统计学

一、统计与数据基础

1.1 • 统计学 statistics　统计数据　分类数据 categorical data（有序/无序）（qualitative）；数值数据 metric ~（离散/连续）（quantitative）

观测数据 observational ~　实验数据 experiemental ~　截面数据 cross-sectional ~

时间序列数据 time series ~　描述统计 descriptive ~　推断统计 inferential ~ $\mu, \sigma, \pi \longleftrightarrow \bar{x}, s, p$

• 总体 population　样本 sample　样本量 sample size　参数 parameter 统计量 statistic

变量 variable（categorical ~/metric ~）

1.2 • 数据来源：间接（二手）、直接 < 调查 / 实验

• 概率抽样 probability sampling：简单随机抽样 simple random ← 抽样框 sampling frame、

分层抽样 stratified、整群抽样 cluster、系统抽样 systematic、多阶段抽样 multi-stage

非概率抽样 non- ~：方便抽样、判断抽样、自愿样本、滚雪球抽样、

配额抽样　　数据搜集方法：自填、面访、电话（CATI）及其特点。

• 实验组 experiment group　对照组 control ~　双盲法　统计方法 → 实验设计（有效性）

• 数据误差：抽样误差 sampling error（仅 probability 中）；非抽样误差 non- ~：内部因果　外部推广

抽样框误差（仅 pro 中）、回答误差、无回答误差、调查员误差、测量误差

1.3 • 数据预处理 data preprocessing：~清洗 cleaning（审核）、~集成 integration、~归约 reduction、

~转换 transformation

• 数据整理与展示：频数分布 frequency distribution　列联表 contingency table

条/柱形图 bar/column chart　帕累托图 Pareto chart　饼图 pie chart　环形图；

数据分组 grouped　组中值 class midpoint　直方图 histogram　箱图 box plot　须线 whiskers

线图 line plot　散点图 scatter diagram　雷达图 radar chart

1.4 ∘ 数据的度量（描述、概括）：集中趋势 central tendency、离散程度、分布形状

平均数/均值 mean $\bar{x}=\frac{\sum x_i}{n}$ simple mean $\bar{x}=\frac{\sum_i m_i f_i}{n}$ weighted mean 加权~

分位数 quantile 中位数 median M_e 四分位数 quartile（SPSS中 $\frac{n+1}{4}$、$\frac{3}{4}(n+1)$）

众数 mode M_o 三种数之选择.

全距 range 四分位距 inter-quartile range 平均差（平均绝对离差）mean deviation

方差 variance 标准差 standard deviation $s^2=\frac{\sum_i(x_i-\bar{x})^2}{n-1}$ $\sigma^2=\frac{\sum_i(x_i-\mu)^2}{N}$

（变异）离散系数 CV coefficient of variation $=\frac{s}{\bar{x}}$ 标准化 standardize $z_i=\frac{x_i-\bar{x}}{s}$

偏度 skewness（偏度系数 SK）$=\sum_i\left(\frac{x_i-\bar{x}}{s}\right)^3$ 左偏/右偏 峰度 kurtosis（~K）$=\sum_i\left(\frac{x_i-\bar{x}}{s}\right)^4-3$

二、概率与抽样分布

2.1 ∘ 试验 事件 随机事件（必然~ 不可能~）random event 基本事件 elementary event

概率 probability 古典定义 统计定义 主观概率定义

随机变量 random variable 概率函数 probability function 离散型/连续型

概率分布 probability distribution （密度）分布函数 density $F(x)$ $P(X)$ $f(X)$ discrete continous $(f(x)\geq 0, \int_{-\infty}^{+\infty}f(x)dx=1)$

期望（值）expected value $E(X)$ (μ) 方差 $D(X)/V(X)$ (σ^2) $(E(X)=\int_{-\infty}^{+\infty}xf(x)dx, D(X)=\int_{-\infty}^{+\infty}(x-\mu)^2f(x)dx)$

均匀分布 uniform~ 二项分布 binomial~ $P(X)=C_n^x p^x q^{n-x}$ $(p+q=1)$, $E(X)=np$, $D(X)=npq$

泊松分布 Poisson~ $P(X)=\frac{\lambda^x e^{-\lambda}}{x!}$, $E(X)=\lambda$, $D(X)=\lambda$

正态分布 normal~ $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $X\sim N(\mu,\sigma^2)$ 标准正态分布 (Z) $N(0,1)$ $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

"6σ 准则"（$6\sigma: 2\times 10^{-9}$, $3\sigma: 0.27\%$）$\phi(x)=1-\phi(-x)$ "1.5σ 漂移"（$\to 3.4\times10^{-6}$）

2.2 ∘ 统计量 $T(X_1,\cdots,X_n)$ 样本k阶距 $m_k=\frac{1}{n}\sum_i X_i^k$ 样本k阶中心距 $v_k=\frac{\sum(X_i-\bar{X})^k}{n-1}$

抽样分布 sampling distribution χ^2分布 Chi-square~ $X_1,\cdots,X_n$ iid. $\sim N(0,1)$ $\sum_i X_i^2\sim\chi^2(n)$

t分布 $X\sim N(0,1)$, $Y\sim\chi^2(n)$, $\frac{X}{\sqrt{\frac{Y}{n}}}\sim t(n)$ $E(t)=0$ $D(t)=\frac{n}{n-2}$ $E(\chi^2)=n$ $D(\chi^2)=2n$

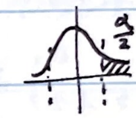
F分布 $Y, Z\sim\chi^2(m)$、$\chi^2(n)$, $X=\frac{Y/m}{Z/n}\sim F(m,n)$ $E(X)=\frac{n}{n-2}$ $D(X)=\frac{2n^2(m+n-2)}{m(n-2)(n-4)}$ $(n>4)$

$X\sim t(n)\leftrightarrow X^2\sim F(1,n)$ 分位数 $F_p(m,n)=\frac{1}{F_{1-p}(n,m)}$

• 马尔科夫不等式 $P(X\geq a)\leq \frac{E(X)}{a}$ 切比雪夫不等式 $P(|X-\mu|\geq k\sigma)\leq\frac{1}{k^2}$

中心极限定理 central limit theorem $\lim_{n\to\infty}\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$ （大样本：30<）

三、参数估计与假设检验



3.1 • 参数估计 parameter estimation 估计量 estimator $\hat\theta$ 估计值 estimated value

点估计 point~ 区间估计 interval~ 置信区间 confidence interval 置信上/下限

置信度（水平、系数）confidence level $1-\alpha$ 显著性水平 significance level α

评价估计量的标准：无偏性 unbiasedness $E(\hat\theta)=\theta$、有效性 efficiency $D(\hat\theta_1)<D(\hat\theta_2)$、一致性 consistency $\lim_{n\to\infty}\hat\theta=\theta$

• 均值估计：1. N总体，σ^2已知，大/小样本

2. 非N总体，大样本 且大样本时 $s^2\to\sigma^2$.

$z=\frac{\bar x-\mu}{\sigma/\sqrt n}\sim N(0,1)$

$\mu\in\bar x\pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt n}$ （$\pm z_{\frac\alpha2}\frac{s}{\sqrt n}$）

2. N总体，σ^2未知，小样本 $t=\frac{\bar x-\mu}{s/\sqrt n}\sim t(n-1)$

$\mu\in\bar x\pm t_{\frac\alpha2}\frac{s}{\sqrt n}$

比例估计：(大样本) $p\to\pi$ $\sigma_p^2=\frac{\pi(1-\pi)}{n}$ $z=\frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}\sim N(0,1)$

$\pi\in p\pm z_{\frac\alpha2}\sqrt{\frac{\pi(1-\pi)}{n}}$ （$\pm z_{\frac\alpha2}\sqrt{\frac{p(1-p)}{n}}$）

方差估计：(N总体) $\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$

$\sigma^2\in[\frac{(n-1)s^2}{\chi^2_{\frac\alpha2}},\frac{(n-1)s^2}{\chi^2_{1-\frac\alpha2}}]$

• 均值差估计：独立样本 independent sample

1. 大样本或σ^2已知 $z=\frac{(\bar x_1-\bar x_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$ $\mu_1-\mu_2\in\bar x_1-\bar x_2\pm z_{\frac\alpha2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ （$\pm z_{\frac\alpha2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$）

2. 小样本、σ^2未知（N总体）

$\sigma_1^2=\sigma_2^2$时：$s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

$\mu_1-\mu_2\in\bar x_1-\bar x_2\pm t_{\frac\alpha2}(n_1+n_2-2)\sqrt{s_p^2(\frac{1}{n_1}+\frac{1}{n_2})}$

$\sigma_1^2\neq\sigma_2^2$时：自由度 $v=\frac{(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2})^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$

$\mu_1-\mu_2\in\bar x_1-\bar x_2\pm t_{\frac\alpha2}(v)\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

匹配样本 matched sample （信息要优于独立下）

$\mu_d=\mu_1-\mu_2\in\bar d\pm z_{\frac\alpha2}\frac{\sigma_d}{\sqrt n}$ （$\pm t_{\frac\alpha2}(n-1)\frac{s_d}{\sqrt n}$）

比例差估计：($p\to\pi$) $\pi_1-\pi_2\in p_1-p_2\pm z_{\frac\alpha2}\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$

方差比估计：$\frac{s_1^2}{s_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}\sim F(n_1-1,n_2-1)$

$\frac{\sigma_1^2}{\sigma_2^2}\in[\frac{s_1^2/s_2^2}{F_{\alpha/2}},\frac{s_1^2/s_2^2}{F_{1-\frac\alpha2}}]$

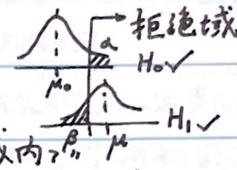
• 确定样本量（反推）：（均值）$n=\frac{(z_{\frac\alpha2})^2\sigma^2}{e^2}$ （或s^2）

（比例）$n=\frac{(z_{\frac\alpha2})^2\pi(1-\pi)}{e^2}$ （或p）$\leq\frac{z_{\frac\alpha2}^2}{4e^2}$

3.2 ◦ 假设检验 hypothesis testing 原假设 null hypothesis 备择假设 alternative~

α错误（弃真错误） β错误（取伪错误）

	H_0	H_a
H_0 ✓	$1-\alpha$	α △
H_0 ✗	β △	$1-\beta$



假设检验的流程："H_0假定，统计量是否落在拒绝域内？"

如 $|z|<|z_{\frac{\alpha}{2}}|$ ✓，$|z|>|z_{\frac{\alpha}{2}}|$ ✗ 或利用P值 ⟶ 伴随（相伴）概率

单侧（尾）/双侧（尾）检验 左/右单侧检验（下/上限检验）

检验统计量的确定：样本量？（N总体）{ 大：z ；小：σ^2？{ 已知：z ；未知：t } } 方差：χ^2（单尾）

两个总体N { 均值差 { z ；小未：t } ；比例差：z ；方差比：F }

检验解释：H_0 ✗ (H_1 ✓)："H_1为真出错的概率不超过α"

H_0 ✓："没有充足证据（在α的显著水平下）拒绝H_0" "α可理解为苛刻的程度，或相同(H_0)显著的程度"。

"H_0常为不轻易否定的（原有、传统的）命题，要证明的常置于备择位(H_1)"。

四、分类数据分析与方差分析

4.1 ◦ χ^2检验 < 拟合优度检验 goodness of fit test（适合性检验） / 独立性检验 independence test（列联分析）

$$\chi^2=\sum\frac{(f_o-f_e)^2}{f_e}$$ "[20%]的单元 $f_e<5$ 不宜使用"。

观察值频数 observed frequency f_o 期望值频数 expected frequency f_e

适合性检验：$df=n-1$ 独立性检验：$df=(R-1)(C-1)$，$f_e=n\cdot\frac{RT}{n}\cdot\frac{CT}{n}=\frac{RT\cdot CT}{n}$

相关系数：1. $\varphi=\sqrt{\frac{\chi^2}{n}}$（$2\times2$时 $\varphi\in[0,1]$，$n\uparrow\varphi\uparrow$） 2. $c=\sqrt{\frac{\chi^2}{\chi^2+n}}$ 列联系数 coefficient of contingency（$n\uparrow c_{max}\uparrow<1$）

3. $V=\sqrt{\frac{\chi^2}{n\cdot\min[R-1,C-1]}}$（$V\in[0,1]$，R or C=2时 $V=\varphi$） 条件百分表之方向。

4.2 ◦ 方差分析 ANOVA analysis of variance 因素（因子）factor 水平（处理）treatment level

误差分解 — 分析检验 基本假定：独立性、正态性、齐方差性

◦ 单因素方差分析 one-way~ 总平方和 SST sum of squares for total $=\sum_i\sum_j(x_{ij}-\bar{\bar{x}})^2$ H_0: $\mu_1=\cdots=\mu_k$.

