Statistics

Latent

Former

① You observe ① outcome (e.g. decision)

② Structure of process (assumption)

③ weight of each Former

Unobservable: ① parameter of process (e.g. utility) Exact one, or weight

→ key approach: reverse thinking [Flipping]

② estimation: ① tailor process

② Conditioning: Fixing parameters

③ Dummy coding

e.g. Var(θ) from Var(n) LL

Max likelihood
FOC: each param
e.g. μ, σ^2
& Fisher

Structure (theory) = ① variables
② how connect
③ interaction rule
④ mean, var, median

③ key for calculation ① Quick

② modular → simplify as much possible on each level

③ use previous modules

④ use for i = 2, 3 and induction for complex calc

Fixed effect → e.g. different groups
↳ $y_{ij} = \mu + a_i + e_{ij}$

- Conditioning → means we write something based on another thing (and we know this new thing)

Holm Bonferroni: $\frac{\alpha_e}{k}$ Stop

→ Fail reject

↑ Reject

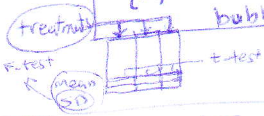
↑ reject

Stop

Fail to reject remaining

Bonferroni: $\frac{\alpha_e}{n}$

ANOVA: Same variance (σ_e)
assump
H_0: nothing going on
H_1: something going on
bubble



→ Reject null when $-2(LL_\omega - LL_\Omega)$ large

≅ $\sum n_i(\bar{x}_{i.} - \bar{x})^2$ is large

$\hat{\mu} = \frac{\sum_i n_i x_{i.}}{n_+}$

$\hat{\sigma}^2 = \frac{\sum_i \sum_j n_i (M_i - \hat{\mu})}{k-1}$

$\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2$

Ω

F-ratio = $\frac{\hat{\sigma}_B^2}{\hat{\sigma}_w^2}$ → Between use mean of each column, & mean of it

→ use all data → $H_0 = H_1$ same σ^2

large = reject

normal .. LL → ANOVA

Contrast vector $\sum c_i = 0$

$H_A: \exists \sum_{i} c_i \mu_i \neq 0$

$n_+ = \sum_i n_i$

$x_{i.} = \frac{\sum x_{ij}}{n_i}$

model: $y_{ij} = \mu + \delta_i + e_{ij}$

$E(\sum_i n_i(\bar{y}_{i.} - \bar{y}_{..}))^2 = \sum n_i \alpha_i^2 + (k-1)\sigma_e^2$

src	df	SS	ms
B/W (σ_B)	$k-1$	$\sum_i n_i(\bar{x}_{i.} - \bar{x})^2$	$SSB/df = \hat{\sigma}_B^2$
within (σ_w)	$\sum_i n_i - k$	$\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2$	$SSW/df = \hat{\sigma}_w^2$
total	$n_+ - 1$	$\sum_i \sum_j (x_{ij} - \bar{x})^2$	

E(ms)
$\sigma^2 + \frac{\sum n_i \alpha_i^2}{n-1}$
$\frac{\sigma^2(n_+ - k)}{(n_+ - k)} = \sigma^2$

Fisher modified LSD (Least significant difference)

$(\bar{x}_{i.} - \bar{x}_{i'.}) \pm t_{\alpha/2} \sqrt{\frac{\hat{\sigma}_w^2}{n_i} + \frac{\hat{\sigma}_w^2}{n_j}}$ df: $\sum_i n_i - k$

$\frac{\sum_i c_i \bar{y}_i}{\sqrt{\sum_i c_i^2 \frac{\hat{\sigma}^2}{n_i}}} \leq \sqrt{(k-1) F_{1-\alpha}(k-1, n_+ - k)}$ $\forall \sum c_i$

$c_i \propto n_i(\bar{y}_{i.} - \bar{y}_{..})$

$\frac{\sum_i n_i(\bar{y}_{i.} - \bar{y}_{..})^2}{\hat{\sigma}^2} \leq (k-1) F_{1-\alpha}(k-1, n_+ - k)$

↳ = Discrim = maximize likelihood of H_A

$y_{ij} = \mu + a_i + e_{ij}$

$e_{ij} \sim N(0, \sigma^2)$

$a_i \sim N(0, \sigma_a^2)$

$Cov(a_i, e_{i'j'}) = 0$

$Cov(e_{ij}, e_{i'j'}) = \begin{cases} 1 & i \neq i', j \neq j' \\ 0 & \text{otherwise} \end{cases}$

$Cov(e_{ij}, e_{i'j'}) = \begin{cases} \sigma_a^2 & \text{if } i \neq i', j \neq j' \\ 0 & \end{cases}$

src	E(ms)
B/W	$\sigma_e^2 + n\sigma_a^2$
within	σ_e^2

define new parameter for new distribution

e.g. ordered & probit: cutoff

Covariate matrix wishart

j = time β_j: fixed
i = person a_i → person, random

$y_{ij} = \mu + a_i + \beta_j + e_{ij}$

$e_{ij} \sim N(0, \sigma^2)$ $Cov(e_{ij}, e_{i'j'}) = \begin{cases} \sigma^2 \\ 0 & \text{other} \end{cases}$

$a_i \sim N(0, \sigma_a^2)$ $Cov(a_i, a_{i'}) = \begin{cases} \sigma_a^2 & i = i' \\ 0 & i \neq i' \end{cases}$

$Cov(y_{ij}, y_{ij'}) = \sigma_a^2$

$V(\bar{y}_{.j} - \bar{y}_{.j'}) = \frac{2\sigma_e^2}{I}$

$(\bar{y}_{.j} - \bar{y}_{.j'}) \pm t_{\alpha/2} \sqrt{\frac{2\hat{\sigma}_e^2}{I}}$

df: $(I-1)(J-1)$

Beta & Drichlet

↳ distribution of P of multinomial & Binomial that themselves have α_i as parameter

src	df	SS	E(ms)
rows (random)	$(I-1)$	$J \sum^I (\bar{y}_{i.} - \bar{y}_{..})^2$	$\sigma_e^2 + J\sigma_a^2$
Column (Fixed)	$(J-1)$	$I \sum^J (\bar{y}_{.j} - \bar{y}_{..})^2$	$\sigma_e^2 + \frac{I}{J-1} \sum \beta_j^2$
Error term $(I-1)(J-1)$	$(I-1)(J-1)$	$\sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	σ_e^2

① when talk about population ⇒ ① talk about mean
↳ or Parameter in interaction: multiply mean

② sleep yesterday, something cool happen today
what button? tools: ① Structure of syst: theory
② what happened today observable = data
③ prior: guess, others (literature) experties

③ distrib fun = structure, machine
Control = parameters (e.g. P of binom → Beta
multinomial → Drichlet

moments: from the time world is created from the mean → variance → kurtosis → population
↳ Could mean moving from distance to population to know it → understand variation = info

Role of multi collin: Relationship does not Exist

Statistics

Considering multiple outcomes ⇑ (set of all outcomes)

- all weighted AVG ⇒ for each outcome
- mean and s^2 is enough most of time
  ↳ parameters of dist if func form clear

like system that has freedom degree
parameters ←

Guitar

parameters (could be mean & variance)
⇓
Calc mean & variance
⇓
Calc variance of mean & variance of variance
⇓
....

$\hat{\theta} = \frac{5x-n}{4n}$ $\quad 2\theta = \ln L(\theta) = \sum_i \delta_i \ln(\frac{1}{5} + \frac{4}{5}\theta) + \sum_i (1-\delta_i)\ln(\frac{4}{5}(1-\theta))$

and from data

$\frac{\partial L(\theta)}{\partial \theta} = 0$

Max Likeli: I see this, so it has highest probability so it has highest prob (weight) ⇒ find params based on model

watching result = outcome of max probability creator $\hat{\theta}$

different language: ① asymptotic approximate
② Error term = variance of limiting dist

Probability power: ① Consider all outcomes
② entity & its complement
Comprehensive systematic view ←

Bays: parameter has distrib (= weight)

- information = variance : shows difference
- independent: Cov = 0 = we have no info
- Var($\bar{x}$) = $E(\frac{1}{n}\sum x_i)^2$ ⇒ you can calculate Var of Var
  on moment you can get moment and on new moment get another moment
  trick: as you power them they will not cancel out each other

- Divide & Conquer: power of Conditioning
⊛ importance of independence in conditioning

$\tilde{x}$ $\tilde{y}$ $\tilde{u}$) & estimation

- Variation comes from model (dist or $y_i = \alpha + \beta x_i + u_i$ & use
⊛ y_i value diff from $P(y_i)$ which is its weight

- estimation: use Tailor series

Standard Error : $\frac{s}{\sqrt{n}} = \sqrt{\frac{\hat{\sigma}^2}{n}}$ ↳ Variance of mean
→ Theory

- Zero mean assumptions

- model ⇒ trick to resolve

Probability world: ① Consider all outcomes
② world is multiplicative
e.g likelihood ←

- Expectation: what I Expect to see from what is unknow

- Statistical models (Beta, Cauchy, Gamma, exponential, Poisson) → Goal → fit data → to stat model
  ↳ Assumptions → src of derivatives
  Variation, and variation of variation

mathematical model ← & not business logic model

Reg u_i, x_i: independent

$|E(\hat{\theta}) - \theta|$ bias
$E(\hat{\theta}-\theta)^2 = V(\hat{\theta}) + Bias^2(\hat{\theta}) = MSE$

- Bayesian Hypothesis is about → mean, variance, moments
  Parameters
  and their distribution ←
  & Covariance

use stat components: relations calculated & proved previously

Likelihood = weight = real likelihood
sample of n
↓
multiply

When single ⇒ have Bias (- or +)
when multiplied to itself ⇒ Convert to distance & become infertile to neutralize others

Components make things Fast & efficient & precise

Conditioning Principles: ① simple: Just on parent & not grand parent at each level, Recursive (order)
② when both subset & set come in condition just keep subset, & attention to independence ⚠
③ set theory: set & Complement [or all subsets]

Bays intuition
⊛ You know something today want to know what happend yesterday, so you reverse

Error term in Reg = Y dist. since its mean is $X\beta$

Stat summary

$i = 1, 2, \dots, I$
$j = 1, 2, \dots, J$
$k = 1, 2, \dots, K$

model

$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$

Constraint: $\sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$

Row, Column, Cell
Row effect
Column effect

Fixed Error

		df
Sum sq rows	$KJ\sum_i(\bar{y}_{i..} - \bar{y}_{...})^2$	$I-1$
SS Col	$KI\sum_j(\bar{y}_{.j.} - \bar{y}_{...})^2$	$J-1$
SS interaction	$K\sum_i\sum_j(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{...})^2$	$(I-1)(J-1)$
SS error	$\sum_i\sum_j\sum_k(y_{ijk} - \bar{y}_{ij.})^2$	$IJ(K-1)$

$\hat{\mu} = \bar{y}_{...}$ $\quad \hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...}$ $\quad \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$ $\quad (\hat{\alpha\beta})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$

highly sensitive to high deg freedom

non central F

Src	E(MS)	Rms
Rows	$\sigma_e^2 + \frac{KJ\sum_i \alpha_i^2}{(I-1)}$	$\frac{Rms}{\text{Err Sumsq}} \sim F((I-1), IJ(K-1))$
Columns	$\sigma_e^2 + \frac{KI\sum_j \beta_j^2}{J-1}$	$\frac{CMS}{\text{Err mean sq}} \sim F(J-1, IJ(K-1))$
interaction	$\sigma_e^2 + \frac{K\sum_i\sum_j(\alpha\beta)_{ij}^2}{(I-1)(J-1)}$	$\frac{\text{Int ms}}{\text{Err}} \sim F((I-1)(J-1), IJ(K-1))$

Model: Random Error $\quad y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk}$ $\quad$ all uncorr

$b_j \sim N(0, \sigma_b^2)$ $\quad a_i \sim N(0, \sigma_a^2)$ $\quad (ab)_{ij} \sim N(0, \sigma_{ab}^2)$ $\quad e_{ijk} \sim N(0, \sigma_e^2)$

Central chi sq

	df	E(MS)	
Rows	$(I-1)$	$\sigma_e^2 + k\sigma_{ab}^2 + Jk\sigma_a^2$	$\frac{\chi^2(I-1)}{I-1}$
Column	$(J-1)$	$\sigma_e^2 + k\sigma_{ab}^2 + Ik\sigma_b^2$	$\frac{\chi^2(J-1)}{J-1}$
Interaction	$(I-1)(J-1)$	$\sigma_e^2 + k\sigma_{ab}^2$	$\frac{\chi^2(I-1)(J-1)}{(I-1)(J-1)}$
Error	$IJ(K-1)$	σ_e^2	$\frac{\chi^2(IJ(k-1))}{IJ(k-1)}$

$H_N: \sigma_{ab}^2 = 0 \Rightarrow$ pooling $\quad H_A: \sigma_{ab}^2 \neq 0$ (bottom up test): by $\frac{MS(Int)}{MS(Error)}$

- Cannot devide error unless fixed

Regression $\quad \frac{y^\lambda - 1}{\lambda} \quad \sqrt{x} \quad \sqrt{y}$ $\quad 0 \le \frac{MRSS}{TSS} \le 1 \Rightarrow 0 \le \underbrace{1 - \frac{MRSS}{TSS}}_{R^2} \le 1$

different plots $\quad \ln x \quad \ln y \quad 1/x \quad 1/y$

$R^2 = 1 - \frac{MRSS}{TSS}$

adjusted $R^2 = 1 - \frac{\frac{RSS}{n-2}}{\frac{TSS}{n-1}}$

intercept only $\quad$ Reg $\quad \Rightarrow 0 \le MRSS \le TSS$

R^2: Prop. of variability in y eleminated by x as predictor

1. outlier removal → (2) then use plot of residual (should be random) scatter $\quad$ NOT Fan
use plot - no influential point should exist

- overflow of computer problem

3 basic steps of Reg
1. backword Elimination
2. add variable forword selection
3. stepwise

- P-Val: unreliable under multicollinearity
- Remove highest p-value untill all lower than .05

$\hat{\sigma}_e$ estimate $\quad \hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2}(n-2)\hat{\sigma}_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$ $\quad$ → value trying predict

Statistic summary ①

- $p(x_i) \geq 0 \quad \sum p(x_i) = 1$ $\quad f(x)$: pdf $\int_{-\infty}^{\infty} f(x)dx = 1$ uncertainty
- heavy tail in business & Economics rather than normal
- discret $P(x \leq a) \neq P(x < a)$ Continuous $P(x \leq a) = P(x < a)$
- $F(a) = P(x \leq a) = \int_{-\infty}^{a} f(x)dx$ $\quad - P(a \leq x \leq b) = F(b) - F(a)$, $f(x) = \frac{\partial F(x)}{\partial x}$

① $\boxed{\text{Param}}$ of distribution: ① location ② Scale ③ Kurtosis : symmetry or asymmetric ④ median : num value sep higher/lower half

heavy tail : platocurtic
light tail : laptocurtic

② Bimodal : Combine two normal

③ Expectation: $E(h(x)) = \int_{-\infty}^{\infty} h(x) f(x) dx$, $\sum_{-\infty}^{\infty} h(x) p(x)$

④ Moment of dist: $\alpha_\alpha = \int_{-\infty}^{\infty} x^\alpha f(x) dx = E(x^\alpha)$

$E(x) = \alpha_1 = \int_{-\infty}^{\infty} x f(x) dx = \mu$

$E(x^2) = \alpha_2 = \int_{-\infty}^{\infty} x^2 f(x) dx = \sigma^2 + \mu^2$

$var(t) = \sigma^2 = E(x^2) - E^2(x)$

⑤ moment general form $\Rightarrow \mu_n = \sum_{k=0}^{m} (-1)^k \binom{m}{k} \alpha_1^k \alpha_{n-k}$

$\mu_2 = \alpha_2 - \alpha_1^2$

$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3$

$\mu_4 = \alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4$

⑥ standard normal: $f(x) = \mu = E(x) = 0$, $E(x^n) = 0$ $\quad x \sim N(0,1)$

$E(x^k) = (k-1)E(x^{k-1})$ for normal

⑦ MGF : $E[e^{tx}] = 1 + \sum_{k=1}^{\infty} \alpha_k \cdot \frac{t^k}{k!}$ not unique

$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\ln x)^2}{2}}$, $f_2(x) = f_1(x)(1 + \sin 2\pi \ln(x))$ e.g

⑧ Charac. func: $E[e^{itx}] = \phi_f(t)$ $\quad i = \sqrt{-1}$ unique $t=0$

$\phi_f(t) = E_f(\cos(tx)) + i E_f(\sin(tx))$ $\quad \alpha_k = \frac{\partial^n}{\partial t^n} \phi_f(t)/i^n$

⑨ standard normal : $E(e^{itx}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{itx - \frac{1}{2}x^2} dx = e^{-t^2/2}$

⑩ Normal $y = \mu + \sigma x$: $E(e^{ity}) = E(e^{t(\mu + \sigma x)}) = e^{t\mu} E_x[e^{t\sigma x}]$

$= e^{t\mu} \phi_x(\sigma t) = e^{t\mu - \frac{\sigma^2 t^2}{2}}$

⑪ Probab. gen. function : $G(t) = E[t^x] = \sum_{n=0}^{\infty} p(x) z^x$

⑫ $z_1, z_2, \ldots, z_p \sim N(0,1) \Rightarrow y = \sum_{i=1}^{p} z_i^2 \sim \chi_p^2$, $f(y) = \frac{1}{2^{p/2}\Gamma(p/2)} e^{-y/2} y^{p/2-1}$ $\quad y > 0$

⑬ $t = \frac{x\sqrt{n}}{\sqrt{\sum_{i=1}^{n} z_i^2}} \sim t$ dist $\quad f(t) = \frac{\Gamma(\frac{p+1}{2})}{\sqrt{p\pi}\,\Gamma(p/2)} (1 + \frac{t^2}{p})^{-\frac{(p+1)}{2}}$ $\quad N > 30 \sim N(0,1)$

⑭ $v \sim \chi_{p_1}^2$, $w \sim \chi_{p_2}^2 \Rightarrow x = \frac{v/p_1}{w/p_2} = \frac{p_2}{p_1} \cdot \frac{v}{w} \Rightarrow x \sim F(p_1, p_2)$ Fisher dist

$P(x) = \frac{(p_1/p_2)^{p_1/2}}{B(p_1/2, p_2/2)} (1 + \frac{p_1}{p_2} x)^{\frac{p_1+p_2}{2}}$ $\quad B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$

