 f_n}{\partial\theta^2} - U^2 \quad ∴$$

$E\left(\frac{1}{f_n}\frac{\partial^2 f_n}{\partial\theta^2}\right) = \int \frac{\partial^2 f_n(\underline{x};\theta)}{\partial\theta^2}\,d\underline{x} = \frac{\partial^2}{\partial\theta^2}\int f_n(\underline{x};\theta)\,d\underline{x} = \frac{\partial^2}{\partial\theta^2}1 = 0$ & ∴

$I(\theta) = -E\left(\frac{\partial^2 l}{\partial \theta^2}\right) = -E\left(\frac{1}{f_n}\frac{\partial^2 f_n}{\partial \theta^2}\right) + E(U^2) = E(U^2)$ QED

\Thm 1.1/ let $\hat{\theta}$ be an unbiased estimator for θ. then under appropriate regularity conditions of f_n, $\mathrm{var}(\hat{\theta}) \geq I(\theta)^{-1}$

this is called Z Cramer-Rao lower bound

\Ex1.2/ $\mathrm{Poi}(\theta)$ mass func $f(x;\theta) = \theta^x \exp(-\theta)/x!$ for $x = 0, 1, \ldots$ find Z Cramér-Rao lower bound for Z var of unbiased estimators of θ

Z likelihood is $L(\theta; \underline{x}) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \theta^{\sum_{i=1}^{n} x_i} e^{-n\theta} \prod_{i=1}^{n} \frac{1}{x_i!}$

Z log likelihood is $l(\theta; \underline{x}) = (\log\theta)\sum_{i=1}^{n} x_i - n\theta - \sum_{i=1}^{n}\log(x_i!)$

Z score is $l'(\theta; \underline{x}) = \frac{1}{\theta}\sum_{i=1}^{n} x_i - n$

& Z observed info is $J(\theta; \underline{x}) = -l''(\theta; \underline{x}) = \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$ ∴ Z expected info is: ~~$J(\theta;\underline{x}) = -l''(\theta;\underline{x}) = \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$~~

$I(\theta) = E[J(\theta; \underline{X})] = \frac{1}{\theta^2}\sum_{i=1}^{n} E(X_i) = \frac{1}{\theta^2}\sum_{i=1}^{n}\theta = \frac{n}{\theta}$ & Z lower bound is $I(\theta)^{-1} = \theta/n$ this equals Z sample mean in Ex1.19 & ∴ $\bar{X}$ is Z minimum variance unbiased estimator for θ in Z poisson distri

\Ex1.21/ $N(\mu, \sigma^2)$ σ is known ∴

$L(\mu; \underline{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$

$l(\mu; \underline{x}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$

& $\frac{\partial l}{\partial \mu} = \frac{n}{\sigma^2}(\bar{x} - \mu)$ & $\frac{\partial^2 l}{\partial \mu^2} = -\frac{n}{\sigma^2}$ ∴ $I(\mu) = -E(\partial^2 l/\partial\mu^2) = n/\sigma^2$ & Z lower bound is $I(\mu)^{-1} = \sigma^2/n$ this equals Z var of Z sample mean in (Ex1.18/ & ∴ $\bar{X}$ is Z minimum var unbi estimator for Z expect of a normal distri

\Def 1.7/ Z <u>efficiency</u> of an unbi estimator $\hat{\theta}$ is $I(\theta)^{-1}/\mathrm{var}(\hat{\theta})$

& if $\mathrm{var}(\hat{\theta}) = I(\theta)^{-1}$ then $\hat{\theta}$ is <u>efficient</u> - if Z efficiency of $\hat{\theta}$ converges to 1 as $n \to \infty$ then $\hat{\theta}$ is <u>asymptotically</u> <u>efficient</u>

if Z earlier regularity conditions hold & we find an unbiased estimator that is efficient then we can stop our search for a minimum var unbiased estimator. Can find minimum var unbi estimators by inspecting Z score func

\Thm 1.2/ under Z previous regularity conditions, $\hat{\theta}$ is unbiased for θ & efficient iff Z score can be written as $U=b(\hat{\theta}-\theta)$ for some const b. if Z score can be written in this way then $b=I(\theta)$

\Proof/ let U be Z score func. Suppose that $\hat{\theta}$ is unbi & efficient so $var(\hat{\theta})=I(\theta)^{-1}=1/var(U)$ & $cov(\hat{\theta},U)=1$ so $var(\hat{\theta})=1/var(U)$ iff ~~corr~~ $|corr(\hat{\theta},U)|=1$ which holds iff $\hat{\theta}$ & U are linearly related that is $U=a+b\hat{\theta}$ for some consts a & b

Since $E(U)=0$ must have $a=-b\theta$ & so $U=b(\hat{\theta}-\theta)$ ∴

$U^2=b(\hat{\theta}-\theta)U$ & taking expectations yields $I(\theta)=E(U^2)=$ $bE(\hat{\theta}U)-b\theta E(U)=b$ since $E(\hat{\theta}U)=1$ & $E(U)=0$. Z converse is also true as we now show if $U=b(\hat{\theta}-\theta)$ then $\hat{\theta}$ is unbiased since $E(U)=0$

Z same argument as above then yields $U=I(\theta)(\hat{\theta}-\theta)$ & ∴

$var(U)=I(\theta)^2 var(\hat{\theta})$ ∴ sub $var(U)=I(\theta)$ ∴ $var(\hat{\theta})=I(\theta)^{-1}$ □

\1.22/ consider (1.20) & show that $\bar{X}$ is Z min var unbi estimator for μ. Found $U=\frac{n}{\sigma^2}(\bar{X}-\mu)$ & ∴ $\bar{X}$ is unbi & efficient with var $I(\mu)^{-1}=\sigma^2/n$

\Ex 1.24/ $Uni(0,\theta)$ then $f_n(x;\theta)=\theta^{-n}$ for $\theta>\max\{x_1,\dots,x_n\}$ st $U=-n/\theta$ & $I(\theta)^{-1}=-\theta^2/n$ this is negative & ∴ is not useful lower bound for Z var ~~of~~ any estimator for θ. Z regularity conditions are not satisfied in this example since Z supp of f_n depends on θ

sometimes Z max likelihood estor is Z min var unbi estor

\Thm 1.3/ if $\tilde{\theta}$ is unbi for θ & efficient, & if Z max likelihood estor $\hat{\theta}$ satisfies $U(\hat{\theta})=0$ then $\tilde{\theta}=\hat{\theta}$

\proof/ (thm 1.2) yields $U(\theta) = I(\theta)(\tilde{\theta}-\theta)$ $\therefore$ $I(\hat{\theta})(\tilde{\theta}-\hat{\theta}) = U(\hat{\theta}) = 0$ & Z result follows since $I(\hat{\theta}) \neq 0$

\Ex1.25/ for (Ex1.22) showed $\bar{x}$ was unbi & efficient & $U(\bar{x}) = 0$ st $\bar{x}$ is Z max likelihood estor

\Ex1.26/ for (Ex1.23) showed $\bar{x}$ was unbi & efficient & $U(\bar{x}) = 0$ st $\bar{x}$ is Z maximum likelihood estor

\1.4 Hypothesis tests/ hypothesis tests asses Z evidence that our data provides for or against param vals. compare a null hypoth $H_0 : \theta \in \Theta_0$ with an alternative hypoth $H_1 : \theta \in \Theta_1$ where $\Theta_0 \cap \Theta_1 = \emptyset$. a hypothesis is simple if it specifies a single val for Z param eg $\Theta_0 = \{\theta_0\}$; otherwise Z hypothesis is composite, eg $\Theta_1 = \{\theta : \theta > \theta_0\}$.

in frequentist inference can't talk about $pr(\theta \in \Theta_0)$. Z prop that H_0 is true. prob is only defined in terms of frequencies under repeated sampling. but params are const which do not vary under repeated sampling so θ is either in Θ_0 or not, independent of any samples we draw & $\therefore$ probs such as $pr(\theta \in \Theta_0)$ are trivial (either one or zero)

use stats test that reject H_0 for H_1 if data fall in some critical region C. two errors possible a associated frequencies under repeated sampling: might reject null when its true (type 1 error) or fail to reject Z null hypoth when its false (type 2 error)

hypoth tests are judged by Z relative frequencies in which these errors are made under repeated sampling. we want frequencies of both error are made under repeated sampling. want frequencies of both errors to be as small as possible now, Z freqency of type 1 errors can be reduced by making C smaller. but then Z frequency of type 2 errors increases. so our choice of critical

$\sqrt{n}(\bar{x}-\mu)$

region involves a compromise. Z classical approach is to seek Z critical region that minimises Z frequency of type 2 errors subject to a limit on Z frequency of type 1 errors. this limit is known as Z size of Z test

write $Pr(\underline{X} \in C; \theta) = \int_C f_n(\underline{x}; \theta)\, d\underline{x}$ for Z prob that a sample will fall in Z critical region when it comes from Z distri with param val θ

Def 1.8/ Z size of test is Z supremum (least upper bound) of Z prob of a type 1 error, that is $\sup_{\theta \in \Theta_0} Pr(\underline{X} \in C; \theta)$

Z power of a test is Z func $w(\theta) = Pr(\underline{X} \in C; \theta)$

for a test of size α ∴ seek a critical region C for which $\sup_{\theta \in \Theta_0} Pr(\underline{X} \in C; \theta) = \alpha$ & for which Z chance of a type 2 error is small, in other words for which $w(\theta)$ is large when $\theta \in \Theta_1$

Ex 1.27 let X have a distri func $Pr(X \le x) = 1 - e^{-x/\theta}$ for $x > 0$

consider Z null hypothesis $H_0: \theta > \theta_0$ & Z alternative hypoth $H_1: \theta \le \theta_0$ for some θ_0 construct critical region test size α & calc power of test distri, $E(X) = \theta$ & ∴ reject H_0 in favour of H_1 when x is small ∴

choosing critical region of form $C = (0, c)$ for critical val $c > 0$

∴ $Pr(X \in C; \theta) = Pr(X \le c; \theta) = 1 - e^{-c/\theta}$ & ∴

for a test size α, require c to satisfy $\alpha = \sup_{\theta > \theta_0} (1 - e^{-c/\theta}) = 1 - e^{-c/\theta_0}$ ∴ $c = -\theta_0 \log(1-\alpha)$ & ∴ a critical region for a test of size α is Z interval $(0, -\theta_0 \log(1-\alpha))$ ∴ sub val of c into for $Pr(X \in C; \theta)$ ∴ power func:

$$w_\alpha(\theta) = 1 - \exp\left[-\frac{-\theta_0 \log(1-\alpha)}{\theta}\right] = 1 - (1-\alpha)^{\theta_0/\theta}$$

can sketch power func for case $\theta_0 = 1$ & $\alpha = 0.1$ shows test has a good chance of rejecting H_0 when data really come from a distri with param val close to zero

& also power func of Z test in (Ex 1.27) when $\theta_0 = 1$, $\alpha = 0.1$

if H_1 is simple st $\Theta_1 = \{\theta_1\}$ then should seek Z test of size α that maximises $w_\alpha(\theta_1)$ called Z Most powerful test of size α. if H_1 is composite then ideally we like a test of size α with a power func $w_\alpha(\theta)$ st if $w'_\alpha(\theta)$ is Z power func of any other test of size α then $w_\alpha(\theta) \geq w'_\alpha(\theta) \ \forall \theta \in \Theta_1$ st a test is called Z uniformly most powerful test of size α

consider uniformly most powerful tests later. first consider Z case where both null & alternative hypoth are simple st we have $H_0 : \theta = \theta_0$ & $H_1 : \theta = \theta_1$ an appropriate critical region for testing these hypoth would be those $\underline{x}$ for which Z likelihood ratio $\frac{f_n(x;\theta_1)}{f_n(\underline{x};\theta_0)} = \frac{L(\theta_1;\underline{x})}{L(\theta_0;\underline{x})}$ is large. in fact this turns out to give Z most powerful test

~~[illegible]~~

$var(\hat{\theta}) \geq I(\theta)^{-1}$ is Cramer-Rao lower bound

efficiency $I(\theta)^{-1}/var(\hat{\theta})$

~~unbiased score~~ unbiased if score $U = b(\hat{\theta} - \theta)$

when you have a simple null & alternative you can construct Z most powerful test (likelihood test)

\Ex 1.21/ $N(\mu, \sigma^2)$ σ known

pdf: $f(x;\mu) = (2\pi)^{-1/2}\sigma^{-1}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$ ∴

$$L(\mu) = \prod_{i=1}^{n} f(x_i;\mu) = (2\pi)^{-n/2}\sigma^{-n}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right] \quad ∴$$

$$l(\mu) = \log L(\mu) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \quad \&$$

$$l'(\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}(n\bar{x} - n\mu) = \frac{n}{\sigma^2}(\bar{x}-\mu) \qquad l''(\mu) = -\frac{n}{\sigma^2}$$

$$∴ \ I(\mu) = E[-l''(\mu)] = E\left(\frac{n}{\sigma^2}\right) = \frac{n}{\sigma^2} \quad ∴$$

the CRLB is $I(\mu)^{-1} = \frac{\sigma^2}{n}$

Ex 1.23 / Some score func is $U(\mu) = \underbrace{\frac{n}{\sigma^2}}_{b}(\underbrace{\bar{x}}_{\hat{\theta}} - \underbrace{\mu}_{\theta})$

{thm 1.2 / $U = b(\hat{\theta} - \theta)$} $\therefore$ by thm 1.2 $\hat{\mu} = \bar{x}$ is unbiased & efficient

Menti / $var(\hat{\theta}_1) = 2$ $var(\theta_2) = 8$ $\therefore$ efficiency is:

$\frac{var(\hat{\theta}_2)}{var(\hat{\theta}_1)} = \frac{8}{2} = 4$

Cramer-Rao lower ~~bounds~~ bounds the variance of unbiased estimators

CRLB of an unbiased estimator of θ is I(theta)

variance are non negative $U(\theta) = -n\log(\theta)$ whats CRLB?

theta/n

$U(\theta) = \frac{n\bar{x}}{\theta} - n\frac{(1-\bar{x})}{(1-\theta)} = \frac{n}{\theta}\bar{x} - \frac{n - n\bar{x}}{1-\theta} = \frac{n}{\theta}\bar{x} - \frac{n}{1-\theta} + \frac{n\bar{x}}{1-\theta}$

$\frac{\partial \bar{x}}{\partial \bar{x}} = \bar{x}\left(\frac{n}{\theta} + \frac{n}{1-\theta}\right) - \frac{n}{1-\theta}$ is there an unbiased ~~&~~ efficient estimator of θ: there is

$\therefore U = \frac{n\bar{x}}{\theta} - \frac{n(1-\bar{x})}{1-\theta} = \frac{n}{\theta(1-\theta)}\left[\bar{x}(1-\theta) - (1-\bar{x})\theta\right] = \frac{n}{\theta(1-\theta)}\left[\bar{x} - \theta\bar{x} - \theta + \theta\bar{x}\right] =$

$\frac{n}{\theta(1-\theta)}(\bar{x} - \theta)$ $\therefore$ estimator is unbiased & efficient

for hypoth test the prob that data fall in the critical region of when H_0 is true is the size is asking the prob of rejecting H_0 when its true is type 1 error & want small

prob data fall in the critical region when H1 is true is the power

Thm 1.4 / Neyman-pearson thm / Z most powerful test of size α for Z null hypoth $H_0: \theta = \theta_0$ against Z alternative hypoth $H_1: \theta = \theta_1$ has a critical region of Z form $C = \{ \underline{x} : \Lambda(\underline{x}) \geq c \}$ where $\Lambda(\underline{x}) = \frac{L(\theta_1; \underline{x})}{L(\theta_0; \underline{x})}$ is Z likelihood ratio

this test is a likelihood ratio test

proof / let C be Z critical region for Z likelihood ratio test of size α & let C' be Z critical region for Z ~~likelihood ratio~~ another test of size α $\therefore$ $Pr(\underline{X} \in C; \theta_0) = \alpha = Pr(\underline{X} \in C'; \theta_0)$ we show Z likelihood ratio test has Z greater power ie that $\Delta = Pr(\underline{X} \in C; \theta_1) - Pr(\underline{X} \in C'; \theta_1)$ write $\bar{C}$ for Z complement of C $\therefore$

$\Delta = Pr(\underline{X} \in C \cap C'; \theta_1) + Pr(X \in C \cap \bar{C}'; \theta_1) - Pr(X \in C' \cap C; \theta_1) - \cancel{Pr(X \in C' \cap \bar{C}} \; Pr(\underline{X} \in C' \cap \bar{C}; \theta_1) = Pr(X \in C \cap \bar{C}'; \theta_1) - Pr(\underline{X} \in C' \cap \bar{C}; \theta_1)$

by def of C, $f_n(\underline{x}; \theta_1) \geq c f_n(\underline{x}; \theta_0) \; \forall \underline{x} \in C \cap \bar{C}'$ & $f_n(\underline{x}; \theta_1) \leq c f_n(\underline{x}; \theta_0) \; \forall \underline{x} \in C' \cap \bar{C}$ $\therefore$

$Pr(\underline{X} \in C \cap \bar{C}'; \theta_1) = \int_{C \cap \bar{C}'} f_n(\underline{x}; \theta_1) d\underline{x} \geq c \int_{C \cap \bar{C}'} f_n(\underline{x}; \theta_0) d\underline{x} = c Pr(\underline{X} \in C \cap \bar{C}'; \theta_0)$ &

$Pr(\underline{X} \in C' \cap \bar{C}; \theta_1) = \int_{C' \cap \bar{C}} f_n(\underline{x}; \theta_1) d\underline{x} \leq c \int_{C' \cap \bar{C}} f_n(\underline{x}; \theta_0) d\underline{x} = c Pr(X \in C' \cap \bar{C}; \theta_0)$ st