组间平方和SSA ~factor A $=\sum_i n_i(\bar{x}_i-\bar{\bar{x}})^2$ 组内平方和/误差平方和SSE ~error $=\sum_i\sum_j(x_{ij}-\bar{x}_i)^2$

SST = SSA + SSE ~均方/方差 MST/A/E

检验统计量 $F=\frac{MSA}{MSE}\sim F(k-1,n-k)$（上限检验） df: $n-1/k-1/n-k$ 方差分析表 ANOVA table

"当因素仅2个水平时，one-way anova 等同于均值差t检验，由于此时 $F=t^2$"

关系强度（判定系数）$R^2=\frac{SSA}{SST}$ "所解释的比例" $\frac{n_1(\mu_1-\frac{n_1\mu_1+n_2\mu_2}{n_1+n_2})^2+n_2\cdots}{s_p^2}=\left(\frac{\mu_1-\mu_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)^2$

多重比较方法 multiple comparison procedures 最小显著差异方法 least significant difference LSD

$|\bar{x}_i-\bar{x}_j|\leqslant LSD=t_{\frac{\alpha}{2}}(n-k)\sqrt{MSE(\frac{1}{n_i}+\frac{1}{n_j})}$

- 双因素方差分析 two-way~ 无/有交互作用 (non-) interaction（无/可重复~）

1. $SST = SSR + SSC + SSE$, $F_R = \frac{MSR}{MSE}$, $F_C = \frac{MSC}{MSE}$, $R^2 = \frac{SSR+SSC}{SST}$

df: $kr-1$　$r-1$　$k-1$　$(k-1)(r-1)$ $\longrightarrow \sum_i \sum_j (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2$ "准确说，难以区分这是error还是interaction."

"如果只对R/C单anova，则C/R被含入残差，故双anova优"

2. $SST = SSR + SSC + SSRC + SSE$ $\longrightarrow \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij})^2$

df: $krm-1$　$r-1$　$k-1$　$(k-1)(r-1)$　$kr(m-1)$　$(k-1)(r-1) \longrightarrow \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2$

$F_{RC} = \frac{MSRC}{MSE}$ "一种联合对因变量的附加效应"

五、一元与多元线性回归

5.1 ◦ 相关分析 correlation 相关系数 (Pearson) correlation coefficient $r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x \sigma_y}$（总体$\rho$）

r的检验：$t = |r|\sqrt{\frac{n-2}{1-r^2}} \sim t(n-2)$　$(S_r = \sqrt{\frac{1-r^2}{n-2}})$　$(H_0: \rho = 0)$ "大样本下r很容易通过检验，但实际相关性可能并不显著"

◦ 回归分析 regression 因变量 dependent value 自变量 independent ~

回归模型 $y = \beta_0 + \beta_1 x + \varepsilon$, $E(y) = \beta_0 + \beta_1 x$, "ε为期望为0，方差保持σ^2的正态误差项".

回归方程、估计的回归方程 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$　最小二乘法 method of least squares

$Q = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$, $\left.\frac{\partial Q}{\partial \beta_0}\right|_{\beta_0 = \hat{\beta}_0} = 0$, $\left.\frac{\partial Q}{\partial \beta_1}\right|_{\beta_1 = \hat{\beta}_1} = 0$ $\rightarrow \hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x^2}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

1. 回归统计 multiple R / r 、 R^2 判定系数 coefficient of determination $= \frac{SSR}{SST}$ 、

$SST = SSR + SSE$（回归平方和+残差平方和）$R^2 \in [0,1]$ $= \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$

df: $n-1$　$(1)k$　$(n-2)\,n-k-1$　　一元时 $R = r$ "$R > R^2$，R^2解释更优"

adjusted R^2 / R_a^2 调整的判定系数 $R_a^2 = 1 - (1-R^2)\cdot\frac{n-1}{n-k-1} = 1 - \frac{MSE}{MST} < R^2$ 、

S_e 估计标准误差 $= \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$　"随着k↑ SSE↓，R^2无论怎样↑趋近1".

（二乘法使S_e取min. or $\sum \hat{y}_i (\hat{y}_i - y_i) = 0$）（一元）

2. 方差分析 df、SS、MS、F、α_F(P)　　3. 回归参数 $\hat{\beta}_0$、$\hat{\beta}_1$、$S_{\hat{\beta}_0}$、$S_{\hat{\beta}_1}$、$t_{\hat{\beta}_0}$、$t_{\hat{\beta}_1}$、P、$\pm t_{\frac{\alpha}{2}} S$

显著性检验 <
- 线性关系检验 $F = \frac{MSR}{MSE} \sim F(1, n-2)$ $((k, n-k-1))$ "回归方程显著"
- 回归系数检验 $S_{\hat{\beta}_1} = \frac{S_e}{\sqrt{n}\sigma_x}$, $t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} \sim t(n-2)$ 、$S_{\hat{\beta}_0} = \frac{S_e\sqrt{\sum x^2}}{n\sigma_x}$

一元时 $t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} = \frac{\hat{\beta}_1 \sqrt{n}\sigma_x}{S_e}$　$F = \frac{n\sigma_x^2 \cdot \hat{\beta}_1^2}{S_e^2} = t^2$ 三者一致，多元时F与t有别.（$F = \frac{\sum_i (\sum_k \hat{\beta}_k (x_{ik} - \bar{x}_k))^2}{k S_e^2}$）

$t = |r|\sqrt{\frac{n-2}{1-r^2}} = \frac{r\sqrt{n}\sigma_y}{S_e} = \frac{\hat{\beta}_1 \sqrt{n}\sigma_x}{S_e}$

$(r^2 = \hat{\beta}_{1\,x/y} \cdot \hat{\beta}_{1\,y/x})$　$S_e^2 = \frac{n\sigma_y^2 \cdot (1-R^2)}{n-2} = \frac{n\sigma_y^2 (1-r^2)}{n-2}$

4. 估计与预测
- 点估计 <
  - 平均值（均值）$E(y_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$
  - 个别值（单值）$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$
- 区间估计 <
  - 均值 $\longrightarrow$ 置信区间 $\pm t_{\frac{\alpha}{2}} S_{\hat{y}_0}$
  - 单值 $\longrightarrow$ 预测区间 prediction $\pm t_{\frac{\alpha}{2}} \sqrt{S_e^2 + S_{\hat{y}_0}^2}$

$y_i \longleftrightarrow \hat{y}_i$　$S_e^2 \Rightarrow \sigma^2$　$E(y_i)$　$S_{\hat{y}_0} = S_e \frac{\sqrt{\sigma_x^2 + (x_0 - \bar{x})^2}}{\sqrt{n}\sigma_x}$

5. 残差分析 residual analysis 残差 { 标准化 standardize / Pearson / 半学生化 semi-studentized $z_i=\frac{e_i}{s_e}=\frac{y_i-\hat{y}_i}{s_e}$; 外学生化 external~ / 学生化删除 ~deleted~ $\frac{e_i}{\sqrt{s_e^2-s_{\hat{y}_i}^2}}$; 内学生化（学生化）$\frac{e_i}{s_{e_i}}$（除i外的拟合）

残差图验证 ε 的正态性.

• 多元回归模型 multiple~ 估计的多元回归方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\cdots+\hat{\beta}_kx_k$ 最小二乘法得 $(\overline{x_jx_k}-\bar{x}_j\bar{x}_k)(\hat{\beta}_k)=(\overline{x_jy}-\bar{x}_j\bar{y})$

R^2（多重~）、R_a^2、s_e 同定义、$F(k,n-k-1)$、$t(n-k-1)$ 同定义.（$F^2=t_1^2+t_2^2+\cdots+Cov(t_1,t_2)+\cdots$）

多重共线性 multicollinearity "回归线性关系显著≠每个 x_i 与 y 都显著"

常见处理：剔除相关的 X（变量选择）；避免用 t 对单个 β 检验；对 y 的估测限定在 X 的范围内.

常见表现：x_i 间显著相关；F 显著但许多 β_i 的 t_i 不显著（或反之）；β_i sign 与预期相反；容忍度 tolerance $1-R_i^2$（以 x_i 为因,剩余 X 为自判定.）

方差扩大因子 VIF variance inflation factor $=\frac{1}{1-R_i^2}$（严重共线：>10）

逐步回归 stepwise regression：x_i F?
向前选择 forward selection + 向后剔除 backward elimination

六、时间序列与指数

6.1 • 时间序列 平稳序列 stationary series 非平稳序列 non-~（复合型~）{ 趋势 trend；季节性 seasonality；周期性 cyclicity；不规则波动 irregular variations

加/乘法模型 additive/multipilcative model $Y_t=T_t\cdot S_t\cdot C_t\cdot I_t$

增长率 growth rate（环比/定基）平均~ average~ $\sqrt[n]{\frac{Y_n}{Y_0}}-1$.

预测方法：（不考虑C）（评估：MSE、MAD、MPE、MAPE...）

T× S× → 平滑预测法 { 简单平均 $F_{t+1}=\frac{1}{t}\sum_{i=1}^{t}Y_i$；移动平均 $F_{t+1}=\bar{Y}_t=\frac{Y_t+\cdots+Y_{t-k+1}}{k}$；指数平滑 $F_{t+1}=\alpha Y_t+(1-\alpha)F_t$ $=\alpha Y_t+\alpha(1-\alpha)Y_{t-1}+(1-\alpha)^2F_{t-1}\cdots$（α 平滑系数）

T✓ S× → 趋势预测法（LS）

S✓ → 时序分解、季节性回归… → 季节指数 seasonal index

6.2 • 指数 < 个体~ / 总~ < 数量指标~ q（$L's$）/ 质量指标~ p（P_s） < 简单~ / 加权~ < 综合 / 平均 — 拉氏/帕氏指数 Laspeyres/Passche

指数体系 $\frac{\sum q_1p_1}{\sum q_0p_0}=\frac{\sum q_1p_0}{\sum q_0p_0}\cdot\frac{\sum q_1p_1}{\sum q_1p_0}$（或平均式） 权数（同度量因素、媒介因素）（基期/报告期） $I=\frac{\sum iw}{\sum w}$

综合评价指数：指标、体系、无量纲化处理、确定权重、计算分析.（z_i、$\frac{x_i}{x_0}$、$\frac{x_i-min}{max-min}$ …）

概率论与数理统计 〈陈希孺〉

一、事件的概率

1.1 • 主观概率　试验　事件　基本事件　随机事件　偶/必然事件

古典概率 $P(E)=\frac{M}{N}$、几何概率 ← "等可能性"（古典定义）

统计概率 $\lim_{n\to\infty}\frac{m}{n}=p(E)$（统计定义）"并不是定义了概率，而只是一种估计的方法"

公理化定义　柯氏公理体系 $(\Omega,\mathcal{F})$：$\mathcal{F}$ 为 Ω 的一个 σ 代数（子集族）

$\forall$ 事件 $A\in\mathcal{F}$，满足 ① $0\leq P(A)\leq 1$　② $P(\Omega)=1$，$P(\phi)=0$　③ $A_i\cap A_j=\phi\ (i\neq j)$ $P(\cup A_i)=\sum_i P(A_i)$

则集函数 P 是 $(\Omega,\mathcal{F})$ 上的概率（测度），$(\Omega,\mathcal{F},P)$ 概率空间

• 古典概率的计算　排列 P_n^r　组合 C_n^r、$\binom{n}{r}$　分堆 $\frac{n!}{r_1!\cdots r_k!}$ $(r_1+\cdots+r_k=n)$

常用公式：$\sum_{i=0}^{n}\binom{n}{i}=2^n$、$\sum_{i=0}^{n}(-1)^i\binom{n}{i}=0$.（二次展开、多项展开）

$\binom{n}{m}+\binom{n}{m+1}=\binom{n+1}{m+1}$、$\sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i}=\binom{m+n}{k}$、$\sum_{i=0}^{m}\binom{n+i}{i}=\binom{n+m+1}{m}$

1.2 • 事件的蕴含、包含、相等 $A\subset B$、$A=B$　事件的互斥、对立 $P(A\cap B)=0$. $B=\overline{A}$

事件的和（并）$C=A+B$　加法定理：A_i 两两互斥，$P(\sum A)=\sum P(A)$.