⑮ Binomial distribution : $\binom{n}{x} p^x (1-p)^{n-x}$ $\quad \phi(x) = (1 - p + pe^{it})^x$

$\mu_1 = np$ $\quad \gamma_1 = \frac{1-2p}{\sqrt{np(1-p)}} \xrightarrow{n \to \infty} 0$

$\mu_2 = np(1-p)$ $\quad \gamma_2 = 3 + \frac{1 - 6p(1-p)}{np(1-p)} \xrightarrow{n \to \infty} 3$

⑯ poisson distr. $\quad p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\quad x = 0, 1, 2, \ldots, \infty$ $\quad \phi(t) = e^{x(it-1)}$

$\mu_2 = \mu_1 = \lambda$ (no flexibility) $\quad \gamma_1 = \frac{1}{\sqrt{\lambda}}$ $\quad \lambda \to \infty$ $\quad \gamma_2 = 3 + \frac{1}{\lambda}$

$P_0(t) = (1 - \lambda dt)^{t/dt}$ $\quad dt = t/N \xrightarrow{\quad} \infty \Rightarrow \lim_{N \to \infty} (1 - \frac{\lambda t}{N})^N = e^{-\lambda t}$

in interval (small) either 0 or 1

$P_N(t+dt) = P_N(t) P_0(dt) + P_{N-1} P_1 dt = P_N(t)(1 - \lambda t) + P_{N-1}(t)\lambda t$ ②

$\Rightarrow P_n(t+dt) = P(t) - \lambda P_n(t)dt + P_{n-1}(t)\lambda dt \Rightarrow$

$P_n(t + \lambda t) - P(t) = \lambda P_n(t) dt + P_{n-1}(t)\lambda dt \Rightarrow \frac{P_n(t+dt) - P(t)}{dt}$

$= -\lambda P_n(t) + \lambda P_{n-1}(t) = -\lambda[P_{n-1}(t) - P_n(t)] = \frac{dP_n(t)}{t}$

$\Rightarrow P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$

⑰ Cumulant func. $\ln(\phi(t/i))$ $\quad \frac{d^k}{dt^k} \ln(\phi(t/i)) = k_n$

$k_1 = \mu \quad k_2 = \sigma^2 \quad k_3 = \sigma^3 \gamma_1$, $\gamma_2 = \frac{k_4}{\sigma^4} + 3$ (kurtosis)

⑱ $\boxed{\text{Karl Pearson}}$ $f(x) \sim$ Normal $\Rightarrow \frac{f'(x)}{f(x)} = \frac{\mu - x}{\sigma^2}$

diff eq. $\frac{f'(x)}{f(x)} = \frac{m - x}{a + bx + cx^2}$

$\Rightarrow f(x) = c\,(a + bx + cx^2)^{-1/2c} \exp[\frac{(b + 2cm)\tan^{-1}[\frac{b + 2cx}{\sqrt{4ac - b^2}}]}{c\sqrt{4ac - b^2}}]$

if $b = c = 0$, $a > 0 \Rightarrow$ Normal dist.

if $b^2 = 4ac$ $\quad c > 0 \Rightarrow t$-dist

⑲ Gamma dis : $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ $\quad \alpha > 0$ $\quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

Gam func $\quad \alpha \in \mathbb{N} \Rightarrow \Gamma(k) = (k-1)!$

small val $\sim$ lognormal

⑳ $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$: $\hat{\mu}_1 = \frac{\sum x_i}{n}$ $\quad \hat{\mu}_2 = \frac{\sum (x_i - \bar{x})^2}{n}$

㉑ $f'(x) = \frac{d f(x)}{f(x)}$ $\quad \frac{f'(x)}{f(x)} = \frac{m - x}{b_0 + b_1 x + b_2 x^2}$

$m = \frac{-\mu_3(\mu_4 + 3\mu_2^2)}{A}$ $\quad \hat{b}_0 = \frac{-\mu_2 c^4 \mu_2 \mu_4 - 3\mu_3^2}{A}$ $\quad \hat{b}_1 = m$

$\hat{b}_2 = \frac{-(2\mu_2\mu_4 - 3\mu_3^2 - 6\mu_2^3)}{A}$ $\quad A = 10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2$

㉒ three param gamma : $f(x/k, a, \lambda, \alpha) = \frac{(x-a)^{\alpha-1}}{\Gamma(\alpha)\lambda^\alpha} e^{-\frac{(x-a)}{\lambda}}$

$x > a$ $\quad a$: start point x $\quad \alpha$: shape param

λ = scale param $\quad E(x) = \mu_1 = \alpha\lambda$

$var(x) = \mu_2 = \alpha\lambda^2$ $\quad \gamma_1 = \frac{2}{\sqrt{\alpha}}$ (skew) $\quad \gamma_2 = 3 + \frac{6}{\alpha}$ kurt

mode (exist $\alpha > 1$) occurs on $\alpha - 1$

$\ln(x-a)^{\alpha-1} - \frac{(x-a)}{\lambda}) = Const - \frac{(x-a)}{\lambda}) + (\alpha-1)\ln(x-a)$

$\frac{\partial}{\partial x} = \frac{(\alpha-1)}{x-a} - \frac{1}{\lambda} = 0$ $\quad \frac{\partial^2}{\partial x^2} = -\frac{(\alpha-1)}{(x-a)^2}$ $\quad \begin{matrix} a = c \\ \gamma = 1 \end{matrix}$

$\Rightarrow x - \lambda = a + \lambda(\alpha-1) \rightarrow$ mod will occur

$\gamma_1 \to 0 \quad \alpha \to \infty$ sym. $\quad \gamma_2 \to 3 \quad \alpha \to \infty$ large $N \sim$ Normal

$\boxed{a = 2, \lambda = 2}$: χ^2 $\quad \alpha = \frac{df}{2}$ $\quad$ df deg frdm

$f(x) = \frac{1}{\Gamma(\frac{df}{2}) 2^{\frac{df}{2}}} \times x^{\frac{df}{2} - 1} e^{-x/2}$ $\quad \chi^2(df)$

$\boxed{a = 0, \alpha = 1}$ Exp. Dist $\quad f(x) = \frac{e^{-x/\lambda}}{\lambda}$ $\quad f(x) = \theta e^{-\theta x}$

measure of disorder: $-\int \ln f(x) f(x) dx$ $\quad$ no memory

$\boxed{\text{Connection Poisson}}$ $x \sim$ poiss(λP)

$\Rightarrow P(x \geq m) = \sum_{k=m}^{\infty} \frac{e^{-\lambda P}(\lambda P)^k}{k!} = \int_0^{\lambda P} \frac{e^{-z} z^{m-1}}{(m-1)!} = P(z < \lambda P \mid \begin{matrix} Ga \\ \alpha = m \\ \lambda = 1 \end{matrix})$

m: int $\quad$, unify concept

㉓ Beta Dist (Pearson Family) $\quad 0 \xrightarrow{\quad x \quad} 1$ $\quad$ Bays

$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$ $\quad f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$ $\quad 0 \leq x \leq 1$

$a \xrightarrow{\quad} a + h$ $\quad f(x) = \frac{1}{h^{\alpha+\beta}} \frac{(x-a)^{\alpha-1}(h + a - x)^{\beta-1}}{B(\alpha, \beta)}$

max at $a + \frac{h(\alpha-1)}{\alpha + \beta - 2}$ $\quad$ if $\alpha > 1 \quad \beta > 1$

Statistics Summary (3)

$E(x) = a + \frac{\beta \alpha}{\alpha+\beta}$ $\quad V(x) = \frac{h^2 \alpha\beta}{(\alpha+\beta)^2} \frac{1}{(\alpha+\beta+1)}$ $\quad r_1 = \frac{2(\beta-\alpha)}{(\alpha+\beta+2)}\sqrt{\frac{\alpha+\beta+1}{\alpha\beta}}$

$r_2 = \frac{6(\alpha-\beta)^2(\alpha+\beta+1)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)} - \frac{6}{\alpha+\beta+3}$

- Sym
- U shape
- right left sk
- key: finite range

$x \sim F$ dist two chisq div & deg. fredom

$U = \frac{(m/n)^n}{1+(\frac{m}{n})^n} \sim Beta(m/2, n/2)$

Gama ~ poiss
Beta ~ Bin

oddity $U \sim Bin(n,p)$ $\quad \alpha = m,\ \beta = n-m+1$

$P(U \geq m) = P_{beta}(x \leq p)$

(24) Burr: $F(x) = \int_{\infty}^{x} f(z)\,dz$ p.d.f.

$\cdot dF = F(x)(1-F(x)) - g(x)dx \Rightarrow dF\left(\frac{1}{F(x)} + \frac{1}{1-F(x)}\right) = g(x)dx$

$\Rightarrow \ln F(x) - \ln(1-F(x)) = \int_{-\infty}^{x} g(t)dt \Rightarrow \ln F(x) - \ln(1-F(x)) = G(x)$

$\frac{dF}{F(x)(1-F(x))} = g(x)dx \Rightarrow \frac{F}{1-F(x)} = e^{G(x)} \Rightarrow F(x) = \frac{e^{G(x)}}{1+e^{G(x)}} = [1+e^{-G(x)}]^{-1}$

uniform Dist: $F(x) = (1-[1+x^c]^{-k})$ Household US income

$F(x) = (x - \sin\frac{2\pi x}{2\pi})^{\alpha}$ close to Bell shape

(25) Johnson Family $\quad Z \sim Normal \quad Z = \gamma + \delta \ln x$ [log normal dist]

Normal $\sim Z_B = \gamma + \delta \ln\left(\frac{x}{1-x}\right)$ $\quad$ Johnson B Family

$Z_u = \gamma + \delta \sinh^{-1} x$ $\quad$ Curve $\quad \sinh(x) = \frac{e^x - e^{-x}}{2} \quad \cosh(x) = \frac{e^x + e^{-x}}{2}$

U-Form $\quad \sinh^{-1}(x) = \ln(x + \sqrt{x^2+1})$

P.d.f B-form: $f_B(x) = \frac{\delta}{\sqrt{2\pi}} \frac{1}{x(1-x)} e^{-\frac{(\gamma + \delta \log\frac{x}{1-x})^2}{2}} \quad 0<x<1$

P.d.f U-form: $f_U(x) = \frac{\delta}{\sqrt{2\pi}} \frac{1}{\sqrt{x^2+1}} e^{-\frac{(\delta + \delta\ln(x+\sqrt{x^2+1}))^2}{2}} \quad -\infty < x < \infty$

(26) $X \sim$ log Normal $\Rightarrow \ln X \sim N(\mu, \sigma^2)$

(22) $g(x)$ ~ monotone invertible Cont. func $g(\cdot)$ of x

$y = g(x) \quad x = g^{-1}(y) \quad$ Cont: no Jump, invert: one-one

$g \nearrow_x$ (increasing) $\quad F_Y(y) = F_X(g^{-1}(y))$ $\quad$ if $g \searrow^x$ $F_Y(y) = 1 - F_X^{-1}(g(y))$

$f_Y(y) = f_X(g^{-1}(x)) \left|\frac{d}{dy} g^{-1}(y)\right|$

(23) Rand var. $x: F(x) \quad y = e^x, \Rightarrow x = \ln(y)$

$f_Y(y) = f_X(\log y)\left|\frac{d}{dy} g^{-1}(x)\right| = \frac{1}{y} f_X(\ln y)$

$f_Y(y) = \frac{1}{y}\frac{1}{\sqrt{2\pi}\sigma} e^{-1/2\left(\frac{\ln -\mu}{\sigma}\right)^2} \quad 0 < x < \infty \quad$ p.d.f of log normal dist

(24) x r.v cdf: $F(x)$ $\quad y = F(x)$ $\quad$ dist y?

$F_Y(y) = P(Y \leq y) \quad 0 \leq y \leq 1 \quad = P_X(F_X(x) \leq y) = P_X(x \leq F^{-1}(y)) = F_X(F_X^{-1}(y))$

$y \sim$ unif range of F $\quad$ CLT - mean: Gen r.v. $F^{-1}(y) = x$, $F^{-1}(u_i) = x_i$ $\quad$ Simul.

(25) $F(F^{-1}(y)) = x = F^{-1}(F(x)) \Rightarrow \frac{dF^{-1}(F(x))}{dx} = \frac{dx}{dx} = 1 \Rightarrow \frac{dF^{-1}(F(x))}{dF(x)}\frac{dF(x)}{dx} = 1$

$y = F(x) \Rightarrow \frac{dF^{-1}(y)}{dy}\cdot\frac{dF(x)}{dx} = 1 \quad \frac{d(F^{-1}(y))}{dy} = \frac{1}{\frac{dF(x)}{dx}}\Big|_{x = F^{-1}(y)}$

$x \sim$ cdf $F(x)$

$y = F(x),\ x = F^{-1}(y)$, p.d.f(y): $f(F^{-1}(y))\left|\frac{d}{dy}F^{-1}(y)\right| dy = f(x)\frac{1}{\left|\frac{dF(x)}{dx}\right|}$

$= f(x)\frac{1}{f(x)} = 1$

(26) Richard Func $\quad \frac{F'(x)}{F(x)} = -\alpha\frac{(F(x)^{\lambda}-1)}{\lambda} - \beta \Rightarrow F(x) = [1 + \gamma\lambda e^{-\alpha x}]^{-\frac{1}{\lambda}}$ $\quad \gamma>0, \alpha>0, \lambda>0$

$\lim_{\lambda\to 0} F(x) = e^{-\gamma e^{-\alpha x}}$ $\quad$ Gompertz $\quad \lambda = -1$ expon Dist $\quad \lambda = 1$ logistic Dist

(27) Extreme value dist. (4)

CLT: iid → max: ① $F_1(x) = e^{-x^{-\alpha}}$

② $F_2(x) = \begin{cases} e^{-(-x)^{\alpha}} & x<0\ \alpha>0 \\ 1 & x>0 \end{cases}$

Goal: get approx. to real data

(28) Exponential family: $f(y|\underline{\theta})$ $\quad \underline{\theta}$: vect of coeff $\begin{pmatrix}\theta_1\\ \theta_2\\ \vdots \\ \theta_n\end{pmatrix}$

$f(y|\underline{\theta}) = e^{[\Sigma_i S_i(y)T_i(\underline{\theta}) + b(\underline{\theta}) + d(y)]}$

Gamma $\quad f(y) = \frac{\beta^{\alpha} y^{\alpha-1} e^{-\beta y}}{\Gamma(\alpha)} = e^{[(\alpha-1)\ln y - \beta y] + \ln\frac{\beta\alpha}{\Gamma(\alpha)} + 0}$

Normal: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{1/2(x^2 - 2\mu x + \mu^2)} = e^{[-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}x] + [-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{\mu^2}{2\sigma^2}]}$

Poisson: $f(x) = \frac{e^{-\theta}\theta^x}{x!} = e^{x\ln\theta - \theta - \ln x!}$

(29) weibull dist: $F_W(x) = 1 - e^{-\left(\frac{x-\alpha}{\gamma}\right)^k}$, $\quad f_W(x) = \frac{k}{\gamma}\left(\frac{x-\alpha}{\gamma}\right)^{k-1} e^{-\left(\frac{x-\alpha}{\gamma}\right)^k}$

$\mu = \alpha + \gamma\Gamma(1 + 1/k), \quad \sigma^2 = \gamma^2\Gamma(1+\frac{2}{k}) - \mu^2$

$r_1 = \frac{\Gamma(1+3/k)\gamma^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3}, \quad r_2 = \frac{\gamma^4\Gamma(1+\frac{4}{k}) - 3\sigma^4 - 4\Gamma(1+\frac{1}{k})\sigma\mu^2 - 6\mu^2\sigma^2 - \mu^4}{\sigma^4}$

Reliability Control $\sim$ exp

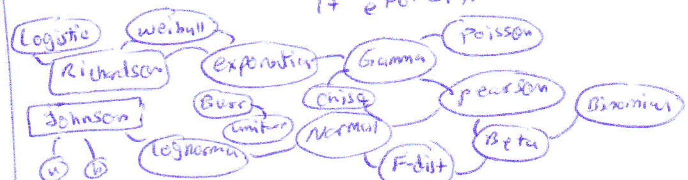
hazard $= \frac{\frac{k}{\gamma}\left(\frac{x-\alpha}{\gamma}\right)^{k-1} e^{\left(\frac{x-\alpha}{\gamma}\right)^k}}{e^{-\left(\frac{x-\alpha}{\gamma}\right)^k}} = \frac{k}{\gamma}\left(\frac{x-\alpha}{\gamma}\right)^{k-1}$

(30) logistic dist: $F(x) = \frac{1}{1+e^{-\left(\frac{x-\mu}{\sigma}\right)}} \quad -\infty<x<\infty \quad = \frac{e^{\left(\frac{x-\mu}{\sigma}\right)}}{1+e^{\left(\frac{x-\mu}{\sigma}\right)}}$

$f(x) = \frac{e^{-\left(\frac{x-\mu}{\sigma}\right)}}{\sigma(1+e^{-\left(\frac{x-\mu}{\sigma}\right)})^2}$ $\quad E(x) = \mu \quad V(x) = \frac{\pi^2\sigma^2}{3}$ Flatter tail

$r_1 = 0 \quad r_2 = 4.2$

dichotomous: $P(x) = \frac{e^{\beta_0 + \Sigma_i \beta_i x_i}}{1 + e^{\beta_0 + \Sigma_i \beta_i x_i}}$



(31) $\Phi_F(t) = E_F(e^{itx}) = \int_{R_x} e^{itx} f(x)dx = E_F(\cos(tx)) + iE_F(\sin(tx))$

if $\int_{-\infty}^{\infty} |\Phi_F(t)|dt < \infty \Rightarrow$ uniq ident. dist: $f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\Phi_F(t)dt$

$y = \alpha + \beta x \quad \Phi_y = e^{i\alpha t}\Phi_x(\beta t)$

$\Sigma_{i=1}^n x_i$, $x_1 \ldots x_n \sim$ indep r.v c.f $\Phi_1(t), \Phi_2(t), \ldots, \Phi_n(t)$

$\Rightarrow$ c.f $(\Sigma_{i=1}^n x_i) = \Pi_{i=1}^n \Phi_i(x)$

(32) $x_1 \ldots x_n \sim$ iid r.v $\Rightarrow \Phi_y(t) = [\Phi_x(t)]^n$

$N(0,1) \Rightarrow \Phi(t) = e^{-t^2/2} \Rightarrow N(\mu, \sigma^2) \quad \Phi(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}$