$\Delta \geq c [Pr(\underline{X} \in C \cap \bar{C}'; \theta_0) - Pr(\underline{X} \in C' \cap \bar{C}; \theta_0)] = c [Pr(\underline{X} \in C \cap \bar{C}'; \theta_0) + Pr(\underline{X} \in C \cap C'; \theta_0) - Pr(\underline{X} \in C \cap C'; \theta_0) - Pr(\underline{X} \in C' \cap \bar{C}; \theta_0)] = c [Pr(\underline{X} \in C; \theta_0) - Pr(\underline{X} \in C'; \theta_0)] = c(\alpha - \alpha) = 0$

Z neyman-pearson thm gives us a way to find Z most powerful test of a simple null against a simple alternative: we use critical region $C = \{ \underline{x} : \Lambda(\underline{x}) \geq c \}$ & choose c st $Pr[\Lambda(\underline{X}) \geq c; \theta_0] = Pr(\underline{X} \in C; \theta_0) = \alpha$ is Z desired size of Z test this is then guaranteed to be Z most powerful test of size α

Z critical region $\{x : \Lambda(x) \geq c\}$ for Z likelihood ratio test can be often be written as $\{x : T(x) \geq d\}$ or $\{x : T(x) \leq d\}$ for some

- simple test stat $T(x)$ & some const d

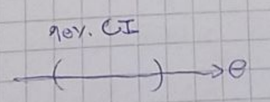
\Ex1.25/ i.i.d exponential density: $f(x;\theta) = \theta^{-1}\exp(-x/\theta)$ for $x > 0$ find Z most powerful test of Z null hypoth $H_0 : \theta = \theta_0$ against Z alternative $H_1 : \theta = \theta_1$ when $\theta_1 < \theta_0$. Z likelihood ratio is

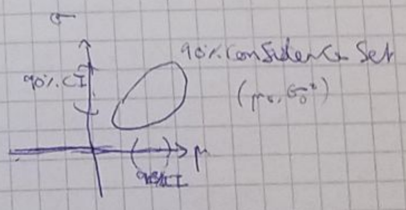
$$\Lambda(x) = \frac{L(\theta_1; x)}{L(\theta_0; x)} = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left[-\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum_{i=1}^n x_i\right]$$

show $\Lambda(x) \geq c$ iff $\sum_{i=1}^n x_i \leq \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)^{-1} \log\left[c\left(\frac{\theta_1}{\theta_0}\right)^n\right]$ Z RHS of this inequality is just a const but Z $\therefore \Lambda(x) \geq c$ is equiv to $\sum_{i=1}^n x_i \leq d$ for some const d $\because \theta_1 < \theta_0$ $\therefore$ from exponential term that $\Lambda(x)$ increases as $\sum_{i=1}^n x_i$ decreases & $\therefore \Lambda(x) \geq c$ is equiv to $\sum_{i=1}^n x_i \leq d$ for some const d) Z most powerful test of size α $\therefore$ rejects H_0 in favour of H_1 is $\sum_{i=1}^n x_i \leq d$ where d satisfies Z eqn $\Pr\left(\sum_{i=1}^n X_i \leq d; \theta_0\right) = \alpha$ since $\sum_{i=1}^n x_i$ has a $Ga(n, \theta)$ distri, d is Z α-quantile of Z $Ga(n, \theta_0)$ distri. In Z special case of $n=1$ have: $\alpha = \Pr(X_1 \leq d; \theta_0) = \int_0^d f(x;\theta_0)\,dx = \int_0^d \theta_0^{-1} e^{-x/\theta_0}\,dx$ $= 1 - \exp(-d/\theta_0)$ when $n=1$ $\therefore d = -\theta_0 \log(1-\alpha)$ & Z most powerful test of size α has critical region $(0, -\theta_0 \log(1-\alpha)]$

eg if $\theta_0 = 1$, $\alpha = 0.1$ & $x_1 = 0.73$ $\therefore d = -\log(1-0.1) = 0.105$ so $x_1 \not\leq d$ & would not reject Z null hypoth at Z 10% level. Z power of this test is $\Pr(X_1 \leq d; \theta_1) = 1 - \exp(-d/\theta_1) = 1 - \exp\left[\frac{\theta_0 \log(1-\alpha)}{\theta_1}\right] =$ $1 - (1-\alpha)^{\theta_0/\theta_1}$

$x_1, \dots, x_n \longrightarrow$





is $N(\mu, \sigma^2)$

$\cancel{X} \sim N(\mu, 1)$ μ is a param X depends on μ

μ is const so isn't a pivot $(X-\mu) \sim N(0,1)$ $\therefore$ is a pivot

$\therefore \exp(X-\mu)$ is also a pivot since a transformation ~~of~~ $N(0,1)$

also ~~monotone~~ doesn't depend on μ

$X \sim N(\mu, \sigma^2)$ $\therefore$ $\frac{X-\mu}{\sigma} \sim N(0,1)$ $\therefore \exp\left(\frac{X-\mu}{\sigma}\right)$

$var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} var(X-\mu) = \frac{1}{\sigma^2} var(X) = \frac{1}{\sigma^2}\sigma^2 = 1$

$var(X-\mu) = var(X) = \sigma^2$

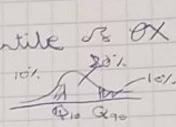
θX be a pivot with p% quantile Q_p & $X > 0$

which is 80% CI for θ?

~~[...]~~

let θX be a pivot let Q_{10} be Z 10% quantile of θX

let Q_{90} be Z 90% quantile of θX



$\therefore Pr(Q_{10} < \theta X < Q_{90}) = 0.8$

$\therefore Pr(Q_{10}/X < \theta < Q_{90}/X) = 0.8$ CI

~~Most~~

moste powerful test of $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$ rejects H_0 if $L(\theta_1; x)/L(\theta_0; x)$ is large since θ_1 is more likely than θ_0

reject H_0 when $\bar{X} < c$ if $P(\bar{X} < c; \theta_0) = 0.1$ & is $P(\bar{X} < c; \theta_1) = 0.8$

size is 0.1 since θ_0 is true and rejecting H_0 $\therefore \bar{X} < C$ $\therefore \bar{X}$ is in CR

H_1 true $\therefore \theta_1$ and reject H_0 $\therefore \bar{X} < C$ $\therefore \bar{X}$ is in $\bar{C}$

\Ex1.29/ $N(\mu, \sigma^2)$ σ known most powerful test of Z null hypoth $H_0: \mu = \mu_0$ against Z alternative hypoth $H_1: \mu = \mu_1$ when $\mu_1 > \mu_0$

$\therefore$ Z likelihood ratio is $\Lambda(\underline{x}) = \frac{L(\mu_1; \underline{x})}{L(\mu_0; \underline{x})} = ~~\exp~~$

$\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\mu_1)^2 - \sum_{i=1}^{n}(x_i-\mu_0)^2\right]\right\} = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[2x_i(\mu_0-\mu_1) + (\mu_1^2-\mu_0^2)\right]\right\} =$

$\exp\left[-\frac{n(\mu_1^2-\mu_0^2)}{2\sigma^2}\right]\exp\left[\frac{n\bar{x}}{\sigma^2}(\mu_1-\mu_0)\right]$ since $\mu_1 > \mu_0$, $\Lambda(\underline{x})$ increases as $\bar{x}$ increases, & $\therefore$ $\Lambda(\underline{x}) \geq c$ is equiv to $\bar{x} \geq d$ for some const d. Z most powerful test of size α $\therefore$ rejects H_0 in favour of H_1 if $\bar{x} \geq d$

d satisfies Z eqn $Pr(\bar{X} \geq d; \mu_0) = \alpha$ when μ_0 is Z param val

$\bar{X}$ has distri $N(\mu_0, \sigma^2/n)$ & ∴ d is Z $(1-\alpha)$- quantile of Z $N(\mu_0, \sigma^2/n)$ distri. this can be written as $d = \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}$ where $z_{1-\alpha}$ is Z $(1-\alpha)$ quantile of Z standard Normal distri. $N(0,1)$ ∴ Z ~~size~~ Z of test size α has critical region $\{\underline{x} : \bar{x} \geq \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}\}$
Z power of this test is $Pr(\bar{X} \geq d; \mu_1) = Pr\left(\frac{\bar{X}-\mu_1}{\sigma/\sqrt{n}} \geq \frac{d-\mu_1}{\sigma/\sqrt{n}}; \mu_1\right) =$
$Pr\left(Z \geq \frac{d-\mu_1}{\sigma/\sqrt{n}}\right)$ {where $Z \sim N(0,1)$} $= Pr\left(Z \geq z_{1-\alpha} - \frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}\right)$

Can approx properties of tests by Monte Carlo simulation. given a simple null hypoth $H_0 : \theta = \theta_0$ sample size n & test stat $T(\underline{x})$, simulate a large number of samples of size n from Z model distri with param val θ_0 & calc Z test stat for each sample. Z frequency distri of Z test stat. is Z critical region is $\{\underline{x} : T(\underline{x}) \leq d\}$ & want a test of size α then can approx d by Z α-quantile of Z simulated test stat similarly given a critical region, can approx Z power of Z test when $\theta = \theta_1$ by simulating a large number of samples from Z model distri with param val θ_1, calcing Z test stat for each sample & then calcing Z proportion of ~~test~~ test stats that fall in Z critical R. Z following code illustrates these ideas for (Ex1.28 when $n=10$, $\theta_0=1$ & $\alpha=0.1$):

```
>>n=10 #sample size#   >>theta0=1 {null hypoth}
>>alpha=0.1 {test size}   >>t=numeric(10000) {vec to store test stats}
>> for(i in 1:10000){ %repeat Z following 10000 times%
>>x=rexp(n, 1/theta0)  <simulate a sample from Exp(theta0)>
>>t[i]=sum(x)  <calc & store test stat>   >>}
>>quantile(t, alpha)  <Monte Carlo approx to d>
>> d=qgamma(alpha, n, 1/theta0)  <true val of d>
>>theta=seq(0.1,2,0.1)  <range of theta vals>
>>power=numeric(length(theta))  <vec of theta vals>
>> for (j in 1:length(theta)) {  <for each theta val>
                                  <repeat 10000 times>
>> for(i in 1:10000) {  <simulate a sample from Exp(theta)>
```

```
>> x= rexp(n,1/theta[j])      >> t[i]=sum(x) <calc & store test stat>   >> }
>> power[j]=mean(t<d) {Monte Carlo approx to power>    >> }
>> plot(theta, power, type="l") <plot power func>
```

\1.5 uniformly most powerful tests/ fewer likhood ratio test is most powerful when both Z null & alternative hypoths are simple. Now consider Z case of a simple null $H_0 : \theta = \theta_0$ & a composite alternat $H_1 : \theta \in \varnothing_1$ & seek uniformly most powerful tests

\Def 1.9/ a test of size α with CR C is <u>uniformly most powerful</u> if any other test of size α with CR C' has lower power, that is $Pr(\underline{X} \in C ; \theta) \geq Pr(\underline{X} \in C' ; \theta)\ \forall \theta \in \varnothing_1$

Z simple alternat $H_1' : \theta = \theta_1$ for some $\theta_1 \in \varnothing_1$ Z Neyman-pearson thm tells us Z most powerful test of H_0 against Z simple H_1' has a critical R of Z form $C = \{\underline{x} : \Lambda(\underline{x}) \geq c\}$ where $\Lambda(\underline{x})$ is Z likhood ratio st $\forall$ other test of size α with CR C' $Pr(\underline{X} \in C ; \theta_1) \geq Pr(\underline{X} \in C' ; \theta_1)$ if thus C remains completely unchanged $\theta_1 \in \varnothing_1$ chose for H_1' then Z inequality would remain true $\forall \theta_1 \in \varnothing_1$ eg is Z composite alternative $H_1 : \theta \in \varnothing_1$ ~~this is~~ Z reasoning:

\thm 1.5/ let H_0 be simple & let C be Z CR for Z most powerful test of size α for H_0 against $H_1' : \theta = \theta_1$ if C is Z same for each $\theta_1 \in \varnothing_1$, then C defines Z uniformly most powerful test of size ~~a for~~ {C is Z same for each $\theta_1 \in \varnothing_1$} ~~for~~ α for H_0 against Z composite hypoth $H_1 : \theta \in \varnothing_1$. otherwise there is no uniformly most powerful test for these hypoth

\Ex 1.30/ let X_1 be an Exponential with density $f(x;\theta) = \theta^{-1}\exp(-x/\theta)$ & consider Z null hypoth $H_0 : \theta = \theta_0$ in (Ex 1.28) found Z most powerful test for H_0 against $H_1 : \theta = \theta_1$ has a CR $[0, -\theta_0 \log(1-\alpha)]$ when $\theta_1 < \theta_0$ now whichever val of $\theta_1 < \theta_0$ were specified by H_1, this CR would remain unchanged & would always be most powerful $\therefore$ Z same CR defines a test that is uniformly most powerful for Z composite alternat hypoth $H_1 : \theta < \theta_0$. A similar argument shows that Z uniformly most powerful test for H_0 against $H_1 : \theta > \theta_0$ has CR $[-\theta_0 \log \alpha, \infty)$

since these two CR disser but there is no unisormly most powersul test sor Z two-sided alternat hypoth $H_1: \theta \neq \theta_0$

\Ex1.31/ $N(\mu, \sigma^2)$ σ is known & consider Z null hypoth $H_0: \mu = \mu_0$ (cf Ex1.29) sound that Z most powersul test sor H_0 against $H_1: \mu = \mu_1$ has CR $\{\underline{x} : \bar{x} \geq \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}\}$ when $\mu_1 > \mu_0$ now whichever val of $\mu_1 > \mu_0$ were specisied by H_1 this CR would remain unchanged & would always be most powersul ∴ Z same CR desines a test that is unisormly most powersul sor Z composite alternat hypoth $H_1: \mu > \mu_0$. A similar argument shows that Z unisormly most powersul test sor H_0 against $H_1: \mu < \mu_0$ has CR $\{\underline{x} : \bar{x} \leq \mu_0 + z_\alpha \sigma/\sqrt{n}\}$ since these two CR disser but there is no UMP test Z two-sided AH $H_1: \mu \neq \mu_0$

~~Harm~~ \1.6 Considence sets/ standard errors & other summary measures of an estimator's sampling distri provide valuable info about Z reliability of Z estimator an alternat approach is to report not a pt val as our esti but to report a set of vals in point esti we have a pt estimator that is a func of our data & specisies a pt in param space intended to be close to Z true param val. set esti is similar: a func of our data specisies a region in param space intended to contain Z true param val note that there <a func of our data specisies a region in param space> is no 'true set' that we're trying to esti. as sor pt estimators we aim sor good srequentist properties but Z sormulations of bias, essiciency ect are unsuitable. Z principle property required of set estimators is that long-run proportion of set estimators containing Z true param val should be sixed. in other words, is $S(\underline{X})$ is a set <long-run proportion of estimates containing Z true param val> estimator sor a param θ where $\underline{X}$ has density $f_n(\underline{x};\theta)$ then $Pr[S(\underline{X}) \ni \theta; \theta] = \int_{\{\underline{x}:\theta \in S(\underline{x})\}} f_n(\underline{x};\theta)\,d\underline{x}$ should be indep of $\underline{x}$

\Def1.10/ is $Pr[S(\underline{X}) \ni \theta; \theta] = \alpha$ $\forall \theta$ then $S(\underline{X})$ is an α-<u>considence set</u> sor θ & α is Z <u>considence level</u> or <u>coverage</u> of $S(\underline{X})$ <α considence set is Z considence level or coverage>

To create confidence sets, in general find func $T(\underline{X};\theta)$ & a set $C(\theta)$ st
$Pr[T(\underline{X};\theta) \in C(\theta);\theta] = \alpha \ \forall\theta$ then $\{\theta : T(\underline{x};\theta) \in C(\theta)\}$ is an α-confidence set for θ.
Some idea works if only one of $T(\underline{X};\theta)$ & $C(\theta)$ depends on θ, so try first
to find $T(\underline{X};\theta)$ & C st $Pr[T(\underline{X};\theta) \in C;\theta] = \alpha \ \forall\theta$

\Def 1.31/ let $\underline{X}$ have a distri that depends on θ. A pivot (pivotal) is a func
$T(\underline{X};\theta)$ whose distri is indep of θ. If $T(\underline{X};\theta)$ is a pivot then can easily find
a set C that is indep of θ for which $Pr[T(\underline{X};\theta) \in C;\theta] = \alpha \ \forall\theta$

\Ex1.32/ $N(\mu,\sigma^2)$ σ known find a pivot & $\therefore$ find a confidence interval for μ
let $\bar{X}$ be Z sample mean st $\bar{X} \sim N(\mu,\sigma^2/n)$ & define $T = (\bar{X}-\mu)/(\sigma/\sqrt{n})$. Z distri of T is
$N(0,1)$ which is indep of μ & $\therefore$ T is a pivot. let z_p define Z p-quantile of Z
$N(0,1)$ distri st $Pr\left(z_\alpha < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{1-\alpha}\right) = 1-2\alpha$ $\therefore$
$Pr\left(\bar{X} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} - z_\alpha\frac{\sigma}{\sqrt{n}}\right) = 1-2\alpha$ $\therefore$ $S(\underline{x}) = \left(\bar{x} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, \bar{x} - z_\alpha\frac{\sigma}{\sqrt{n}}\right)$ is a $(1-2\alpha)$-
confidence set for μ if σ is known

\Ex1.33/ find a pivot when σ is known in (Ex 1.32) & $\therefore$ find a confidence interval for μ
$\therefore$ let $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ be Z sample var. Z test stat $\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}}$ has a Student's
t-distri with n-1 degrees of freedom, denoted $Stu(n-1)$ & $\therefore$ is a pivot $\therefore$ if t_p
denotes Z p-quantile of $Pr\left(t_\alpha < \frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} < t_{1-\alpha}\right) = 1-2\alpha$ $\therefore$
$Pr\left(\bar{X} - t_{1-\alpha}\frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} - t_\alpha\frac{\hat{\sigma}}{\sqrt{n}}\right) = 1-2\alpha$ & $\therefore$ $\left(\bar{x} - t_{1-\alpha}\frac{\hat{s}}{\sqrt{n}}, \bar{x} - t_\alpha\frac{\hat{\sigma}}{\sqrt{n}}\right)$ is a $(1-2\alpha)$-confidence
~~interval~~ set for μ. eg if $\bar{x} = 1.35$ & $\hat{\sigma} = 2.73$ for sample size $n=30$ then Z 90%
confidence interval for μ is $1.35 \pm 1.70(2.73)/\sqrt{30} = (0.50, 2.20)$ since $t_{0.95} = qt(0.95,29) =$
$1.70 = -t_{0.05}$