事件的积（交）、差 $C=AB$、$A-B=A\overline{B}$　特别地，$P(\overline{A})=1-P(A)$.

（布尔）运算律（De Morgan 律）

条件概率　$P(B)\neq 0$ 时，$P(A|B)=\frac{P(AB)}{P(B)}$　事件的独立、乘法定理：

$(P(A)=P(A|B)$ $\Rightarrow P(A)=P(A|\overline{B}))$　$P(AB)=P(A)P(B)$、一般地，

▷ 一组独立事件任一部分，或其组合、对立均独立.　A_i 相互独立，

• 全概率公式　$B_iB_j=\phi\ (i\neq j)$，$\sum_i B_i=\Omega$.（完备事件群）　△不是"两两"!　$P(\prod A)=\prod P(A)$.

$P(A)=\sum_i P(B_i)P(A|B_i)$.

贝叶斯公式　$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)}$ ← "由结果推原因"

二、随机变量及概率分布

2.1 • 随机变量　离散型、连续型　概率分布（离散）$p_i=P(X=a_i)$

分布函数 $F(x)=P(X\leq x)$　$X\sim F$　↗　概率密度函数 $f(x)=F'(x)$

▷ ① $f(x)\geq 0$　② $\int_{-\infty}^{\infty}f(x)dx=1$　③ $\int_a^b f(x)dx=P(b\leq X\leq a)$ $=F(b)-F(a)$

- 重要分布：二项分布 $p_i=\binom{n}{i}p^i(1-p)^{n-i}$ $(X\sim B(n,p))$，泊松分布 $p_i=\dfrac{e^{-\lambda}\lambda^i}{i!}$ $(X\sim P(\lambda))$ ← 二伯，$n\to\infty$，$np\to\lambda$ 的推导，超几何分布 $\dfrac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$

  负二项分布、几何分布：$\binom{i+r-1}{r-1}p^r(1-p)^i$、$p(1-p)^i$（令 $r=1$）

  正态分布 $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $(X\sim N(\mu,\sigma^2))$、标准正态分布 $N(0,1)/Z$ $f(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

  指数分布 $f(x)=\lambda e^{-\lambda x}$ $(x>0)$、威布尔分布 $f(x)=\lambda\alpha x^{\alpha-1}e^{-\lambda x^{\alpha}}$ $(x>0)$ $(\alpha>1)$

  均匀分布 $f(x)=\dfrac{1}{b-a}$ $(x\in[a,b])$ $(X\sim R(a,b))$ |▷ $X\sim R(0,1)\Rightarrow F^{-1}(X)\sim F$

2.2 - n 维随机变量（向量）$X=(X_1,\cdots,X_n)$，$f(x_1,\cdots,x_n)$

  多项分布 $p_{(k_1\cdots k_n)}=\dfrac{N!}{k_1!\cdots k_n!}p_1^{k_1}\cdots p_n^{k_n}$ $(\sum_i k_i=N)$，二维正态分布 $f(x_1,x_2)=\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{\left[\frac{(x_1-a)^2}{\sigma_1^2}-\frac{2\rho(x_1-a)(x_2-b)}{\sigma_1\sigma_2}+\frac{(x_2-b)^2}{\sigma_2^2}\right]}{2(1-\rho^2)}}$

  边缘分布 $P(X_1=a_{1k})=\sum_{j_2\cdots j_n}p(k,j_2,\cdots,j_n)$

  $f_1(x_1)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,x_2,\cdots,x_n)\,dx_2\cdots dx_n$ ←"派生而并不包含耦合（相互作用）的信息"

- 条件概率分布 $P(X_1=a_i|X_2=b_j)=\dfrac{p_{ij}}{\sum_k p_{kj}}$　$f(x_1|a\le X_2\le b)=\dfrac{\int_a^b f(x_1,t)\,dt}{\int_a^b f_2(x_2)\,dx_2}$；$f(x_1|x_2)=\dfrac{f(x_1,x_2)}{f_2(x_2)}$

  （n 维类似有 $f(x_1,\cdots,x_n)=g(x_1,\cdots,x_k)h(x_{k+1},\cdots,x_n|x_1,\cdots,x_k)$）

  （连续全概率公式）$f_1(x_1)=\int_{-\infty}^{\infty}f(x_1|x_2)f_2(x_2)\,dx_2$　（若 $f(x_1|x_2)$ 不是 x_2 函数（独立）则能提出，$=f(x_1)$）

  随机变量的独立性 $f(x_1,\cdots,x_n)=f_1(x_1)\cdots f_n(x_n)$

  |▷ $X_1,\cdots,X_n$ 独立，$\forall a_i<b_i$，$A_i=\{a_i\le X_i\le b_i\}$，$A_1,\cdots,A_n$ 也独立；反之亦成立。

  $X_1,\cdots,X_n$ 独立，$Y_1=g_1(X_1,\cdots,X_m)$，$Y_2=g_2(X_{m+1},\cdots,X_n)$ 则 Y_1Y_2 独立。

  若 $f(x_1,\cdots,x_n)$ 能分表为 $g_1(x_1),\cdots,g_n(x_n)$ n 个函数之积，$X_1,\cdots,X_n$ 独立且 $f_i(x_i)$ 与 $g_i(x_i)$ 仅差一因子（倍数）。

  事件的指示变量 $X_i=\begin{cases}1 & A_i\text{发生}\\0 & A_i\text{不发生}\end{cases}$　（A与X独立性一致）

2.3 - 随机变量函数的概率分布：（离散）对Y不同的取值情况求和。

  $Y=g(X_1,\cdots,X_n)$

  （一元）$y=g(x)$　　$l(y)=f(h(y))|h'(y)|$；不单调时可能要加和多个同值，如

  假设 g 严格单调，反函数 $g^{-1}=h$。　　$Y=X^2$ 时 $l(y)=\dfrac{1}{2\sqrt{y}}\cdot[f(\sqrt{y})+f(-\sqrt{y})]$

  （二/多元）$X_1=h_1(Y_1,Y_2,\cdots)$，$X_2=h_2(Y_1,Y_2,\cdots)\cdots$　（坐标变换）

  $J_{X/Y}=\left|\dfrac{\partial h_i}{\partial y_j}\right|=\dfrac{1}{J_{Y/X}}$，$l(y_1,\cdots,y_n)=|J(y_1,\cdots,y_n)|f(h_i(y_1,\cdots,y_n)\cdots)$

  |▷ $2X\sim f(x)$，则 $X\sim\frac{1}{2}f(\frac{x}{2})$ ✗

  （注意系数的正倒）$X\sim 2f(2x)$ ✓

$$Y=X_1+X_2:\ l(y)=\int_{-\infty}^{\infty}f(y-x,x)\,dx \overset{i.}{=} \int_{-\infty}^{\infty}f_1(y-x)f_2(x)\,dx$$

- 和的密度函数：理解 / $P(Y\le y)$ 写出积分区域求导 / 引入变量为新坐标系

如 $(X_1,X_2)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，$Y=X_1+X_2\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$

卡方分布 $f_n(x)=\dfrac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}e^{-\frac{x}{2}}x^{\frac{n}{2}-1}$ $(X\sim\chi_n^2)$　$X_1,\cdots,X_n\overset{iid.}{\sim}N(0,1)$，则 $Y=X_1^2+\cdots+X_n^2\sim\chi_n^2$

▷ $X_1\sim\chi_n^2$，$X_2\sim\chi_m^2$，则 $X_1+X_2\sim\chi_{n+m}^2$.

$X_1,\cdots,X_n\overset{iid.}{\sim}\lambda e^{-\lambda x}$，则 $2\lambda(X_1+\cdots+X_n)\sim\chi_{2n}^2$.

商的密度函数：$\cdots\ Y=\dfrac{X_2}{X_1}:\ l(y)=\int_0^{\infty}x_1 f(x_1,x_1y)\,dx_1 \overset{i.}{=} \int_0^{\infty}x_1 f_1(x_1)f_2(x_1y)\,dx_1$ ▲不要落 $|x|$!!

(令 $X_1\geq0$)

t分布（学生(氏)分布）$f_n(x)=\dfrac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}\left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$ $(X\sim t_n)$　$X_1\sim\chi_n^2$，$X_2\sim N(0,1)$，则 $Y\overset{(i.)}{=}\dfrac{X_2}{\sqrt{\frac{X_1}{n}}}\sim t_n$

F分布 $f_{mn}(x)=m^{\frac{m}{2}}n^{\frac{n}{2}}\dfrac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}x^{\frac{m}{2}-1}(mx+n)^{-\frac{m+n}{2}}$ $(X\sim F_{mn})$　$(F_{m,n}(\alpha)=F_{n,m}(1-\alpha))$　$X_1\sim\chi_n^2$，$X_2\sim\chi_m^2$，则 $Y\overset{(i.)}{=}\dfrac{\frac{X_2}{m}}{\frac{X_1}{n}}\sim F_{mn}$

三大分布重要结论：① $X_1,\cdots,X_n\overset{iid.}{\sim}N(\mu,\sigma^2)$.

$\bar{X}=\dfrac{\sum_i X_i}{n}$，$S^2=\dfrac{\sum_i(X_i-\bar{X})^2}{n-1}$，则 $\dfrac{\sum_i(X_i-\bar{X})^2}{\sigma^2}=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi_{n-1}^2$.

② 同①假设，则 $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{n-1}$.　③ 同①假设，且 $Y_1,\cdots,Y_m\overset{iid.}{\sim}N(\mu_2,\sigma_2^2)$，则 $\dfrac{S_Y^2}{\sigma_2^2}\Big/\dfrac{S_X^2}{\sigma_1^2}\sim F_{m-1,n-1}$.