$y = \Sigma_i x_i \quad \Phi_y(t) = [e^{it\mu - \frac{\sigma^2 t^2}{2}}]^n = e^{itn\mu - \frac{n\sigma^2 t^2}{2}}$

$\Rightarrow y \sim N(n\mu, n\sigma^2)$

x_p^2 d.f $\quad \Phi(t) = (1-2it)^{-p/2} \Rightarrow$

$y = \Sigma_{i=1}^n x_i \Rightarrow \Phi_y(t) = (1-2it)^{-\frac{np}{2}} \Rightarrow y \sim x^2_{np}$

(33) Binom dist. $x_i \sim B(n,p) \quad \Phi(t) = (1-p+pe^{it})^n$

$\Rightarrow y = \Sigma_i x_i$, $x_i \sim Bin(n_i, p) \Rightarrow \Phi_y(t) = (1-p+pe^{it})^{\Sigma_i n_i}$

$\Rightarrow y \sim B(\Sigma n_i, p)$

(34) Poisson: $\Phi(t) = e^{\lambda(e^{it}-1)} \Rightarrow y = \Sigma x_i \quad \Phi_y(t) = e^{n\lambda(e^{it}-1)}$

$\Rightarrow y \sim Poisson(n, \lambda)$

(35) Gamma: $\Phi(t) = \frac{1}{(1-it)^{\alpha}}$ $\quad x_i \sim \Gamma(\alpha) \quad y = \Sigma_i x_i \quad \Phi_y(t) = \frac{1}{(1-it)^{n\alpha}}$

$\Rightarrow \Sigma_i x_i \sim \Gamma(n\alpha)$

Statistics summary

(36) Cauchy Dist: $f(x) = \frac{1}{\pi(1+x^2)}$ $-\infty < x < \infty$ heavy tail

$E(x) = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} d(\ln(1+x^2)) dx = \frac{1}{2\pi} \ln(1+x^2)\Big|_{-\infty}^{\infty} = ?$

not converge ⇒ no mean   median = mode = 0

$\Phi(t) = e^{-|t|}$ $x \sim$ Cauchy $f(x) = \frac{1}{\pi(1+x^2)}$

$y = \frac{1}{x}$ $x = \frac{1}{y}$ $dx = \left|-\frac{1}{y^2}\right| dy \Rightarrow dx = \frac{1}{y^2} dy$

$\Rightarrow f(y) = f(\frac{1}{y}) f(\frac{1}{y^2}) = \frac{1}{\pi(1+\frac{1}{y^2})} \times \frac{1}{y^2} = \frac{y^2 \times \frac{1}{y^2}}{\pi(1+y^2)}$

$\frac{1}{\pi(1+y^2)}$ ⇒ same

heavy tail / overwhelm with tail
invariant to look at $1/x$ or x

(37) Two param Cauchy $f(x) = \frac{1}{\gamma\pi} \frac{1}{[1+(\frac{y-\alpha}{\gamma})^2]}$ α: location param
γ: scale

$F(x) = \frac{1}{\pi} \tan^{-1}(\frac{y-\alpha}{\gamma}) + \frac{1}{2}$, $F^{-1}(p) = \alpha + \gamma \tan[\pi(p-\frac{1}{2})]$

$\Phi(t) = e^{i\alpha t + \gamma|t|}$

(38) $x_1 - x_n \sim$ iid Cauchy (α, γ)

$y = \Sigma y_i \sim \Phi_y(t) = e^{i\alpha n t - n\gamma t} \Rightarrow \Sigma_i y_i \sim$ Cauchy$(\alpha\alpha, n\gamma)$

p.d.f $(\Sigma_i y_i) = \frac{1}{\pi n\gamma} \frac{1}{[1+[\frac{n\bar{y}-n\alpha}{n\gamma}]^2]} = \frac{1}{n\pi\gamma} \frac{1}{[1+(\frac{\bar{y}-\alpha}{\gamma})^2]}$

$\bar{y} = \frac{\Sigma_i y_i}{n}$ $d\bar{y} = d\frac{(\Sigma y_i)}{n}$ $f(\bar{y}) = \frac{1}{\pi\gamma} \frac{1}{[1+(\frac{\bar{y}-\alpha}{\gamma})^2]}$

≡ No CLT   sum exact dist of components
No moment

(39) $x_1 - x_n \sim$ iid r.v $\alpha_j \sim N(\mu_j, \sigma_j^2)$ $\Phi_j(t) = e^{it\mu_j - \sigma_j^2 t^2/2}$

$y = \Sigma x_i$ $\Phi_y(t) = e^{it\Sigma_j \mu_j - \Sigma_j \sigma_j^2 t^2/2} \Rightarrow y \sim N(\Sigma \mu_j, \Sigma_j \sigma_j^2)$

(40) $x_j \sim \chi^2(p_j) \Rightarrow \Phi_{x_j} = (1-2it)^{-p_j/2}$ $y = \Sigma_j x_j$

$\Phi_y(t) = (1-2it)^{-\Sigma_j p_j/2} \Rightarrow y \sim \chi^2(\Sigma_i p_i)$ still chi-sq family

(41) $x_j \sim$ poisson $\Phi_{x_j}(t) = e^{\lambda_j(e^{it}-1)}$

intuition: Quick writing like professor

$y = \Sigma_i x_i \Rightarrow \Phi_y(t) = e^{\Sigma \lambda_i(e^{it}-1)} \Rightarrow y \sim$ poisson$(\Sigma_j \lambda_j)$

(42) Binomial $x_j \sim B(n_j, p_j)$

$\Phi_{x_j}(t) = (1-p_j+p_j e^{it})^{n_j} \Rightarrow y = \Sigma_i x_i \Rightarrow \Phi_y(t) = \Pi(1-p_j+p_j e^{it})^{n_j}$

$\neq$ Binom $\boxed{\text{Not Binomial but close}}$

intuition: when in c.f. things are upstairs mean (exp) then close form

intuition: everything has story behind & to recall it you need to recall story, and story should be intuitive

intuitition: Sum of most r.v are close in their family

(43) Gamm $x_j \sim \Gamma(\alpha_j)$ $\Phi_{\alpha_j}(t) = \frac{1}{(1-it)^{\alpha_j}}$ $y = \Sigma_j x_j$

$\Phi_y(t) = \frac{1}{(1-it)^{\Sigma\alpha_i}} \sim \Gamma(\Sigma_j \alpha_j)$

(44) Cauchy: (α_j, γ_j) $\Phi(t) = e^{\alpha_j it - \gamma_j|t|}$

$y = \Sigma_i x_i \Rightarrow \Phi_y(t) = e^{(\Sigma_j \alpha_j)it - (\Sigma_j \gamma_j)|t|} \Rightarrow y \sim C(\Sigma_j \alpha_j, \Sigma_j \gamma_j)$

intuition: ① characteristic function: key for summation in family
② summation in family first by iid, 2nd by diff param
③ excep Bin   diff param same family still in family

(45) $\Sigma_j \alpha_j x_j \sim$ Dow Jones Indust. Avg.

intuition ① $\Sigma_i x_i$ iid ② $\Sigma_j x_j$ indep ⇒ ③ $\Sigma_j a_j x_j$ indep r.v

(46) $x_j \sim N(\mu_j, \sigma_j^2)$ $a_j x_j \sim N(a_j\mu_j, a_j^2\sigma_j^2)$

$\Phi_{a_j x_j}(t) = e^{it\mu_j a_j - a_j^2 \sigma_j^2 t^2/2}$

$y = \Sigma_j a_j x_j \Rightarrow \Phi_y = e^{it(\Sigma_j a_j \mu_j) - (\Sigma_j a_j^2 \sigma_j^2) t^2}$

$y \sim N(\Sigma_j a_j \mu_j, \Sigma_j a_j^2 \sigma_j^2)$

(47) χ^2 $x_j \sim \chi^2(p_j)$ c.f $(1-a_j it)^{-p_j t}$

$\alpha_j x_j$ is $a_j \chi^2(p_j)$ c.f $(1-a_j it)^{-p_j t/2}$

$y = \Sigma_j a_j x_j = \Pi_{i=1}^{n} (1-2a_j ti)^{-p_j/2} \neq \chi^2$

(48) $H_0: \mu_1 = \mu_2$ $x_1 - x_{n_1}$ $t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ Not t

$H_A: \mu_1 \neq \mu_2$ $x_2 - x_{n_2}$

need assump: $\sigma_1^2 = \sigma_2^2$ pooled t-test   (since t special case of χ^2)

(49) Gamma dist: $x_j \sim \Gamma(\alpha_j) \frac{1}{(1-it)^{\alpha_j}}$

$a_j x_j \sim a_j \Gamma(\alpha_j) \frac{1}{(1-ia_j t)^{\alpha_j}}$ $y = \Sigma_j a_j x_j$

$\Pi_{i=1}^{n} \frac{1}{(1-ia_j t)^{\alpha_j}} \neq \Gamma$

(50) Cauchy $x_j \sim C(\alpha_j, \gamma_j)$ $e^{i\alpha_j t - \gamma_j|t|}$

$a_j x_j \sim a_j C(\alpha_j, \gamma_j)$: $e^{i a_j \alpha_j t - a_j \gamma|t|}$

$y = \Sigma_j a_j x_j$ $e^{i(\Sigma_j a_j \alpha_j)t - (\Sigma_j a_j \gamma_j)|t|}$ ~ Cauchy like Normal

intuition: when weighted sum of r.v of diff param is calculated the result only in same family for Cauchy & Normal

(51) class of Dist: Class of Stable Dist.

r.v y_1, y_2 $y_1 = \alpha_1 + \gamma_1 x_1$ $e^{i\alpha_1 t}\Phi_{x_1}(\gamma_1 t)$

$y_2 = \alpha_2 + \gamma_2 x_2$ $e^{i\alpha_2 t}\Phi_{x_2}(\gamma_2 t)$

c.f $(y_1 + y_2) = e^{i(\alpha_1+\alpha_2)t} \Phi_{x_1}(\gamma_1 t)\Phi_{x_2}(\gamma_2 t)$

$\exists c, d$: $e^{i(\alpha_1+\alpha_2)t}\Phi_{x_1}(\gamma_1 t)\Phi_{x_2}(\gamma_2 t) = e^{idt}\Phi(ct) \Rightarrow$ Stable dist

(52) Stable dist:

① Normal: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ $-\infty < x < \infty$ $\alpha = 2$

② Cauchy: $f(x) = \frac{1}{\pi} \frac{\gamma(x-\alpha)}{\gamma^2 + (x-\alpha)^2}$ $\alpha < x < \infty$ $\alpha = 0$

③ levy: $f(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{1}{(x-\alpha)^{3/2}} e^{-\frac{\gamma}{2(x-\alpha)}}$ $\alpha = -\frac{1}{2}$
$\alpha < x < \infty$

(53) charact func:
$\Phi_N(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$ ~ Normal   (gets moment only)

$\Phi_c(t) = e^{i\alpha t - \gamma|t|}$ ~ Cauchy

$\Phi_l(t) = e^{i\alpha t - \sqrt{-2i\alpha t}}$ ~ levy

(54) Multivar: $x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ $F(x_1, x_2, \dots, x_p) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_p} f(x)dx$

Can't order

$f(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_2$ $f(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_1$

marginal dist   Conditional dist ≡ Cut sand pile

Statistics summary

(2)

$f(x_2|x_1) = \frac{f(x_1,x_2)}{f(x_1)}$

(55) MVN: Multivar Normal: $x_1 \ldots x_p$ $x_j \sim N(\xi_j, \tau_j^2)$ indep

$\text{p.d.f}(x) = \prod_{j=1}^{m} f_j(x_j) = \frac{1}{(2\pi)^{p/2} \prod_{j=1}^{t} \tau_j} e^{-1/2 \sum_j \left(\frac{x_j - \xi_j}{\tau_j}\right)^2}, \tau_i^2 > 0$

$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$ $\underset{p\times p}{D_{\tau^2}} = \begin{pmatrix} \tau_1^2 & & 0 \\ & \tau_2^2 & \\ 0 & & \ddots \ \tau_p^2 \end{pmatrix}$ $D_{\tau^2}^{-1} = \begin{pmatrix} 1/\tau_1^2 & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\tau_p^2} \end{pmatrix}$

$|D_\lambda| = \prod_{i=1}^{p} \lambda_i$ $\sqrt{D_\lambda} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \sqrt{\lambda_2} & \\ 0 & & \ddots \ \sqrt{\lambda_p} \end{pmatrix}$ $\lambda_i > 0$ $\sqrt{D_\lambda}\sqrt{D_\lambda} = D_\lambda$

$\Rightarrow \text{pdf}(x) = \frac{1}{(2\pi)^{p/2} |D_{\tau^2}|^{1/2}} e^{-1/2 (x-\xi)' D_{\tau^2}^{-1} (x-\xi)}$

↳ Quadratic form $x'Ax = \sum_i \sum_j x_i a_{ij} x_j$

$x' = (x_1, x_2, \ldots, x_p)$ $A = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ & \ddots & \\ a_{p1} & a_{p2} \cdots & a_{pp} \end{pmatrix}$

intuition: Generalize From independent Case

$\text{pdf}(x) = \frac{1}{(2\pi)^{p/2} |D_{\tau^2}|^{1/2}} e^{-1/2 (x-\xi)' D_{\tau^2}^{-1} (x-\xi)}$

intuition: One orthogonal transform

$y_i = \sum_{j=1}^{p} \Gamma_{ij} x_j$ $i = 1, 2, \ldots$ Γ orthogonal: $\Gamma\Gamma' = \Gamma'\Gamma = I$

$y = \Gamma x \Rightarrow x = \Gamma' y$ $E(y) = \Gamma E(x) = \Gamma \xi = \mu$

$\text{Cov}(y) = \Gamma D_{\tau^2} \Gamma' = \Sigma$ $\Sigma^{-1} = \Gamma D_{1/\tau^2} \Gamma'$ $\Sigma^{-1}\Sigma = \Gamma \underbrace{D_{1/\tau^2} \overbrace{\Gamma'\Gamma}^{I} D_{\tau^2}}_{I} \Gamma' = I$

inverse make

$\xi = \Gamma' \mu \Rightarrow f(y) = \frac{1}{(2\pi)^{p/2} \underbrace{|\Gamma D_{\tau^2} \Gamma'|^{1/2}}_{\Sigma}} \exp(-\tfrac{1}{2} (\Gamma' y - \Gamma' \mu)' D_{\tau^2}^{-1} (\Gamma' y - \Gamma' \mu))$

$\underbrace{(y-\mu)' \Gamma D_{\tau^2}^{-1} \Gamma' (y-\mu)}_{\Sigma^{-1}}$

$\frac{dx}{dy} = |\Gamma'| = 1 \Rightarrow f(y) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp(-\tfrac{1}{2}(y-\mu)' \Sigma^{-1} (y-\mu))$

- Factor analysis $y \sim MVN \Rightarrow \exists x_1 \ldots x_p$ Nuncor

Rotate in space will not change anything: Verimax

(56) properts of multivar: $\text{Cov}(y_i, y_j) = \sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$

$\text{Cov}(Ay, By) = A \Sigma B'$

$[k\times p][p\times 1]$ $[k\times p][p\times 1]$

$\Phi_{MVN}(t) = e^{i t' \mu^{(p\times 1)} - 1/2 t' \Sigma t}$

$\underset{[k\times p][p\times 1]}{A y} \sim MVN(A\mu, A\Sigma A')$

(57) Multivar Dist $n_1, n_2 \cdots, n_k$ $\underset{p_1}{\overset{n_1}{CAT_1}}, \underset{p_2}{\overset{n_2}{CAT_2}}, \ldots, \underset{p_k}{\overset{n_k}{CAT_k}}$ $\sum n_i = n$ $\sum_{}^{n} p_i = 1$

$\Rightarrow p(n_1, \ldots, n_k) = \frac{n!}{\prod_{i=1}^{k} n_i!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$ $E(n_i) = np_i$ $V(n_i) = np_i(1-p_i)$ $\text{Cov}(n_i, n_j) = -np_ip_j$

(58) Drichlet Dist: Generalize of Beta dist

observe $x_1, x_2, \ldots, x_n$ $x_i \geq 0$ $\sum_{i=1}^{n} x_i = 1$

$P(x_1, x_2, \ldots, x_k) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} x_i^{\alpha_i - 1}$ Bayesian Binom prior → Drichlet

$E(x_i) = \frac{\alpha_i}{\sum_{i=1}^{k} \alpha_i}$ $\text{Var}(x_i) = \frac{\alpha_i(\alpha_+ - \alpha_i)}{\alpha_+^2(\alpha_+ + 1)}$ $\alpha_+ = \sum_{i=1}^{n} \alpha_i$

$\text{Cov}(x_i, x_j) = \frac{-\alpha_i \alpha_j}{(\alpha_+)^2(\alpha_+ + 1)}$ $\underset{\text{mode}}{x_{md}} = \left\{ \frac{\alpha_i - 1}{\alpha_+ - k} \right\}$ $\alpha_i > 1$

if $Y_1, Y_2, \ldots, Y_k \sim \Gamma(\alpha_i, 1) \Rightarrow \frac{Y_1}{\sum_i Y_i}, \frac{Y_2}{\sum_i Y_2}, \ldots, \frac{Y_i}{\sum Y_i}$ ~ drichlet

intuition: Selection of time

(8)

(59)

① Convergence $\text{pdf}: F_1, F_2, \ldots, F_n$

$S_1, S_2, \cdots, S_n$ Seq r.v

S_i AVG i gamma dist r.v

$S_n \xrightarrow{P} S$ $\forall \varepsilon > 0$ $p(|S_n - S| > \varepsilon) \xrightarrow[n\to\infty]{} 0$

② measure theory $S_1, S_2, \cdots, S_n$

$S_n \xrightarrow{a.s} S$ $P(\{S_n\}$ don't Converge$) = 0$

③ $S_1, \ldots S_n \xrightarrow{a.s} S \Rightarrow S_1, \ldots, S_n \xrightarrow{P} S$

(60) WLLN $x_1 \ldots x_n$ iid r.v $\exists E(x_i)$

$\bar{x}_n = \sum_i \frac{x_i}{n}$ $(S_n = \bar{x}_n) \Rightarrow \bar{x}_n \xrightarrow{P} \mu$

$S = cte$, $S_n \xrightarrow{P} S \Rightarrow h(S_n) \xrightarrow{P} h(S)$ h: Cont func

(61) I_j: r.v $\begin{cases} I_j = 1 & \pi \\ I_j = 0 & 1-\pi \end{cases}$ $E(I_j) = \pi$ $V(I_j) = \pi(1-\pi)$

$\hat{\pi}_n \to \sum_{j=1}^{n} I_j/n$ $\hat{\pi}_m \xrightarrow{P} \pi$ $\text{var}(\hat{\pi}_n) = \frac{\pi(1-\pi)}{n}$

$\frac{\widehat{\pi(1-\pi)}}{n}\Big|_m = \frac{\hat{\pi}_m(1-\hat{\pi}_m)}{m}$ $(1-x)x$ Cont.