\Ex1.34/ X Exponential distri func $Pr(X \le x) = 1-\exp(-x/\theta)$ show $T = X/\theta$ is a pivot $\therefore$
find a confidence interval for θ $\therefore$ $Pr(T \le t) = Pr(X \le \theta t) = 1-\exp(-t)$ this distri
is indep of θ & $\therefore$ T is a pivot. Z p-quantile t_p of Z distri of T satisfies
$p = Pr(T \le t_p) = 1-\exp(-t_p)$ $\therefore$ $t_p = -\log(1-p)$ $\therefore$ $Pr[X/\theta \le -\log(1-\alpha)] = \alpha$ $\therefore$
$Pr[-X/\log(1-\alpha) \le \theta] = \alpha$ $\therefore$ $S(x) = [-x/\log(1-\alpha), \infty)$ is a one-sided α CI for θ
for a two-sided interval: $Pr[-\log(1-\alpha) \le X/\theta \le -\log\alpha] = 1-2\alpha$ $\therefore$

$Pr[-X/\log\alpha \le \theta \le -X/\log(1-\alpha)] = 1-2\alpha$ ∴ is an equi-tailed $(1-2\alpha)$ CI for θ

alternat approach: finding a stat $T(\underline{x})$ & sets $C(\theta)$ st $Pr[T(\underline{X}) \in C(\theta);\theta] = \alpha \ \forall \theta$

one possibility to choose $C(\theta_0)$ to be Z CR for a test ~~for~~ of size α of Z simple null hypoth $\theta=\theta_0$ st $Pr[\underline{X} \in C(\theta_0);\theta_0] = \alpha$ now consider all of Z null hypoth that would not be rejected for a particular sample of data $\underline{x}$. that is all θ for which $\underline{x} \notin C(\theta)$ this set $S(\underline{x}) = \{\theta : \underline{x} \notin C(\theta)\}$ is a set of 'good' param vals & ∴ $S(\underline{X})$ is a possible set esti for Z true val of θ ∴ $Pr[S(\underline{X}) \ni \theta;\theta] = Pr[\underline{X} \notin C(\theta);\theta] = 1-\alpha$ by def of $C(\theta)$ & ∴ $S(\underline{X})$ is a $(1-\alpha)$ confi set for θ this proves result known as <u>inverting Z hypoth test</u>

\thm1.6/ let $C(\theta_0)$ be Z critical region for a test of size α of $H_0: \theta=\theta_0$ then $\{\theta : \underline{X} \in C(\theta)\}$ is a $(1-\alpha)$ confi set for θ

if θ is scalar & $C(\theta)$ is constructed for a two-sided alternat hypoth then Z corres confi set is usually a finite interval; one-sided alternats usually lead to intervals with one infinite end-pt

Ex1.35/ $N(\mu,\sigma^2)$ σ known consider tests of size α for $H_0: \mu=\mu_0$ showed in (Ex1.31) for $H_1: \mu>\mu_0$ Z CR of Z uniformly most powerful test of size α is $C(\mu_0) = \{\underline{x} : \bar{x} \ge \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}\}$ invert this test to find a CI for μ. A $(1-\alpha)$ CI for μ is $S(\underline{X}) = \{\mu : \bar{X} < \mu + z_{1-\alpha}\sigma/\sqrt{n}\} = (\bar{X} - z_{1-\alpha}\sigma/\sqrt{n}, \infty)$

{inverting a hypoth test grey area in sample space $\mathbb{X}$: CR is $C(\theta_0)$ for $H_0: \theta=\theta_0$ for three vals of θ_0 Z sample $\underline{x}$ leads to rejection of $\theta=\theta_0^{(1)}$ but not $\theta=\theta_0^{(2)}$ & $\theta=\theta_0^{(3)}$ blue area in param space Θ: Z set $S(\underline{x})$ containing all param vals θ for which $C(\theta)$ doesn't exclude $\underline{x}$}

for $H_1: \mu<\mu_0$ Z CR is $C(\mu_0) = \{\underline{x} : \bar{x} \le \mu_0 + z_\alpha\sigma/\sqrt{n}\}$ & Z corres $(1-\alpha)$ CI is $S(\underline{X}) = \{\mu : \bar{X} > \mu + z_\alpha\sigma/\sqrt{n}\} = (-\infty, \bar{X} - z_\alpha\sigma/\sqrt{n})$ for $H_1: \mu \ne \mu_0$ there is no uniformly most powerful test but a possible CR is $C(\mu_0) =$
$\{\underline{x} : \bar{x} \le \mu_0 + z_{\alpha/2}\sigma/\sqrt{n} \text{ or } \bar{x} \ge \mu_0 + z_{1-\alpha/2}\sigma/\sqrt{n}\}$ Z corres $(1-\alpha)$ CI is
$S(\underline{X}) = \{\mu : \mu + z_{\alpha/2}\sigma/\sqrt{n} < \bar{X} < \mu + z_{1-\alpha/2}\sigma/\sqrt{n}\} = (\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}, \bar{X} - z_{\alpha/2}\sigma/\sqrt{n})$

\Ex1.36/ X Exponential density func $f(x;\theta)=\theta^{-1}\exp(-x/\theta)$ & consider tests of size α for $H_0:\theta=\theta_0$ showed in (Ex1.30) for $H_1:\theta<\theta_0$ Z CR of Z UMP test is $C(\theta_0)=(0,-\theta_0\log(1-\alpha)]$ invert this test to find a CI for θ A $(1-\alpha)$ CI for θ is $S(X)=\{\theta: X>-\theta\log(1-\alpha)\}=(0,-X/\log(1-\alpha))$ for $H_1:\theta>\theta_0$ Z CR is $C(\theta_0)=[-\theta_0\log\alpha,\infty)$ & Z corres $(1-\alpha)$ CI is $S(X)=\{\theta: X<-\theta\log\alpha\}=(-X/\log\alpha,\infty)$

\Ex1.37/ $X\sim Bin(n,\theta)$ & consider a test of size α for $H_0:\theta=\theta_0$ against $H_1:\theta\neq\theta_0$ a CR is $C(\theta_0)=\{x: Pr(X\le x;\theta_0)<\alpha/2 \text{ or } Pr(X\ge x;\theta_0)<\alpha/2\}$ & Z corres $(1-\alpha)$ CS: $S(X)=\{\theta: Pr(X\le x;\theta)\ge\alpha/2 \text{ \& } Pr(X\ge x;\theta)\ge\alpha/2\}$ now let θ_U be st $Pr(X\le x;\theta_U)=\alpha/2$ & note $Pr(X\le x;\theta)$ increases as θ decreases $\therefore Pr(X\le x;\theta)\ge\alpha/2\ \forall\theta\le\theta_U$ similarly let θ_L satisfy $Pr(X\ge x;\theta_L)=\alpha/2$ & note $Pr(X\ge x;\theta)$ increases as θ increases $\therefore Pr(X\ge x;\theta)\ge\alpha/2\ \forall\theta\ge\theta_L$ $\therefore S(X)=[\theta_L,\theta_U]$ where

$$\sum_{r=0}^{x}\binom{n}{r}\theta_U^r(1-\theta_U)^{n-r}=\frac{\alpha}{2} \qquad \sum_{r=x}^{n}\binom{n}{r}\theta_L^r(1-\theta_L)^{n-r}=\frac{\alpha}{2}$$

can approx Z coverage of CI by Monte Carlo simul given formula for a CI choose true val θ_0 for Z param simu from a model distri $\theta=\theta_0$ & calc CI for each sample Z proportion of intervals containing θ_0 is an approx ~~for each sample~~ to Z coverage of Z interval when Z true param val is θ_0 $\therefore$ is equal-tailed 90% CI in (Ex1.34) $\theta_0=1$:

```
>>n=1 <sample size>  >>theta=1 <true param val>
>>alpha=0.05 <coverage is 1-2*alpha>
>>lo=up=numeric(10000) <vecs to store confi limits>
>>for (i in 1:10000) { <repeat Z following 10000 times>
>>x=rexp(n,1/theta) <simulate a sample from Exp(theta)>
>>lo[i]=-x/log(alpha) <calc & store lower limit>
>>up[i]=-x/log(1-alpha) <calc & store upper limit>  >>}
>>mean(theta<lo) <proportion of times theta is below lower limit>
>>mean(theta>up <theta above upper limit
```

$\{x : x > \mu_0 + a\}$ is critical region for $H_0: \mu = \mu_0$

$\therefore \{x : x < \mu_0 + a\} \therefore (-\infty, \mu_0 + a)$ X

CI for μ : $(x - a, \infty)$

$\{x : x - a > \mu_0\} \therefore (x - a, \infty)$

$\{x : x < a\mu_0$ or $x > b\mu_0\}$ is CR $H_0 : \mu_0 = \mu_0 \therefore$

CI μ : $\therefore \{x : \frac{x}{a} < \mu_0$ or $\frac{x}{b} > \mu_0\} \therefore$

$(\frac{x}{a}, \frac{x}{b})$ X

$\therefore b > a \therefore \frac{x}{b} < \frac{x}{a} \therefore \{x : \frac{x}{a} < \mu_0$ or $\frac{x}{b} > \mu_0\} \therefore (\frac{x}{b}, \frac{x}{a})$ is CI

$X_1, \ldots, X_n \quad \underset{\frac{1}{n+1}}{X_{(1)}} \le \underset{\frac{1}{n+1}}{X_{(2)}} \le \underset{\frac{1}{n+1}}{X_{(3)}} \le \ldots \le \underset{\frac{1}{n+1}}{X_{(n-1)}} \le \underset{\frac{1}{n+1}}{X_{(n)}}$

$X_1, \ldots, X_n$ & $X_0 \therefore$

$\therefore P_r(X_{(1)} < X_0 < X_{(n)}) = \frac{n-1}{n+1} \quad \therefore$ ~~$P_r(X_{(k)}$~~ $P_r(X_{(2)} < X_0 < X_{(n-1)}) = \frac{n-3}{n+1}$

$\therefore P_r(X_{(k)} < X_0 < X_{(n+1-k)}) = \frac{n-2k+1}{n+1}$ this is nicer since not lop sided

in notes: $P_r(X_{(k)} < X_0 < X_{(n+k)}) = \frac{n-2k}{n+1} \therefore$

\1/ $X_0, X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$ 90% CI μ be wider than the 90% prediction interval for X_0

$\therefore$ prediction intervals will be wider than CI since must first do things like predict μ

since variability model is presenting has effect

if $\frac{X - \alpha}{\beta} \sim N(0,1)$ then $X \sim N(\alpha, \beta^2)$

\mention1/ as n gets big $1/I$ asymptotic distri of the MLE for μ is $N(\mu, 1/I)$

\2/ what is asymptotic distri of the score func is $N(0, I)$

since expectation of the score is zero

\3/ what is bias of MLE for μ is: you can't tell

\4/ is asymptotic bias of the MLE for μ is: 0 $\quad \hat{\mu} \sim N(\mu, I(\mu)^{-1})$

$E(\hat{\mu}) \to \mu$ as $n \to \infty$

\5/ what is Z efficiency of Z MLE of μ is: Can't tell

\6/ what is Z asymptotic efficiency of Z MLE of μ is 1 since use Cramer lower bound

\7/ MLE is $\hat{\mu} = 800$ $\quad I(\mu) = n/\mu$ & $n = 200$ $\quad 800/200 = 4$

$I^{-1} = \mu/n$ $\quad \hat{\mu} \sim N(\mu, I(\mu)^{-1})$ $\quad \therefore var(\hat{\mu}) \approx \frac{\mu}{n}$

$sd = \sqrt{var(\hat{\mu})} \approx \sqrt{\frac{\mu}{n}} \approx \sqrt{\frac{\hat{\mu}}{n}}$ $\quad \therefore \sqrt{\frac{\hat{\mu}}{n}} = \sqrt{\frac{800}{200}} = \sqrt{4} = 2$

$\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ $\quad I^{-1} = \begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$

\1-7 Prediction Intervals/

suppose sample data $\underline{x} = (x_1, \dots, x_n)$ would model as realisations of n indep & ident distried r.v. $\underline{X} = (X_1, \dots, X_n)$ with common distri func $F(x;\theta)$ & density $f(x;\theta)$ suppose want to predict Z val of a further datum x_0 that we model as a realisation of a r.v X_0 also has distri $F(x;\theta)$ & is indep of $X_1, \dots, X_n$ a common type of prediction is a predictn interval, a set of possible data vals intended to contain Z true val of x_0. Z main frequentist property required of predictn intervals is Z long-run proportn of predictn intervals containing Z true val of x_0 should be fixed. in other words $R(\underline{X})$ is a predictn interval for X_0 then

$$Pr[X_0 \in R(\underline{X});\theta] = \iint_{\{x : x_0 \in R(\underline{x})\}} f_n(\underline{x};\theta)\, dx\, f(x_0;\theta)\, dx_0$$

should be indep of θ, where $f_n(\underline{x};\theta)$ is Z joint density of $\underline{x}$

\Defi.12/ if $Pr[X_0 \in R(\underline{X});\theta] = \alpha$ $\forall \theta$ then $R(\underline{X})$ is an α-<u>prediction</u> <u>interval</u> for X_0 & α is Z coverage of $R(\underline{X})$

if we knew Z val of θ then an equal-tailed $(1-2\alpha)$-predictn interval for X_0 would be $(q_\alpha, q_{1-\alpha})$ where q_p denotes Z p-quantile of $f(x;\theta)$ how can we form a predictn if we dont know θ? we might think about using quantiles of $f(x;\hat{\theta})$ where $\hat{\theta}$ is an estimate of θ

\Def 1.13/ z equal tailed plug-in $(1-2\alpha)$-prediction interval (or estimative prediction interval) is $(\hat{q}_\alpha, \hat{q}_{1-\alpha})$ where

- $F(\hat{q}_p; \hat{\theta}) = p$ & $\hat{\theta}$ is an estimate of θ

z coverage frequencies of plug-in PIs are typically lower than they should be as z following examples illustrates

\Ex 1.38/ $X_0, X_1, \dots, X_n$ indep $N(\mu, \sigma^2)$ r.v σ known const consider estimator $\bar{X} = \sum_{i=1}^{n} X_i / n$ for μ. z p-quantile of $N(\bar{X}, \sigma^2)$ is $\bar{X} + \sigma z_p$ where z_p is z p-quantile of z $N(0,1)$ distri & $\therefore$ z plug-in $(1-2\alpha)$-PI for X_0 is $(\bar{X} - \sigma z_\alpha, \bar{X} + \sigma z_{1-\alpha})$

$X_0 - \bar{X}$ follows a Normal-distri with expectatn $E(X_0 - \bar{X}) = E(X_0) - E(\bar{X}) = \mu - \mu = 0$ & var $var(X_0 - \bar{X}) = var(X_0) + var(\bar{X}) = (1 + 1/n)\sigma^2$ $\therefore$ $Z = (X_0 - \bar{X}) / \sqrt{(1+1/n)\sigma^2} \sim N(0,1)$ $\therefore$ z coverage of this plug-in predictn Interval is:

$Pr(\bar{X} + \sigma z_\alpha < X_0 < \bar{X} + \sigma z_{1-\alpha}) = Pr(z_\alpha < \frac{X_0 - \bar{X}}{\sigma} < z_{1-\alpha}) =$

$Pr(\frac{z_\alpha}{\sqrt{1+1/n}} < Z < \frac{z_{1-\alpha}}{\sqrt{1+1/n}}) < Pr(z_\alpha < Z < z_{1-\alpha}) = 1 - 2\alpha$

z coverage of z plugin 90%. PI is plotted against sample size n

z plug-in PI's contain X_0 less often than they claim is that

- they fail to account for z uncertainty in z esti of z param. Can improve on plug-in predictn intervals if can find a stat $T(X_0, \underline{X})$ whose distri doesn't depend on θ

\Def 1.14/ let X_0 & $\underline{X}$ have distris that depend on θ. an ancillary stat is a func $T(X_0, \underline{X})$ whose distri is indep of θ

ancillary stat is like a pivot but is a func of $\underline{X}$ & X_0 instead of $\underline{X}$ & θ

- \Ex 1.39/ z quantity $T(X_0, \underline{X}) = \frac{X_0 - \bar{X}}{\sigma\sqrt{1+1/n}}$ is an ancillary stat $\because$ it has a $N(0,1)$ distri $\therefore$ $1 - 2\alpha = Pr(z_\alpha < \frac{X_0 - \bar{X}}{\sigma\sqrt{1+1/n}} < z_{1-\alpha}) =$

$Pr(\bar{X} + z_\alpha \sigma\sqrt{1+1/n} < X_0 < \bar{X} + z_{1-\alpha}\sigma\sqrt{1+1/n})$ & $\therefore$ $(1-2\alpha)$ PI for X_0:

$(\bar{X} + z_\alpha \sigma\sqrt{1+1/n}, \bar{X} + z_{1-\alpha}\sigma\sqrt{1+1/n})$ this is wider than z plug-in interval which omits z $1/n$ terms