$X_i\sim N(\mu_1,\sigma_1^2)$

④ 同③假设，若 $\sigma_1^2=\sigma_2^2$，则

$$\sqrt{\frac{nm(n+m-2)}{n+m}}\,\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sum_i(X_i-\bar{X})^2+\sum_j(Y_j-\bar{Y})^2}}\sim t_{n+m-2}.$$

<2.4>

- $\Gamma(x)$ 函数　$\Gamma(x)=\int_0^{\infty}e^{-t}t^{x-1}\,dt$　$\Gamma(x)=(x-1)\Gamma(x-1)$，$\Gamma(x)=(x-1)!$

$B(x,y)$ 函数　$B(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}\,dt=\dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$　余元 $\Gamma(p)\Gamma(1-p)=\dfrac{\pi}{\sin p\pi}$

$G(n)$ 函数　$G(n)=\int_0^{\infty}x^n e^{-\lambda x^2}\,dx=\lambda^{-\frac{n+1}{2}}\begin{cases}\frac{\sqrt{\pi}}{2}\cdot\frac{1}{2}\cdot\frac{3}{2}\cdots & n\ \text{even}\\ \frac{1}{2}\cdot\frac{2}{2}\cdot\frac{4}{2}\cdots & n\ \text{odd}\end{cases}=\lambda^{-\frac{n+1}{2}}\cdot\frac{1}{2}\Gamma(\frac{n+1}{2})$　$\to(n-1)!!$

- 斯特林公式 $\lim\limits_{n\to\infty}n!=\sqrt{2\pi n}\left(\dfrac{n}{e}\right)^n$

▷ 证明：Wallis 公式（$I_{2k+1}<I_{2k}<I_{2k-1}$ 夹逼得）$\lim\limits_{n\to\infty}\left(\dfrac{(2n)!!}{(2n-1)!!}\right)^2\dfrac{1}{2n+1}=\dfrac{\pi}{2}$.

$\Sigma\to\int$ ln 的面积，鞍点法 $n!=\int_0^{\infty}e^{-t}t^n\,dt=\int_0^{\infty}e^{g(t)}\,dt=e^{g(n)}\int_0^{\infty}e^{\frac{1}{2}g''(n)(t-n)^2}dt$

$g(t)=-t+n\ln t$

$=n^n e^{-n}\sqrt{2\pi n}$

▷ 注意 $\dfrac{n!}{(n-i)!}$ 在 $n\to\infty$ 时 $i\sim o(1)$ $\sim n^i$ 不带 e^i!

$g'(t)=-1+\dfrac{n}{t}$，$g''(t)=-\dfrac{n}{t^2}$，$g'(n)=0$.

- ① 的证明：证 $\overset{(1)}{\bar{X}}$ 与 $\overset{(n-1)}{\sum(X_i-\bar{X})^2}$ 独立.　$Y_1=\dfrac{\sum X}{\sqrt{n}}$　$\sum Y_i^2=\sum X_i^2$

(②③④ 顺推)　$Y=QX$，$Q=\begin{pmatrix}\frac{1}{\sqrt{n}}&\frac{1}{\sqrt{n}}&\frac{1}{\sqrt{n}}&\cdots&\frac{1}{\sqrt{n}}\\ \vdots&&&&\vdots\\ &&&&0\end{pmatrix}$，故 $\sum_{i=2}Y_i^2=\sum(X_i-\bar{X})^2$

$(X_1,\cdots,X_n)\sim\dfrac{1}{(\sqrt{2\pi}\sigma)^n}e^{-\frac{\sum(X_i-\mu)^2}{2\sigma^2}}\longrightarrow\sum Y_i^2-\mu\sqrt{n}\,y_1+n\mu^2$ 则 $\begin{cases}y_1\sim N(\sqrt{n}\mu,\sigma^2)\\ y_2\cdots y_n\underset{iid.}{\sim}N(0,\sigma^2)\end{cases}$

• 二项分布推出泊松分布与正态分布(可称为中心极限)

$\binom{n}{i}p^i(1-p)^{n-i} \xrightarrow[np=\lambda]{i \quad n\to\infty, p\to 0} \frac{e^{-\lambda}\lambda^i}{i!}$ (代入e定义) $\xrightarrow[\lambda=i+\varepsilon \quad \varepsilon]{i\to\infty,\lambda\to\infty}$ (Stirling公式)

$\xrightarrow[i\to\infty, n\to\infty]{i=np+\varepsilon \quad \varepsilon}$ $\frac{1}{\sqrt{2\pi}\sqrt{np(1-p)}} \cdot \frac{1}{(1+\frac{\varepsilon}{np})^{np+\varepsilon}(1-\frac{\varepsilon}{n(1-p)})^{n(1-p)-\varepsilon}}$

$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\varepsilon^2}{np(1-p)}+\frac{\varepsilon^2}{2np(1-p)}}$

|▷ e的定义式中 $(1+x)^{\frac{1}{x}}$ 与e存在 $e^{\frac{x}{2}}$ $(-\frac{e}{2}x)$ 的一阶差.

这里必须考虑!

• 重要分布总结: Bernoulli (two-point)、binomial、multinomial、Poisson、(Pascal) negative binomial, geometric, hypergeometric, exponential, Weibull, normal (Gauss), ln-normal, logarithmic, Rayleigh, Chi-square, student, Fisher, Cauchy, Laplace, gamma, beta, multivariate beta (Dirichlet), uniform, maxmin...

对数正态分布: $Y\sim N(\mu,\sigma^2)$, $X=e^Y\sim LN$, $f(x)=\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$. $x_m=e^{\mu-\sigma^2}$,

σ^2↑ 偏态越明显. $E(x)=e^{\mu+\frac{\sigma^2}{2}}$, $V(x)=e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$

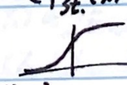
对数级数分布: $-\frac{(1-p)^i}{i\ln p}$. |▷ $\binom{n}{m}=(-1)^m\binom{m-n-1}{m}$. NB名称来自于 $(1-x)^{-r}$ $(=(1+x+x^2+\dots)^r)$

NB(负二项)分布 (r,p) $=\sum_{iid.}^{r}$ geometric. dis., $E(nb)=\frac{r(1-p)}{p}$, $V(nb)=\frac{r(1-p)}{p^2}$.

$\xrightarrow[\frac{r(1-p)}{p}=\lambda \quad r\to\infty, p\to 1]{}$ Poisson. (e定义) $\frac{}{i}$

Logistic 逻辑斯谛分布(增长分布): $L(\mu,\gamma)$. $F(x)=\frac{1}{1+e^{-\frac{x-\mu}{\gamma}}}$, $f(x)=\frac{e^{-\frac{x-\mu}{\gamma}}}{\gamma(1+e^{-\frac{x-\mu}{\gamma}})^2}$

$(F_{st.}(x)=\frac{1}{1+e^{-x}})$ $V(X)=\frac{\pi^2\gamma^2}{3}$ $(E(X)=0)$



$m(x)=\pi x\sec(\pi x)$.

瑞利分布: 二维独立高斯分布的模长

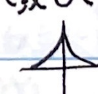
$x>0$. $f(x)=\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}$ $(\sqrt{\chi_2^2}=X)$

$E(X)=\sqrt{\frac{\pi}{2}}\sigma$, $V(X)=(2-\frac{\pi}{2})\sigma^2$.

柯西分布: $C(x_0,\gamma)$ $f(x)=\frac{1}{\pi}\cdot\frac{\gamma}{(x-x_0)^2+\gamma^2}$ $(\frac{1}{\pi}\cdot\frac{1}{1+x^2})$ 不存在期望.(各阶矩均不存在). 估计x_0适合用中位数.

$X_1, X_2 \overset{iid.}{\sim} N(0,1)$, 则 $Y=\frac{X_1}{X_2}\sim C(0,1)$ (自由度为1的t分布)(或$U(-\frac{\pi}{2},\frac{\pi}{2})$上$y=\tan x$的分布)

拉普拉斯分布: $f(x)=\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{2\lambda}}$ $E(X)=\mu$, $V(X)=2\lambda^2$ $(K(X)=3)$



(或标准正态商的分布$\frac{X_1}{X_2}$)

伽马Γ分布: $f(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$. $Ga(\alpha,\lambda)+Ga(\beta,\lambda)=Ga(\alpha+\beta,\lambda)$.

$\alpha=1$, 为指数分布; $\alpha=\frac{n}{2}$, $\lambda=\frac{1}{2}$, 为卡方分布 $\chi^2\sim Ga(\frac{1}{2},\frac{1}{2})$ $E(X)=\frac{\alpha}{\lambda}$, $V(X)=\frac{\alpha}{\lambda^2}$

$\chi_n^2\sim Ga(\frac{n}{2},\frac{1}{2})$

α: 形状参数 λ: 尺度参数 $(\lambda X\sim Ga(\alpha,1))$

Multinomial ~ Dirichlet

Binomial ~ Beta

为共轭先验.

贝塔B分布: $f(x)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ $E(X)=\frac{\alpha}{\alpha+\beta}$, $V(X)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

狄利克雷分布: $f(x)=\frac{1}{B(\vec{\alpha})}\prod_i x_i^{\alpha_i-1}$, $B(\vec{\alpha})=\frac{\prod_i\Gamma(\alpha_i)}{\Gamma(\sum_i\alpha_i)}$ 且$\|X\|=1$. Beta为其marginal.

最大最小值: $F_{max}(x)=F^n(x)$, $F_{min}(x)=1-(1-F(x))^n$ 再求导.

三、随机变量的数字特征

3.1 · 数学期望（均值） $E(\Sigma X)=\Sigma E(X)$. X_i $id.$, $E(\Pi X)=\Pi E(X)$.

要求绝对收敛 $\int_{-\infty}^{\infty}|x|f(x)dx<\infty$. $E(g(X))=\int_{-\infty}^{\infty}g(x)f(x)dx$ （故 $E(cX)=cE(X)$）

▷ $E(\chi_n^2)=n$. $E(\frac{1}{\chi_n^2})=\frac{1}{n+2}$. $E(F_{mn})=\frac{n}{n+2}$. $E(Z^n)=(n-1)!!$ 偶/0 奇

$E(X^n)$, N分布，只需对 $(X-\mu+\mu)^n$ 展开.

条件均值 $E(Y|x)=\int_{-\infty}^{\infty}yf(y|x)dy$. $E(Y)=E_y(E_x(Y|X))$ 中位数

· 方差 标准差 $V(X)=E(X^2)-E^2(X)$ （平行轴定理） $V(X+c)=V(X)$.

X_i $id.$, $V(\Sigma X)=\Sigma V(X)$. $E((\hat{\theta}-\theta)^2)=V(\hat{\theta})+(E(\hat{\theta})-\theta)^2$ $V(cX)=c^2V(X)$.

标准化 $Z=\frac{X-\mu}{\sigma}$ ▷ $V(\chi_n^2)=2n$. $V(t_n)=\frac{n}{n-2}$. $V(F_{mn})=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$.

$(V(S^2)=2(n-1)\sigma^4)$

矩 原点矩 中心矩 偏度 $\frac{\mu_3}{\mu_2^{3/2}}$ 峰度 $\frac{\mu_4}{\mu_2^2}(-3)$ 左偏、右偏分布

· 协方差 $Cov(X,Y)=E((X-E(X))(Y-E(Y)))=E(XY)-E(X)E(Y)$

X,Y id, $Cov(X,Y)=0$. $Cov^2(X,Y)\leq\sigma_X^2\sigma_Y^2$, $E(XY)\in E(X)E(Y)\pm D(X)D(Y)$

相关系数 $Corr(X,Y)=\rho_{XY}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}\in[-1,1]$ ▷ Cov or Corr $=0$ 仅能称"不相关"，

并不一定"独立"！

线性关系、 $(f(x,y)\neq f(x)f(y))$

（二维正态，是"独立"的。）

最小二乘法 $E(Y-a-bX)^2\to min$

$=\sigma_Y^2+b^2\sigma_X^2-2b\sigma_{XY}^2+(a-E(Y)+bE(X))^2$

$\xrightarrow{min}\sigma_Y^2(1-\rho^2)$

$\llcorner$ 取$=0$ ，$a_m=E(Y)-b_mE(X)$.