$\Rightarrow \frac{\hat{\pi}(1-\hat{\pi})}{n} \xrightarrow{P} \frac{\pi(1-\pi)}{n}$

(62) S_n estimator(θ) Consistent if $S_n \xrightarrow[n\to\infty]{P} \theta$ $\forall\theta$

≡ unbiased as $n \to \infty$ $E(S_n|\theta) \underset{n\to\infty}{\text{bias}} = 0$

(63) Rand sample $x_1, x_2, \ldots, x_n$ $E(x_i) = \mu$ $V(x_i) = \sigma^2$

$\hat{\mu}^2 = \bar{x}^2$ $\text{var}(\bar{x}) = E(\bar{x}^2) - E^2(\bar{x}) \Rightarrow E(\bar{x}^2) = V(\bar{x}) + E^2(\bar{x})$

$E(\bar{x}^2) = \mu^2 + \boxed{\frac{\sigma^2}{n}}$ → Biased asymptotically unbiased $(n\to\infty)$

(64) $S_1, S_2, \ldots, S_n$ Converge in dist to S

$S_n \xrightarrow{d} S \Rightarrow p(S_n \leq t) \xrightarrow[n\to\infty]{} p(S \leq t)$

if $E(|S_n - E(S_n)|^2) < \infty \Rightarrow$ Speed of Con: $\frac{1}{\sqrt{n}}$

(65) CLT (Central limit theorem)

$x_1 \ldots x_n$ iid $(\mu, \sigma^2) \Rightarrow \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} \Rightarrow \bar{x} \xrightarrow{d} N(\mu, \frac{\sigma^2}{n})$

if $E(|x_i - \mu|^2) < \infty \Rightarrow$ Convergence with speed $\frac{1}{\sqrt{n}}$

proof $\sum(x_i) \sim [\Phi_x(t)]^n$

$y = \alpha + \beta x \Rightarrow y: e^{\alpha} \Phi_x(\beta t)$

$E(x) = 0, V(x) = 1 \Rightarrow \Phi_x(t) = \sum_{j=1}^{\infty} \frac{a_j(t)}{j!\, i^j} = 1 + \frac{0 t}{1 \cdot j} + \frac{1 \cdot t^2}{2 j^2} + o(t^2)$

$O(g(t))$ if $\lim_{t\to\infty} \frac{f(t)}{g(t)} = 0$

$f(t): O(g(t))$ if $\lim_{t\to\infty} \frac{f(t)}{g(t)} = k \neq 0$

$\Phi_x(t) = [1 - \frac{t^2}{2} + o(t^2)]$

$\Rightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}}{\sigma} \sum_{i=1}^{n} \frac{1}{n}(x_i - \mu) = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \left(\frac{x_i - \mu}{\sigma}\right)$

$E\left(\frac{x_i - \mu}{\sigma}\right) = 0$ $V\left(\frac{x_i - \mu}{\sigma}\right) = 1$

$\Phi_{\left(\frac{x_i - \mu}{\sigma}\right)}(t) = [1 - t^2/2 + o(t^2)]$

c.f: $\frac{1}{\sqrt{n}}\left(\frac{x_i - \mu}{\sigma}\right)$ $[1 - \frac{t^2}{2n} + o(\frac{t^2}{n})]$

$\frac{t}{\sqrt{n}} \to t \Rightarrow \sum_i \frac{1}{\sqrt{n}} \left(\frac{x_i - \mu}{\sigma^2}\right)$ $[1 - \frac{t^2}{2n} + o(\frac{t^2}{n})]^n$

$\simeq \lim_{n\to\infty} [1 - \frac{t^2}{2n}]^n = e^{-t^2/2}$

(66) Converge in mean: look at $|S_n - S|^r$ $E(x_i) = \mu_i$ & $V(x_i) = \sigma_i^2$

$\lim_{n\to\infty} E(|S_n - S|^r) = 0$ $S_n \xrightarrow{L^r} S$ $S_n^2 = \sum_{i=1}^{n} \sigma_i^2$

$r_n = \sum_{i=1}^{n} E(|x_i - \mu|^r)$

$\lim_{n\to\infty} \frac{r_n}{S_n} \to 0 \Rightarrow \sum_i x_i \xrightarrow{L} x \sim N(\sum_i \mu_i, \sum_i \sigma_i^2)$

statistics summary

(67) $E(x_i) = \mu_i = \pi_i$ $\quad x_i = \begin{cases} 1 & \pi_i \\ 0 & 1-\pi_i \end{cases}$

$\hookrightarrow B(1, \pi_i)$

$V(x_i) = \sigma_i^2 = \pi_i(1-\pi_i)$

$E(|x_i - \mu_i|^3) = (1-\pi_i)^3 \pi_i + |(0-\pi_i)^3| (1-\pi_i) = (1-\pi_i)^3 \pi_i + \pi_i^3 (1-\pi_i)$

$= \pi_i(1-\pi_i)[(1-\pi_i)^2 + \pi_i^2] \leq \pi_i(1-\pi_i)$

intuition: use higher moment to calculate

$r_n = [\sum_{i=1}^n \pi_i(1-\pi_i)[(1-\pi_i)^2 + \pi_i^2]]^{1/3}$

$\frac{r_n}{s_n} \leq \frac{[\sum_i \pi_i(1-\pi_i)]^{1/3}}{[\sum_i \pi_i(1-\pi_i)]^{1/2}} = \frac{1}{[\sum_i \pi_i(1-\pi_i)]^{1/6}} \to 0$

(68) application: sample $x_1 \dots x_n$

$I_i(y) = \begin{cases} 1 & \text{if } x_i \leq y \\ 0 & \text{otherwise} \end{cases}$ $\quad \hat{F}_n(y) = \frac{\sum_{i=1}^n I_i(y)}{n}$ $\quad \pi = F(y)$

Bernoulli

$\Rightarrow \hat{F}_n(y) \xrightarrow{L} N(F(y), \frac{F(y)(1-F(y))}{n})$

(69) $\boxed{\text{Generalized CLT:}}$ $x_1 \dots x_n$ iid rv $\exists$ seq. a_n, b_n

and non degenerated rv Z: $a_n(\sum_i x_i) - b_n \xrightarrow{d} Z$, Z: α stable, $0 < \alpha \leq 2$

charac func stable distr: $\Phi_Z = e^{it\mu - |ct|^\alpha (1 - i\beta \operatorname{sgn}(t) \Phi)}$

(Cauchy α ..)

$\Phi = \tan(\pi \alpha/2)$ if $\alpha \neq 1$

$= -(2/\pi) \ln|t|$ $\quad \alpha = 1$

c: scale

α: shape

μ: shift

β: skewness

tail magnitude $= \frac{1}{|x|^{\alpha+1}}$

(70) have stat: $\frac{(x_n - \mu)}{\sqrt{V(x_n)}} \to Z$ $\quad n\operatorname{Var}(x_n) \xrightarrow{L} \tau^2 > 0$

(71) $h(x)$: ① cont. ② at zero derivative ($h'(\mu)$) not zero

$\Rightarrow h(x_n) \xrightarrow{d} N(h(\mu), \operatorname{Var}(x_n)[h'(\mu)]^2)$

(72) $\bar{y} \sim N(\mu, \frac{\sigma^2}{n})$ $\quad h(\mu) = e^\mu$ $\quad h'(\mu) = e^\mu > 0$

$h(\bar{y}) \xrightarrow{d} N(e^\mu, \frac{\sigma^2}{n} e^{2\mu})$

$\boxed{\text{log Normal}}$ $\mu_i = e^{\mu + \sigma^2/2n}$ $\quad \operatorname{Var}_i = (e^{\frac{\sigma^2}{n}} - 1) e^{2\mu + \frac{\sigma^2}{n}}$

$e^{\sigma^2/n} - 1 = \frac{\sigma^2}{n} + \frac{\sigma^4}{2n} + \dots = \frac{\sigma^2}{n} + O(\frac{1}{n})$

(73) $\underset{[p\times1]}{x} \sim MVN(\underset{[p\times1]}{\mu}, \underset{[p\times p]}{\Sigma})$ $\quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$

$h(x)$ cont. func of x s.t $\frac{\partial h(x)}{\partial x_i}\Big|_\mu \neq 0$ $\quad i = 1, 2, \dots, p$

$\Rightarrow h'(\mu) = \begin{bmatrix} \frac{\partial h(x)}{\partial x_1} \times \frac{1}{\Sigma} \\ \vdots \end{bmatrix} \Rightarrow h(x) \xrightarrow{d} N(h(\mu), \underset{[1\times p]}{[h'(\mu)]} \underset{[1\times p]}{\Sigma} \underset{[p\times1]}{h'(\mu)})$

(74) Propagation of Error:

$\binom{x}{y}$ $\quad E(x) = \mu_1$ $\quad E(y) = \mu_2$ $\quad \operatorname{Corr}(x, y) = \rho$

$V(x) = \sigma_1^2$ $\quad V(y) = \sigma_2^2$ $\quad h(\mu) = \frac{\mu_2}{\mu_1}$

$\frac{\partial h(\mu)}{\partial \mu_1} = -\frac{\mu_2}{\mu_1^2}$ $\quad \frac{\partial h(\mu)}{\partial \mu_2} = \frac{1}{\mu_1}$

$\frac{\hat{\mu}_2}{\hat{\mu}_1} = \frac{\hat{\mu}_1}{\hat{\mu}_2} = \frac{\bar{y}}{\bar{x}}$ $\quad \operatorname{Cov}\binom{\bar{x}}{\bar{y}} = \frac{1}{n}\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

$\Rightarrow \frac{\bar{y}}{\bar{x}} \xrightarrow{d} N(\frac{\mu_2}{\mu_1}, [-\frac{\mu_2}{\mu_1}, \frac{1}{\mu_1}] \frac{1}{n} \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} -\mu_2/\mu_1^2 \\ 1/\mu_1 \end{pmatrix}]$

$= \frac{1}{n\mu_1^2}[(\frac{\mu_2}{\mu_1})^2 \sigma_1^2 - 2\rho\sigma_1\sigma_2(\frac{\mu_2}{\mu_1}) + \sigma_2^2]$

$\widehat{\operatorname{Var}}(\bar{y}/\bar{x}) = \frac{1}{n\bar{x}^2}[(\frac{\bar{y}}{\bar{x}})^2 S_1^2 - 2\rho S_1 S_2 (\frac{\bar{y}}{\bar{x}}) + S_2^2]$

(75) $\boxed{\text{Randomization \& Simulation}}$ pseudo random

mathematica $\rightarrow$ mersenne Twister

(76) $F(y) \sim U(0,1)$ $\quad F^{-1}(u)$

(77) Exponential Distribution $\quad u \sim \operatorname{unif}(0,1)$

$F(x) = 1 - e^{-x/\lambda} = u$, $\quad e^{-x/\lambda} = 1 - u \Rightarrow x = \lambda \ln(1-x)$

(78) Cauchy distribution: $F(x) = \frac{\gamma}{\pi(\gamma^2 + (x-\alpha)^2)}$

$F(x) = \frac{1}{\pi} \tan^{-1}(\frac{x-\alpha}{\gamma}) + 1/2 = u$

$\Rightarrow x = \gamma \tan[\pi(\alpha - 1/2)] + \alpha$

(79) weibull: $F(x) = 1 - e^{-(\frac{x-\alpha}{k})^k} \Rightarrow x = \alpha + \gamma(-\ln u)^{1/k}$

(80) $\boxed{\text{Box muller}}$ Gen $u_1, u_2 \sim U(0,1)$ indep

pseudo rand if little dep exist

$\begin{cases} x_1 = \mu + \sigma\sqrt{-2\ln u_1} \cos(2\pi u_2) \\ x_2 = \mu + \sigma\sqrt{-2\ln u_1} \sin(2\pi u_2) \end{cases}$ $\quad Z_1, Z_2 \dots Z_n \sim N(0,1)$ Normal

simulate χ_n^2: $\sum_{i=1}^n z_i^2$

levy: $x = \frac{\gamma}{z^2} + \alpha$ $\quad Z \sim N(0,1)$

Gamma k deg f: Gamma $= \sum_{i=1}^k$ indep Expon.

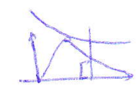
$u_1, u_2, \dots, u_k \sim U(0,1)$ indep

$= \sum_{i=1}^k \lambda \ln(u_i)$

(81) $\boxed{\text{Accep rejection method}}$

① easy to sim. from $g(x)$, but want $f(x)$

② find c, $f(y) \leq c(g(x))$



$\boxed{\text{method}}$ ① gen rand val y from $g(y)$

② gen another rand u

③ if $u \cdot c \cdot g(y) \leq f(y) \Rightarrow x = y$

Prob accepting value $= \int \frac{f(y)}{c\,g(y)} g(y) dy = \frac{1}{c}$

$p(y \leq x \mid \text{accepted}) = \int_{-\infty}^{\infty} \frac{f(y)}{c} dy / \frac{1}{c} = \int_{-\infty}^{x} f(y) dy = F(x)$

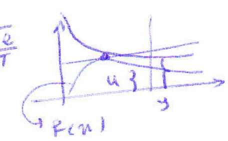
$f(x) = \frac{2e^{-x^2/2}}{\sqrt{2\pi}}$ $\quad g(x) = e^{-x} \Rightarrow \frac{f(x)}{g(x)} = \frac{2e^{-x^2/2}/\sqrt{2\pi}}{e^{-x}} = \frac{2}{\sqrt{2\pi}} e^{x - x^2/2}$

$\frac{d}{dx} \frac{f(x)}{g(x)} = \frac{2}{\sqrt{2\pi}} e^{x - x^2/2} (1-x) = 0 \Rightarrow x = 1$

$\frac{d^2}{dx^2} \frac{f(x)}{g(x)} = \frac{2}{\sqrt{2\pi}} [-e^{x - x^2/2} + (1-x)^2 e^{x - x^2/2}]$

$= \frac{2e^{x - x^2/2}}{\sqrt{2\pi}} [-1 + 1 - 2x + x^2]|_x < 0$

$\max \frac{f(x)}{g(x)} = \frac{2e^{1-1/2}}{\sqrt{2\pi}} = \frac{2\sqrt{e}}{\sqrt{2\pi}} = \frac{\sqrt{2e}}{\pi}$



(82) Gen values in multidim. space:

$MVN(\mu, \Sigma)$ $\quad p\times1$ $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$

$\Sigma > 0$ means $x'\Sigma x > 0$ $\quad \forall x \neq 0$ positive definite

$\Rightarrow \exists$ orthogonal matrix P & Diagonal matrix D_λ

$\Sigma = P D_\lambda P'$ $\quad \Sigma$ symmetric $\Rightarrow$ diagonal D_λ: Eigen values of Σ $(\lambda_i; x)$ + Columns of P are eigen vectors of Σ

i.e. $\Sigma x_j = \lambda_j x_j$

column of P

statistics summary

(11)

(83) - To generate $z_1, z_2, \dots, z_p$ $\sim$ indep $N(0,1)$

$x = \mu + \Gamma D_{\sqrt{\lambda}} z$ $\quad D_{\sqrt{\lambda}} = \begin{pmatrix} \sqrt{\lambda_1} & \dots & 0 \\ 0 & & \sqrt{\lambda_p} \end{pmatrix}$

x_1 $\quad \mu_1 = \mu_2 = 0$ $\quad \bar{\mu} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ $\quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ $\quad \rho = Corr(x_1, x_2)$

x_2 $\quad \sigma_1 = \sigma_2 = 1$ $\quad$ eigen values: $1+\rho$ $\ 1-\rho$

$$x_2 = \begin{pmatrix} \sqrt{\frac{1+\rho}{2}} z_1 + \sqrt{\frac{1-\rho}{2}} z_2 \\ \sqrt{\frac{1+\rho}{2}} z_1 - \sqrt{\frac{1-\rho}{2}} z_2 \end{pmatrix}$$

(84)

Gibbs (physisist)

Gibbs Sampler

$x_1/x_2, \dots, x_p$ hard, but x_i others is easy

start with arbitrary $\Rightarrow$ seq. of vectors

$X \sim N(20, 10^{50})$ $\quad Y \sim N(15, 5)$ $\quad \rho = Corr(x,y)$

$y|x \sim N(9 + .3x, 4)$ $\quad x|y \sim N(2 + 1.2y, 8)$

$\beta_{y|x} = \rho \frac{\sigma_y}{\sigma_x}$ $\quad \alpha_{y|x} = \mu_y - \beta_{y|x}\mu_x$

- no need for matrix inversion and eigen value
- recursive process

(85)

p-p plot $\quad$ q-q plot

$x_{(1)}, x_{(2)}, \dots, x_{(n)}$ $\quad$ ordered stat

$F^{-1}(i/_{n+1})$

$E(x_{(i)})$ as $n \to \infty$

q-q plot: $x_{(i)}$ vs. $F^{-1}(i/_{n+1})$

p-p plot: $F(x_{(i)})$ vs. $i/_{n+1}$

$x_1, x_2, \dots, x_n$ $\quad \mu + \sigma F^{-1}(\frac{i}{n+1})$

- look at tails

(86) numerical quadratics $\quad$ (numerical method)

$\boxed{\text{Simpson Rule}}$ $\int_a^b f(x)dx = \frac{(b-a)}{6}[f(a) + 4f(\frac{a+b}{2}) + f(b)]$

$f(x)=1$ $\quad \int_0^1 1\,dx = 1/6(1+4+1) = 1$

$f(x)=x$ $\quad \int_0^1 x\,dx = 1/2 x^2 = 1/2$ $\quad 1/6[0 + \frac{1}{2}\times 4 + 1] = 3/6 = .5$

$f(x)=x^2$ $\quad \int_0^1 x^2 dx = 1/3 x^3 = 1/3$ $\quad 1/6[0 + 4\cdot 1/4 + 1] = 2/6 = 1/3$