\Ex1.40/ $X_0, X_1, \ldots, X_n$ indep $N(\mu, \sigma^2)$ r.v μ & σ unknown

know: $(X_0 - \bar{X})/\sqrt{1+1/n}$ has a $N(0, \sigma^2)$ distri $\therefore$ if write $S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$ for z sample var, z stat $T(X_0, \underline{X}) = \frac{X_0 - \bar{X}}{S\sqrt{1+1/n}}$ has a $Stu(n-1)$ distri & so is ancillary. if t_p denotes z p-quantile of z $Stu(n-1)$ distri then $1-2\alpha) = Pr\left(t_\alpha < \frac{X_0 - \bar{X}}{S\sqrt{1+1/n}} < t_{1-\alpha}\right)$ $= Pr(\bar{X} + t_\alpha S\sqrt{1+1/n} < X_0 < \bar{X} + t_{1-\alpha} S\sqrt{1+1/n})$ & $\therefore$ a $(1-2\alpha)$ PI for X_0: $(\bar{X} + t_\alpha S\sqrt{1+1/n}, \bar{X} + t_{1-\alpha} S\sqrt{1+1/n}]$ eg z 90% PI based on z data in (Ex1.33) is $(-3.37, 6.07)$

\Ex1.41/ let $X_0, X_1, \ldots, X_n$ be indep Exponential distri func $1-\exp(-x/\theta)$ for $x>0$ recall: sum of n indep Exponential r.v has a gamma distri in particular. $\sum_{i=1}^{n} X_i$ has density $f(s) = \frac{s^{n-1}}{\theta^n \Gamma(n)} e^{-s/\theta}$ for $s>0$ $\therefore$ $T = X_0/\bar{X}$ is ancillary & $\therefore$ find a predictn interval for X_0 $\therefore$ writing $\phi = \theta/(1+t/n)$ have $Pr(T \le t) = Pr(X_0 \le t\bar{X}) = Pr(X_0 \le \frac{t}{n}\sum_{i=1}^{n} X_i)$ $= \int_0^\infty Pr(X_0 \le ts/n \mid \sum_{i=1}^{n} X_i = s) f(s)\,ds = \int_0^\infty Pr(X_0 \le ts/n) f(s)\,ds$ by indep $= \int_0^\infty \{1 - e^{-ts/(n\theta)}\} \frac{s^{n-1}}{\theta^n \Gamma(n)} e^{-s/\theta}\,ds = 1 - \int_0^\infty \frac{s^{n-1}}{\theta^n \Gamma(n)} e^{-(1+t/n)s/\theta}\,dS =$

$1 - (1+t/n)^{-n} \int_0^\infty \frac{s^{n-1}}{\phi^n \Gamma(n)} e^{-s/\phi}\,dS = 1 - (1+t/n)^{-n}$ since z last integral is z density of a gamma distri $\therefore$ z distri func of T is indep of θ & $\therefore$ T is ancillary solving $Pr(T \le t) = p$ for t fi z p-quantile of T is $t_p = n[(1-p)^{-1/n} - 1]$ $\therefore$
$Pr(n[(1-\alpha)^{-1/n} - 1] < X_0/\bar{X} < n[\alpha^{-1/n} - 1]) = 1-2\alpha$ $\therefore$ z $(1-2\alpha)$ PI:
$(n[(1-\alpha)^{-1/n} - 1]\bar{X}, n[\alpha^{-1/n} - 1]\bar{X})$

third way of forming predict interv used when a suitable ancillary stat cannot be found since $X_0, X_1, \ldots, X_n$ are iid sort in order then X_0 is equally likely to be in any positn $\therefore$ there are $n+1$ positns & $\therefore$ z prob of X_0 being in any particular positn is $1/(n+1)$ $\therefore$

if $X_{(1)}, \dots, X_{(n)}$ denote Z other stats of $X_1, \dots, X_n$ then $Pr(X_0 < X_{(1)}) = 1/(n+1)$, $Pr(X_{(k)} < X_0 < X_{(k+1)}) = 1/(n+1)$ for $k=1,\dots,n-1$ & $Pr(X_0 > X_{(n)}) = 1/(n+1)$ $\therefore$ $Pr(X_{(k)} < X_0 < X_{(n-k)}) = \frac{n-2k}{n+1}$ $\forall 1 \le k \le n/2$ & $\therefore$ $(X_{(k)}, X_{(n-k)})$ is a $(n-2k)/(n+1)$ PI. this interval makes no use of a parametric model for Z distri of Z r.v. & $\therefore$ is called a nonparametric PI.

can approx Z coverage of PI's by MonteCarlo simulatn, similar to its use for CI's containing a new data val simulated from Z model distri. eg: code for equal-tailed 90% PI in (Ex 1.41) when $\theta = 1$:

```
>> n=10   [[sample size]]   >>theta=1   #true param val #
>>alpha=0.05   #coverage is 1-2*alpha #
lo=up=numeric(10000)   #vecs to store PIs # predictn limits #
>>for (i in 1:10000){ # repeat Z following 1000 times #
x=rexp(n,1/theta) # simulate a sample from Exp(theta) #
>>lo[i]=((1-alpha)^(-1/n)-1)*sum(x)   #calc lower limit #
>>up[i]=(alpha^(-1/n)-1)*sum(x)   # calc upper limit #   >>}
>>x0=rexp(10000,1/theta) # simulate 10000 new data vals #
>>mean(x0<lo) # proportn of x0 is below Z lower limit #
>>mean(x0>up) #proportn of times x0 is above Z upper limit #
```

week 5 / 2.1 Maximum likelihood estimators

~~Maximum~~ maximum likelihood estimators. Let $\underline{X} = (X_1, \dots, X_n)$. Z maximum likelihood estimator $\hat{\theta}(\underline{X})$ maximises Z log-likelihood $l(\theta; \underline{X})$ & can find Z sampling properties of Z MLE by examining Z sampling properties of $l(\theta; \underline{X})$ when $X_1, \dots, X_n$ are iid, Z likelihood: $L(\theta; \underline{X}) = \prod_{i=1}^{n} f(X_i; \theta)$ st Z log-likelihood, score & observed info are all sums of (funcs of) iid r.v. $l(\theta; \underline{X}) = \sum_{i=1}^{n} \log f(X_i; \theta)$

$U(\theta; \underline{X}) = \sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}$ $J(\theta; \underline{X}) = -\sum_{i=1}^{n} \frac{\partial^2 \log f(X_i; \theta)}{\partial \theta^2}$ use Z shorthand

$\hat{\theta} = \hat{\theta}(\underline{X})$, $U = U(\theta; \underline{X})$ & $J(\theta) = J(\theta; \underline{X})$. also need to know what is meant by var of a vec of r.v. if Y is a vec of r.v. with ith element Y_i then Z expectatn of Y is defined to be Z vec with ~~ith~~ ith element $E(Y_i)$. & Z var of Y is defined to be Z matrix with (i,j)th element $Cov(Y_i, Y_j)$ this can be:

$var(Y) = E\{[Y - E(Y)][Y - E(Y)]^T\} = E(YY^T) - E(Y)E(Y)^T$

note: $Cov(Y_i, Y_i) = var(Y_i)$ st Z diagonal elements of $var(Y)$ are Z vars of Z elements of Y note: that Z $var(Y)$ is a symmetric matrix b. $Cov(Y_i, Y_j) = Cov(Y_j, Y_i)$. going to derive Z sampling distri of Z MLE in Z limit as Z sampling size n increases to infinity. Z idea is that Z limit will be good approx to Z sampling distri of Z MLE for a large but finite n. to find this asymptotic distri of Z MLE find Z asymptotic distri of Z score then use a taylor expansion of $U(\theta)$ about $\hat{\theta}$ to link Z score to Z MLE. we consider Z general case in which θ is a vec st Z score is also a vec & Z info is a matrix. in our derivation use Z strong law of nums & Z central limit thm

\ ~~theo~~ Strong law of nums / let $X_1, ..., X_n$ be iid with expectatn $\mu < \infty$ ∴ $\bar{X} \to \mu$ with probab 1 as $n \to \infty$ iff ~~E E~~ $E|X_1| < \infty$

\ ~~theo~~ central limit thm / let $X_1, ..., X_n$ be iid expectation $\mu < \infty$ & var matrix Σ then Z distri of $n^{-1/2}(\sum_{i=1}^{n} X_i - n\mu)$ converges to Z $N(0, \Sigma)$ distri as $n \to \infty$ if Σ is non-negative definite & ~~det~~ $det(\Sigma) < \infty$

write $U(\theta) = \sum_{i=1}^{n} U_i$ where $U_i = \partial \log f(X_i; \theta)/\partial\theta$ are iid r.v. & $J(\theta) = \sum_{i=1}^{n} J_i$ where $J_i = -\partial^2 \log f(X_i; \theta)/\partial\theta^2$ ∴ under suitable regularity condins on ~~gl~~ f, $E(U_i) = 0$ & $var(U_i) = E(J_i)$ now since Z J_i are identi distried, all have Z same expectation which denotes as I_1 ∴ Z expected info is $I(\theta) = E[J(\theta)] = \sum_{i=1}^{n} E(J_i) = nI_1$.

Z Central limit thm tells $n^{-1/2}U(\theta) \to N(0, I_1)$ as $n \to \infty$ &
∴ $I(\theta)^{-1/2}U(\theta) = (nI_1)^{-1/2}U(\theta) \to N(0, \mathcal{I})$ as $n \to \infty$ where $I(\theta)^{-1/2}$
satisfies $I(\theta)^{-1/2}I(\theta)^{-1/2} = I(\theta)^{-1}$ & $\mathcal{I}$ is Z identity matr

having found Z asymptotic distri of Z score, can now do Z same for Z MLE, $\hat{\theta}$, shall know $\hat{\theta}$ is consistent for θ but consistency does hold under suitable regularity conditions. if this case $\hat{\theta} - \theta$ is small for large n & a Taylor expansion of Z score yields $U(\hat{\theta}) \approx U(\theta) + \frac{\partial U(\theta)}{\partial \theta}(\hat{\theta} - \theta) = U(\theta) - J(\theta)(\hat{\theta} - \theta)$ if Z MLE occurs at a local max of Z log-likelihood then $U(\hat{\theta}) = 0$ ∴ $\hat{\theta} - \theta \approx J(\theta)^{-1}U(\theta)$

Z observed info is $J(\theta)$ is a sum of n iid r.v. ∴ by strong law of large Nums: $J(\theta) \approx I(\theta)$ when n is large ∴ $\hat{\theta} - \theta \approx I(\theta)^{-1}U(\theta)$
∴ $I(\theta)^{1/2}(\hat{\theta} - \theta) \approx I(\theta)^{-1/2}U(\theta) \to N(0, \mathcal{I})$ ∴ when n is large Z distri of Z MLE is approx Normal with expectatn θ & var $I(\theta)^{-1}$ ∴ MLE is asymptotically ~~distrib~~ unbiased & asymptotically efficient
∴ $\hat{\theta} \sim N(\theta, I(\theta)^{-1})$ when n is large
∴ use to construct hypothesis tests & confidence sets

note: if n is large then Z square roots of Z diagonal elements of $I(\theta)^{-1}$ approx Z standard errors of Z MLEs in $\hat{\theta}$ ∴ standard errors estied by replacing θ with $\hat{\theta}$ in $I(\theta)^{-1}$

\Ex 2.1/ Poi(θ) MLE $\hat{\theta} = \bar{x}$ & $I(\theta)^{-1} = \theta/n$ ∴ Z asymptotic distri of Z MLE is $N(\theta, \theta/n)$ & its estimated standard error is $\sqrt{\hat{\theta}/n}$

\Ex 2.2/ Exp(θ) MLE is $\hat{\theta} = 1/\bar{x}$ & $I(\theta)^{-1} = \theta^2/n$ st Z asymptotic distri of Z MLE is $N(\theta, \theta^2/n)$ & its estimated standard error is $\hat{\theta}/\sqrt{n}$

\Ex 2.3/ for Z $N(\mu, \sigma^2)$ MLE $(\hat{\mu}, \hat{\sigma})$ where $\hat{\mu} = \bar{x}$ & $\hat{\sigma}^2 = \sum_{i=1}^{n} x_i^2/n - \bar{x}^2$ & its asymptotic distri is Normal with expectatn (μ, σ) & var matr $I(\mu, \sigma)^{-1} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & \sigma^2/2n \end{bmatrix}$ if have a sample of $n = 30$ & $\sum_{i=1}^{n} x_i = 37.4$ & $\sum_{i=1}^{n} x_i^2 = 269.6$ then our esties are $\hat{\mu} = 1.25$ & $\hat{\sigma} = 2.73$ with standard errors $\hat{\sigma}/\sqrt{n} = 0.50$ & $\hat{\sigma}/\sqrt{2n} = 0.25$ respectively

2.2 Likelihood-based tests

uniformly most powerful tests rarely exist for two-sided alternative hypotheses. instead, can consider tests that are generally applicable & that typically have good, if not optimal properties. testing $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$ where θ is a vec of length p & let $T(\underline{X})$ be z test stat. in order to construct a test with z desired size, must determine z distri of $T(\underline{X})$ under z null hypoth. that is $Pr[T(\underline{X}) \leq t; \theta_0]$

let $\hat{\theta}$ denote z maximum likelihood estimator based on r.v. $\underline{X} = (X_1, \ldots, X_n)$ shall present three tests: z wald score & likelihood ratio tests, shall need z following result in order to obtain z null distri of their stats

prop let Z be a $N(0,1)$ r.v. with density $(2\pi)^{-1/2}\exp(-z^2/2)$ z moment generating func of z r.v. Z^2 is $E(e^{tZ^2}) =$

$$\int_{-\infty}^{\infty} e^{tz^2}(2\pi)^{-1/2}e^{-z^2/2}\,dz = \int_{-\infty}^{\infty}(2\pi)^{-1/2}e^{-(1-2t)z^2/2}\,dz = \sigma\int_{-\infty}^{\infty}(2\pi\sigma^2)^{-1/2}e^{-z^2/(2\sigma^2)}\,dz$$

$\{$where $\sigma = (1-2t)^{-1/2}\}$ $= \sigma$ ∵ z last integrand is z density of a $N(0,\sigma^2)$ distri & ∴ integrates to 1. now let $Y = \sum_{i=1}^{p} Z_i^2$ be z sum of p squared indep $N(0,1)$ r.v. $Z_1, \ldots, Z_p$. z moment generating func of Y is $E(e^{tY}) = E(e^{t\sum_{i=1}^{p} Z_i^2}) = \prod_{i=1}^{p} E(e^{tZ_i^2}) =$
$\prod_{i=1}^{p}(1-2t)^{-1/2} = (1-2t)^{-p/2}$ this is z moment generating func of z χ^2_p distri so $Y \sim \chi^2_p$

z wald test stat $W = (\hat{\theta}-\theta_0)^T I(\hat{\theta})(\hat{\theta}-\theta_0)$ measures z dist betw θ_0 & $\hat{\theta}$ eg if θ is scalar then $W = (\hat{\theta}-\theta_0)^2 / I(\hat{\theta})^{-1}$ st z wald test stat is z squared dist betw θ_0 & $\hat{\theta}$ standardized by an esti of z asymptotic var of $\hat{\theta}$. if $\hat{\theta}$ is far from θ_0 (relative to its asymptotic var) then our data are saying θ_0 is an unlikely param val & ∴ should reject H_0 if W is large. z critical region ∴ has form $\{\underline{x}: W \geq d\}$ to find z critical val, d, for a test of size α we need to know Z null distri of W. ∴ know when θ_0 is Z true param val, $I(\theta_0)^{1/2}(\hat{\theta}-\theta_0) \to N(0, I)$ as Z sample size $n \to \infty$. Since $\hat{\theta}$ is consistent, can replace $I(\theta_0)$ with $I(\hat{\theta})$ to show Z p-vec $Z = I(\hat{\theta})^{1/2}(\hat{\theta}-\theta_0)$ also converges to Z $N(0, I)$ distri ∴ each of Z p elements of Z is asymptotically indep & standard Normal. $W = Z^T Z$ is Z sum of Z squared elements of Z & ∴ W converges to Z χ_p^2 distri when H_0 is true ∴ Z null distri of W is approx χ_p^2 when Z sample size is large, irrespective of Z underlying probab model for $\underline{X}$. Z Wald test ∴ rejects H_0 at level α if W exceeds Z $(1-\alpha)$-quantile of Z χ_p^2 distri

Z <u>Score test stat</u> $S = U(\theta_0)^T I(\theta_0)^{-1} U(\theta_0)$ measures Z difference betw $U(\theta_0)$ & $U(\hat{\theta}) = 0$ eg if θ is scalar then $S = U(\theta_0)^2 / I(\theta_0)$ st Z score test stat is Z squared diff betw $U(\theta_0)$ & $U(\hat{\theta})$ standardized by its asym var when H_0 is true. if $U(\theta_0)$ is far from zero then θ_0 is an unlikely param val & ∴ we reject H_0 if S is large. Z critical region ∴ has form $\{\underline{x} : S \geq d\}$ Z score test stat is asym equal to Z wald test stat & ∴ Z null distri of S is also approx χ_p^2 when n is large, consider taylor expansion:

$U(\theta_0) \approx U(\hat{\theta}) + J(\hat{\theta})(\hat{\theta}-\theta_0) \approx I(\hat{\theta})(\hat{\theta}-\theta_0) \approx I(\theta_0)(\hat{\theta}-\theta_0)$ when n is large & H_0 is true ∴ $S \approx (\hat{\theta}-\theta_0)^T I(\theta_0)^{-1} I(\theta_0)(\hat{\theta}-\theta_0) I(\theta_0)$

$(\hat{\theta}-\theta_0)^T I(\theta_0)(\hat{\theta}-\theta_0) \approx W$ since Z expected info is symmetric matr ∴ Z score test rejects H_0 at level α if S exceeds Z $(1-\alpha)$-quantile of Z χ_p^2 distri