$b_m=\frac{\sigma_{XY}^2}{\sigma_X^2}=\frac{\sigma_Y}{\sigma_X}\rho_{XY}$.

▷ 柯西不等式的推论：$E(\frac{1}{X})\geq\frac{1}{E(X)}$. $E(X)\leq\sqrt{E(X^2)}$

另一个推论：$V(X_1)=V(X_2)$，则 $Cov(X_1+X_2,X_1-X_2)=0$.

"均值趋于期望"

3.2 · 大数定理 $X_1,\cdots,X_n$ $iid.$ 且均值为μ，则 $\lim_{n\to\infty}P(|\bar{X}_n-\mu|\geq\varepsilon)=0$ 对$\forall\varepsilon>0$.

（存在方差（为σ^2））

"依概率收敛"

马尔科夫不等式 $P(Y\geq\varepsilon)\leq\frac{E(Y)}{\varepsilon}$, $Y\geq0$, $\forall\varepsilon>0$. （积分放缩）

证明

契比雪夫不等式 若$V(Y)$存在，则 $P(|Y-EY|\geq\varepsilon)\leq\frac{V(Y)}{\varepsilon^2}$

"依分布收敛"

· 中心极限定理 X_i $iid.$ $\forall x$有 $\lim_{n\to\infty}P(\frac{1}{\sqrt{n}\sigma}(X_1+\cdots+X_n-n\mu)\leq x)=\phi(x)$

Lindeberg-lévy CLT

其中ϕ是标准正态分布函数 $\phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$.

特殊地，de Moivre-Laplace Th: X_i二项分布，$\lim_{n\to\infty}P(\frac{\Sigma X-np}{\sqrt{np(1-p)}}\leq x)=\phi(x)$.

（有 $\phi(-x)=1-\phi(x)$）

▷ 正态分布（同二项分布、泊松分布一样）是再生分布，叠加、分割仍为正态分布.

（线性变换稳定性：$X\sim N(\mu,\Sigma)$，$Y=CX\sim N(C\mu,C\Sigma C^T)$.）

指数分布有无后效性，且 $n\min_{i\in[1,n]}X_i$ 与 X 同指数分布.

⟨3.3⟩ • **矩母函数 MGF、特征函数** 类似于 Fourier、Laplace 变换，可以把 $\sum_{iid}$ 的卷积化为乘积.

$m(t)=E(e^{tx})=\int_{-\infty}^{\infty}e^{tx}f(x)\,dx$ $\left(\sum e^{tx}f(x)\right)$ 一个 MGF 决定了一个分布.(全部信息)

$=E(1)+tE(x)+\frac{t^2}{2!}E(x^2)+\cdots$ 故 n 阶矩 $E(x^n)=\frac{d^n}{dt^n}m(0)$

指数分布 $m(t)=\frac{\lambda}{\lambda-t}$. 正态分布 $m(t)=e^{t\mu+\frac{t^2\sigma^2}{2}}$ (多元 $e^{t^T\mu+\frac{t^T\Sigma t}{2}}$)

卡方分布 $m(t)=\frac{1}{(1-2t)^{\frac{n}{2}}}$. ($Z$: $m(t)=e^{\frac{t^2}{2}}$, $\varphi(t)=e^{-\frac{t^2}{2}}$)

二项分布 $m(t)=(1-p+pe^t)^n$.

伽马分布 $m(t)=\left(\frac{\lambda}{\lambda-t}\right)^{\alpha}$. $\varphi(t)=\left(\frac{\lambda}{\lambda-it}\right)^{\alpha}$.

泊松分布 $m(t)=e^{\lambda(e^t-1)}$.

$\varphi(t)=E(e^{itx})$ (以上 $m(t)$ 换为 $m(it)$ 即可)

▷ $X\sim N(\mu,\Sigma)$. $E(X^TAX)=tr(A\Sigma)+\mu^TA\mu$.

$m_{X^TAX}(t)=|K|^{-\frac{1}{2}}e^{-\frac{\mu^T(I-K^{-1})\Sigma^{-1}\mu}{2}}$.

其中 $K=I-2tA\Sigma$.

$V(X^TAX)=2tr(A\Sigma)^2+4\mu^TA\Sigma A\mu$.

$Cov(BX,X^TAX)=2B\Sigma A\mu$.

$(X-\mu)^T\Sigma^{-1}(X-\mu)\sim\chi^2(n)$.

△ $Y=aX+b$, $m_Y(t)=e^{bt}m_X(at)$. X_i iid, $Y=\sum_i X_i$, $m_Y(t)=\prod_i \varphi_{X_i}(t)$.

• **正态分布相关来源**: 1. central limit Th. $Y=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\frac{X_1+\cdots+X_n}{\sqrt{n}\sigma}-\frac{\sqrt{n}}{\sigma}\mu$.

$\varphi_Y(t)=e^{-i\mu t\frac{\sqrt{n}}{\sigma}}\cdot\varphi_X^n\left(\frac{t}{\sqrt{n}\sigma}\right)$ 证明其形式为 $e^{-\frac{t^2}{2}}$: 令 $\eta=\frac{t}{\sqrt{n}\sigma}$, $\lim_{n\to\infty}\ln\varphi_Y(t)=$

$\lim_{\eta\to 0}\frac{t^2}{\sigma^2}\cdot\frac{\ln\varphi_X(\eta)-i\mu\eta}{\eta^2}\overset{L'}{=}\lim_{\eta\to 0}\frac{t^2}{\sigma^2}\cdot\frac{\varphi''\varphi-\varphi'^2}{2\varphi^2}=-\frac{t^2}{2}$. 得证.

(已知 $\varphi_X(0)=1$, $\varphi_X'(0)=i\mu$, $\varphi_X''(0)=-(\mu^2+\sigma^2)$.)

2. Gauss (1809) error dis. $L(\theta)=f(x_1-\theta)\cdots f(x_n-\theta)$ ($f(e)$ 为误差函数)

极大似然估计 $\hat{\theta}=\bar{x}\to L_{max}(\theta)$: 令 $g=\frac{f'}{f}$, 则求导为0要求 $\sum_i g(x_i-\hat{\theta})=0$.

▷ $(e_1,\cdots e_n)$ 概率极大, $\frac{1}{(\sqrt{2\pi}\sigma)^n}e^{-\frac{\sum e_i^2}{2\sigma^2}}$ 有 $\sum_i e_i^2$ 极小, 解释了最小二乘法.

由 $\hat{\theta}=\bar{x}$ 发现 $g(x)$ 的性质 ($n=2$, $x_1-\bar{x}=-(x_2-\bar{x})$ 得 $g(x)=-g(-x)$.

只能为 $g(x)=cx$. $n=m+1$, $mg(-x)+g(mx)=0$ 得 $g(cx)=cg(x)$)

故 $f(x)=C'e^{cx^2}$. 规一化后得 $N(0,\sigma^2)$

3. Herschel & Maxwell symmetry of sp.

正交独立+旋转对称: $f(x)f(y)=p(x,y)=g(r,\theta)=g(r)=g(\sqrt{x^2+y^2})$

由于 $y=0$ 有 $g(x)=f(x)\cdot f(0)$. 令 $h(x)=\ln\frac{f(x)}{f(0)}$, 则 $h(x)+h(y)=h(\sqrt{x^2+y^2})$, $h(x)=cx^2$.

(Maxwell 分布律类似) 故 $f(x)=\sqrt{\frac{\alpha}{\pi}}e^{-\alpha x^2}$.

4. Landon stable noise. 累加微小的随机噪声不改变分布模式 (仅层级/方差)

$X\sim p(x,\sigma^2)$, $\varepsilon\sim q(e)$, $X'=X+\varepsilon$ 应有 $X'\sim p(x,\sigma^2+V(\varepsilon))$:

X' 的 pdf $f(x)=\int p(x-e,\sigma^2)q(e)de\ \overset{\text{对}p\text{展开}}{=\!=}\ p-\frac{\partial p}{\partial x}\int eq(e)de+\frac{1}{2}\frac{\partial^2 p}{\partial x^2}\int e^2q(e)de+o(\bar{\varepsilon^2})$.

随机, $\bar{\varepsilon}$ 认为$=0$

把 p 对方差展开 $f(x)=p+\frac{\partial p}{\partial\sigma^2}\cdot\bar{\varepsilon^2}$

扩散方程 $\frac{1}{2}\frac{\partial^2 p}{\partial x^2}=\frac{\partial p}{\partial\sigma^2}\to p(x,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}$

5. maximum entropy. $H(p)=-\int p(x)\log p(x)dx$ 给定分布的 μ, σ^2，熵最大的为正态分布：

一个分布的熵总是小于相对熵 $H(p)\le -\int p(x)\log q(x)dx$ （$\int p(x)\log\frac{q(x)}{p(x)}dx\le 0$，利用 $\log x\le 1-x$）

代入 $q(x)=N(\mu,\sigma^2)$ 得 $H(p)\le\frac{1}{2\sigma^2}\int p(x)(x-\mu)^2dx+\log\sqrt{2\pi}\sigma=\frac{1}{2}+\log\sqrt{2\pi}\sigma$.（可在 $p(x)=N(\mu,\sigma^2)$ 时取等(max)）.

6. Galton board. $\frac{\binom{n}{k}}{2^n}\xrightarrow[n\to\infty]{k=\frac{n}{2}+x}\frac{2}{\sqrt{2\pi n}}e^{-\frac{x^2}{2n}}$ 与 binomial/poisson 逼近类似. 略.

$(\frac{1}{2})-1 \quad (\frac{1}{2})+1$

四、参数估计 　△参数/变量的范围！

4.1 • 参数估计　假设检验　总体"无限总体"　样本　独立随机样本

统计量　样本方差　样本矩　样本均值　估计量

点估计　$X_1\cdots X_n\overset{iid.}{\sim}f(x;\theta_1,\cdots,\theta_k)$ 估计 θ_i 或 $g(\theta_1,\cdots,\theta_k)$. $\hat{\theta}_i(X_1,\cdots,X_n)\longrightarrow\theta_i$.

• 矩估计法　样本矩→总体矩　$\alpha_m(\theta_1,\cdots,\theta_k)=a_m$ 解得 $\hat{\theta}_i=\hat{\theta}_i(X_1,\cdots,X_n)$.

修正有 s 替代 $\sqrt{m_2}$ 等. 一般能用低阶矩不用高阶. $\hat{g}=\hat{g}(X_1,\cdots,X_n)=g(\hat{\theta}_1,\cdots,\hat{\theta}_k)$.

(MLE)

极大似然估计法　样本的分布 $L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)=f(x_1;\theta_1,\cdots\theta_k)\cdots f(x_n;\theta_1,\cdots,\theta_k)$

"似然函数"看作 θ_i 的函数. $L(x_1,\cdots x_n;\theta_1^*,\cdots,\theta_k^*)=\max L(X_i;\theta_i)$

$\frac{\partial\ln L}{\partial\theta_i}=0$，$\ln L=\sum_i\ln f(X_i;\theta_1,\cdots\theta_k)$.　一般更优良但个别情况会给出很不理想的结果. θ_i^*

$N(\mu,\sigma^2)$ 矩和似然均给出 $\bar{X}, m_2$.　要求有分布具体的参数形式.

$R(0,\theta)$ 矩 $2\bar{X}$，似然 $\max_i(X_i)$.

贝叶斯法　先验分布 $h(\theta)$　后验分布 $h(\theta|X_1,\cdots,X_n)=\frac{h(\theta)f(X_1,\theta)\cdots f(X_n,\theta)}{p(X_1,\cdots,X_n)}$.