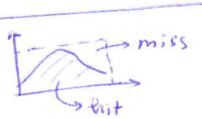
$f(x)=x^3$ $\quad \int_0^1 x^3 dx = 1/4 x^4 = 1/4$ $\quad 1/6[0 + 4\times 1/8 + 1] = \frac{1.5}{6} = 1/4$

$f(x)=x^4$ $\quad \int_0^1 x^4 dx = 1/5 x^5 = 1/5$ $\quad 1/6[0 + 4\times 1/16 + 1] = 1.25$

doesnt work but prety close

$\boxed{\text{Intuition}}$ for any integral Simpson Rule Simplifies Calculation, by puting 3 times more weight on midle than tail

(87) Monte Carlo integration:



$\int_0^1 g(x)dx$ $\quad \max_{0<x<1} g(x) = 1$ $\quad \min_{0<x<1} g(x) = 0$

Count a hit if $y' < g(x')$ $\quad$ - Do n times
otherwise Count a miss $\quad$ - x = # hit

$\widehat{\int_0^1 g(x)dx} = \frac{x}{n}$ $\quad E(\frac{x}{n}) \to p$ area

$Var(\frac{x}{n}) = \frac{p(1-p)}{n}$ binomial

(88) $\theta = \int_0^1 g(x)dx = E(g(x))$ $\quad x \sim U(0,1)$ $\quad \hat{\theta} = \sum_{i=1}^n \frac{g(u_i)}{n}$

$Var(\hat{\theta}) = \sum_{i=1}^n \frac{[g(x_i) - \hat{\theta}]^2}{n(n-1)}$

(82) $E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx$ $\quad f(x) \geq 0$ $\quad \int_{-\infty}^{\infty} f(x)dx = 1$

① Generate $x_1, x_2, \dots, x_n$ from $f(x)$ $\quad$ ② Compute $g(x_1), g(x_2), \dots, g(x_n)$

(12)

③ Compute $\hat{\alpha} = \sum_{i=1}^{\infty} \frac{g(x_i)}{n}$ of $\int_{-\infty}^{\infty} f(x) g(x) dx$

(88) fit data - gen - q-q plot

$\boxed{\text{Estimation}}$: Given data (Random sample) estimate param (r, θ) of the model

(89) oldest method: $\boxed{\text{method of moments}}$ (MM)

estimate moments from your data

use as many samples as posible

$\hat{x} = \sum_i \frac{x_i}{n}$ $\quad s^2 = \sum_i \frac{(x_i - \bar{x})^2}{n-1}$ $\quad \hat{\mu}_3 = \sum_i \frac{(x_i - \bar{x})^3}{n}$ $\quad \hat{\mu}_4 = \sum_i \frac{(x_i - \bar{x})^4}{n}$

(90) Gamma: $E(x^k) = \int_0^\infty \frac{\theta^r x^{k+r-1} e^{-\theta x}}{\Gamma(r)} dx = \frac{\Gamma(r+k)}{\theta^k \Gamma(r)}$

$\int_0^\infty \frac{\theta^{r+k} x^{r+k-1} e^{-\theta x}}{\theta^k \Gamma(r+k)} dx = \frac{\Gamma(r+k)}{\theta^k \Gamma(r)}$ $\Rightarrow$ $E(x) = \frac{\Gamma(r+1)}{\Gamma(r)\theta} = \frac{r}{\theta}$

$E(x^2) = \frac{\Gamma(r+2)}{\Gamma(r)\theta^2} = \frac{r(r+1)}{\theta^2}$ $\quad Var(x) = \frac{r^2}{\theta^2} + \frac{r}{\theta^2} - (\frac{r}{\theta})^2 = \frac{r}{\theta^2}$

two unknown params & two Equations

(MM) set $\bar{x} = \hat{r}/\hat{\theta}$ $\quad s^2 = \hat{r}/\hat{\theta}^2$

$\frac{s^2}{\bar{x}} = \frac{1}{\hat{\theta}} \Rightarrow \hat{\theta} = \frac{\bar{x}}{s^2}$ $\quad \bar{x} = \frac{\hat{r}}{\hat{\theta}} \Rightarrow \hat{r} = \bar{x}.\hat{\theta} = \frac{\bar{x}^2}{s^2}$

method of moment gives: point estimates $\Rightarrow$ Confidence interval?

(91) $\hat{\theta}, \hat{r} \sim f(\bar{x}, s^2)$

$\Rightarrow f(x,y) = f(\mu_1, \mu_2) + \frac{\partial f}{\partial x}\Big|_{\xi_1 \text{ b/w } x, \mu_1} (x - \mu_1) + \frac{\partial f}{\partial y}\Big|_{\xi \text{ b/w } y, \mu_2} (y - \mu_2)$

x close to μ_1, y close to μ_2

$E[f(x,y)] \simeq f(\mu_1, \mu_2)$

$E(\bar{x}^2) = \mu^2 + \frac{\sigma^2}{n}$ $\quad E(\bar{x})^2 = \mu^2$

$Var(f(x,y)) = (\frac{\partial f}{\partial x}\Big|_{\mu_1})^2 Var(x) + (\frac{\partial f}{\partial y}\Big|_{\mu_2})^2 Var(y) + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y} Cov(x,y)$

$\hat{\theta} = \frac{\bar{x}}{s^2}$ $\quad Var(\bar{x}) = \frac{\sigma^2}{n} = (\frac{r}{\theta^2})/n = \frac{r}{\theta^2 n}$ $\quad \mu_4$

$\boxed{Var(y) = Var(s^2) = \frac{2\sigma^2}{n-1} + \left[\frac{E(x-\mu)^4 - 3\sigma^4}{n}\right]}$ at least sample size = 5

$Cov(\bar{x}, s^2) = \frac{\mu_3}{n} = E(x-\mu)^3$

$E(x^3) = \frac{\Gamma(r+3)}{\Gamma(r)\theta^3} = \frac{(r+2)(r+1)r}{\theta^3}$

$E(x-\mu)^3 = E(x^3) - 3E(x^2)\mu + 3E(x)\mu^2 - \mu^3$

$= \frac{(r+2)(r+1)r}{\theta^3} - 3\frac{(r+1)r}{\theta^2} + 3(\frac{r}{\theta})(\frac{r}{\theta})^2 - (\frac{r}{\theta})^3$

$= \frac{1}{\theta^3}[(r+2)(r+1)\cdot r - 3r^2(r+1) + 2r^3] = \frac{r}{\theta^3}[(r+2)(r+1) - 3(r+1) + 2r^2] = \frac{2r}{\theta^2}$

$Cov(\bar{x}, s^2) = \frac{2r}{\theta^2 n}$ $\quad E(x-\mu)^4 = \frac{6r}{\theta^4}$ $\quad Var(\bar{x}) = \frac{r}{n\theta^2}$

$E(x) = \frac{r}{\theta}$ $\Rightarrow \hat{\theta} = \frac{\bar{x}}{s^2}$ $\quad \frac{\partial \hat{\theta}}{\partial \bar{x}} = 1/s^2$ $\quad \frac{\partial \hat{\theta}}{\partial s^2} = \frac{-\bar{x}}{s^4}$

$Var(\bar{x}) = \frac{r}{n\theta^2}$ $\quad Var(s^2) = [2(\frac{r}{\theta^2})^2 + [\frac{6r}{\theta^4} - 3(\frac{r}{\theta^2})^2]]\frac{1}{n}$

$\Rightarrow Var(s^2) = \frac{2r^2}{(n-1)\theta^4} + \frac{1}{\theta^4 n}(6r - 3r^2)$

$\frac{1}{s^2} = \frac{\hat{\theta}^2}{\hat{r}}$ $\quad \hat{r} = (\frac{-\bar{x}}{s^2})\frac{1}{s^2}$ $\quad \frac{-\bar{x}}{s^4} = \frac{-\hat{\theta}^3}{\hat{r}}$

$\frac{\partial \theta}{\partial \bar{x}} = \frac{\hat{\theta}^2}{\hat{r}}$ $\quad \frac{\partial \theta}{\partial s^2} = \frac{-\hat{\theta}^3}{\hat{r}}$

$Var(f(x,y)) = [\frac{\partial f}{\partial x} \ \frac{\partial f}{\partial y}] \begin{bmatrix} V(x) & Cov(x,y) \\ Cov(x,y) & V(y) \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$

$\boxed{\text{intuition}}$ Same as Delta method, variance is second moment, and power two in matrix form is sandwitch

statistics summary

13

$$\widehat{Var}(\hat\theta) = \begin{pmatrix} \frac{\hat\theta^2}{\hat r} & \frac{\hat\theta^3}{\hat r}\end{pmatrix} \begin{pmatrix} \frac{\hat r}{n\hat\theta^2} & \frac{2\hat r}{n\hat\theta^3} \\ \frac{2\hat r}{n\hat\theta} & \frac{\left[\frac{6r}{\theta^4} - 3\left(\frac{r^2}{\theta^4}\right)\right]}{n} + \frac{2r^2}{(n-1)\theta^4} \end{pmatrix} \begin{pmatrix} \frac{\hat\theta^2}{\hat r} \\ -\frac{\hat\theta^3}{\hat r}\end{pmatrix}$$

$$= \frac{3\hat\theta^2}{n\hat r} + \frac{2\hat\theta^2}{n-1}$$

$$\widehat{Var}(\hat r) = (2\hat\theta \quad -\hat\theta^2) \begin{pmatrix} \frac{\hat r}{n\hat\theta^2} & \frac{2\hat r}{n\hat\theta^3} \\ \frac{2\hat r}{n\hat\theta^3} & \frac{2\hat r^2}{(n-1)\hat\theta} + \left[\frac{6\hat r}{\hat\theta^4} - 3\left(\frac{\hat r}{\hat\theta}\right)^2\right] \end{pmatrix} \times \begin{pmatrix} \hat\theta^2 \\ \hat\theta^3 \end{pmatrix}$$

$$= \frac{2\hat r}{n}\left[1 + \frac{n\hat r}{n+1}\right] \simeq \frac{2\hat r(\hat r + 1)}{n}$$

$\hat\beta_1 = f_1(x_1, x_2)$ $\quad \hat\beta_2 = f_2(x_1, x_2)$ with $x_1 \to \bar x$, $x_2 \to s^2$

$$\hat\beta_1 = \frac{\partial f_1}{\partial x_1}(x_1 - \mu_1) + \frac{\partial f_1}{\partial x_2}(x_2 - \mu_2) + f_1(\mu_1, \mu_2)$$

$$\hat\beta_2 = \frac{\partial f_2}{\partial x_1}(x_1 - \mu_1) + \frac{\partial f_2}{\partial x_2}(x_2 - \mu_2) + f_2(\mu_1, \mu_2)$$

$$Cov(\hat\beta_1, \hat\beta_2) = \left[\frac{\partial f_1}{\partial x_1} \quad \frac{\partial f_1}{\partial x_2}\right]\begin{bmatrix} Var(x_1) & Cov(x_1,x_2) \\ Cov(x_1,x_2) & Var(x_2)\end{bmatrix}\begin{bmatrix}\frac{\partial f_2}{\partial x_1} \\ \frac{\partial f_2}{\partial x_2}\end{bmatrix}$$

$$= \left(\frac{\partial f_1}{\partial x_1}\cdot\frac{\partial f_2}{\partial x_1}\right)Var(x_1) + \left(\frac{\partial f_1}{\partial x_2}\cdot\frac{\partial f_2}{\partial x_2}\right)Var(x_2) + \left[\frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_2}Cov(x_1,x_2) + \frac{\partial f_1}{\partial x_2}\cdot\frac{\partial f_2}{\partial x_1}Cov(x_1,x_2)\right]$$

intuition: Covariance will send with different Breads

92 Coefficient of Variation: $c_v = \frac{\sigma}{\mu}$

e.g. Data ~ Normal $Cov(\bar x, s^2) = 0$

$$\hat c_v = \frac{\sqrt{s^2}}{\bar x} \qquad \frac{\partial \hat c_v}{\partial \bar x} = -\frac{\sqrt{s^2}}{\bar x^2} \qquad \frac{\partial \hat c_v}{\partial s^2} = \frac{1}{2\bar x\sqrt{s^2}} = \frac{1}{2\bar x s}$$

$$V(\hat c_v) = \left(-\frac{s}{\bar x^2}\right)^2 V(\bar x) + \left(\frac{1}{2\bar x s}\right)^2 Var(s^2) = \frac{s^2}{\bar x^4}\cdot\frac{\sigma^2}{n} + \frac{1}{4\bar x^2 s^2}\frac{2\sigma^4}{n}$$

$$= (\sigma^2/n)\left(s^2/\bar x^4 + \frac{1}{2\bar x^2}\frac{\sigma^2}{s^2}\right) = \frac{1}{n}\left(\frac{\sigma^2}{\bar x^2}\right)\left(\frac{s^2}{\bar x^2} + \frac{1}{2}\frac{\sigma^2}{s^2}\right)$$

$\boxed{\hat\sigma = s}$ $\Rightarrow$ $= \frac{1}{n}\left(\frac{s}{\bar x}\right)^2\left[\left(\frac{s}{\bar x}\right)^2 + \frac{1}{2}\right] = \frac{1}{n}(\hat c_v)^2 \times \left[(c_v)^2 + \frac{1}{2}\right]$

93 application: Crime: k: Crime $\quad x_i \sim Bin(k, p)$

p: report prob. $\quad E(x_i) = kp$

$SD(x_i) = kp(1-p)$ $\quad$ goal: estimate k & p

$\bar x = \sum_i \frac{x_i}{n}$ $\quad E(\bar x) = \sum_{i=1}^n \frac{E(x_i)}{n} = \frac{nkp}{n} = kp$

$s^2 = \frac{1}{n-1}\sum_i (x_i - \bar x)^2 = \frac{1}{n-1}\left[\sum_i (x_i)^2 - n\bar x^2\right]$

$E(s^2)$: need $E(x_i^2)$

$E(x_i^2) = V(x_i) + E^2(x_i) = kp(1-p) + k^2p^2$

$E(\bar x^2) = Var(\bar x) + E^2(\bar x) = \frac{kp(1-p)}{n} + k^2p^2$

$$E(s^2) = \frac{1}{n-1}\left[\sum_i E(x_i^2) - nE(\bar x^2)\right] = \frac{1}{(n-1)}\left[nkp(1-p) + nk^2p^2 - n\left[\frac{kp(1-p)}{n} + k^2p^2\right]\right] = \frac{1}{n-1}\left[nkp(1-p) + nk^2p^2 - kp(1-p) - nk^2p^2\right]$$

$$= \frac{kp(1-p)(n-1)}{(n-1)} = kp(1-p)$$

$\bar x = kp \quad s^2 = kp(1-p) \Rightarrow \frac{s^2}{\bar x} = 1 - p \Rightarrow \hat p = 1 - \frac{s^2}{\bar x} \quad 0 < \hat p < 1$

$$\hat k = \frac{\bar x}{1 - s^2/\bar x}$$

14

94 one sample t-test

$H_N: \mu = \mu_0 \quad x_1, x_2, \dots, x_n$

$H_A: \mu \ne \mu_0 \quad t_{obs} = \frac{\bar x - \mu}{s/\sqrt n} \sim$ t-dist ($n-1$ deg frdm)

95 two samples: $\quad$ t-dist ~ deg frdm

$x_{11} \quad x_{21}$ $\quad$ 1-deg frdm → Cauchy

$\vdots$ $\quad$ 2-deg frdm

$x_{1n_1} \quad x_{2n_2}$ $\quad$ Small sample size → t-dist

$H_0: \mu_1 = \mu_2 \quad t_{obs} = \frac{\bar x_1 - \bar x_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ Not A t-dist

$H_N: \mu_1 \ne \mu_2$ $\quad$ Behrens-Fisher problem

need assumption: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\hat\sigma = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} \qquad \frac{\bar x_1 - \bar x_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s^2}{n_2}}} \sim t(n_1 + n_2 - 2)$$

$H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 \ne \sigma_2^2$

Saither white, welch, Huzurbazur: $t_{obs} = \frac{\bar x_1 - \bar x_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

$$f = \frac{\frac{s_1^2}{n_1}}{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}} \Rightarrow df = \frac{1}{\frac{f^2}{(n_1 - 1)} + \frac{(1-f)^2}{(n_2 - 1)}}$$

two sample t-test with an eq variance ← $\quad s_1^2 = s_2^2$

$\min(n_1 - 1, n_2 - 1) \le df \le (n_1 + n_2 - 2)$

deg frdm from method of moment (MM)

96 $Y_1, Y_2, \dots, Y_N$ ~ iid

$Y_i \sim \chi^2(f_i) \quad \sum a_i Y_i$ Not χ^2, but assume close to $c\chi^2(v)$

find c, v approx? $\quad$ mean of chisq: v d.f. $\quad$ variance $2v$

$E(\sum_i a_i Y_i) = \sum_i a_i E(Y_i) = \sum_i a_i f_i = cv$ (MM)

$V(\sum_i a_i Y_i) = \sum_i a_i^2 V(Y_i) = 2\sum_i a_i^2 f_i = 2c^2 v$

$$\hat c = \frac{\sum_i a_i^2 f_i}{\sum a_i f_i} = \frac{\hat c^2 \hat v}{\hat c \hat v}$$

$$\hat v = \frac{(\sum_i a_i f_i)^2}{\sum_i a_i^2 f_i}$$

⟹ method of moment gives us approx.