<u>Likelihood ratio test</u>: Z likelihood ratio for $H_0 : \theta = \theta_0$ & $H_1 : \theta \neq \theta_0$ is $\Lambda = \frac{L(\theta_0)}{\sup_\theta L(\theta)} = \frac{L(\theta_0)}{L(\hat{\theta})}$ & Z <u>likelihood ratio test stat</u>: $-2\log\Lambda = 2[l(\hat{\theta}) - l(\theta_0)]$ measures Z distance betw $l(\theta_0)$ & $l(\hat{\theta})$ if $l(\hat{\theta})$ is much greater than $l(\theta_0)$ then θ_0 is a unlikely param val & ∴ reject H_0 if $-2\log\Lambda$ is large. Z critical

region $\therefore$ has form $\{\underline{x} : -2\log\Lambda \geq d\}$ Z likelihood ratio test is asym equal to Z Wald test stat $\therefore$ Z null distri of $-2\log\Lambda$ approx χ^2_p when n is large. taylor approx:

$l(\theta_0) \approx l(\hat{\theta}) + (\theta_0 - \hat{\theta})^T U(\hat{\theta}) - \frac{1}{2}(\theta_0 - \hat{\theta})^T J(\hat{\theta})(\theta_0 - \hat{\theta}) = l(\hat{\theta}) - \frac{1}{2}(\theta_0 - \hat{\theta})^T J(\hat{\theta})(\theta_0 - \hat{\theta})$

when n is large since $U(\hat{\theta}) = 0 \therefore$

$-2\log\Lambda \approx (\theta_0 - \theta)^T J(\hat{\theta})(\theta_0 - \hat{\theta}) \approx (\theta_0 - \hat{\theta})^T I(\hat{\theta})(\theta_0 - \hat{\theta}) = W$

Z likelihood ratio test $\therefore$ rejects H_0 at level α if $-2\log\Lambda$ exceeds Z $(1-\alpha)$-quantile of Z χ^2_p distri

\Ex2.4/ $N(\mu, \sigma^2)$ r.v. σ is known & consider testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu$ Z likelihood is $L(\mu) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$

$\therefore$ Z log-likelihood: $l(\mu) \doteq \text{const} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$

Z Score is $U(\mu) = \frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = \frac{n(\bar{x} - \mu)}{\sigma}$ Z maximum likelihood esti is $\hat{\mu} = \bar{x}$, Z observed info is $J(\mu) = -\frac{\partial^2 l}{\partial \mu^2} = \frac{n}{\sigma^2}$ & Z expected info is $I(\mu) = E[J(\mu)] = n/\sigma^2 \therefore$ Z Wald test stat is

$W = (\hat{\mu} - \mu_0)^2 I(\hat{\mu}) = \frac{n(\bar{x} - \mu_0)^2}{\sigma^2}$ Z Score test stat is: $S = \frac{U(\mu_0^2)}{I(\mu_0)} =$

$\frac{n^2(\bar{x} - \mu_0)^2/\sigma^4}{n/\sigma^2} = \frac{n(\bar{x} - \mu_0)^2}{\sigma^2}$ & Z likelihood ratio test stat is:

$2[l(\hat{\mu}) - l(\mu_0)] = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[(x_i - \mu_0)^2 - (x_i - \hat{\mu})^2\right] = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[2x_i(\hat{\mu} - \mu_0) - (\hat{\mu}^2 - \mu_0^2)\right] =$

$\frac{\hat{\mu} - \mu_0}{\sigma^2}\sum_{i=1}^{n}\left[2x_i - (\hat{\mu} - \mu_0)\right] = \frac{\bar{x} - \mu_0}{\sigma^2}\sum_{i=1}^{n}\left[(x_i - \bar{x}) + (x_i - \mu_0)\right] = \frac{\bar{x} - \mu_0}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)$

$= \frac{n(\bar{x} - \mu_0)^2}{\sigma^2}$ in this ex $\therefore$ all three test stats are Z same since μ is scalar, Z tests reject H_0 in favour of H_1 at level α is Z test stats exceed Z $(1-\alpha)$-quantile of Z χ^2_1 distri. eg if want to test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ for Z data in (Ex2.3) suppose know $\sigma = 3 \therefore n(\bar{x} - \mu_0)^2/\sigma^2 = 36(1.25^2)/9 = 5.18$ with p-val

>> 1-pchisq(5.18, 1) $\Rightarrow$ 0.023

\Ex2.5/ $X \sim Bin(n, \theta)$ & $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ Z likelihood: $L(\theta) \propto \theta^x(1-\theta)^{n-x}$ & $\therefore$ log-likelihood: $l(\theta) =$ ~~const + X log~~ const + X $\log\theta + (n - X)\log(1-\theta)$ with score $U(\theta) = \frac{\partial l}{\partial \theta} = \frac{X}{\theta} - \frac{n-X}{1-\theta} = \frac{X - n\theta}{\theta(1-\theta)}$

Z info: $I(\theta) = -E\left(\frac{\partial^2 l}{\partial \theta^2}\right) = E\left[\frac{X}{\theta^2} + \frac{n-X}{(1-\theta)^2}\right] = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}$ since

$E(X) = n\theta$ $\therefore$ Z MLE is $\hat{\theta} = X/n$, Z Wald test stat:

$W = (\hat{\theta} - \theta_0)^2 \, I(\hat{\theta}) = \frac{n(\hat{\theta} - \theta_0)^2}{\hat{\theta}(1-\hat{\theta})}$ Z score test stat:

$S = \frac{U(\theta_0)^2}{I(\theta_0)} = \frac{n(\hat{\theta} - \theta_0)^2}{\theta_0(1-\theta_0)}$ & Z likelihood ratio test stat:

$-2\log\Lambda = 2[l(\hat{\theta}) - l(\theta_0)] = 2n\left[\hat{\theta}\log\left(\frac{\hat{\theta}}{\theta_0}\right) + (1-\hat{\theta})\log\left(\frac{1-\hat{\theta}}{1-\theta_0}\right)\right]$ since θ is a scalar in this ex, Z three tests would reject H_0 in favour of H_1 at level α if Z test stat exceeds Z $(1-\alpha)$-quantile of Z χ^2_1 distri

Further generalisation of Z likelihood ratio test to Z case where both Z null & alternative hypoth are composite

Consider $H_0: \theta \in \Theta_0$ & $H_1: \theta \notin \Theta_0$ where Θ_0 specifies vals of q of Z p elements of θ. Z likelihood ratio:

~~$\Lambda = \sup_{\theta \in \Theta_0} L(\theta)$~~ $\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_\theta L(\theta)}$ & Z asympto null distri of Z test stat $-2\log\Lambda$ is χ^2_q

\Ex2.6/ indep $N(\mu, \sigma^2)$ $X_1, \dots, X_n$ consider $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$ are composite hypoth because no val is specified for σ^2. Z likehood: $L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right]$

& Z maximum lihood under H_0 is $L(\mu_0, \hat{\sigma}_0^2)$ where $\hat{\sigma}_0^2$ maximises $L(\mu_0, \sigma^2)$ $\therefore$ ~~l~~ $l(\mu_0, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2$

& $\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i - \mu_0)^2$ st $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu_0)^2$

& $l(\mu_0, \hat{\sigma}_0^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\hat{\sigma}_0^2 - \frac{n}{2}$

Z unrestricted maximum lihood is $L(\hat{\mu}, \hat{\sigma}^2)$ where $\hat{\mu} = \bar{x}$ & $\hat{\sigma}^2 = \sum_{i=1}^n (x_i - \bar{x})^2/n$ are Z usual maximum likehood estirs $\therefore$

$l(\hat{\mu}, \hat{\sigma}^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\hat{\sigma}^2 - \frac{n}{2}$ &

$-2\log\Lambda = -2[l(\mu_0, \hat{\sigma}_0^2) - l(\hat{\mu}, \hat{\sigma}^2)] = 2n\log(\hat{\sigma}_0/\hat{\sigma})$ Z test reject H_0 at level α if $-2\log\Lambda$ exceeds Z $(1-\alpha)$-quantile of Z χ^2_1 distri

eg if test $H_0: \mu = 0$ $H_1: \mu \neq 0$ for (ex2.3) then $n = 30$, $\hat{\sigma} = 2.73$ & $\hat{\sigma}_0^2 = \sum_{i=1}^n x_i^2/n = 8.99$ st $-2\log\Lambda = 5.61$ with p-val >> 1-pchisq(5.61,1)

$\Rightarrow$ 0.018

Likelihood-based Confidence Sets 2.3

2.3: let θ be a vec of length p & let $\hat{\theta}$ be Z maximum lihood estr for θ based on a sample of size n. Z distri of $I(\theta)^{1/2}(\hat{\theta}-\theta)$ is approx $N(0, I)$ where n is large $\therefore$ Z distri of Z rv $T(\underline{X};\theta) = (\hat{\theta}-\theta)^T I(\theta)(\hat{\theta}-\theta)$ is approx χ_p^2 when n is large. this χ_p^2 distri is indep of θ & $\therefore$ $T(\underline{X};\theta)$ is approx pivot $\therefore$ if c denotes Z $(1-\alpha)$ quantile of Z χ_p^2 distri then an approx $(1-\alpha)$ confidence set for θ based of Z MLE is

$$S(\underline{X}) = \left\{\theta : (\hat{\theta}-\theta)^T I(\theta)(\hat{\theta}-\theta) < c\right\}$$

Can also derive confidence sets by inverting lihood-based tests of $H_0: \theta=\theta_0$ $H_1: \theta \neq \theta_0$. Z Wald test of level α has critical region $C(\theta_0) = \left\{\underline{x} : (\hat{\theta}-\theta_0)^T I(\hat{\theta})(\hat{\theta}-\theta_0) \geq c\right\}$ & Z corres ~~$(1-\alpha)$~~ quantile $(1-\alpha)$-confidence set for θ is $S(\underline{X}) = \left\{\theta : (\hat{\theta}-\theta)^T I(\hat{\theta})(\hat{\theta}-\theta) < c\right\}$

or Z score test of level α has critical region

$C(\theta_0) = \left\{\underline{x} : U(\theta_0)^T I(\theta_0)^{-1} U(\theta_0) \geq c\right\}$ & Z corres $(1-\alpha)$-confidence set for θ is $S(\underline{X}) = \left\{\theta : U(\theta)^T I(\hat{\theta})^{-1} U(\theta) < c\right\}$

Z lihood ratio test of level α has CR: $C(\theta_0) = \left\{\underline{x} : 2[l(\hat{\theta}) - l(\theta_0)] \geq c\right\}$ & Z corres $(1-\alpha)$-confidence set for θ is $S(\underline{X}) = \left\{\theta : 2[l(\hat{\theta}) - l(\theta)] < c\right\}$

Ex 2.7: $X_1, \dots, X_n$ indep $N(\mu, \sigma^2)$ rv σ known from (Ex 2.4) Z max lihood esti of μ is $\hat{\mu} = \bar{x}$, Z expected info is $I(\mu) = n/\sigma^2$ & Z Wald score & lihood ratio test stats are all equal to $n(\bar{x}-\mu)^2/\sigma^2$ $\therefore$ Z approx α-confidence interval for μ based on Z MLE is $\left\{\mu : n(\bar{x}-\mu)^2/\sigma^2 < c\right\} = \left(\bar{x} - \sigma\sqrt{c/n}, \bar{x} + \sigma\sqrt{c/n}\right)$ where c is Z ~~of~~ α-quantile of Z χ_1^2 distri since $c = z_{\alpha/2}^2$. this is Z same interval as derived in (Ex 1.35) & is also Z interval that corres to Z Wald, score, lihood ratio tests. for dat in (Ex 2.3) known is $\sigma = 3$ then Z 90% CI for μ is $1.25 \pm 3\sqrt{2.71/30} = (0.35, 2.15)$

Ex 2.8 $X \sim Bi(n, \theta)$ from (Ex 2.5) Z max likelihood esti of θ is $\hat\theta = x/n$ & Z expected info is $I(\theta) = n/\{\theta(1-\theta)\}$ ∴ Z approx α-CI for θ based on Z MLE is

$$\{\theta : (\hat\theta - \theta)^T I(\theta)(\hat\theta - \theta) < c\} = \left\{\theta : \frac{n(\hat\theta-\theta)^2}{\theta(1-\theta)} < c\right\}$$ inequality satisfied iff

$(n+c)\theta^2 - (2n\hat\theta + c)\theta + n\hat\theta^2 < 0$ holds iff θ lies betw Z roots of Z quadratic: $\dfrac{(2n\hat\theta + c) \pm \sqrt{(2n\hat\theta+c)^2 - 4n\hat\theta^2(n+c)}}{2(n+c)}$ these roots ∴ define Z endpts of an approx α-confidence interval for θ. also:

an confidence set based on Z wald test:

$$\{\theta : (\hat\theta - \theta)^T I(\hat\theta)(\hat\theta - \theta) < c\} = \left\{\theta : \frac{n(\hat\theta-\theta)^2}{\hat\theta(1-\hat\theta)} < c\right\} = \left(\hat\theta - \sqrt{\frac{c\hat\theta(1-\hat\theta)}{n}}, \hat\theta + \sqrt{\frac{c\hat\theta(1-\hat\theta)}{n}}\right)$$

Z score test yields Z same confidence set as Z MLE in this ex, while Z approx α-CS based on Z likelihood ratio test,

$\{\theta : x\log(\hat\theta/\theta) + (n-x)\log[(1-\theta)/(1-\hat\theta)] < 2c\}$ has no simple form

Ex 2.9 $X_1, \ldots, X_n$ be indep $N(\mu, \sigma^2)$ rv μ & σ unknown from (Ex 1.7) Z max likelihood esti of $\theta = (\mu, \sigma)$ is $\hat\theta = (\bar x, s)$ & Z expected info is

$I(\mu, \sigma) = \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & 2n/\sigma^2 \end{pmatrix}$ ∴ Z approx α-CS based of Z MLE is

$$\{\theta : (\hat\theta - \theta)^T I(\theta)(\hat\theta - \theta) \le c\} = \left\{\begin{pmatrix}\mu\\ \sigma\end{pmatrix} : \begin{pmatrix}\bar x - \mu \\ s - \sigma\end{pmatrix}^T \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & 2n/\sigma^2\end{pmatrix}\begin{pmatrix}\bar x - \mu \\ s-\sigma\end{pmatrix} \le c\right\}$$

$$= \left\{\begin{pmatrix}\mu\\ \sigma\end{pmatrix} : (\bar x - \mu)^2 + 2(s-\sigma)^2 \le \frac{c\sigma^2}{n}\right\}$$ Z approx α-CS based on Z wald test is $\{\theta : (\hat\theta - \theta)^T I(\hat\theta)(\hat\theta - \theta) \le c\} =$

$$\left\{\begin{pmatrix}\mu\\ \sigma\end{pmatrix} : \begin{pmatrix}\bar x - \mu \\ s - \sigma\end{pmatrix}^T \begin{pmatrix} n/s^2 & 0 \\ 0 & 2n/s^2\end{pmatrix}\begin{pmatrix}\bar x - \mu \\ s-\sigma\end{pmatrix} \le c\right\} = \left\{\begin{pmatrix}\mu\\ \sigma\end{pmatrix} : (\bar x - \mu)^2 + 2(s-\sigma)^2 \le \frac{cs^2}{n}\right\}$$

this defines an ellipse centred at $(\bar x, s)$ with axes whose squared lengths are $4cs^2/n$ & $2cs^2/n$ for Z data in (Ex 2.3) Z 90% CS centred at (1.25, 2.73) & its axes have lengths 2.14 & 1.51. Z CS's based on Z score & likelihood ratio tests are different again but their formulae are slightly complicated

2.4 Models with covariates

Ex 2.10 if X_i denote sea level in year z_i then can assume X_i has a Normal distri with expectatn $\alpha + \beta z_i$ & var σ^2 ∴ $X_i \sim N(\alpha + \beta z_i, \sigma^2)$ ∴ three params (α, β, σ) ∴ like to make inferences eg esti params or test Z hypoth $\beta = 0$ which corres to being no trend in expected sea level. Also want to predict Z sea level in a future year. Z known consts are $z_1, \ldots, z_n$ in Z model are called Z **covariates** ∵ Z data x_i vary with Z val of z_i

recall: Z lihood is Z joint density of mass func of Z data: $L(\theta : x) = f_n(\underline{x}; \theta)$ where $\underline{x} = (x_1, \ldots, x_n)$ if Z rV's are indep then Z lihood factorizes as $L(\theta; \underline{x}) = \prod_{i=1}^{n} f(x_i; \theta)$ this is true even if Z density or mass funcs differ for each x_i, & we use Z lihood exactly Z same way to construct pt esti's & CIs, & to conduct hypoth tests

Ex 2.11 in our sea level Ex, Z density func for x_i is

$$f(x_i; \theta) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{(x_i - \alpha - \beta z_i)^2}{2\sigma^2}\right]$$ where $\theta = (\alpha, \beta, \sigma)$ is assume Z annual max sea levels are indep of another ∴ lihood is

$$L(\theta; \underline{x}) = \prod_{i=1}^{n} f(x_i, \theta) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \alpha - \beta z_i)^2\right]$$ Z log-lihood ∴

$$l(\theta; \underline{x}) = -\frac{n}{2}\log(2\pi) - n\log n - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \alpha - \beta z_i)^2$$ & can differentiate to find Z max lihood esti's ∴ $\frac{\partial l}{\partial \alpha} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \alpha - \beta z_i) = \frac{n(\bar{x} - \alpha - \beta\bar{z})}{\sigma^2}$

$\bar{z} = \sum_{i=1}^{n} z_i / n$ $\quad \frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n} z_i(x_i - \alpha - \beta z_i) = \frac{n(\overline{xz} - \alpha\bar{z} - \beta\overline{z^2})}{\sigma^2}$

where $\overline{xz} = \sum_{i=1}^{n} x_i z_i / n$ & $\overline{z^2} = \sum_{i=1}^{n} z_i^2 / n$ & $\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \alpha - \beta z_i)^2$

setting these derivs equal to zero & solving for α, β, σ yields Z max lihood esti's $\hat{\alpha} = \bar{x} - \hat{\beta}\bar{z}$ $\quad \hat{\beta} = \frac{\overline{xz} - \bar{x}\bar{z}}{\overline{z^2} - \bar{z}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2}$ &