（广义先验密度 $\int h(\theta)d\theta\ne 1$）　$p(X_1,\cdots,X_n)=\int h(\theta)f(X_1,\theta)\cdots f(X_n,\theta)d\theta$.

先验分布的选取："同等无知"原则　之后常采用均值作为估计 $\tilde{\theta}=\int\theta h(\theta|X_1,\cdots,X_n)d\theta$.

（争议）（$h(\mu)=1$，$h(\sigma)=\frac{1}{\sigma}$ 等）（MLE 可以想为 $h(\theta|X)_{max}$ 且 $h(\theta)=1$）

共轭先验分布　$h(\theta|X_1,\cdots,X_n)$ 与 $h(\theta)$ 同一分布族，则 $h(\theta)$ 共轭于似然函数 $f(X|\theta)$

如 Bin: Beta, Poisson: Gamma, Uni: Pareto,　或称 θ 的共轭先验分布为 $h(\theta)$.)

(μ) Normal: Normal, (σ^2) Normal: IGamma...

4.2 • 点估计的优良性准则　　无偏性　　($X\sim Ga$, $\frac{1}{X}\sim IGa$)

（结合大数定理，可以理解为 $\forall(\theta_1,\cdots,\theta_k)$, $E_{\theta_1,\cdots,\theta_k}(\hat{g}(X_1,\cdots,X_n))=g(\theta_1,\cdots,\theta_k)$ （$\forall\theta$, $E(\hat{g})=g$）

$\frac{\sum_{i=1}^{N}\hat{g}_i}{N}$ 依概率收敛于 g.）如 $E(\bar{X})=\mu$, $E(S^2)=\sigma^2$, $E(S)<\sigma$ (可通过修正系数 c_nS，对正态总体，$c_n=\frac{\sqrt{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}$）

均方误差 $M_\theta(\hat{\theta})=E_\theta(\hat{\theta}(X_1,\cdots,X_n)-\theta)^2\overset{E(\hat{\theta})=\theta\text{时}}{=\!=\!=}V_\theta(\hat{\theta})$.

最小方差无偏估计 (MVU)　$\forall\theta,\theta_1$, $V_\theta(\hat{\theta})\le V_\theta(\theta_1)$ 则 $\hat{\theta}$ 为 MVUE.

- Fisher信息量 $0=\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x,\theta)dx \overset{\Delta}{=} \int_{-\infty}^{\infty}\frac{\partial f(x,\theta)}{\partial\theta}dx = E(\frac{\partial \ln f(x,\theta)}{\partial\theta})$.

$I(\theta)=V(\frac{\partial \ln f(x,\theta)}{\partial\theta})=\int_{-\infty}^{\infty}(\frac{\partial \ln f(x,\theta)}{\partial\theta})^2 f(x,\theta)dx \overset{(有条件)}{=} \int_{-\infty}^{\infty}\frac{(\frac{\partial f}{\partial\theta})^2}{f}dx$.

(由 $0=\frac{d^2}{d\theta^2}\int_{-\infty}^{\infty}f(x,\theta)dx$ 得 $\int_{-\infty}^{\infty}\frac{\partial^2\ln f}{\partial\theta}f\,dx+\int_{-\infty}^{\infty}(\frac{\partial \ln f}{\partial\theta})^2 f\,dx=0$，故 $I(\theta)=-E(\frac{\partial^2\ln f(x,\theta)}{\partial\theta^2})$.

样本的总信息量 $nI(\theta)$ ($V(\Sigma\frac{\partial\ln f}{\partial\theta})\overset{iid}{=}\Sigma V(\frac{\partial \ln f}{\partial\theta})$) 亦即衡量似然函数的方差或 $\log L$ 的峰聚程度.

克拉美-劳不等式 $V_\theta(\hat{g})\geq\frac{(g'(\theta))^2}{nI(\theta)}$ ($\forall\hat{g}\to g(\theta)$) (简写为 $\Delta^2\hat{\theta}\geq\frac{1}{I(\theta)}$)

(CRLB: Cramer-Rao Lower Bound)

|▷ 证明: $Cov(\hat{g},\Sigma\frac{\partial\ln f}{\partial\theta})^2\leq V(\Sigma\frac{\partial \ln f}{\partial\theta})V(\hat{g})$

$E(\hat{g}\cdot\Sigma\frac{\partial\ln f}{\partial\theta})=\int_{-\infty}^{\infty}\frac{\partial\ln\Pi f}{\partial\theta}\hat{g}\,\Pi f\,dx=\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\hat{g}\,\Pi f\,dx$

无偏有 $=\frac{\partial g}{\partial\theta}$. 得证.

- 相合性 (一致性) $\lim_{n\to\infty}P_{\theta_i}(|\hat{g}(X_1,\cdots,X_n)-g(\theta_i)|\geq\varepsilon)=0$

(样本大小无限扩大时, 估计依概率收敛于待估值.)

如样本的 m_2, $S^2\to\sigma^2$.

|▷ 正态有 $I(\mu)=\frac{1}{\sigma^2}$, $I(\sigma^2)=\frac{1}{2\sigma^4}$. (已知 μ 下)

渐近正态性 $n\to\infty$ 时 $\hat{g}$ 的分布 $\to$ Normal

如 $\bar{X}$ (CLT) 即为.

大/小样本性质/方法 ($\to\infty$ 与否)

4.3 • 区间估计 $[\hat{\theta}_1(X_1,\cdots,X_n),\hat{\theta}_2(X_1,\cdots,X_n)]$ 置信系数 $1-\alpha$ 置信区间/界 (上.下)

(置信水平 $\geq 1-\alpha$, $\forall\theta$)

枢轴变量法: 1. 找一个与待估 $g(\theta)$ 有关的统计量 T (常为点估计). 2. 找一枢轴变量 $S=S(T,g(\theta))$, 其分布 F 应与 θ 无关. 3. $S(T,g(\theta))$ 能进行改写 $a\leq S\leq b\to A(a,b,T)\leq g(\theta)\leq B(a,b,T)$. 4. F 的上/下分位点、

贝伦斯-费歇尔 Behrens-Fisher问题 (见5.1)

$P(w_{1-\frac{\alpha}{2}}\leq S(T,g(\theta))\leq w_{\frac{\alpha}{2}})=1-\alpha$ 改写为 $g(\theta)\in[A,B]$ ($1-\alpha$下).

大样本法: 利用 n 很大时极限分布 (正态)

(一般 $n>30$, 其实太小区间估计本就意义不大) (离散型枢轴法不易使用)

如 $\frac{X-np}{\sqrt{np(1-p)}}$ 近似 $\sim N(0,1)$, 近似得 $\hat{p}\pm u_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. ($n\gg1$)

贝叶斯法: "后验信度" $\int_{\theta_1}^{\theta_2}h(\theta|X_1,\cdots,X_n)d\theta=1-\alpha$ (上下界即改为 $-\infty/+\infty$)

常选择使 $\hat{\theta}_2-\hat{\theta}_1$ 最小 (另一种为 $\int_{\theta_2}^{+\infty}=\frac{\alpha}{2}$, $\int_{-\infty}^{\theta_1}=\frac{\alpha}{2}$ 选择)

求解容易但先验分布的确定是问题, 而奈曼法确定分布很复杂.

(大样本法有时无法做到很大难以控制误差)

奈曼理解: "抽取出大量的区间, 有 $(1-\alpha)$ 的部分的区间里 (Neyman) 包含有真值. $\overset{\times}{\longleftrightarrow}$ 真值有 $(1-\alpha)$ 的概率落在这一区间里" (Bayes)

|▷ MVUE的唯一性证明: 若 $\hat{\theta}_1$、$\hat{\theta}_2$ 均为MVUE, 则 $\frac{\hat{\theta}_1+\hat{\theta}_2}{2}$ (同无偏) 有

$V(\frac{\hat{\theta}_1+\hat{\theta}_2}{2})\geq V_{min}=V(\hat{\theta}_1)=V(\hat{\theta}_2)$. $V(\frac{\hat{\theta}_1+\hat{\theta}_2}{2})=\frac{1}{4}E(\hat{\theta}_1-\theta+\hat{\theta}_2-\theta)^2$

$=\frac{1}{4}E(\hat{\theta}_1-\theta)^2+\frac{1}{4}E(\hat{\theta}_2-\theta)^2+\frac{1}{2}E((\hat{\theta}_1-\theta)(\hat{\theta}_2-\theta))$

$\leq\frac{1}{2}V_{min}+\frac{1}{2}\sqrt{V_{min}V_{min}}=V_{min}$.

故必须取等且 $\hat{\theta}_1$ 与 $\hat{\theta}_2$ 的任意线性组合也为MVUE.

取等有 $\hat{\theta}_1=a\hat{\theta}_2+b$. 而 $E(\hat{\theta}_1)=E(\hat{\theta}_2)$. $b=(1-a)\theta$.

故其同族均为 $a\hat{g}+(1-a)\theta$ 的带真值的平移,

不存在两个仅用样本的MVUE.

五、假设检验

5.1 • 原假设（零假设）H_0 对立假设（备择假设）H_1

"统计上的显著性（差异不能被随机误差解释）≠ 现实中的重要性."

检验统计量　接受/否定域　临界值

简单/复合假设　赘余参数

"检验水平/显著性水平/奇刻度(H_0)/弃真错误"...

• 功效函数 $\phi: \bar{X} \geq C$ ✓ oth. X. $\beta_\phi(\lambda) = P_\lambda(\bar{X} < C)$

第一/二类错误 $\begin{cases} \beta_\phi(\theta), \ \theta \in H_0 \\ 1-\beta_\phi(\theta), \ \theta \in H_1 \end{cases}$　检验水平 α s.t. $\beta_\phi(\theta) \leq \alpha$, $\forall \theta \in H_0$.

（α尽量 min）

一致最优检验　对 $\forall\alpha$水平的检验 g, $\beta_\phi(\theta) \geq \beta_g(\theta)$, $\forall \theta \in H_1$, 以 ϕ 为 α水平 UBT.

"一致即处处的条件很高，仅单参且单边或一些特例存在".

• 重要参数检验方法：$1°\ H_0: \theta \geq \theta_0$　$2°\ H_0: \theta \leq \theta_0$　$3°\ H_0: \theta = \theta_0$

①基于点估计 $T(\theta) \leftrightarrow \theta_0 . C$（计算功效函数，由水平确定C）

$\mu; \sigma^2$已知：$1°: \phi: \bar{X} \geq C = \theta_0 - \sigma\frac{u_\alpha}{\sqrt{n}}$ ✓ oth. x. $\beta_\phi(\theta) = \Phi(\frac{\sqrt{n}(\theta_0-\theta)}{\sigma} - u_\alpha)$.（2°.3°略）

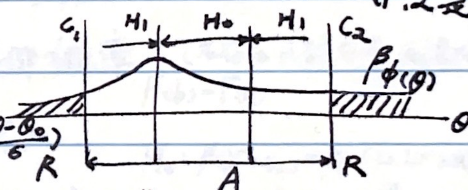
（3°: $|\bar{X}-\theta_0| \leq C$ ✓）

"问题提法的倾向性"　3°是双侧检验而非一致最优（不存在）.（1°.2°是）

（α. n均可以调整以适应'要求'）

$\mu; \sigma^2$未知：$\psi: \frac{\sqrt{n}(\bar{X}-\theta_0)}{S} \geq -t_{n-1}(\alpha)$ ✓ oth. X

（t检验）$\beta_\psi(\theta,\sigma) = P_{\theta,\sigma}(\frac{\sqrt{n}(\bar{X}-\theta_0)}{S} < -t_{n-1}(\alpha)) = f(\frac{\theta-\theta_0}{\sigma})$　且 $f\downarrow$.