97 $s_1^2 = \frac{\sum_i (x_{1i} - \bar x_1)^2}{n-1} = \frac{\sum_i (x_{1i} - \bar x_1)^2}{f_1}$

$$\frac{\sum_i (x_{1i} - \bar x_1)^2}{\sigma_1^2} = \frac{\sum (x_{1i} - \bar x_1)^2}{\sigma_1} \longrightarrow \chi^2_{f_1}(n_1 - 1)$$

$\frac{f_1 s_1^2}{\sigma_1^2} \sim \chi^2(f_1) \qquad \frac{f_2 s_2^2}{\sigma_2^2} \sim \chi_2^2(f_2)$

$s_1^2 \sim \sigma_1^2\frac{\chi^2(f_1)}{f_1} \qquad s_2^2 \sim \sigma_2^2\frac{\chi^2(f_2)}{f_2}$

$$\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \sim \sigma_1^2\frac{\chi^2(f_1)}{n_1 f_1} + \sigma_2^2\frac{\chi^2(f_2)}{n_2 f_2} \qquad a_1 = \frac{\sigma_1^2}{n_1 f_1} \quad a_2 = \frac{\sigma_2^2}{n_2 f_2}$$

$$\hat v = \frac{\left(\frac{\sigma_1^2}{n_1 f_1}\cdot f_1 + \frac{\sigma_2^2}{n_2 f_2}\cdot f_2\right)^2}{\left(\frac{\sigma_1^4}{n_1^2 f_1^2}\cdot f_1 + \frac{\sigma_2^4}{n_2^2 f_2^2}\cdot f_2\right)} = \frac{(\sigma_1^2/n_1 + \sigma_2^2/n_2)^2}{\frac{\sigma_1^4}{f_1 n_1^2} + \frac{\sigma_2^4}{f_2 n_2^2}}$$

98 **Max likelihood** $x_1, x_2, \dots, x_n$ (reasonable model)

density $p(x, \underline{\theta})$: a step is choosing model (depend on model choice)

$\text{Prob}(x_1, x_2, \dots, x_n | \underline{\theta}) = \prod_{i=1}^{n} p(x_i, \underline{\theta})$

philosophy: Nature is indiff to you (no bias in sample)

set of param's pop up (one most likely to happen)

$\equiv$ no systematic bias Exists

Fisher: $L(\underline{\theta}) = \prod_{i=1}^{n} p(x_i, \theta)$ $\quad \underline{\theta}$ that maximizes $L(\underline{\theta})$

$\theta \in [0, \infty)$ if variance $\quad \hat{\theta} \in [0,1]$ if prob

- max in boundary, Relative max (check convergence on multiple points)

99 e.g. Extra Sensory Perception (ESP)

$\text{Prob} \begin{cases} 0 & \text{if No ESP} \\ 1 & \text{if perfect} \end{cases} \quad \theta = \frac{p - 1/5}{4/5} \quad p$: proportion of success

Beauty of max likelihood: write way you want (defined, since may frame sense in some field)

obsrv random variable $S_i = \begin{cases} 1 & \pi \\ 0 & 1-\pi \end{cases} \quad \theta = \frac{\pi - 1/5}{4/5}$

$\pi = 1/5 + 4/5\,\theta$

$S_i = \begin{cases} 1 & 1/5 + 4/5\,\theta \\ 0 & \frac{4(1-\theta)}{5} \end{cases} \quad 1 - \theta = \frac{4(1-\theta)}{5}$

$P(S_i) = [\frac{1}{5} + \frac{4}{5}\theta]^{S_i} [\frac{4(1-\theta)}{5}]^{1-S_i}$

$L(\theta) = \prod_{i=0}^{n} [\frac{1}{5} + \frac{4}{5}\theta]^{S_i} [\frac{4(1-\theta)}{5}]^{1-S_i}$

$L(\theta) = [\frac{1}{5} + \frac{4}{5}\theta]^{\sum_i S_i} [\frac{4(1-\theta)}{5}]^{\sum_i 1-S_i}$

Statistics: additive word is multiplicative log is used to convert to additive

$L(\underline{\theta}) = \ln L(\underline{\theta}) = \sum_i S_i \ln(\frac{1}{5} + \frac{4}{5}\theta) + \sum_i (1-S_i) \ln(\frac{4(1-\theta)}{5})$

$= \sum_i S_i \ln(1+4\theta) - \sum S_i \ln(5) + \sum_i (1-S_i) \ln(1-\theta) + \sum_i (1-S_i) \ln 4$

$- \sum_i (1-S_i) \ln 5 = C + \sum_i S_i \ln(1+4\theta) + \sum_i (1-S_i) \ln(1-\theta)$

$\Rightarrow \frac{\partial L(\theta)}{\partial \theta} = \sum_i \frac{4 S_i}{1+4\theta} - \sum_i \frac{(1-S_i)}{1-\theta} \quad x = \sum_i S_i, \quad n - x = \sum_i (1-S_i)$

$= \frac{4x}{1+4\theta} - \frac{(n-x)}{1-\theta}$

$\frac{\partial^2 L(\theta)}{\partial \theta^2} = \frac{-16x}{(1+4\theta)^2} - \frac{(n-x)}{(1-\theta)^2} < 0 \Rightarrow \frac{4x(1-\hat{\theta}) - (1-4\hat{\theta})(n-x)}{(1+4\hat{\theta})(1-\hat{\theta})} = 0$

$\Rightarrow 4x - 4x\hat{\theta} - (n-x) - 4\hat{\theta} n + 4\hat{\theta} x = 0 \Rightarrow 4x - n + x + n \cdot 4\hat{\theta}$

$\Rightarrow (5x - n) + n \cdot 4(\hat{\theta}) = 0 \Rightarrow \hat{\theta} = \frac{5x - n}{4n} = \frac{5}{4}(\frac{x}{n}) - \frac{1}{4}$

100 a density, $\underline{\theta}$: r dim vector of param

Define ("score function") $S_i(x, \underline{\theta}) = \frac{\partial}{\partial \theta_i} \ln(x, \underline{\theta})$

$S_{ij}(x, \underline{\theta}) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(x, \underline{\theta})$

$p(x, \underline{\theta})$ regular as w.r.t first derivative

$E_x(S_i(x, \underline{\theta})) = 0 \quad i = 1, 2, \dots, r \quad$ w.r.t 2nd derivative regular

if $E(S_i(x, \underline{\theta}) S_j(x, \underline{\theta})) = -E(S_{ij}(x, \underline{\theta}))$ then:

1. a max likelihood estimator Exists
2. max likelihood estimator satisfies equation $\sum_{i=1}^{n} S_i(x_i, \hat{\underline{\theta}}_{ML}) = 0$
3. If $\hat{\underline{\theta}}_{ML}$ unique for $n >$ some $n_0 \Rightarrow \hat{\underline{\theta}}_{ML}$ is asymptotically multivar Normal with mean vector $\underline{\theta}$ and cover matrix $B^{-1}(\underline{\theta}, \underline{\theta})/n$ where $B_{ij} = E[S_i(x, \underline{\theta}), S_j(x, \underline{\theta})] = -E(S_{ij}(x, \underline{\theta}))$

product of first partials

mixed partial $\equiv$ Information matrix

101 you can Estimate $E(S_{ij}(B, \underline{\theta})) = \frac{\sum_{k=1}^{n} S_{ij}(x_k, \hat{\underline{\theta}}_{ML})}{n}$

$\hat{B}_{ij}(\hat{\theta}_{ML}) = -\frac{1}{n} \sum_{k=1}^{n} S_{ij}(x_k, \hat{\theta}_{ML})$

$\hat{\theta} \sim MVN(\underline{\theta}, (-\sum_{k=1}^{N} S_{ij}(x_k, \hat{\theta}_{ML})^{-1}))$

$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

102 $\theta = \frac{\pi - 1/5}{4/5} \quad \hat{\pi} = \frac{x}{n}$

$\pi = 1/5 (1 + 4\theta) \quad \hat{\theta} = \frac{p - 1/5}{4/5} \quad 1 - \pi = 4/5 (1-\theta)$

$\text{Var}(\hat{\theta}) = (\frac{5}{4})^2 \text{Var}(\hat{p}) = (\frac{5}{4})^2 \frac{\pi(1-\pi)}{n} = (\frac{5}{4})^2 \frac{1}{5}(1+4\theta) \frac{4(1-\theta)}{5n}$

$= \frac{(1+4\theta)(1-\theta)}{4n}$

$S_i = \begin{cases} 1 & \frac{(1+4\theta)}{5} \\ 0 & \frac{4(1-\theta)}{5} \end{cases} \quad E(S_i) = \frac{1+4\theta}{5}, \quad E(1-S_i) = \frac{4(1-\theta)}{5}$

$f(S_i) = (\frac{1+4\theta}{5})^{S_i} (\frac{4(1-\theta)}{5})^{1-S_i} \times \frac{1}{5}$

$L(\theta) = \ln f(S_i) = S_i \ln(1+4\theta) + (1-S_i) \ln(1-\theta)$

$+ (1-S_i) \ln 4 - \ln 5$

$\Rightarrow \frac{\partial L(\theta)}{\partial \theta} = \frac{4 S_i}{1+4\theta} - \frac{(1-S_i)}{1-\theta}$

$E(\frac{\partial L(\theta)}{\partial \theta}) = \frac{4 E(S_i)}{1+4\theta} - \frac{E(1-S_i)}{(1-\theta)} = \frac{4(1+4\theta)}{5(1+4\theta)} - \frac{4(1-\theta)}{5(1-\theta)}$

$= 4/5 - 4/5 = 0$ regular w.r.t 1st derivative

$E(\frac{\partial L(\theta)}{\partial \theta})^2 = E(\frac{16 S_i^2}{(1+4\theta)^2} + \frac{(1-S_i)^2}{(1-\theta)^2} - \frac{8 S_i(1-S_i)}{(1+4\theta)(1-\theta)})$

$= \frac{16 E(S_i^2)}{(1+4\theta)^2} + \frac{E(1-S_i)^2}{(1-\theta)^2} = \frac{16 E(S_i)}{(1+4\theta)^2} + \frac{E(1-S_i)}{(1-\theta)^2}$

$= \frac{16(1+4\theta)}{5(1+4\theta)^2} + \frac{4(1-\theta)}{5(1-\theta)^2} = 4/5 [\frac{4}{1+4(\theta)} + \frac{1}{1-\theta}] =$

$4/5 [\frac{4 - 4\theta + 1 + 4\theta}{(1+4\theta)(1-\theta)}] = \frac{4}{(1+4\theta)(1-\theta)} \longrightarrow \text{Var}(\hat{\theta})$

$E(\frac{\partial L(\theta)}{\partial \theta^2}) = E[\frac{-16 S_i}{(1+4\theta)^2} - \frac{(1-S_i)}{(1-\theta)^2}] = \frac{-16}{5}(\frac{1+4\theta}{(1+4\theta)^2} -$

$\frac{4(1-\theta)}{5(1-\theta)^2}] = -4/5[\frac{4}{1+4\theta} + \frac{1}{1-\theta}] \Rightarrow -E[S_{ij}(x, \underline{\theta})] =$

$E(S_i(x, \underline{\theta}), S_j(x, \underline{\theta}))$

103 $f(x) = \frac{\theta^r e^{-\theta x} x^{r-1}}{\Gamma(r)} \quad \underline{\theta} = (\begin{smallmatrix} r \\ \theta \end{smallmatrix})$

$\ln f(x) = r \ln(\theta) - \theta x + (r-1) \ln(x) - \ln(\Gamma(r)) \quad x_1, x_2, \dots, x_n$

$L(r, \theta) = nr \ln\theta - \theta \sum_i x_i + (r-1) \sum_i \log x_i - n \ln \Gamma(r)$

$= n[r \ln\theta - \theta \bar{x} + (r-1) \overline{\ln x} - \ln \Gamma(r)]$

intuition:
1. get density & get log
2. sum over all $x_1 \dots x_n$ of $\log(f) =$ likelihood
3. summarize data and use mean $= \bar{x}$
4. get derivatives

$\overline{\ln x} = \sum_{i=1}^{n} \frac{\ln x_i}{n} \quad$ Gamma func. Regular

1. $\frac{\partial L(r, \theta)}{\partial r} = n[\ln\theta + \overline{\ln x} - \underbrace{\frac{\Gamma'(x)}{\Gamma(r)}}_{\text{DIGamma func}}] = 0$
2. $\frac{L(r, \theta)}{\partial \theta} = n[\frac{r}{\theta} - \bar{x}] = 0 \Rightarrow \hat{\theta} = \frac{\hat{r}}{\bar{x}}$
3. 2nd deriv. $\frac{\partial^2 L(r, \theta)}{\partial r^2} = -n[\frac{\Gamma(r)\Gamma''(r) - [\Gamma'(r)]^2}{\Gamma(r)^2}]$

$\frac{\partial^2}{\partial \theta \partial r} L(r, \theta) = \frac{n}{\theta} \quad \frac{\partial^2}{\partial r \partial \theta} L(r, \theta) = \frac{n}{\theta} \quad \frac{\partial^2 L(r, \theta)}{\partial \theta^2} = \frac{-nr}{\theta^2}$

Statistics Summary

at $\hat{r}$

$$L(r,\theta) = L(r, \frac{r}{\bar{x}}) = n[r\ln(\frac{r}{\bar{x}}) - \frac{r}{\bar{x}}\cdot\bar{x} + (r-1)\overline{\ln x} - \ln\Gamma(r)] =$$
$$n[r\ln r - r\ln\bar{x} - r + (r-1)\overline{\ln x} - \ln\Gamma(r)]$$

(104) use newton method to evaluate:

e.g $x=\sqrt{k}$ $\quad f(x) = x^2 - k = 0$ (*)

$f(x) = f(\mu) + f'(\xi)(x-\mu) \Rightarrow$ (*): $0 = f(\mu) + f'(\xi)(x-\mu)$ idiosyncracy, neighbor of solution

Step: Guess solution x_k not sol but close

$\Rightarrow 0 = f(x_k) + f'(x_k)(x_{k+1} - x_k) \Rightarrow x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$

for e.g: $x_{k+1} = x_k - (\frac{x_k^2 - K}{2x_k}) = (x_k + \frac{K}{x_k}) \times \frac{1}{2}$

Convergance is important, and derivative should Exist

intuition: ① when don't know solution don't just rely on memory and calculation but Guess and check method is more effective technique

(105) e.g. $f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_k(x) \end{pmatrix}$ $\quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}$ $\quad f_i(x) = f_i(\mu) \sum_{j=1}^{r} \frac{\partial f_i(x)}{\partial x_j}(x_j - \mu_j)$

$\mu = \hat{\theta}_k$ $\quad f(x) = \begin{pmatrix} \frac{\partial g_1(\theta)}{\partial\theta_j} \\ \vdots \\ \frac{\partial g_i(\theta)}{\partial\theta_j} \end{pmatrix}$ $\quad 0 = f(\mu) + \{\frac{\partial^2 g(\theta)}{\partial\theta_i\partial\theta_j}\}(\theta - \mu)$

$k \times k$ matrix

$\hat{\theta}_{k+1} = \hat{\theta}_k - [\frac{\partial g(\theta)}{\partial\theta_i\partial\theta_j}]^{-1} f(x_k)$

(106) max likelihood of transformation:

family of transfo: $Z_i(\lambda) = \frac{x_i^\lambda - 1}{\lambda}$ $\begin{cases} -\infty < \lambda < \infty \\ \lambda = 0 \\ Z_i(\lambda) = \ln x_i \end{cases}$

way to fit data but atlast needs to be checked by p-p q-q

$\exists\lambda$ s.t. $Z_i(\lambda) \sim N(\mu(\lambda), \sigma^2(\lambda))$

Step ① Transform ② check with p-p or q-q plot

- random sample $x_1, x_2, \dots, x_n \longrightarrow Z_1(\lambda), Z_2(\lambda), \dots, Z_n(\lambda)$

joint density: $\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma(\lambda)} e^{-\frac{1}{2}(\frac{Z_i(\lambda) - \mu(\lambda)}{\sigma(\lambda)})^2} dZ_1(\lambda) dZ_2(\lambda) \dots dZ_n(\lambda)$

$x_i = (1 + \lambda Z_i(\lambda))^{1/\lambda}$ $\quad \lambda = 0 \quad x_i = e^{Z_i}$

$\frac{dZ_i(\lambda)}{\partial x_i} = \frac{\lambda x_i^{\lambda-1}}{\lambda} = x_i^{\lambda-1}$ $\quad [\lambda = 0 \quad \frac{1}{x_i}]$

pdf $\pi(x_1, x_2, \dots, x_n | \lambda, \mu(\lambda), \sigma(\lambda)) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma(\lambda)} e^{-\frac{1}{2}[\frac{\frac{x_i^\lambda - 1}{\lambda} - \mu(\lambda)}{\sigma(\lambda)}]^2} \cdot x_1^{\lambda-1} x_2^{\lambda-1} \dots x_n^{\lambda-1} \times dx_1 dx_2 \dots dx_n$

$dZ_i(\lambda) = x_i^{\lambda-1} dx_i$

$\Rightarrow L(\lambda, \mu(\lambda), \sigma^2(\lambda)) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_i [\frac{\frac{x_i^\lambda - 1}{\lambda} - \mu(\lambda)}{\sigma(\lambda)}]^2$

$- \frac{n}{2}\ln\sigma^2(\lambda) + (\lambda - 1)\sum_i \ln x_i$

How estimate $\mu(\lambda), \sigma^2(\lambda), \lambda$? Given λ:

$\hat{\mu}(\lambda) = \frac{\sum_{i=1}^{n}}{n}(\frac{x_i^\lambda - 1}{\lambda}) = \bar{Z}(\lambda)$ $\quad \hat{\sigma}^2(\lambda) = \frac{\sum_{i=1}^{n}}{n}(\frac{x_i^\lambda - 1}{\lambda} - \bar{Z}(\lambda))^2$
$= \frac{1}{n}\sum_{i=1}^{n}(Z_i(\lambda) - \bar{Z}(\lambda))^2$

$L(\lambda, \hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)) = -\frac{n}{2}\ln 2\pi - \frac{\sum_i}{2}\frac{(Z_i(\lambda) - \bar{Z}(\lambda))^2}{\hat{\sigma}^2(\lambda)} - \frac{n}{2}\ln\hat{\sigma}^2(\lambda)$

$+ \cancel{\dots} + (\lambda - 1)\sum_i \ln x_i$

$= -\frac{n}{2}\ln 2\pi - \frac{n}{2} - \frac{n}{2}\hat{\sigma}^2(\lambda) + (\lambda - 1)\sum_i \ln x_i$

Numerically Steps
① pick λ
② Compute for each x_i: $Z_i(\lambda) = \frac{x_i^\lambda - 1}{\lambda}$
③ Compute $\bar{Z}(\lambda)$ and $\hat{\sigma}^2(\lambda)$
④ Compute $-\frac{n}{2}\hat{\sigma}^2(\lambda) + (\lambda - 1)\sum_i \ln x_i$
⑤ plot ④ against λ
⑥ Go back to 1 after increasing λ

Start from large range & then narrow down

(107) e.g.