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\alpha} - \hat{\beta} z_i)^2$ & Z 2nd derivs are:

$$\frac{\partial^2 l}{\partial \alpha^2} = -\frac{n}{\sigma^2}, \quad \frac{\partial^2 l}{\partial \alpha \partial \beta} = -\frac{n\bar{z}}{\sigma^2}, \quad \frac{\partial^2 l}{\partial \alpha \partial \sigma} = -\frac{2n(\bar{x} - \alpha - \beta\bar{z})}{\sigma^3} \quad \frac{\partial^2 l}{\partial \beta^2} = -\frac{n\overline{z^2}}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial \beta \partial \sigma} = -2n(\overline{xz} - \alpha\bar{z} - \beta\overline{z^2})/\sigma^3 \quad \frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(x_i - \alpha - \beta z_i)^2$$

to take Z expectatns of Z 2nd derivs remember z_i are consts ∴

$E(X_i) = \alpha + \beta z_i$ & $E[(X_i - \alpha - \beta z_i)^2] = \mathrm{var}(X_i) = \sigma^2$ then Z expected info: $I(\theta) = \frac{n}{\sigma^2}\begin{pmatrix} 1 & \bar{z} & 0 \\ \bar{z} & \bar{z^2} & 0 \\ 0 & 0 & 2 \end{pmatrix}$ as earlier in this section, Z asymp distri of Z max lihood esti is $N(\theta, I(\theta)^{-1})$ & such results can still be used to obtain standard errors for Z MLEs, Wald, Score & lihood ratio test stats, & associated CS's eg, Z Wald test is still $W = (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0)$ with asymp null distri χ^2_3 in this Ex ∵ $\theta = (\alpha, \beta, \sigma)$ has three dims & Z corres CS still has Z same formula as in Section 2.3 note: Z expected info in this ex is diagonal if $\bar{z} = 0$, in which case $I(\theta)^{-1} = \frac{\sigma^2}{n}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/S_z^2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}$ where S_z^2 is Z sample var of Z covariates, z_i. this means Z MLEs are asymp uncorrelated. Makes Z inferences simpler & is one reason why covariates are often centred to have mean zero eg if define z_i to be year − 1956 ∴ $\bar{z} = 0$ & Z estied standard errors for $\hat{\alpha} = 120$cm, $\hat{\beta} = 0.57$ cm/year & $\hat{\sigma} = 18$cm are $\hat{\sigma}/\sqrt{n} = 2.6$cm, $\hat{\sigma}/(S_z\sqrt{n}) = 0.17$cm/year & $\hat{\sigma}/\sqrt{2n} = 1.8$cm respectively

Ex 2.12 data two groups ∴ first m data from one distri & rest from another but related distri. $X_1, \ldots, X_m$ indp $\mathrm{Exp}(\theta)$ rv density $f_1(x;\theta) = (1/\theta)\exp(-x/\theta)$ & expectation θ

Let $X_{m+1}, \ldots, X_n$ indp $\mathrm{Exp}(\theta/3)$ rv density func $f_2(x;\theta) = (3/\theta)\exp(-3x/\theta)$ & expectation $\theta/3$ Z lihood: $L(\theta; \underline{x}) = \prod_{i=1}^{m} f_1(x_i;\theta) \prod_{i=m+1}^{n} f_2(x_i;\theta)$

$= \prod_{i=1}^{m} \theta^{-1} e^{-x_i/\theta} \prod_{i=m+1}^{n} 3\theta^{-1} e^{-3x_i/\theta} \propto \theta^{-m} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{m} x_i\right) \theta^{-(n-m)} \exp\left(-\frac{3}{\theta}\sum_{i=m+1}^{n} x_i\right)$

$= \cancel{\exp}\ \theta^{-n} \exp\left[-\frac{1}{\theta}\left(\sum_{i=1}^{m} x_i + 3\sum_{i=m+1}^{n} x_i\right)\right]$ Z log-lihood: ~~l(θ) = const~~

$l(\theta) = \text{const} - n\log\theta - \frac{1}{\theta}\left(\sum_{i=1}^{m} x_i + 3\sum_{i=m+1}^{n} x_i\right)$ with deriv:

$l'(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\left(\sum_{i=1}^{m} x_i + 3\sum_{i=m+1}^{n} x_i\right)$ & solving $l'(\hat{\theta}) = 0$ yields Z max lihood esti $\hat{\theta} = \frac{1}{n}\left(\sum_{i=1}^{m} x_i + 3\sum_{i=m+1}^{n} x_i\right)$ ∴ $l''(\theta) = \frac{n}{\theta^2} - \frac{2}{\theta^3}\left(\sum_{i=1}^{m} x_i + 3\sum_{i=m+1}^{n} x_i\right)$

st $l''(\hat{\theta}) = -n/\hat{\theta}^2 < 0$ & $I(\theta) = -E[l''(\theta)] = -\frac{n}{\theta^2} + \frac{2}{\theta^3}\left[\sum_{i=1}^{m} E(X_i) + 3\sum_{i=m+1}^{n} E(X_i)\right]$

$= -\frac{n}{\theta^2} + \frac{2}{\theta^3}\left[\sum_{i=1}^{m}\theta + 3\sum_{i=m+1}^{n}\frac{\theta}{3}\right] = -\frac{n}{\theta^2} + \frac{2}{\theta^3} n\theta = \frac{n}{\theta^2}$ & $\hat{\theta} \sim N(\theta, \theta^2/n)$

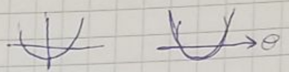
Test stat $-2\log\Lambda$ is χ^2_q where q is degrees freedom for params specified in $H_0: \theta = \theta_0$

\Ex 2.8/ $\hat{\theta} = x/n \quad I(\theta) = n/\{\theta(1-\theta)\}$ ∴

$$\{\theta : (\hat{\theta}-\theta)^T I(\theta)(\hat{\theta}-\theta) < c\} = \left\{\theta : \frac{n(\hat{\theta}-\theta)^2}{\theta(1-\theta)} < c\right\}$$

$$\frac{n(\hat{\theta}-\theta)^2}{\theta(1-\theta)} < c \quad \therefore n(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2) < c\theta(1-\theta) \therefore$$

$$(n+c)\theta^2 - (c+2n\hat{\theta})\theta + n\hat{\theta}^2 < 0 \quad \therefore$$



∴ roots are $\left((c+2n\hat{\theta}) \pm \sqrt{(c+2n\hat{\theta})^2 - 4n\hat{\theta}^2(n+c)}\right) / (2(n+c))$

$$\text{wald test}: \{\theta : (\hat{\theta}-\theta)^T I(\hat{\theta})(\hat{\theta}-\theta) < c\} = \left\{\theta : \frac{n(\hat{\theta}-\theta)^2}{\hat{\theta}(1-\hat{\theta})} < c\right\} =$$

$$\left(\hat{\theta} - \sqrt{\frac{c\hat{\theta}(1-\hat{\theta})}{n}}, \hat{\theta} + \sqrt{\frac{c\hat{\theta}(1-\hat{\theta})}{n}}\right) \therefore$$

$$\frac{n(\hat{\theta}-\theta)^2}{\hat{\theta}(1-\hat{\theta})} < c \quad \therefore (\hat{\theta}-\theta)^2 < \frac{c\hat{\theta}(1-\hat{\theta})}{n} \therefore$$

$$\cancel{-\sqrt{\frac{c\hat{\theta}(1-\hat{\theta})}{n}} < \hat{\theta} - \theta} \quad -\sqrt{\frac{c\hat{\theta}(1-\hat{\theta})}{n}} < \hat{\theta} - \theta < \sqrt{\frac{c\hat{\theta}(1-\hat{\theta})}{n}} \therefore$$

$$\hat{\theta} - \sqrt{\frac{c\hat{\theta}(1-\hat{\theta})}{n}} < \theta < \hat{\theta} + \sqrt{\frac{c\hat{\theta}(1-\hat{\theta})}{n}}$$

Menti:

\1/ $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$ & $H_0: \mu = 0$ which adjective describes H_0? composite since only μ is specified so σ can take a range of vals

\2/ let $\theta = (\alpha, \beta, \gamma, \delta, \varepsilon)$ & $H_0: (\alpha, \beta) = (0, 1)$ what is the asym distr of the LR test stat (linear regression)? Chi(2) eg 2 degrees of freedom since only defining 2 params

\3/ let $W(\mu)$ be the wald test stat & Q_p the p% quantile of Chi(1) which inequality defines a 90% CI for μ? $W(\mu) < Q_{90}$

\Menti/ \1/ let L be a likelihood for params a, b, c how do we find the profile likelihood for a? let p|q denote fix q & maximise L wrt p? ∴ fix a & maximise out b, c ∴ b, c|a

\2/ how find profile likood for b, c? a|b,c

\3/ derived Z profile likood for 2 params of intrest you use Z
profile likood like a standard

\4/ X pdf $f(x;\theta)$ $g(\theta, X)$ be an estimating func for θ
what eqn d we solve to esti θ? $g=0$

\5/ an esti func $g(\theta, X)$ is unbiased if $E[g(\theta,X)]$
equals? 0

\6/ if an esti func is unbiased then Z resulting
of estimators? consistent

\preparatory sheet 2 / \7/ $X_1, \dots, X_n$ iid
$E(X_i) = \theta$ $var(X_i) = \theta^2$ $G = \mu_\theta^T \Sigma^{-1}(X-\mu)$

$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ $\mu = \begin{pmatrix} \theta \\ \vdots \\ \theta \end{pmatrix} = \theta 1$, $\Sigma = \begin{pmatrix} \theta^2 & & 0 \\ & \ddots & \\ 0 & & \theta^2 \end{pmatrix}$ {since X idep} $= \theta^2 I$

$\mu_\theta = 1$, $\Sigma^{-1} = \theta^{-2} I$ $\therefore G = 1^T(\theta^{-2} I)(X - \theta 1) = \theta^{-2} 1^T \begin{pmatrix} X_1 - \theta \\ \vdots \\ X_n - \theta \end{pmatrix} =$

$\theta^{-2} \sum_{i=1}^{n} (X_i - \theta) = \theta^{-2} n(\bar{X} - \theta)$ Z quasi-likood estimator is
$\hat\theta = \bar{X}$ {since set $\theta^{-2} n(\bar{x} - \theta) = 0$ $\therefore (\bar{X} - \hat\theta) = 0 \therefore \hat\theta = \bar{X}$}

$\therefore$ Z asymp distri $\therefore K = \mu_\theta^T \Sigma^{-1} \mu_\theta = 1^T \theta^{-2} I 1 = \theta^{-2} \underbrace{1^T 1}_{n} = \theta^{-2} n = \frac{n}{\theta^2} \therefore$
asymp distri of $\hat\theta$ is $N(\theta, \frac{\theta^2}{n})$

\Ex/ if $X_1, \dots, X_n$ iid $N(\mu, \mu^2)$

\2.5 profile likelihood/ suppose our probab model still has
distri func $F(x;\theta)$ but only intrested in some eles of θ $\therefore$ let
$\theta = (\psi, \eta)$ be a partition into two vecs, ψ & η, suppose only intrested
in ψ $\therefore \eta$ is a nuisance param

can show Z usual likood procedures are effective if we
first replace Z likood with Z profile likood

\Def 2.1/ $L(\psi, \eta; \underline{x})$ be Z likood based on data $\underline{x}$. Z profile
likood for ψ is $L_P(\psi; \underline{x}) = \sup_\eta L(\psi, \eta; \underline{x})$ $\therefore$
$L_P(\psi; \underline{x}) = L(\psi, \hat\eta(\psi); \underline{x})$ where $\hat\eta(\psi)$ is Z val of η that maxs $L(\psi, \eta; \underline{x})$

when ψ is fixed

Z profile l'hood removes Z nuisance param η by replacing it with Z 'best' esti $\hat{\eta}(\psi)$ for each val of ψ. geometrically Z profile l'hood is Z skyline of profile of Z l'hood surface $L(\psi, \eta)$ looked at from Z ψ-axis. Z log-l'hood, score & info are defined: Z profile l'hood: $l_p(\psi; \underline{x}) = \log L_p(\psi; \underline{x})$

Z score is: $U_p(\psi; \underline{x}) = \frac{\partial}{\partial \psi} l_p(\psi; \underline{x})$

Z expected info is $I_p(\psi) = -E\left[\frac{\partial^2}{\partial \psi^2} l_p(\psi; \underline{X})\right]$

\Ex 2.13/ $x_1, \dots, x_n$ $N(\mu, \sigma^2)$ Z param vec is $\theta = (\mu, \sigma^2)$ Z l'hood:

$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right]$

Z log-l'hood: $l(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$

to find Z profile log-l'hood for μ, first find $\hat{\sigma}^2(\mu)$ have:

$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2$ & setting this equal to zero & solving for σ^2 shows that, when μ is fixed $l(\mu, \sigma^2)$ is maxed at

$\hat{\sigma}^2(\mu) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$ $\therefore$ $l_p(\mu) = l(\mu, \hat{\sigma}^2(\mu)) = -\frac{n}{2} \log\left[\frac{2\pi}{n} \sum_{i=1}^{n} (x_i - \mu)^2\right] - \frac{n}{2}$

to find Z profile log-l'hood for σ^2, first find $\hat{\mu}(\sigma^2)$ $\therefore$

$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)$ $\therefore$ when σ^2 is fixed; solving for μ: $l(\mu, \sigma^2)$ maxed at $\hat{\mu}(\sigma^2) = \bar{x}$ $\therefore$ $l_p(\sigma^2) = l(\hat{\mu}(\sigma^2), \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2$

profile l'hoods are not real l'hoods: they do not usually correspond to Z joint density or mass func for a prob model but can use same way a l'hood for use esti's & hypoth tests

first show maximising Z profile l'hood yields Z same esti for ψ that obtain is maximise Z full l'hood

\thm 2.1/ Z profile l'hood $L_p(\psi)$ is maximised by Z maximum l'hood esti of ψ

\proof/ let $(\hat{\psi}, \hat{\eta})$ be Z max l'hood esti of (ψ, η). by def of Z profile l'hood: $L_p(\hat{\psi}) = L(\hat{\psi}, \hat{\eta}(\hat{\psi})) \geq L(\hat{\psi}, \eta)$ $\forall \eta$ $\therefore$

setting $\eta = \hat{\eta}$ yields $L_p(\hat{\psi}) \geq L(\hat{\psi}, \hat{\eta}) \geq L(\psi, \eta)$ $\forall \psi \, \& \, \eta$ ∴
$L_p(\hat{\psi}) \geq L(\psi, \hat{\eta}(\psi)) = L_p(\psi)$ $\forall \psi$ $\square$

\Ex 2.14/ consider Z profile lhoods from (Ex 2.13) have:
$\frac{\partial}{\partial\mu} l_p(\mu) = \frac{n \sum_{i=1}^{n}(x_i - \mu)}{\sum_{i=1}^{n}(x_i-\mu)^2} = \frac{n^2(\bar{x}-\mu)}{\sum_{i=1}^{n}(x_i-\mu)^2}$ & setting this equal to zero &
solving for μ ∴ $l_p(\mu)$ is maxed at $\hat{\mu} = \bar{x}$ &
$\frac{\partial}{\partial\sigma^2} l_p(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{n}{2\sigma^4}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \sigma^2\right]$ & ∴ solving
for σ^2 ∴ $l_p(\sigma^2)$ is maxed at $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$. these are Z
usual maximum lhood ests of μ & σ^2

hypoth tests based of profile lhood: $H_0: \psi = \psi_0$ & $H_1: \psi \neq \psi_0$
define Z profile lhood ratio: $\Lambda_p = \frac{L_p(\psi_0)}{\sup_\psi L_p(\psi)}$ if $\Theta_0 = \{\theta : \psi = \psi_0\}$
where $\theta = (\psi, \eta)$ then $L_p(\psi_0) = \sup_\eta L(\psi_0, \eta) = \sup_{\theta \in \Theta_0} L(\theta)$ while
$\sup_\psi L_p(\psi) = \sup_\psi \sup_\eta L(\psi, \eta) = \sup_\theta L(\theta)$ ∴ $\Lambda_p = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_\theta L(\theta)}$

this equivalence is appropriate ∵ $H_0: \psi = \psi_0$ can be thought of as
Z composite hypoth $H_0: \theta \in \Theta_0$ where $\Theta_0 = \{\theta : \psi = \psi_0\}$. Z lhood
ratio test stat is $-2\log\Lambda_p = -2\left[l_p(\psi_0) - l_p(\hat{\psi})\right]$ where $\hat{\psi}$ is Z MLE
for ψ & its asymp null distri is χ^2_q where q is Z dimension
of ψ. Z test ∴ rejects $H_0: \psi = \psi_0$ in favour of $H_1: \psi \neq \psi_0$
at level α is $-2\log\Lambda_p$ exceeds Z $(1-\alpha)$-quantile of Z χ^2_q distri
\Ex 2.15/ profile log-lhood $l_p(\mu)$ for μ from (Ex 2.13) & Z hypoth
$H_0: \mu = \mu_0$ & $H_1: \mu \neq \mu_0$ ∴ $-2\log\Lambda_p = -2\left[l_p(\mu_0) - l_p(\hat{\mu})\right] =$
$n\left\{\log\left[\sum_{i=1}^{n}(x_i-\mu_0)^2\right] - \log\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]\right\}$

Z profile Wald & score test stats: $W_p = (\hat{\psi}-\psi_0)^T I_p(\hat{\psi})(\hat{\psi}-\psi_0)$
& $S_p = U_p(\psi_0)^T I_p(\psi_0)^{-1} U_p(\psi_0)$ where U_p & I_p defined as before
Z asymp distri of both W_p & S_p is χ^2_q st Z tests reject
$H_0: \psi = \psi_0$ in favour of $H_1: \psi = \psi_0$ at level α is Z test stat
exceeds Z $(1-\alpha)$-quantile of Z χ^2_q distri

construct sets for Ψ in exactly Z same way from Z profile likhood if c denotes Z $(1-\alpha)$-quantile of Z χ^2_q distri then Z (profile)