Norm.

β 性质较 u 检验复杂，

1. $\forall \theta_1 < \theta_0$，不存在 n s.t. $\theta \leq \theta_1$ 时 ψ ✓ 的 $P <$ 充分小的 β.（$\sigma \to \infty$!）　△（单边时可以指定一个 $\theta_1 < \theta_0$

2. 除非 $\alpha \geq \frac{1}{2}$，oth. ψ 1°.2° 也不是一致最优.　1° 要求 $\beta_\phi(\theta) \geq 1-\beta$, $\theta \leq \theta_1$（n）以约束二类错误.）

$\Delta\mu$; σ^2相近：$U = \sqrt{\frac{nm}{n+m}} \cdot \frac{\bar{X}-\bar{Y}-\theta_0}{\sigma_0}$; $T = \sqrt{\frac{nm}{n+m}} \frac{\bar{X}-\bar{Y}-\theta_0}{S}$, $S^2 = \frac{\sum^n (X-\bar{X})^2 + \sum^m (Y-\bar{Y})^2}{n+m-2}$.

σ^2: $1°: \varphi: \sum^n (X-\bar{X})^2 \geq \sigma_0^2 \chi^2_{n-1}(\alpha)$ ✓ oth. X（$1°\ H_0: \sigma^2 \geq \sigma_0^2$）; $1°: \varphi: \frac{S_1^2}{S_2^2} \geq a F_{n-1,m-1}(1-\alpha)$ $\left(\frac{S_2^2}{S_1^2} \leq \frac{1}{a} F_{m-1,n-1}(\alpha)\right)$ ✓ oth. X（$1°\ H_0: \frac{\sigma_1^2}{\sigma_2^2} \geq a$）

exp. λ: $3°: \varphi'': \frac{\chi^2_{2n}(1-\frac{\alpha}{2})}{2n\lambda_0} \leq \bar{X} \leq \frac{\chi^2_{2n}(\frac{\alpha}{2})}{2n\lambda_0}$ ✓ oth. X（1°.2°一致最优而3°不是且没有 见5.3）
（φ φ' 　 φ'' 　下同.）

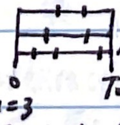
截尾寿命检验（由于部分X过长）：

1. 定数截尾法　$r < n$，至 r 个失效为止. $T = Y_1 + \cdots + Y_r + (n-r)Y_r$

$2\lambda T \sim \chi^2_{2r}$　$H_0: \lambda \leq \lambda_0$ 可由 $\varphi: T \geq \frac{\chi^2_{2r}(1-\alpha)}{2\lambda_0}$.

2. 定时截尾法　至 T_0 为止的总工作时间（未失效以 T_0 计）　$2\lambda T'$ 近似 $\sim \chi^2_{2r+1}$（r为失效个数）

3. 接替定时截尾



X为总失效数.（X=7）　$X \sim Poisson(nT_0\lambda)$

对泊松分布参数进行检验（见下）

Bin. p: $1°\ H_0: p \leq p_0$. $\varphi: X \leq C$ ✓ oth. X. $\beta_\varphi(p) = 1 - \sum_{i=0}^{C} \binom{n}{i} p^i (1-p)^{n-i}$，要求 $\sum_{i=0}^{C} \binom{n}{i} p_0^i (1-p_0)^{n-i} = 1-\alpha$ 得C.

抽样验收　操作特征（OC）（↗）函数 $L_\varphi(p) = \sum_{i=0}^{C} \binom{n}{i} p^i (1-p)^{n-i}$

$C_0 <\quad < C_0+1$ 可略放缩 α（n很大时修改很小）或随机化（检验）

$L_\varphi(p_0) = 1-\alpha$. $L_\varphi(p_1) = \beta$（双约束）　以（p_0时）$\frac{\beta_\phi(C_0) - \alpha}{\beta_\phi(C_0) - \beta_\phi(C_0+1)}$ 的比例通过 C_0 的产品.

1\. 符号检验（如众人给产品打分 $+-+\cdots$）检验 $H_0: p_0=\frac{1}{2}$. $\varphi: c_1\le X\le c_2$ ✓, $\sum_{i=0}^{c_1-1}=\frac{\alpha}{2}$, $\sum_{i=c_2+1}^{n}=\frac{\alpha}{2}$.

非参数检验（如果给分尺度无大差别 t 检验信息或许更多）

总统大选 $\hat p=\frac{m}{n}$. $\approx \hat p \pm u_{\frac{\alpha}{2}}\sqrt{\hat p(1-\hat p)/n}$. 也是二选一的估计, $n\to\infty$ 了.

2\. 民调:

Poisson. λ: 1° $H_0: \lambda\le\lambda_0$. $\varphi: X\le C$ ✓. $\beta_\varphi(\lambda)=1-\sum_{i=0}^{C}\frac{e^{-\lambda}\lambda^i}{i!}=K_{2C+2}(2\lambda)$（K为 χ^2 cdf.）

C_0 满足 $2\lambda_0=\chi^2_{2C_0+2}(1-\alpha)$.

▷ $F_{Poi}(x)=\sum_{i=0}^{x}e^{-\lambda}\frac{\lambda^i}{i!}=\frac{1}{x!}\int_\lambda^\infty t^x e^{-t}dt \overset{t=\frac{t}{2}}{=\!=} \frac{\int_{2\lambda}^\infty t^x e^{-\frac{t}{2}}dt}{2^{x+1}\Gamma(x+1)}$

$=1-F_{\chi^2_{2x+2}}(2\lambda)$.

② 大样本检验: 如 Behrens-Fisher Problem:

$X_1,\cdots,X_n\sim N(\theta_1,\sigma_1^2)$, $Y_1,\cdots,Y_m\sim N(\theta_2,\sigma_2^2)$. $H_0:\theta_1=\theta_2$.

$\varphi: \frac{|\bar X-\bar Y|}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}}\le u_{\frac{\alpha}{2}}$ ✓. oth.x. $n,m>>1$（使用极限分布）

但无法确知与 α 的偏差.

统计推断

③ 贝叶斯法: $P(H_0|X_1,\cdots,X_n) \gtrless P(H_1|X_1,\cdots,X_n)$（若=之则悬而不决）（不一定是之: 统计决策）

当 H_0 为单点时, 可以给该点 θ_0 一个先验概率 p_0, 剩下 $1-p_0$ 以某种分布位于两边.

如 $X_1,\cdots,X_n\sim N(\theta,\sigma^2)$. $H_0: a\le\theta\le b$. θ 广义先验密度 1, σ 广 $\sim\frac{1}{\sigma}$, 得 θ 的边缘后验密度

Bayes:

$P(H_0|X)=T_{n-1}\left(\frac{\sqrt n(b-\bar X)}{s}\right)-T_{n-1}\left(\frac{\sqrt n(a-\bar X)}{s}\right) \gtrless \frac{1}{2}$?

▷ 贝叶斯的分母可以先不算而确定后验分布的形式, 由归一自然猜出系数 〇. $〇\cdot\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}$, $t=\sqrt n\,\frac{\theta-\bar X}{S}$.

Neyman: $t\sim t_{n-1}$. 有 α. 构造否定域.

Fisher: $\sqrt n\frac{(\bar X-\theta)}{s}\sim t_{n-1}$ 反过来决定了 θ 的信仰分布 F 信仰概率 $F(b)-F(a)$（不存在先验分布, 无法则可循）

5.2 ◦ 拟合优度检验 $Z=\sum\frac{(f_e-f_o)^2}{f_e}$ （$H_0: f_e$）

已知离散分布.有限: $H_0: P(X=a_i)=p_i\ (i=1,\cdots,k)$

$Z=\sum_{i=1}^{k}\frac{(np_i-\nu_i)^2}{np_i}$ （$\sum_{i=1}^{k}\nu_i=n$）若 H_0 ✓, $n\to\infty$ 时 $Z\sim\chi^2_{k-1}$.

则 $\varphi: Z\le\chi^2_{k-1}(\alpha)$ ✓ oth.x. 拟合优度 $p(Z_0)=P(Z\ge Z_0|H_0)\approx1-K_{k-1}(Z_0)$

（注意样本量 n 对分辨率的影响）　$p(Z_0)$ 小于 α 则差异愈稀奇, 拒绝 H_0.

半未知离散分布.有限（带参）: $H_0: P(X=a_i)=p_i(\theta_1,\cdots,\theta_r)$ $(i=1,\cdots,k)$（θ 满足一定范围）

对某组 $(\theta_1^0,\cdots,\theta_r^0)$ 成立.

多一步用极大似然估计 θ. ① $L=p_1^{\nu_1}(\theta_1,\cdots,\theta_r)\cdots p_k^{\nu_k}(\theta_1,\cdots,\theta_r)$

$\frac{\partial\ln L}{\partial\theta_i}=0$ 即 $\sum_{i=1}^{k}\frac{\nu_i}{p_i(\theta)}\frac{\partial p_i(\theta)}{\partial\theta_j}=0$ $(j=1,\cdots,k)$ 解出 $(\hat\theta_1,\cdots,\hat\theta_r)$.

② $p_i=p_i(\hat\theta_1,\cdots,\hat\theta_r)$ 计算 Z.

若 H_0 ✓, $n\to\infty$ 时 $Z\sim\chi^2_{k-1-r}$　③ 计算 $p(Z)=1-K_{k-1-r}(Z)$ 比较 α 以决定.

特例: 列联表　齐一/独立性检验　▷ 高维列联表的压缩.分层（边缘表.部分表）

A \ B	ν_j, $j=1,\cdots,c$			
u_i	⋮	⋮	⋮	$n_{1\cdot}$
$i=1,\cdots,r$				$n_{2\cdot}$
	$n_{\cdot1}$	$n_{\cdot2}$	$n_{\cdot3}$	n

（$r=c=2$ 时称四格表 $Z=\frac{n(n_{11}n_{22}-n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot1}n_{\cdot2}}$）

似然估计得 $\hat u_i=\frac{n_{\cdot i}}{n}$, $\hat\nu_j=\frac{n_{j\cdot}}{n}$

$\hat p_{ij}=\hat u_i\hat\nu_j=\frac{n_{\cdot i}n_{j\cdot}}{n^2}$

$\chi^2_{rc-1-(r-1+c-1)=(r-1)(c-1)} \leftarrow Z=\sum_i\sum_j\frac{(n_{ij}-n\hat p_{ij})^2}{n\hat p_{ij}}$.

（$\hat u_i,\hat\nu_j\ne0$, 若为 0 说明样本中不含 i, j 水平, 应视为没有）

高维独立性: （辛普森悖论）

$r\times c\times t$: $rct-r-c-t+2$

A, B, C $\begin{cases}A, BC\ (r-1)(ct-1) \\ B, AC\ \cdots\end{cases}$ $\begin{cases}BA, BC \\ CA, CB\end{cases}$ $c(r-1)(t-1)$

…（→ 推出）

一般（连续、无限）分布，已知或半未知（待估）：H_0：$F(x)$ 或 $F(x|\theta)$ 对一组 θ^0.

多步划分区间. $-\infty = a_0 < a_1 < \cdots < a_{k-1} < a_k = \infty$ 　　并统计各区间频数 ν_i.

$(a_0, a_1] \cdots (a_{k-1}, a_k)$ k 个区间，$p_i(\theta) = F(a_i;\theta) - F(a_{i-1};\theta)$ $(i=1,\cdots,k)$

（$k\uparrow$ 更接近但每个区间 ν 太小（应 ≥ 5）；$k\downarrow$ 使 $n(\nu_i)\uparrow$ 与 χ^2 更接近；

一般 $50 \leq n \leq 100$, $k = 6\sim8$；$100 \leq n \leq 200$, $k = 9\sim12$；$200 < n$, $k < 20$）

有时 $\hat\theta$ 很难解，可用如 $\hat\mu = \bar{X}$ 或 $\frac{1}{n}\sum m_i \nu_i$，$\hat\sigma = S$ 或 $(\frac{1}{n}\sum \nu_i (m_i - \hat\mu)^2)^{\frac{1}{2}}$ 近似计算之.