$x \sim U(0,\theta) \Rightarrow f(x) = \frac{1}{\theta}$ $\quad 0 \le x \le \theta$

$x_1, x_2, \dots, x_n$ $\quad f(x) = 2\theta$

$\prod_{i=1}^{n} f(x_i) = \frac{1}{\theta^n}$ $\quad x_n = \max(x_1, x_2, \dots, x_\theta) \le \theta$

$L(\theta|x) = \frac{1}{\theta^n}$ on $\theta \ge x_{max}$

$L(\theta|x)$ monotone decreasing & maximized in $\hat{\theta} = x_{max}$

$\hat{\theta}_{max} = x_m = \max(x_1, \dots, x_n)$

$Prob(x_1, x_2, \dots, x_n \le z) = (\frac{z}{\theta})^n \Rightarrow$ pdf of $x_m = \frac{nz^{n-1}}{\theta^n}$ $\quad 0 \le z \le \theta$

$E(z) = \int_0^\theta \frac{n}{\theta^n} z^{n-1} z\,dz = \frac{n}{\theta^n}\int_0^\theta z^n dz = \frac{n}{(n+1)\theta^n}\int_0^\theta (n+1)z^n dz$

$= z^{n+1}\Big|_0^\theta \frac{n}{(n+1)\theta^n} = \frac{\theta^{n+1}\cdot n}{(n+1)\theta^n} = \frac{n}{n+1}\theta$

intuition: take new variable Z, and calculate probability that the definition holds for (definition of parameter: e.g θ), calculate likelihood based on parameter and Estimate, and calculate $E(Z)$ by integrating on the range of real estimate, use sample size in your calculation [likelihood]

(108) Bias of estimate: $\hat{\theta}$ of θ: $\quad Bias(\hat{\theta}) = |E(\hat{\theta}) - \theta|$

e.g: Bias of $\hat{\theta}_{ML} = |\frac{n}{n+1}\theta - \theta| = \theta|\frac{n-n-1}{n+1}| = \frac{\theta}{n+1}$ asymptotically unbiased

$E(z^2) = \int_0^\theta \frac{nz^{n-1}}{\theta^n} z^2 dz = \frac{n}{\theta^n(n+2)}\int_0^\theta (n+2)z^{n+1}dz = \frac{nz^{n+2}}{\theta^n(n+2)}\Big|_0^\theta$

$= \frac{n\theta^2}{n+2} \Rightarrow Var(Z) = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} = n\theta^2[\frac{1}{n+2} - \frac{n}{(n+1)^2}]$

$= \frac{n\theta^2}{(n+2)(n+1)^2}[(n+1)^2 - n(n+2)] = \frac{n\theta^2}{(n+2)(n+1)}[n^2 + 2n + 1 - n^2 + 2n]$

$= \frac{n\theta^2}{(n+2)(n+1)^2}$

intuition: ① you can find max likelihood value by looking at function and guessing
② if max likelihood function based on sample param: e.g $ml(\dots) = x_m \in U(0,\theta)$ you can use param definition to construct density function to calculate $E(\hat{\theta})$, based on your estimation

(109) $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + Bias^2(\hat{\theta})$

- Trick use $E(\hat{\theta})$ in

(110) $MSE(\hat{\theta}_{ML}) = Var(\hat{\theta}_{ML}) + Bias^2(\hat{\theta}_{ML}) = \frac{n\theta^2}{(n+2)(n+1)^2} + \frac{\theta^2}{(n+1)^2}$

$= \frac{2\theta^2}{(n+2)(n+1)}$

(111) $\hat{\theta}_u = \frac{(n+1)}{n}x_n \Rightarrow E(\hat{\theta}_u) = \frac{(n+1)}{n}E(x_n) = \theta$

$Bias(\hat{\theta}_u) = 0$ $\quad Var(\hat{\theta}_u) = \frac{(n+1)^2}{n^2}Var(x_n)$

$Var(\hat{\theta}_u) > Var(\hat{\theta}_{ML})$ $\quad MSE(\hat{\theta}_{ML}) \ge MSE(\hat{\theta}_u)$

(112) Blue Estimate: Best linear unbiased Estimator

MVUE = minimum variance unbiased Estimator

Statistics Summary

(113) Find c : $\hat{\theta}_c = cx_m$ $Var(\hat{\theta}_c) = c^2 Var(x_n)$ $E(\hat{\theta}_c) = cE(x_n)$

$MSE(\hat{\theta}) = E(cx_m - \theta)^2 = E(cx_m - \theta)^2 = E(c^2 x_n^2 - 2x_n\theta + \theta^2)$

Best c that minimizes MSE : $\frac{\partial}{\partial c} MSE(\hat{\theta}_c) = 2cE(x_n^2) - 2E(x_n)\cdot\theta$

$\frac{\partial^2}{\partial c^2} MSE(\hat{\theta}_c) = 2E(x_n^2) > 0 \Rightarrow c = \frac{\theta E(x_n)}{E(x_n^2)} = \frac{\frac{\theta \cdot n}{n+1}\theta}{\frac{n\theta^2}{n+2}} = \frac{n+2}{n+1}$

$\Rightarrow Bias(\hat{\theta}_c) = |cE(x_n) - \theta| = [\frac{n+2}{n+1}\frac{n}{n+1}\theta - \theta] = \frac{\theta}{(n+1)^2}$

$\Rightarrow MSE(\hat{\theta}_c) = \frac{(n+2)^2}{(n+1)^2} Var(x_n) + (\frac{\theta}{(n+1)^2})^2 = \frac{(n+2)^2}{(n+1)^2}\frac{n\theta^2}{(n+1)(n+2)^2}$

$+ \frac{\theta^2}{(n+1)^4} = \frac{\theta^2}{(n+1)^4}(n(n+2)+1) = \frac{\theta^2 (n+1)^2}{(n+1)^4} = \frac{\theta^2}{(n+1)^2}$

$\Rightarrow MSE(\hat{\theta}_c) \le MSE(\hat{\theta}_u) \le MSE(\hat{\theta}_{ML})$

intuition : to get MVUE you can multiply your estimator to a value and calculate F.O.C of MSE and find appropriate c, then calculate its bias and variance

(114) Bays estimation — Belif, unknown not unknown

intuition : ① three step procedure ② Variance of $\hat{\theta} \neq Var(\theta)$ data

$\underline{\theta}$ vector of prams

① $p(x|\underline{\theta})$ pdf $\underline{x}|\underline{\theta}$

② $g(\underline{\theta})$ pdf of θ : $\int_{R_{\underline{\theta}}} g(\underline{\theta})d\theta = 1 \Rightarrow$ Prior dist θ (expert judgment)

Bays : $p(\underline{\theta}|\underline{x}) = \frac{p(\underline{x}|\underline{\theta})\cdot g(\underline{\theta})}{h(\underline{x})}$ $h(\underline{x}) = \int_{R_{\underline{\theta}}} p(x|\underline{\theta})g(\underline{\theta})d\theta$

③ Loss function : $L(\underline{\theta}, \hat{\underline{\theta}})$ obj func → minimize estimator

ⓐ Quadratic loss : $L(\underline{\theta}, \hat{\underline{\theta}}) = \min_{\hat{\theta}} E(\hat{\underline{\theta}} - \underline{\theta})^2$ $\boxed{mean}$

$E(\theta - m)^2 = E_\theta(\theta - E(\theta) + E(\theta) - m)^2 = E(\theta - E(\theta))^2 + (E(\theta) - m)^2$ $MSE(\hat{\theta})$

ⓑ linear loss : $L(\theta, \hat{\theta}) = E|\hat{\theta} - \theta|$ $\hat{\theta} = \boxed{median}$ of posterior

if $L(\theta - \hat{\theta}) = \begin{cases} 0 & \text{if } \hat{\theta} = \theta \\ k & \text{if } \hat{\theta} \neq \theta \end{cases}$ $\hat{\theta} = \boxed{mode}$ of posterior

(115) e.g. ① $S_i = \begin{cases} 1 & \pi \\ 0 & 1-\pi \end{cases}$ $S_1, S_2, \ldots, S_n$ Estimate π

$\frac{\partial LL}{\partial \pi} = 0 : \hat{\pi} = \frac{\sum_i S_i}{n} = \frac{x}{n}$: MVUE (min variance unbiased Estimator)

$E(\hat{\pi}) = \pi$ $Var(\hat{\pi}) = \frac{\pi(1-\pi)}{n}$

Natural prior : Pick $\pi \sim Unit(0,1) \Rightarrow g(\pi) = 1$ $0 \le \pi \le 1$

② $A = p(x|\pi)g(\pi) = \frac{n}{x!(n-x)!}\pi^x(1-\pi)^{n-x}$ — Beta dist

$h(x) = \int_0^1 A = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}\int_0^1 \pi^x(1-\pi)^{n-x} d\pi = \frac{\Gamma(n+1)\cdot B(x+1, n-x+1)}{\Gamma(x+1)\Gamma(n-x+1)}$

$\Rightarrow \pi(\theta|x) = \frac{\frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}\cdot\pi^x(1-\pi)^{n-x}}{\frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}B(x+1, n-x+1)} = Beta(x+1, n-x+1)$

③ Pick quadratic loss function :

mean: $\frac{\alpha}{\alpha+\beta} = \frac{x+1}{(x+1)+(n-x+1)} = \frac{x+1}{n+2} \Rightarrow \hat{\pi}_b = \frac{x+1}{n+2}$ $\hat{\pi}_{ML} = \frac{x}{n}$

$\hat{\pi}_b = \frac{n\hat{\pi}_{ML} + 1}{n+2}$ $E(\hat{\pi}_b) = \frac{n\pi + 1}{n+2} = \pi + \underbrace{\left(\frac{1-2\pi}{n+2}\right)}_{Bias}$

intuition : use max likelihood

$Var(\hat{\pi}_b) = \frac{n}{(n+2)^2} Var(\hat{\pi}_{ML}) = \frac{n^2}{(n+2)^2}\cdot\frac{\pi(1-\pi)}{n} = \frac{n\pi(1-\pi)}{(n+2)^2}$

$MSE(\hat{\pi}_b) = \frac{n\pi(1-\pi)}{(n+2)^2} + \frac{(1-2\pi)^2}{(n+2)^2} = \frac{n\pi - n\pi^2 + 1 - 4\pi + 4\pi^2}{(n+2)^2}$

$\Rightarrow MSE(\hat{\pi}_b) \le MSE(\hat{\pi}_{ML})$ $\pi \in \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{n+1}{2n+1}}$

intuition : everything is around
① plug variables back to Equations you used in Econ, or microEcon e.g. $u(P_{rw})$
② use Price as signal of opportunity cost in Erdem's paper, although it is used for another reason somewhere else

(115) Conjugate priors

Jeffrey prior or uniform $(0, \infty)$

$X \sim U(0, \theta)$ $f(x|\theta) = \frac{1}{\theta}$ $0 \le x \le \theta \Rightarrow f(\underline{x}|\theta) = \frac{1}{\theta^n}$ $0 \le x_n \le \theta$

$\theta \sim U(x_{max}, N) \Rightarrow g(\theta) = \frac{1}{N - x_{max}}$ $x_{max} \le \theta \le N$

① $p(\theta|x)\cdot g(\theta) = \frac{1}{\theta^n}\frac{1}{(N - x_{max})}$ $x_{max} \le \theta \le N$

$h(x) = \int_{x_{max}}^N p(\theta|x)g(\theta)d\theta = \frac{1}{(N - x_m)}\frac{1}{n-1}\int_{x_n}^N \frac{(n-1)}{\theta^n}d\theta$

$= \frac{1}{N - x_m}\frac{1}{n-1}\left[\frac{1}{\theta^{n-1}}\right]_{x_m}^N = \frac{1}{N - x_n}\cdot\frac{1}{(n-1)}\left[\frac{1}{x_n^{n-1}} - \frac{1}{N^{n-1}}\right]$

② $P(\theta|x) = \frac{n-1}{\left[\frac{1}{x_n^{n-1}} - \frac{1}{N^{n-1}}\right]}\frac{1}{\theta^n}$ $x_n \le \theta \le N$

Quadratic loss ≡ $\hat{\theta}_n$: mean of posterior dist.

$\hat{\theta}_b = c\int_{x_m}^N \theta\frac{1}{\theta^n}d\theta = \frac{c}{n-2}\left[-\frac{1}{\theta^{n-1}}\Big|_{x_n}^N\right] = \frac{c}{n-2}\left[-\frac{1}{\theta^{n-2}}\Big|_{x_n}^N\right]$

③ $= \frac{c}{n-2}\left[\frac{1}{x_n^{n-2}} - \frac{1}{N^{n-2}}\right]$

$\hat{\theta}_b = \frac{n-1}{n-2}\frac{\left[\frac{1}{x_n^{n-2}} - \frac{1}{N^{n-2}}\right]}{\left[\frac{1}{x_n^{n-1}} - \frac{1}{N^{n-1}}\right]} = \frac{n-1}{n-2}\frac{x_m^{n-1}}{x_m^{n-2}}\frac{\left[1 - \left(\frac{x_m}{N}\right)^{n-2}\right]}{\left[1 - \left(\frac{x_m}{N}\right)^{n-1}\right]}$

$\Rightarrow \lim_{N\to\infty}\hat{\theta}_b = \frac{(n-1)}{(n-2)}x_m$ max likelihood est

$\hat{\theta}_b = \frac{(n-1)}{(n-2)}x_m \Rightarrow E(\hat{\theta}_b) = \frac{n-1}{n-2} \times \overline{E(x_m)} = \frac{n(n-1)}{(n-2)(n+1)}\cdot\theta$

intuition : ① calculate estimator based on data and data based on another estimator to compare estimators
② for bayesian follow 3 step procedure

$Bias(\hat{\theta}_b) = \theta\left|\frac{n(n-1)}{(n-2)(n+1)} - 1\right| = \frac{2\theta}{(n-2)(n+1)} \Rightarrow Bias(\hat{\theta}_{ML}) > Bias(\hat{\theta}_b)$

also $Bias(\hat{\theta}_{ML}) = \frac{\theta}{n+1}$

$MSE(\hat{\theta}_b) = \frac{(n-1)^2}{(n-2)^2}\cdot\frac{n\theta^2}{(n+2)(n+1)^2} + \frac{4\theta^2}{(n-2)^2(n+1)^2}$

$= \frac{\theta^2}{(n-2)(n+1)^2(n+2)}[n^3 - 2n^2 + 5n + 8]$

$MSE(\hat{\theta}_c) \le MSE(\hat{\theta}_u) \le MSE(\hat{\theta}_b) \le MSE(\hat{\theta}_{ML})$

$\underbrace{\frac{(n+1)}{n}x_m}_{\hat{\theta}_u}$ $\underbrace{x_m}_{\hat{\theta}_{ML}}$ $\underbrace{\frac{n-1}{n-2}x_m}_{\hat{\theta}_b}$ $\underbrace{\frac{n+2}{n+1}x_m}_{\hat{\theta}_c}$

Statistics summary (21) (22)

(16) $g(\theta)$ pdf θ

$$p(\theta|\underline{x}) = \frac{p(\underline{x}|\theta)\, g(\theta)}{h(x)}$$ $p(x|\theta)$ shared with max likelihood

$$h(x) = \int_{R(\theta)} p(x|\theta) \cdot g(\theta)\, d\theta$$

$$\Rightarrow \ln p(\theta|x) = \ln p(x|\theta) + \ln g(\theta) - \ln h(\theta)$$

$$\Rightarrow \ln p(\theta|\underline{x}) = \Sigma_{i=1}^{n} \ln p(x_i|\theta) + \ln g(\theta) - \ln(h(\underline{x}))$$

$$= \left(n \frac{\Sigma_{i=1}^{n} \ln p(x_i|\theta)}{n}\right) + \ln g(\theta) + \ln h(x)$$

$$= \underbrace{n\, E[\ln p(x_i|\theta)]}_{\nearrow n} + \ln g(\theta) - \ln h(x)$$

$\equiv$ dominate as sample big enough

---

1. frequentist
2. peragmatist : use prior if you have it
3. Bays : Can handle complicated loss function
   - no need for analytical solution
   - don't have to check for regularity

(1) Freq

observe $x_1 \sim B(n_1, \pi)$ later do it again $x_2 \sim B(n_2, \pi)$

first x_1 later x_2

$$x_1 + x_2 \sim B(n_1+n_2, \pi) \qquad \hat{\pi}_1 = \frac{x_1+x_2}{n_1+n_2}$$

$$E(\hat{\pi}_1) = \frac{E(x_1)+E(x_2)}{n_1+n_2} = \frac{n_1\pi + n_2\pi}{n_1+n_2} = \pi$$

$$V(\hat{\pi}_1) = \frac{1}{(n_1+n_2)^2}\left(\underset{\text{indep}}{Var(x_1) + Var(x_2)}\right) = \frac{1}{(n_1+n_2)^2}\left(n_1\pi(1-\pi) + n_2\pi(1-\pi)\right) = \frac{\pi(1-\pi)}{n_1+n_2}$$

$$\Rightarrow \hat{\pi}_1 = \frac{x_1+x_2}{n_1+n_2} = \frac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1+n_2} = \left(\frac{n_1}{n_1+n_2}\right)\hat{P}_1 + \left(\frac{n_2}{n_1+n_2}\right)\hat{P}_2 \qquad \hat{P}_i = \frac{x_i}{n_i}$$

(2) Bayesian

$$f(x_1, x_2|n_1, n_2, \pi) \propto \pi^{x_1}(1-\pi)^{n_1-x_1}\, \pi^{x_2}(1-\pi)^{n_2-x_2}$$

$$\propto \pi^{x_1+x_2}(1-\pi)^{n_1+n_2-x_1-x_2}$$

$$g(\pi) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}$$

$$g(\pi|x_1, x_2) \propto \pi^{x_1+x_2+\alpha-1}(1-\pi)^{n_1+n_2-x_1-x_2+\beta-1}$$