Wald test of level α has critical Region

$C(\Psi_0)=\{x:(\hat{\Psi}-\Psi_0)^T I_p(\hat{\Psi})(\hat{\Psi}-\Psi_0)\geq c\}$

Z corres $(1-\alpha)$ CS for Ψ is $S(\underline{X})=\{\Psi:(\hat{\Psi}-\Psi)^T I_p(\hat{\Psi})(\hat{\Psi}-\Psi)<c\}$

Z (profile) score test of level α has CR: ~~$C(\Psi_0)=\{\underline{x}: U_p(\Psi_0)^T I(\Psi_0)^{-1} U_p(\Psi_0$~~

$C(\Psi_0)=\{\underline{x}:U_p(\Psi_0)^T I_p(\Psi_0)^{-1} U_p(\Psi_0)\geq c\}$ Z corres $(1-\alpha)$ CS for Ψ is

$S(\underline{X})=\{\Psi:U_p(\Psi)^T I(\Psi)^{-1} \Psi_p(\Psi)<c\}$

Z (profile) likhood ratio test of level α has CR

$C(\Psi_0)=\{x:2[L_p(\hat{\Psi})-L_p(\Psi_0)]\geq c\}$ & Z corres $(1-\alpha)$ CS for Ψ is

$S(\underline{X})=\{\Psi:2[L_p(\hat{\Psi})-L_p(\Psi)]<c\}$

\Ex2.16/ Z likhood ratio test stat from (Ex2.15) an approx α-CS for μ is $\{\mu: n\log[\sum_{i=1}^n (x_i-\mu)^2]-n\log(nS^2)<c\}$ where $S^2=\sum_{i=1}^n (x_i-\bar{x})^2/n$ & c is Z α-quantile of Z χ^2_1 distri. Z inequality is satisfied iff $\sum_{i=1}^n (x_i-\mu)^2<nS^2e^{c/n}$ & $\therefore$ $n\mu^2-2n\bar{x}\mu+(\sum_{i=1}^n x_i-nS^2e^{c/n})<0$ $\therefore$ roots of quadratic when μ lies in interval $(\bar{x}-S\sqrt{e^{c/n}-1}, \bar{x}+S\sqrt{e^{c/n}-1})$ when n is large, $S\approx\sigma$ & $\exp(c/n)\approx 1+c/n$ st this interval approxs Z one in (Ex2.7)

for Z data in (Ex2.3) Z 90% CI for μ is

$(1.25-2.73\sqrt{e^{2.71/30}-1}, 1.25+2.73\sqrt{2^{2.71/30}-1})=(0.41, 2.09)$

\2.8 Quasi-likhood/ let $\underline{X}=(X_1,\dots,X_n)$ have MLE satisfies $U(\hat{\theta};\underline{x})=0$

\Def2.2/ an eqn of Z form $g(\hat{\theta};\underline{X})=0$ that is used to define Z esti $\hat{\theta}$ is an <u>estimating eqn</u> & g is called an <u>estimating func</u>

if θ is a vec then g is also a vec with ith elem $g_i(\theta;\underline{X})$

if no param model available can use estimating eqns to make inferences

\Ex2.17/ $X_1,\dots,X_n$ indep $E(X_i)=\mu$ & $var(X_i)=\sigma^2$ but doesn't write more about their distri $\therefore$ cannot write down a likhood func but consider Z estimating func $g(\theta;\underline{X}) = \begin{pmatrix} \mu - \frac{1}{n}\sum_{i=1}^{n} X_i \\ \sigma^2 + \mu^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{pmatrix}$ where $\theta = (\mu, \sigma)$

$\therefore$ solving $g(\hat{\theta};\underline{X}) = 0$ yield method of moments estimators for μ & σ

since $E(X_i^2) = var(X_i) - E(X_i)^2 = \sigma^2 + \mu^2$

for best estimating func. write $E[g(\theta;\underline{X});\theta]$ for Z expectation of $g(\theta;\underline{X})$ when $\underline{X}$ has a distri with param val θ. Z first requirement is for Z estimating func to be <u>unbiased</u>

\Def 2.3/ an esti func $g(\theta;\underline{X})$ is <u>unbiased</u> if $E[g(\theta;\underline{X});\theta] = 0 \ \forall \theta$

if g is a vec then we require each elem of g to have expectatn zero

\Ex 2.18/ if $g(\theta;\underline{X})$ is Z estimating func for Z Method of Moments, that is Z vec with kth elem $g_k(\theta;\underline{X}) = E(X_1^k) - \sum_{i=1}^{n} X_i^k / n$ $\therefore E[g_k(\theta;\underline{X});\theta] = E(X_1^k) - \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = E(X_1^k) - E(X_1^k) = 0$ & $\therefore g$ is unbiased. Z Score, U is also an unbiased estimating func

an unbiased esti func does not necessarily imply Z resulting estimator is unbiased but does imply Z estimator is consistent, for some regularity conditns. theres usually several unbiased esti funcs, choose the one that yields Z estimator with Z lowest asymp var. suppose specify a model for just, expectatn vars & covars of rv's $X_1, \dots, X_n$ in terms of param θ $\therefore$ $E(\underline{X}) = \mu(\theta) = \mu$ & $var(\underline{X}) = \Sigma(\theta) = \Sigma$, ; θ is Z vec with ith elem θ_i, μ is Z vec with ith elem $\mu_i = E(X_i)$ & Σ is Z matrix with $(i,j)^{th}$ elem $cov(X_i, X_j)$ $\therefore$ consider esti funcs of form $g(\theta;\underline{X}) = c^T(\underline{X} - \mu)$ for some const vec $c = c(\theta)$ & let $\hat{\theta}$ be Z corres estimator that solves $g(\hat{\theta};\underline{X}) = 0$ such esti funcs are unbiased ($\therefore$ yield consistent estimators) $\because E[g(\theta;\underline{X});\theta] = c^T[E(\underline{X}) - \mu] = c^T(\mu - \mu) = 0$

choice of c minimizes Z asymp var of $\hat{\theta}$: is $c^T = \mu_\theta^T \Sigma^{-1}$; $\mu_\theta = \partial\mu / \partial\theta$ is Z matrix with $(i,j)^{th}$ elem $\partial\mu_i / \partial\theta_j$ Z resulting esti func $G(\theta;\underline{X}) = \mu_\theta^T \Sigma^{-1}(\underline{X} - \mu)$ is Z <u>quasi-score-func</u> $\therefore$

it shares some properties with Z score func, U

recall: $E(U)=0$ & $var(U) = -E(\partial U/\partial\theta) = I(\theta)$. for Z quasi-score:

$E(G)=0$ & $var(G) = -E(\partial G/\partial\theta)$ note: $E(G) = \mu_\theta^T \Sigma^{-1}[E(\underline{X})-\mu] =$

$\mu_\theta^T \Sigma^{-1}(\mu-\mu) = 0$ & $\because E(G)=0 \therefore var(G) = E(GG^T) =$

$E[\mu_\theta^T \Sigma^{-1}(\underline{X}-\mu)(\underline{X}-\mu)^T \Sigma^{-1}\mu_\theta] = \mu_\theta^T \Sigma^{-1} E[(\underline{X}-\mu)(\underline{X}-\mu)^T]\Sigma^{-1}\mu_\theta =$

$\mu_\theta^T \Sigma^{-1} var(\underline{X})\Sigma^{-1}\mu_\theta = \mu_\theta^T \Sigma^{-1}\Sigma\Sigma^{-1}\mu_\theta = \mu_\theta^T \Sigma^{-1}\mu_\theta$ &

$\frac{\partial G}{\partial\theta} = \left(\frac{\partial}{\partial\theta}\mu_\theta^T\Sigma^{-1}\right)(\underline{X}-\mu) - \mu_\theta^T\Sigma^{-1}\mu_\theta$ st

$E\left(\frac{\partial G}{\partial\theta}\right) = \left(\frac{\partial}{\partial\theta}\mu_\theta^T\Sigma^{-1}\right)[E(\underline{X})-\mu] - var(G) = -var(G)$ Z esti that

satisfies $G(\hat{\theta};\underline{X})=0$ is known as Z quasi-lihood estimator for θ.

Z asymp distri of Z quasi-lihood esti is $\hat{\theta} \sim N(\theta, K(\theta)^{-1})$:

$K(\theta) = var(G) = \mu_\theta^T\Sigma^{-1}\mu_\theta$ $\therefore$ Z is a probab model specifies

only $E(\underline{X}) = \mu(\theta)$ & $var(\underline{X}) = \Sigma(\theta)$ then Z quasi-lihood estimator,

$\hat{\theta}$, for θ solves ~~for~~ $G(\hat{\theta};\underline{X})=0$ & has asymp distri $N(\theta, K(\theta)^{-1})$

Ex 2.19/ $X_1,\ldots,X_n$ be indep r.v with expectatn $E(X_i)=\theta z_i$ for some

covariate z_i & $var(X_i)=\theta z_i$ $\therefore \mu = \theta(z_1,\ldots,z_n)$ & $\Sigma = \theta\, diag(z_1,\ldots,z_n)$

$\therefore G(\theta;\underline{X}) = \mu_\theta^T\Sigma^{-1}(\underline{X}-\mu) = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}^T \theta^{-1}\begin{pmatrix} z_1^{-1} & & \\ & \ddots & \\ & & z_n^{-1}\end{pmatrix}\left[\begin{pmatrix} X_1 \\ \vdots \\ X_n\end{pmatrix} - \theta\begin{pmatrix} z_1 \\ \vdots \\ z_n\end{pmatrix}\right] =$

$\theta^{-1}\begin{pmatrix} 1 \\ \vdots \\ 1\end{pmatrix}^T\left[\begin{pmatrix} X_1 \\ \vdots \\ X_n\end{pmatrix} - \theta\begin{pmatrix} z_1 \\ \vdots \\ z_n\end{pmatrix}\right] = \theta^{-1}\sum_{i=1}^n X_i - \sum_{i=1}^n z_i$ st Z quasi-lihood estimator

is $\hat{\theta} = \bar{X}/\bar{z}$ & $K(\theta) = \mu_\theta^T\Sigma^{-1}\mu_\theta = \begin{pmatrix} z_1 \\ \vdots \\ z_n\end{pmatrix}^T\theta^{-1}\begin{pmatrix} z_1^{-1} & & \\ & \ddots & \\ & & z_n^{-1}\end{pmatrix}\begin{pmatrix} z_1 \\ \vdots \\ z_n\end{pmatrix} = \frac{n\bar{z}}{\theta}$ st Z

asymp distri of $\hat{\theta}$ is $N(\theta, \theta/(n\bar{z}))$

Z square roots of Z diagonal elems of $K(\hat{\theta})^{-1}$ are estimated standard

errors for Z pt estis in $\hat{\theta}$. Can also use to construct hypoth ~~tests~~

tests & confidence sets eg replace U & I with G & K in defs

of Wald & score test stats & in defs of Z lihood based

confidence sets. Often, $\Sigma(\theta) = \sigma^2\Sigma_0(\theta)$ for a param σ that does

not ~~appear~~ appear in $\mu(\theta)$. in such cases, $\hat{\theta}$ does not provide

an estimate of σ so it has to be estimated seperately. a common

esti is $\hat{\sigma}^2 = \frac{1}{n-p}(\underline{X}-\hat{\mu})^T\hat{\Sigma}_0^{-1}(\underline{X}-\hat{\mu})$, p is Z dimensn of $\hat{\theta}$, $\hat{\mu} = \mu(\hat{\theta})$ &

$\hat{\Sigma}_0 = \Sigma_0(\tilde{\theta})$.

\Ex 2.20/ $X_1, \dots, X_n$ indep expectation $E(X_i) = \alpha + \beta z_i$ for some covariate z_i & var: $var(X_i) = \sigma^2$ $\therefore \mu = (\alpha + \beta z_1, \dots, \alpha + \beta z_n)$ & $\Sigma = \sigma^2 I$ st

$\Sigma_0 = I$ $\therefore$ $\mu_\theta = \begin{pmatrix} \partial\mu_1/\partial\alpha & \partial\mu_1/\partial\beta \\ \vdots & \vdots \\ \partial\mu_n/\partial\alpha & \partial\mu_n/\partial\beta \end{pmatrix} = \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix}$ $\quad \Sigma^{-1} = \sigma^2 I$ &

$G(\theta; \underline{X}) = \mu_\theta^T \Sigma^{-1}(\underline{X} - \mu) = \frac{1}{\sigma^2}\begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix}^T \begin{pmatrix} X_1 - \alpha - \beta z_1 \\ \vdots \\ X_n - \alpha - \beta z_n \end{pmatrix} =$

$\frac{1}{\sigma^2}\begin{pmatrix} \sum_{i=1}^n (X_i - \alpha - \beta z_i) \\ \sum_{i=1}^n z_i(X_i - \alpha - \beta z_i) \end{pmatrix} = \frac{n}{\sigma^2}\begin{pmatrix} \bar{X} - \alpha - \beta\bar{z} \\ \overline{Xz} - \alpha\bar{z} - \beta\overline{z^2} \end{pmatrix}$ solving $G(\hat{\theta}; \underline{X}) = 0$ shows the

quasi-likelihood estimators for α & β are the same as the MLE's in (Ex 2.11)

also $K(\theta) = \mu_\theta^T \Sigma^{-1} \mu_\theta = \frac{1}{\sigma^2}\begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix}^T \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix} = \frac{n}{\sigma^2}\begin{pmatrix} 1 & \bar{z} \\ \bar{z} & \overline{z^2} \end{pmatrix}$ ~~shows in (Ex 2.11)~~ &

~~$\hat{\sigma}^2 = \frac{1}{n-2}(\underline{X} - \hat{\mu})$~~ if $\bar{z} = 0$, the asymp var of $\hat{\theta}$ is $K(\theta)^{-1} = \frac{\sigma^2}{n}\begin{pmatrix} 1 & 0 \\ 0 & 1/S_z^2 \end{pmatrix}$

as in (Ex 2.11) & $\hat{\sigma}^2 = \frac{1}{n-2}(\underline{X} - \hat{\mu})^T \hat{\Sigma}_0^{-1}(\underline{X} - \hat{\mu}) = \frac{1}{n-2}\sum_{i=1}^n (X_i - \hat{\alpha} - \hat{\beta} z_i)^2$ $\therefore$

$\hat{\Sigma}_0 = I$

__Bootstrap:__ ~~X~~ $X = (X_1, \dots, X_n)$ of iid r.v.'s $\sim F$ want properties of

$T = T(X, F)$ eg $T = \hat{\theta} - \theta$

Estimate by properties of $T^* = T(X^*, \hat{F})$ where X^* iid $\hat{F}$.

approx properties of T^* by • simulate x^* from $\hat{F}$

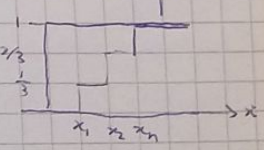
• compute $t^* = T(x^*, \hat{F})$ • repeat many times

the distri of the t^* vals approx the distri of T^*, which esti the

distri of T.

parametric case: X_i iid poi(μ) esti μ by $\bar{x}$ $\therefore$ $\hat{F}$ is poi($\bar{x}$)

non-parametric e.d.f $\hat{F}(x) =$ 'proportion of $(x_1, \dots, x_n)$ less than or equal to x'




```
>> x = c(...)        >> B = 1000        >> n = length(x)
>> theta_hat = 1/mean(x)
>> t_star = numeric(B)
>> x_star = sample(x, replace = True)        >> x_star
>> 1/mean(x_star) - theta_hat        >> for (b in 1:B){...
>> hist(t_star )  >> mean(t_star)  >> sd(t_star)
```

\menti/ \1/ X_1, X_2 be iid $Poi(\mu)$ & $T = \bar{X} - \mu$ & let $x_1 = 1$ & $x_2 = 2$

\2/ if $T^* = A - B$ what is A? $\bar{X}^*$

\3/ what is B? 1.5 {replace μ with $\hat{\mu}$}

\4/ what is the distri of X_1^* & X_2^* for parametric resampling?

$Poi(1.5)$ {>> rpois(1.5)}

>> what distri of X_1^* & X_2^* for nonparametric resampling?

Unif{1, 2}

\5/ let $T^* = \bar{X}^* - 1.5$ for parametric resampling, X_1^*, X_2^* are iid $Poi(1.5)$ what is $Pr(T^* = -1.5)$? e^{-3}

\6/ let $T^* = \bar{X}^* - 1.5$ for nonparametric resampling X_1^*, X_2^* are iid unif{1, 2} what's $Pr(T^* = -1.5)$? 0

$X_1, X_2, \dots, X_n \rightarrow \hat{\theta}$ $\quad X_2, \dots, X_n \rightarrow \hat{\theta}_{-1}$ $\quad \frac{X_2 + \dots + X_n}{n-1} = \frac{1}{n-1} \sum_{j=2}^{n} X_j$

\menti/ \1/ $\hat{\theta} = \sum_{i=1}^{n} x_i$ $\quad n = 3, \; x_1 = 1, \; x_2 = 2, \; x_3 = 3$ $\quad \hat{\theta}_{-2} = ?$

$x_{-2} =$ $\quad x_3 = 3$ $\quad n = 3$ $\quad$ X $\hat{\theta}_{-2} = 4$

take out $x_2 = 2$ $\therefore x_1 + x_3 = 1 + 3 = 4$

\2/ $\hat{\theta} = n^{-1} \sum_{i=1}^{n} x_i$ $\quad n = 3$ $\quad \hat{\theta}_{-2}$? $\quad x_1 = 1 \quad x_2 = 2 \quad x_3 = 3$

$\hat{\theta}_{-2} = \frac{1}{2} (x_1 + x_3) = \frac{1}{2}(1 + 3) = 2 = \hat{\theta}_{-2}$

\3/ $n / \sum_{i=1}^{n} x_i^{-1}$ $\quad n = 3$ $\quad x_1 = 1, \; x_2 = 2, \; x_3 = 3$

$\hat{\theta}_{-2} = 2 / (\frac{1}{1} + \frac{1}{3}) = 2 / (\frac{4}{3}) = \frac{3 \cdot 2}{4} = \frac{3}{2} = \hat{\theta}_{-2}$

\4/ $\hat{\theta} = \prod_{i=1}^{n} x_i^{1/n}$ $\quad n = 3 \; \dots \; \hat{\theta}_{-2}$?