$\hookrightarrow$ since as you integ. cancel out

For Quadratic loss

$$\Rightarrow \hat{\pi}_3 = E(\pi|\text{posterior}) = \frac{x_1+x_2+\alpha}{n_1+n_2-x_1-x_2+\beta+x_1+x_2+\alpha}$$

$$= \frac{x_1+x_2+\alpha}{n_1+n_2+\alpha+\beta} = \frac{(n_1+n_2)\hat{\pi}_1 + \alpha}{n_1+n_2+\alpha+\beta}$$

intuition: if you do not have close dist. for Baysian integrate manually to get $E(\hat{\theta})$ otherwise use the distrib. formula mean for Quadratic Loss

$$Bias = E(\hat{\pi}_3 - \pi) = E\left(\frac{(n_1+n_2)\hat{\pi}_1 + \alpha}{n_1+n_2+\alpha+\beta} - \pi\right) = \left|\frac{(n_1+n_2)\pi + \alpha}{n_1+n_2+\alpha+\beta} - \pi\right|$$

$$= \left|\frac{\alpha - \beta\pi - \alpha\pi}{n_1+n_2+\alpha+\beta}\right| = \left|\frac{\alpha(1-\pi) - \beta\pi}{n_1+n_2+\alpha+\beta}\right|$$

$$Var(\hat{\pi}_3) = \frac{(n_1+n_2)^2}{(n_1+n_2+\alpha+\beta)^2} \times Var(\hat{\pi}_1) = \frac{\pi(1-\pi)(n_1+n_2)}{(n_1+n_2+\alpha+\beta)^2}$$

intuition: put x based on θ to calculate Bias MSE and Variance of estimator, and for Variance just look at coeff. since intercept will cancel out

e.g. $\alpha = 1, \beta = 1$ $\quad \hat{\pi}_3 = \frac{x_1+x_2+1}{n_1+n_2+2}$ $\quad (Bias)^2 = \frac{(1-2\pi)^2}{(n_1+n_2+2)^2}$

$$MSE(\hat{\pi}_3) = \frac{\pi(1-\pi)(n_1+n_2)}{(n_1+n_2+2)^2} + \frac{(1-2\pi)^2}{(n_1+n_2+2)^2}$$

Peragmatist

$$\hat{P}_1 = \frac{x_1}{n_1}$$

$$\hat{P}_1 = \frac{\alpha}{\alpha+\beta} \qquad V = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}$$

(I) $\hat{\alpha} = \hat{P}_1(n_1-1)$
$\hat{\beta} = (1-\hat{P}_1)(n_1-1)$ $\Rightarrow$ Prior $\propto \pi^{\hat{\alpha}-1}(1-\pi)^{\hat{\beta}-1}$

$$\pi | n_2, x_2, \hat{\alpha}, \hat{\beta} \propto \pi^{x_2+\hat{\alpha}-1}(1-\pi)^{n_2-x_2+\hat{\beta}-1}$$ resulted (I)

$$\Rightarrow \hat{\pi}_2 = \frac{x_2+\hat{\alpha}}{x_2+x_2+\hat{\alpha}-x_2+\hat{\beta}} = \frac{x_2+\hat{\alpha}}{n_2+\hat{\alpha}+\hat{\beta}} = \frac{x_2+\hat{\alpha}}{n_1+n_2-1}$$

$$\Rightarrow \hat{\pi}_2 = \frac{x_2 + \left(\frac{x_1}{n_1}\right)(n_1-1)}{n_1+n_2-1} = \frac{n_2\hat{P}_2 + (n_1-1)\hat{P}_1}{n_1+n_2-1}$$

$$= \frac{(n_1-1)\hat{P}_1}{(n_1+n_2-1)} + \frac{n_2\hat{P}_2}{(n_1+n_2-1)} \qquad \hat{P}_i = \frac{x_i}{n_i}$$

$\hat{\pi}_2$ is unbiased $E(\hat{\pi}_2) = \pi$

$$\hat{\pi}_1 = \frac{n_1\hat{P}_1}{n_1+n_2} + \frac{n_2\hat{P}_2}{n_1+n_2}$$

intuition: get estimate based on first sample and plug in to Bayes update, rather than use all data in Bayes

(17) Class of Estimators : $\hat{\hat{\pi}} = w\hat{P}_1 + (1-w)\hat{P}_2 \quad (w \geq 0)$

$$\text{indep} \Rightarrow Var(\hat{\pi}) = w^2 Var(\hat{P}_1) + (1-w)^2 Var(\hat{P}_2)$$

$$\frac{\partial Var(\hat{\pi})}{\partial w} = 2w\, Var(\hat{P}_1) - 2(1-w) Var(\hat{P}_2) = 0$$

$$\frac{\partial^2 Var(\hat{\pi})}{\partial w^2} = 2Var(\hat{P}_1) + 2Var(\hat{P}_2) > 0$$

$$\Rightarrow w \cdot Var(\hat{P}_1) - (1-w) Var(\hat{P}_2) = 0 \Rightarrow w(Var(\hat{P}_1) + Var(\hat{P}_2)) = Var(\hat{P}_2)$$

$$\Rightarrow w = \frac{Var(\hat{P}_1)}{Var(\hat{P}_1) + Var(\hat{P}_2)} = \frac{\frac{1}{n_2}}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{n_1}{n_1+n_2}$$

$\Rightarrow$ weights $\frac{n_1}{n_1+n_2}$ $\quad \frac{n_2}{n_1+n_2}$ $\Rightarrow$ best estimator: $\hat{\pi}_1 = \frac{n_1}{n_1+n_2}\hat{P}_1 + \frac{n_2\hat{P}_2}{n_1+n_2}$ $\{Var(\hat{\pi}_2)$

---

non parametric statistics

- remove assump of know something (gamma, beta, ...)
- Good for hypoth testing & not Estimation { Jack knife, Boot strap

(18) $x_1, x_2, \ldots, x_n \sim$ iid F(cdf)

$\hat{\theta} = (x_1, \ldots, x_n)$ $\quad Bias(\hat{\theta}) = E_F(\hat{\theta} - \theta)$

remove data point i : $\hat{\theta}_{(i)} = \hat{\theta}(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$

n-times: $\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(n)}$

$$\hat{\theta}_{(\cdot)} = \Sigma_{i=1}^{n} \frac{\hat{\theta}_{(i)}}{n} \qquad Bias(\hat{\theta}) = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta})$$

Correct by subtracting Bias: $\hat{\theta} - \widehat{Bias} = \hat{\theta} - (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta})$
$= n\hat{\theta} - (n-1)\hat{\theta}_{(\cdot)}$

Estimate μ^2, $x_1, x_2, \ldots, x_n$

(1) $\hat{\mu}^2 = \bar{x}^2$ (2) $Var(\bar{x}) = E(\bar{x}^2) - E^2(\bar{x})$

(3) $\frac{\sigma^2}{n} = E(\bar{x}^2) - \mu^2$ (4) $E(\hat{\mu}^2) = E(\bar{x}^2) = \mu^2 + \frac{\sigma^2}{n} \Rightarrow$ Biased

Statistics summary

(119) $\theta = \mu \quad \hat{\theta} = \bar{x}$

$$\hat{\theta}_{(i)} = \frac{\sum_{j\neq i} x_j}{(n-1)} = \frac{\sum_{j=1}^{n} x_j - x_i}{n-1}$$

$$\hat{\theta}_{(\cdot)} = \frac{\sum_{i=1}^{n} \hat{\theta}_{(i)}}{n} = \frac{\sum_{i=1}^{n} (\sum_{j=1}^{n} x_j - x_i)}{n(n-1)} = \frac{\sum_{i=1}^{n}(n\bar{x} - x_i)}{n(n-1)}$$

$$= \frac{\sum_{i=1}^{n} \bar{x} - \sum_i x_i}{n-1} = \frac{n\bar{x}}{n-1} - \frac{\bar{x}}{n-1} = \frac{(n-1)\bar{x}}{(n-1)} = \bar{x} \quad \text{Bias} = 0$$

Correction = 0

(120) Variance $\quad \hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} = \frac{1}{n}\left[\sum_{i=1}^{n} x_j^2 - \frac{(\sum_j x_j)^2}{n}\right] = \hat{\theta}$

$$\Rightarrow \hat{\theta}_{(i)} = \frac{1}{n-1}\left[\sum_{j\neq i} x_j^2 - \frac{(\sum_{j\neq i} x_j)^2}{n}\right]$$

$$\hat{\theta}_{(\cdot)} = \frac{\sum_i \hat{\theta}_{(i)}}{n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\sum_{j\neq i} x_j^2 - \frac{(\sum_{i\neq j} x_j)^2}{n}\right)$$

$$\sum_{i=1}^{n} \sum_{i\neq j} x_j^2 = \sum_{i=1}^{n} \left(\sum_{i=1}^{n} x_j^2 - x_i^2\right) = n\sum_{j=1}^{n} x_j^2 - \sum_{i=1}^{n} x_i^2$$

$$= (n-1)\sum_i x_i^2 \Rightarrow n\bar{x}^2 = \left(\sum_{j=1}^{n} x_j\right)^2 = \left(\sum_{i\neq j} x_j + x_i\right)^2$$

$$= \left(\sum_{i\neq j} x_j\right)^2 + 2x_i \sum_{j\neq i} x_j + x_i^2 = \left(\sum_{j\neq i} x_j\right)^2 + 2x_i\left(\sum_{j=1}^{n} x_j - x_i\right)$$

$$+ x_i^2 = \left(\sum_{j\neq i} x_j\right)^2 + 2n x_i \bar{x} - x_i^2$$

$$\Rightarrow \left(\sum_{j\neq i} x_j\right)^2 = n\bar{x}^2 - 2\bar{x}x_i - x_i^2$$

$$\hat{\theta}_{(\cdot)} = \frac{1}{n(n-1)}\left[(n-1)\sum_i x_i^2 - \frac{n^2(n-2)}{n-1}\bar{x}^2 - \frac{\sum_{i=1} x_i^2}{n-1}\right]$$

$$= \frac{\sum x_i^2}{n} - \frac{n(n-2)\bar{x}^2}{(n-1)^2} - \frac{\sum_i x_i^2}{n(n-1)^2}$$

$$\Rightarrow \hat{\theta} = \frac{\sum_i^2}{n} - \bar{x} \Rightarrow \hat{\theta}_{(\cdot)} - \hat{\theta} = \bar{x}^2 - \frac{n(n-2)}{(n-1)^2}\bar{x}^2 - \frac{\sum_i x_i^2}{n(n-1)^2}$$

$$= \frac{\bar{x}^2[n^2 - 2n + 1 - n^2 + 2n]}{(n-1)^2} - \frac{\sum_i x_i^2}{n(n-1)^2} = \frac{-1}{(n-1)^2}\left[\frac{\sum_i x_i^2}{n} - \bar{x}^2\right]$$

$$= -\frac{1}{n(n-1)}\sum_i (x_i - \bar{x})^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}\left[1 + \frac{1}{n-1}\right] = \frac{\sum_i (x_i - \bar{x})^2}{n-1}$$

(120) for any Quadratic func $\quad$ Bias $= (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta})$ is ~~unbiased~~ unbiased

$$\operatorname{Var}(\bar{x}) = \frac{\sigma^2}{n}, \quad \widehat{\operatorname{Var}}(\bar{x}) = \frac{\hat{s}^2}{n} = \frac{\sum_i (x_i - \bar{x})^2}{n(n-1)}$$

$$\bar{x}_{(i)} = \frac{\sum_{j=1} x_j}{(n-1)} = \frac{\sum_j x_j - x_i}{n-1} = \frac{n\bar{x} - x_i}{(n-1)}$$

$$\bar{x}_{(\cdot)} = \frac{\sum_i \bar{x}_{(i)}}{n} = \bar{x}$$

$$\sum_{i=1}^{n} (\bar{x}_{(i)} - \bar{x}_{(\cdot)})^2 = \sum_{i=1}^{n}\left[\frac{n\bar{x} - x_i}{n-1} - \bar{x}\right]^2 = \sum_{i=1}^{n} \frac{[n\bar{x} - x_i - n\bar{x} + \bar{x}]^2}{(n-1)^2}$$

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

for any stat $\hat{\theta}$ estimate θ:

$$V(\hat{\theta}) = \frac{(n-1)}{n}\sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2$$

$$\widehat{\operatorname{Cov}}(\hat{\theta}) = \frac{(n-1)}{n}\sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta})(\hat{\theta}_{(i)} - \hat{\theta})'$$

(121) Factorization theorem:

$f(\underline{x} \mid \theta)$ pdf $\underline{x} \Rightarrow T(\underline{x})$ suff stat of θ $\quad \forall x, \theta$

if $\exists\, g(\theta \mid \theta) + h(x) \quad$ s.t $\quad f(x, \theta) = g(T(x) \mid \theta) \cdot h(x)$

(122) e.g Exponential family of dist:

$$f(\underline{x} \mid \underline{\theta}) = h(\underline{x}) \cdot c(\underline{\theta}) \cdot \exp\left[\sum_i w_i(\underline{\theta}) * t_i(\underline{x})\right]$$

$$\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)' \quad T(x) = \left(\sum_{i=1}^{n} t_1(x_j), \sum_{j=1}^{n} t_2(x_j), \ldots, \sum_{j=1}^{n} t_p(x_j)\right)$$

$\boxed{\text{Intuition}}$ math is word of Equality → e.g indiff curve Game theory & Micro

(123) Variance of MLE:

$$\operatorname{var}(\theta) = [I(\theta)]^{-1} \Rightarrow [I(\theta)] = -E[H(\theta)]$$

$$H(\theta) = \frac{\partial^2 \ln L(\theta)}{\partial\theta\,\partial\theta'} = \text{hessian}$$

$$\operatorname{Var}(\theta) = (I(\theta))^{-1} = (-E(H(\theta)))^{-1} = \left(-E\left[\frac{\partial^2 \ln L(\theta)}{\partial\theta\,\partial\theta'}\right]\right)^{-1}$$

e.g $\quad L(\beta) = \prod^{n} \frac{e^{-y_i/\beta x_i}}{\beta x_i} \Rightarrow \log L(\beta) = -\frac{1}{\beta}\sum \frac{y_i}{x_i} - n\log\beta$

$$-\sum_{i=1}^{n} \log x_i \Rightarrow \frac{\partial L}{\partial \beta} = 0 : \hat{\beta}_{MLE} = \sum_{i=1}^{n} \frac{y_i}{x_i} \quad ①$$

$$V(\beta) = \left[-\frac{\partial^2 L}{\partial \beta^2}\right]^{-1} = \left(\frac{-2}{\beta^3}\sum \frac{y_i}{x_i} + \frac{n}{\beta^2}\right)^{-1} \overset{①}{=} -\left(\frac{2n\beta}{\beta^3} + \frac{n}{\beta^2}\right)^{-1} = \frac{\beta^2}{n}$$

① $y_1 - y_n \sim \operatorname{Unif}(0, \theta) \quad \max(y_i) = y_m$ sufficient

Conf. interval $\left(y_n, \frac{y_n}{(1-\gamma)^{1/n}}\right)$

$$\Pr(y_n < \theta) = \prod_{i=1}^{n} \Pr(x_i < c) = \left(\frac{c}{\theta}\right)^n$$

Conf interval: P% e.g %95 $\quad \left(y_n, \frac{y_n}{P^{1/n}}\right)$

reason: $1 - P_n(x_n < \theta(1-\gamma)^{\frac{1}{n}}) = \Pr(\theta(1-\gamma)^{\frac{1}{n}} < y_n)$

$$= \Pr\left(\theta < \frac{y_n}{(1-\gamma)^{\frac{1}{n}}}\right)$$

$\boxed{\text{intuition}}$: think in reverse and after calculating θ based on x_m then calculate x_n based on θ ① $\underline{x_{max}}$ based on $\underline{\theta}$ ② $\underline{\theta}$ based on x_{max}

$\boxed{\equiv \text{Endogeneity}}$

$$P(x_{max} < \theta < f(x_{max})) = 1 - \gamma$$

$\boxed{\text{math word of Equality look at C, not } x_m}$

$$\Rightarrow P(x_{max} < C) = \left(\frac{C}{\theta}\right)^n = \gamma \Rightarrow C = \gamma^{\frac{1}{n}} \times \theta$$

since we want reverse

$$\Rightarrow P(x_m < \gamma^{\frac{1}{n}}\theta) = \gamma \Rightarrow 1 - P\left(\frac{x_m}{\gamma^{\frac{1}{n}}} < \theta\right) = P\left(\theta < \frac{x_m}{\gamma^{\frac{1}{n}}}\right) = 1 - \gamma$$

② to calculate ML, and bayesian estimator for θ param unif

① ML: $\hat{\theta}_{ML} = x_{max} \quad F(x_{max}) = F(x_1 - x_n < x_{max} = z) = \left(\frac{z}{\theta}\right)^n$

$$\Rightarrow f(x_{max} = z) = n\left(\frac{z}{\theta}\right)^{n-1} \Rightarrow E(z) = \int z f(z)\,dz = \frac{n}{n+1}\theta$$

$$V(z) = \frac{n\theta^2}{(n+2)(n+1)^2} \Leftarrow E(z^2) = \int z^2 f(z)\,dz = \frac{n\theta^2}{(n+2)}$$

$$\Rightarrow \text{Bias} = \frac{\theta}{n+1} \Rightarrow MSE = \frac{2\theta^2}{(n+2)(n+1)}$$

② Bayesian: ① prior: $\frac{1}{N - x_{max}}$

② $P(x \mid \theta) \cdot g(\theta) = \frac{1}{N - x_{max}} \times \frac{1}{\theta^n}$

$$h(x) = \int_{x_{max}}^{N} \frac{1}{N - x_{max}} \frac{1}{\theta^n}\,d\theta = \frac{1}{(N - x_m)(n-1)}\left(\frac{1}{x_m^{n-1}} - \frac{1}{N^{n-1}}\right)$$

$$\Rightarrow P(\theta \mid x) = \frac{P(x \mid \theta) \cdot g(\theta)}{h(x)} = \frac{n-1}{\left(\frac{1}{x_m^{n-1}} - \frac{1}{N^{n-1}}\right)} \frac{1}{\theta^n}$$

$$\Rightarrow \theta_b = E(\theta) = c\int_{x_{max}}^{N} \frac{1}{\theta^n}\theta\,d\theta = \frac{n-1}{n-2} \frac{\left[\frac{1}{N^{n-2}} - \frac{1}{x_{max}^{n-2}}\right]}{\left[\frac{1}{x_n^{n-1}} - \frac{1}{N^{n-1}}\right]}$$

$$\Rightarrow \underset{N \to \infty}{\theta_n} = \frac{n-1}{n-2} x_m$$

From MLE calc

$$E(\theta_b) = \frac{n-1}{n-2} E(x_m) = \frac{n-1}{n+1} \times \frac{n}{n-2}\theta$$

$\boxed{\text{intuition}}$ $\boxed{\equiv \theta / x_m}$

$$\operatorname{Var}(\theta_b) = \frac{n-1}{n-2} V(x_m) = \left(\frac{n-1}{n+1}\right)^2 \frac{n\theta^2}{(n+1)(n+1)^2}$$

$\boxed{\text{intuition}}$ Just like Econometric that you can write Error based on depen, indep var, in stat you can write param based on data & vice versa $\boxed{\text{Math = word of Equity}}$