$x_1^{1/2} * x_3^{1/2} = 1^{1/2} * 3^{1/2} = 1 * \sqrt{3} = \sqrt{3}$

>> ~~estimate~~ mymean <- function(x) >> { sum(x)/length(x) >> }

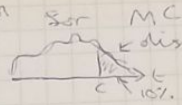
% $\frac{1}{n} \sum_{i=1}^{n} x_i$ % $\quad$ >> x <- c(. . .) $\quad$ >> mymean(x) $\Rightarrow$ 3.5

>> mymean(x[-1]) $\Rightarrow$ 4 $\quad$ >> mymean(x[-2]) $\Rightarrow$ 3.8

\Menti / \5/ $Poi(\mu)$ $\quad H_0: \mu = 1$ $\quad H_1: \mu = 3$ $\quad \hat{\mu} = 2$ $\quad$ MC

test simulates sample from which distri? $Poi(1)$

\6/ $t_1^* < \dots < t_{100}^*$ stat simulated for MCt reject H_0 at 10% level is $t > c$ what is c? t^*_{90}



\7/ iid $N(\mu, \sigma^2)$ $H_0: \sigma = 1$, $H_1: \sigma = 3$ $\bar{x} = 2$ a bootstrap test simulates from which distri? $N(2,1)$

\Section 3/ <u>bootstrap</u>: Consider pt estimati: Suppose Model our data $\underline{x} = (x_1, \dots, x_n)$ as indep r.v. $\underline{X} = (X_1, \dots X_n)$ with common distri func $F(x;\theta)$ & let $\hat{\theta}$ be an estimator for θ. like to know Z sampling distri of $\hat{\theta}$ in to obtain standard errors etc. Can use MonteCarlo simulation to do this, to approx Z sampling distri of $\hat{\theta}$ when Z data are from $F(x;\theta_0)$ & calc Z esti for each sample. these esti's are a sample from Z sampling distri of $\hat{\theta}$ when $\theta = \theta_0$ & can be used to approx properties of $\hat{\theta}$ like its standard error. like to choose θ_0 to be true val of θ for Z situation we modelling as true val is unknown. Can use our esti $\hat{\theta}(\underline{x})$ instead to give following algorithm. for $b = 1, \dots, B$ where B is large number

1) simulate a sample $\underline{x}_b^* = (x_{b1}^*, \dots, x_{bn}^*)$ from $F(x; \hat{\theta}(\underline{x}))$;
2) calc Z esti $\hat{\theta}_b^* = \hat{\theta}(\underline{x}_b^*)$ ∴ our esti of $E(\hat{\theta})$ is Z mean $\bar{\theta}^* = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b^*$

our esti of Z standard error of $\hat{\theta}$ is Z square root of Z variance $B^{-1} \sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\theta}^*)^2$ & our esti of Z p-quantile of $\hat{\theta}$ is Z sample p-quantile $\hat{\theta}^*_{(pB)}$

\Ex 3.1/ $X_1, \dots, X_n$ indep $\mathrm{Exp}(\theta)$ r.v. let $\hat{\theta} = 1/\bar{X}$ data $\underline{x}(x_1, \dots, x_n)$ our pt esti is $\left\{ \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i \right\}$ $\hat{\theta}(\underline{x}) = 1/\bar{x}$ to esti Z standard error of $\hat{\theta}$ we simulate samples from $\mathrm{Exp}(\hat{\theta}(\underline{x}))$, calc Z esti for each sample & compute Z standard deviation of Z estis. Z following code for data Z pt esti is $\hat{\theta} = 1.43$ with estimated standard error 0.55. Z estied sampling distri shown:

```
>> x = c(...)   >> B = 1000   >> n = length(x) %samplesize%
>> theta_hat = 1/mean(x)  %esti theta%   >> theta_star = numeric(B) %vec to store estis%
>> for (b in 1:B) {  %for each b=1,...,B%   >> x_star = rexp(n, theta_hat)
%simulate a new sample%   >> theta_star[b] = 1/mean(x_star) # calc & store estis%   >> }
>> hist(theta_star) %estied samplig distri%   >> sd(theta_star) %estied SE%
```

more generally may want to esti Z sampling distri of some func $T(\underline{X}, F)$ of $\underline{X}$ & F. ~~eg~~ to esti Z bias of $\hat{\theta}$ need to esti Z expectation of $T(\underline{X}, F) = \hat{\theta} - \theta$ is a func of both $\underline{X}$ & F ∵ $\hat{\theta} = \hat{\theta}(\underline{X})$ depends on $\underline{X}$ & $\theta = \theta(F)$ depends on F. Z sampling distri of $T(\underline{X}, F)$ is determined by F: depends on F. Z sampling distri of $T(\underline{X}, F)$ is determined by F: it is both Z 2nd argument of T & Z distri of Z r.v.'s $\underline{X}$ ∵ F is unknown, replace with an esti $\hat{F}$ eg we esti sampling of ~~$\hat{\theta}$ in (Ex3.1)~~ $T(\underline{X}, F)$ by Z sampling distri of $T(\underline{X}^*, \hat{F})$: $\underline{X}^* = (X_1^*, \ldots, X_n^*)$ & $X_1^*, \ldots, X_n^*$ are indep r.v. with distri func $\hat{F}$. ∴ replace F with $\hat{F}$ & replace Z r.v. $\underline{X}$ (which have distri F) with r.v. X^* that have distri $\hat{F}$. eg if $F(x) = F(x; \theta)$ & $\hat{F}(x) = F(x; \hat{\theta}(\underline{x}))$ then we would esti Z expectation of $T(\underline{X}, F) = \hat{\theta}(\underline{X}) - \theta$ by Z expectation of $T(\underline{X}^*, \hat{F}) = \hat{\theta}(X^*) - \hat{\theta}(\underline{x})$. ∵ know $\hat{F}$, can find Z sampling distri distri of $T(\underline{X}^*, \hat{F})$ by simulation ∴ find an esti of Z sampling distri of $T(\underline{X}, F)$. shall write $T = T(\underline{X}, F)$ & $T^* = T(\underline{X}^*, \hat{F})$ hereafter.

this general idea is known as <u>bootstrapping</u> & Z algorithm is: for $b = 1, \ldots, B$ 1): simulate $\underline{x}_b^* = (x_{b1}^*, \ldots, x_{bn}^*)$ from $\hat{F}$;
2): calc $t_b^* = T(x_b^*, \hat{F})$ ∴ $t_1^*, \ldots, t_B^*$ form a sample from Z distri of T^* & can use them to approx properties of T^*. eg Z mean $\bar{t}^* = \frac{1}{B} \sum_{b=1}^{B} t_b^*$ approxs $E(T^*)$ & ∴ estis $E(T)$; Z variance $\frac{1}{B} \sum_{b=1}^{B} (t_b^* - \bar{t}^*)^2$ approxs $var(T^*)$ & ∴ estis $var(T)$; & $t^*_{(pB)}$ approxs Z p-quantile of T^* & ∴ estis Z p-quantile of T. Such estis are called <u>Bootstrap estis</u>. T^* is Z bootstrap version of T, & Z sample $\underline{x}_b^*$ are called <u>bootstrap samples</u> or <u>resamples</u>

\Ex 3.2/ Compute bootstrap esti of Z bias in (Ex 3.1) ∴ bias 0.14 ∴ subtract estied bias from pt esti gives $\hat{\theta} = 1.43$ yields a bias-corrected esti 1.29.

```
>>t_star = numeric(B) %.vec to store stats %.
>>for (b in 1:B) { %.for each b=1,...,B
>>x_star = rexp(n, theta_hat) %.simulating bootstrap sample
>>t_star[b] = 1/mean(x_star) - theta_hat %.calc stat %.
>>}
>>mean(t_star) >>estied bias
```

if we esti F using a param model, $F(x;\hat{\theta}(x))$ like above, then this procedure is called Z parametric bootstrap & Z simulation is Z param bootstrap resampling. alternative is to esti F without assuming a param model. a common choice of esti in this case is Z empirical distri func $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(x_i \le x)$ $x_1, \dots, x_n$ is data. is a discrete distri of which Z probab mass func has probab $1/n$ on each x_i eg $Pr(X_i^* = x_i) = \frac{1}{n}$ for $i=1,\dots,n$ is called a nonparametric esti of F & Z resulting bootstrap procedure is Z nonparam bootstrap. simulating a sample of size n from Z empirical distri func is equiv to sampling at random with replacement n vals from $\{x_1, \dots x_n\}$. this is known as nonparam bootstrap resampling.

\Ex 3.3/ compute nonparam bootstrap esti of Z bias & Z standard error in (Ex 3.1) yield an estied bias of 0.19 & estied SE of 0.67.

```
>>t_star = numeric(B) % vec to store stats >>for (b in 1:B) { % for each
b=1, ... B%   >> x_star = sample(x, replace=TRUE) % simulate bootstrap sample%
>>t_star[b] = 1/mean(x_star) - theta_hat % calc stat% >> }
>>mean(t_star) % estied bias   >> sd(t_star) % estied SE%
```

There are three sources of error in bootstrap estis: model error, simulation error, estimation error. each can sometimes be avoided or reduced. Model error occurs when esti F using a param model, $F(x;\theta)$ but Z true F doesn't belong to this family of distris. Model error can be reduced by ensuring $F(x;\theta)$ is a good model for Z data, or avoided entirely by estimating F nonparametrically. Simulation errors from being unable to simulate all possible samples from $\hat{F}$, that is from using finite B. simulation error can be made as small as need by increasing B. in practice, increase B until results stabilise.

\Ex 3.4/ effect of increasing B on Z bootstrap esti of Z bias in (Ex 3.3) note how bias stabilises as B increases: >>bias = numeric(B)

```
>> for (b in 1:B) bias[b] = mean(t_star[1:b]) % bias from b resamples
```

>> plot(1:B, bias, type = "l", xlab = "B", ylab = "Bias")

Simulation (& simulation error) can be avoided completely if can calc properties of T^* analytically:

Ex 3.5/ $X_1, ..., X_n$ indep iid r.v. $T = \sum_{i=1}^{n} X_i/n$ be an esti for their expectation

Calc Z nonparam bootstrap esti of Z var of T from $x_1, ..., x_n$ data.

we esti Z var of T be Z var of $T^* = \sum_{i=1}^{n} X_i^*/n$: $X_1^*, ..., X_n^*$ is a nonparam bootstrap sample. ∴ $Pr(X_1^* = x_i) = 1/n$ for $i = 1, ..., n$ have:

$E(X_1^*) = \sum_{i=1}^{n} x_i Pr(X_1^* = x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ & $var(X_1^*) = E\{[X_1^* - E(X_1^*)]^2\} = \sum_{i=1}^{n} (x_i - \bar{x})^2 Pr(X_1^* = x_i) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. ∴ Z nonparam bootstrap esti of Z var of T is $var(T^*) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i^*) = \frac{1}{n} var(X_1^*) = \frac{1}{n^2}\sum_{i=1}^{n}(x_i - \bar{x})^2$

estimation error from replacing F by $\hat{F}$ is avoided entirely if T is a pivot.

Suppose F is a param model, then Z distris of T & T^* differ only ∴ of Z difference betw Z (unknown) param val in $F(x;\theta)$ & Z (estied) param val in $\hat{F}(x) = F(x; \hat{\theta}(x))$. if T is a pivot when its distri is indep of Z param val & ∴ T & T^* have Z same distri

Ex 3.6/ $X_1, ..., X_n$ be indep r.v. distri func F & expectation μ

let $\bar{X} = \sum_{i=1}^{n} X_i/n$ & $V = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ if assume F is a normal distri ∴ $T = \frac{\bar{X} - \mu}{\sqrt{V/n}}$ has a Stu(n-1) distri & ∴ is a pivot. if our param esti of F Z $N(\hat{\mu}, \hat{\sigma}^2)$ distri then Z bootstrap version of T is

$T^* = \frac{\bar{X}^* - \hat{\mu}}{\sqrt{V^*/n}}$: $\bar{X}^* = \sum_{i=1}^{n} X_i^*/n$, $V^* = \sum_{i=1}^{n}(X_i^* - \bar{X}^*)^2/(n-1)$ & $X_1^*, ..., X_n^*$ are indep $N(\hat{\mu}, \hat{\sigma}^2)$ r.v. st T^* also is a Stu(n-1) distri

pivots are typically unavailable in Z non-param case & ∴ esti error cannot be avoided entirely. & nonparam estis of F are often less efficient than param estis & ∴ esti error can be greater in Z nonparam case than in Z param case. but its still beneficial to choose T to be a pivot from an approx param model ∵ this should make Z distris of T & T^* as similar as possible.

Ex 3.7/ if dont want to assume F is normal in Z previous ex then our nonparam esti of F is Z empirical distri func, $\hat{F}$, of Z data $x_1, ... x_n$ &

Z bootstrap version of T is $T^* = \frac{\bar{X}^* - \bar{x}}{\sqrt{V^*/n}}$ where $\bar{x}$ is Z expectation of $\hat{F}$ (see Ex 3.5) & $X_1^*, \ldots, X_n^*$ are indep r.v. with distri $\hat{F}$. Z estimation error should be small if Z normal distri is a reasonable ~~expected~~ approx of F ∴ T will be an approx pivot & Z distri of T & T^* will be similar

have described param & non param bootstrap resampling for Z case of indep iid r.v. resampling can be more complicated in Z case of covariate models. $X_1, \ldots, X_n$ be indep r.v. with X_i distri $F_i(x;\theta)$ for $i = 1, \ldots, n$. if $\hat{\theta}(\underline{x})$ is an esti of θ then Z param bootstrap forms resamples $(x_1^*, \ldots, x_n^*)$ by simulating x_i^* from $F_i(x; \hat{\theta}(\underline{x}))$ for $i = 1, \ldots, n$.

\Ex 3.8/ use param bootstrap resampling to esti Z bias & SE of $\hat{\beta}$ in (Ex 2.10) yields $\hat{\beta} = 0.57$ cm/year with estied bias 0.00 cm/year & SE 0.17 cm/year

```
>> x = c(...)  %sealevels%   >> z = (1931:1981) - 1956 % covariates %
>> B = 1000  >> n = length(x) % sample size %
>> beta-hat = cov(x, z) / var(z)  % esti beta %
>> alpha-hat = mean(x) - beta-hat * mean(z)  % esti α %
>> mu-hat = alpha-hat + beta-hat * z  % esti expectations %
>> sigma-hat = sqrt(mean((x - mu-hat)^2))  % esti σ %
>> t-star = numeric(B)  % vec for stats %   >> for (b in 1:B) { % for b = %
>> x_star = rnorm(n, mu-hat, sigma-hat)  % simulate bootstrap sample %
>> beta_star = cov(x-star, z / var(z)  % esti beta %
>> t_star[b] = beta-star - beta-hat  % calc stat   >> }
>> mean(t-star)  % estied bias %   >> sd(t-star)  % estied SE %
```

for non param bootstrap resampling like to have a nonparam esti for each F_i. if Z F_i differ for each i but ∵ ∃ only one data val x_i for each F_i & ∴ Z empirical distri func esti of F_i will be Z distri with probab 1 on x_i ∴ non param bootstrap resamples $(x_1^*, \ldots, x_n^*)$ will be identical to Z original data, T^* will be a const & Z distri of T^* will be a bad esti of Z distri of T.

is $F_i(x;\theta)=F(x;z_i,\theta)$ for a covariate z_i then an alternative nonparam approach is to resample Z vecs $y_i=(z_i,x_i)$ instead of only Z x_i. Z idea here is to define $Y_i=(Z_i,X_i)$ for $i=1,\dots,n$ & to let F be Z common distri func of Z r.v.s $\underline{Y}=(Y_1,\dots,Y_n)$. Z distri of $T(\underline{Y},F)$ is then estied by Z distri of $T(\underline{Y}^*,\hat{F})$ where $\underline{Y}^*=(Y_1^*,\dots,Y_n^*)$ are indep r.v.s with distri $\hat{F}$. if $\hat{F}$ is Z empirical distri func of Z data $y=(y_1,\dots,y_n)$ then resamples $(y_1^*,\dots,y_n^*)$ are formed as before by sampling at random with replacement from $\{y_1,\dots,y_n\}$. this approach is called case resampling & is appropriate when Z covariates z_i are sampled rather than fixed (in previous ex, Z covariates are years & ∴ are fixed. 1931 would still be 1931 if a second set of sea level measurements were available.) there are other, semiparam approaches to bootstrap resampling for covariate models, eg model for sea levels is equiv to assuming $X_i=\alpha+\beta z_i+W_i$; where Z 'errors' W_i are indep $N(0,\sigma^2)$ r.v. Z param bootstrap would esti Z error distri by $N(0,\hat{\sigma}^2)$ & let $X_i^*=\hat{\alpha}+\hat{\beta}z_i+W_i^*$, W_i^* have distri $N(0,\hat{\sigma}^2)$. a simple semiparam bootstrap would differ by esting Z error distri with Z empirical distri func of $(w_1,\dots,w_n)$, $w_i=x_i-\hat{\alpha}-\hat{\beta}z_i$. is called error resampling.

3.2 Jackknife / alternative bootstrap resampling. let $\hat{\theta}$ be an estimator based on a sample of size n. if had several such samples could compute Z esti for each one & plot their hsto to get an idea of Z sampling distri. only have one sample but can use it to create new samples in various ways. Z bootstrap creates new samples by simulating vals from an esti of Z model distri. another way is to take subsamples of Z original data. known as jackknifing. consider Z n subsamples of size $n-1$ formed by omitting each datum in turn & let $\hat{\theta}_{-i}$ be Z esti obtained when Z ith datum is omitted