（分组数据，m_i 为各区间中位数，头尾可参考相邻区间长.）

5.3 • UMPT 简单检验 $H_0: f_0(x)$，$H_1: f_1(x)$ 　（$x = (x_1, \cdots, x_n)$，$f(x_1)\cdots f(x_n)$ 记为 $f(x)$.）

奈-皮 N-P 基本引理 　α 的 UMPT φ 的否定域 Q 应找一 C，使

$$Q = \{x \mid \frac{f_1(x)}{f_0(x)} > C\}，且 \int_Q f_0(x)dx = \alpha.$$

复合检验 　在 H_0:H_1 中各取一值 H_0': θ_0；H_1': θ_1，由 N-P 引理求出其 φ.

若 1. φ 也是 H_0:H_1 的水平 α 检验 2. φ 与 θ_1 无关 则 φ 为 H_0:H_1 的 $\text{UMPT}_{(\alpha)}$.

（对于 $\theta \lesseqgtr a$ 形单侧原假设，θ_0 总取 a） 如 $N(\theta, \sigma^2)$ σ^2 已知，H_0: $\theta \leq a$.

$$Q = \{x \mid \sum_i x_i \cdot (\theta - a) > C\}$$

同理可证 Norm. Exp. Bin. Poisson. 的单侧检验均为 UMPT. 　则 $\sum_i x_i > C$ 此下 α 应有 $C = na + \sqrt{n}\sigma u_\alpha$，与 θ_1 无关.

f_0 双侧可见 $\theta_1 > a$ 时 $\sum_i x_i > C$；$\theta_1 < a$ 时 $\sum_i x_i < C$. X.

证明一个 T 为 UMPT：

1. φ（即 Q 或 A）与 θ_1 无关 2. $\frac{f_1(x)}{f_0(x)} > \frac{f_1(y)}{f_0(y)}$ 　$\forall x \in Q, y \in A$.

• 非中心 t 分布 $Z = \frac{X + \delta}{\sqrt{\frac{Y}{n}}} \sim t_{n,\delta}$ 　$E(t_{n,\delta}) = \delta \cdot \sqrt{\frac{n}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}$. 　（具体分布较复杂）

$\delta_2 > \delta_1$，则 $F_{n,\delta_2}(x) \leq F_{n,\delta_1}(x)$. 　$V(t_{n,\delta}) = \frac{n}{n-2}(1+\delta^2) - E^2(t_{n,\delta})$.

5.1 中 μ；σ^2 未知的 t 检验 $\beta_\varphi(\theta, \sigma) = F_{n-1, \frac{\sqrt{n}(\theta - \theta_0)}{\sigma}}(-t_{n-1}(\alpha)) = f(\frac{\theta - \theta_0}{\sigma})$ 可得其单调性及性质 1.

非中心 χ^2 分布 $(i=1,\cdots,n)$ $X_i \sim N(\mu_i, \sigma^2)$ 　$E(\chi^2_{n,\lambda}) = n + \lambda$

$$\lambda = \sum_i \mu_i^2. \quad \chi^2_{n,\lambda} = \sum_i X_i^2 \quad V(\chi^2_{n,\lambda}) = 2(n + 2\lambda)$$

$X \sim \lambda e^{-\lambda x}$ 时

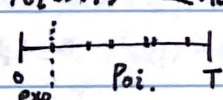
• 截尾法证明：1.: $n \cdot \min_n X \sim X$. 　递推：$T_{k+1} = T_k + (n-k)(Y_{k+1} - Y_k)$

△注意参数可取范围！ 　无后效性使 $Y_{k+1} - Y_k = \min_{n-k} X$. 由此 $\chi^2_{2k} \to \chi^2_{2k+2}$.

max. min! 如 $R(0, \theta)$，θ 下限应为 $\max_i X_i$. 　（Norm. Chi-sq. Bin. Poisson. 均有可加性（再生性））

（Exp.↑）

3.: 设从单个开始替换 X 记 T 时替换次数. 　$X \sim P_{oi}(\lambda T)$ 　递推：$\int_0^T \lambda e^{-\lambda(T-t)} \cdot \frac{e^{-\lambda t}(\lambda t)^n}{n!} dt$



$$= e^{-\lambda T} \frac{(\lambda T)^{n+1}}{(n+1)!}$$

• 卡方检验的证明：1. 一些验证(E.V)：$E(Z)=\sum_i \frac{E(np_i-\nu_i)^2}{np_i} = \sum_i (1-p_i) = k-1$

$V(Z)$ 需 $E(Z^2)$，$E(np_i-\nu_i)^4$ 归为 $\underline{Bin}$：$E(X^4)$，$\triangleright$ 二项、多项分布的构造(由此计算 $E(X^k)$)
而交叉项归为 $\underline{Multi.}(p_i,p_j,1-p_i-p_j)$．　$\underline{Bin}$：$E\left(\frac{X(X-1)\cdots(X-i+1)}{n(n-1)\cdots(n-i+1)}\right)=p^i$ ($\underline{Multi.}$同理)
得 $\underset{(n\to\infty)}{V(Z)} = 2(k-1)$．　(方便化出求和!)

2. 简单的理解：每格均为 Poisson．$\frac{(X-EX)^2}{VX^2} \xrightarrow{D} N$，
加和为 χ^2，由于 $\sum_i \nu_i = n$ 而失去1个自由度．

3. 记 $Y_i = \frac{a_i - np_i}{\sqrt{np_i}}$．　矢量 Y 为多项分布．$\Sigma = \mathrm{Cov}\,Y = \begin{pmatrix} 1-p_1 & -\sqrt{p_1p_2} & \cdots \\ -\sqrt{p_1p_2} & 1-p_2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$
(Y_i 的边缘为二项分布)
$(E(a_ia_j)=n(n-1)p_ip_j)$

有 $\Sigma = I - pp^T$ 其中 $p=(\sqrt{p_1},\cdots,\sqrt{p_k})^T$ 且 $\|p\|=1$．

Σ 为投影阵(幂等)，特征值 1个0，$k-1$个1．($\Sigma p = 0$．$\Sigma^2=\Sigma$．注意 Σ 不可逆)

$n\to\infty$ 时 $Y \xrightarrow{D} N(0,\Sigma)$．现在通过一个旋转 $A\Sigma A^T = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$

则 $X = AY \xrightarrow{D} N(0, A\Sigma A^T)$　(把 $\sum_i p_i = 1$ 的缺失一维抓出来了)

即 $X = (0, X_1, \cdots, X_{k-1})^T$，$X_i \overset{iid}{\sim} N(0,1)$．$Y_i$ 的形式无非是 a_i (多项 → 正态 的归一)

$Z = \sum_i Y_i^2 = \|Y\|^2 = \|X\|^2 \sim \chi^2_{k-1}$．得证．

• 线代在概统中的简单应用：　协方差矩阵 $\Sigma = E((X-\bar{X})(X-\bar{X})^T)$

Σ 为正定(半正定)、对称阵．若 $|\Sigma|=0$ 说明样本中有非独立变量．　$= \begin{pmatrix} V(X_1) & \mathrm{Cov}(X_1,X_2) & \cdots \\ \mathrm{Cov}(X_1,X_2) & V(X_2) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$

Σ 的变换　$Y=AX$，$\Sigma_Y = A\Sigma_X A^T$．Σ 的对角化：找到一组独立的方向，$Y=Q^TX$．
(如二维正态下 $\sigma_2X+\sigma_1Y$，$(a,b,\sigma_1^2,\sigma_2^2,\rho)$　$\sigma_2X-\sigma_1Y$ 独立，但对角化并非此特征向量)　但非唯一！$X\Lambda_1X^T=\Lambda_2 \nRightarrow X$ 为对角阵．
(准确说为不相关)

Σ 的 PCA：找到方差的最大解释正交方向．(见下)

六、回归、相关与方差分析

6.1 • 理论回归方程(函数)　$y = f(X_1,\cdots,X_p)+e$，$E(e)=0$．线性回归分析
自变量的选取、回归函数的形式 ⟶ σ^2 的控制　"模型"
应用：1.描述．2.估计回归函数．3.预测．4.控制　自变量的随机性 回归设计

• 一元线性回归　中心化　$Y_i = \beta_0 + \beta_1(X_i-\bar{X}) + e_i \quad (i=1,\cdots,n)$
e_i iid．$E(e_i)=0$，$V(e_i)=\sigma^2 \quad (i=1,\cdots,n)$

最小二乘法 $\sum_i \delta_i^2 = \sum_i (Y_i-\hat{Y}_i)^2 \to \min$
(亦即极大似然时)

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1(X_i-\bar{X})$
$E(Y_i) = \beta_0 + \beta_1(X_i-\bar{X}) = \hat{Y}_i$

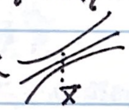
∂偏导、配方或线代(见后)得 $\hat{\beta}_0=\bar{Y}$, $\hat{\beta}_1=\frac{\sum(X_i-\bar{X})Y_i}{\sum(X_i-\bar{X})^2}$ ($\sum(X_i-\bar{X})Y_i=\sum(X_i-\bar{X})(Y_i-\bar{Y})$, $\sum(X_i-\bar{X})X_i=\sum(X_i-\bar{X})^2$)

线性、无偏 $E(\hat{\beta}_0)=\beta_0$, $E(\hat{\beta}_1)=\beta_1$. $V(\hat{\beta}_0)=\frac{\sigma^2}{n}$, $V(\hat{\beta}_1)=\frac{\sigma^2}{\sum(X_i-\bar{X})^2}=\frac{\sigma^2}{S_x^2}$

最小方差 ($I(\beta_0)$、$I(\beta_1)$ 可得)、不相关 $Cov(\hat{\beta}_0,\hat{\beta}_1)=0$ 当$e_i\sim N$时两者独立.

残差方和(估计方差): (利用$E(e_ie_j)=\sigma^2\delta_{ij}$ (或1与$X-\bar{X}$正交) |←两个正态分布的联合分布只要不存在映射(如$Y=X$)则为二维(只有计算)正态分布.

$$\sum_i\hat{\delta}_i^2=\sum_i(Y_i-\hat{Y}_i)^2=\sum_i(Y_i-\bar{Y})^2-\hat{\beta}_1\sum_i(X_i-\bar{X})Y_i$$
$$=\sum_i Y_i^2-n\bar{Y}^2-\hat{\beta}_1\sum_i(X_i-\bar{X})Y_i \qquad E\left(\frac{\sum\hat{\delta}_i^2}{n-2}\right)=\sigma^2$$

区间估计: $\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}/S_x}\sim t_{n-2}$ 或对函数  $\frac{\sum\hat{\delta}_i^2}{\sigma^2}\sim\chi^2_{n-2}$. 回归诊断 ($\rho=0$ iid.)

$\hat{m}(x)=\hat{\beta}_0+\hat{\beta}_1(x-\bar{X})\to m(x)\in\hat{m}(x)\pm\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x-\bar{X})^2}{S_x^2}}\,t_{n-2}(\frac{\alpha}{2})$ (不能"叠加"起来说直线都在此范围里)

$Y_0\in\hat{m}(x_0)\pm\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{X})^2}{S_x^2}}\,t_{n-2}(\frac{\alpha}{2})$ ($n\to\infty$时$m(x)$可以$\to\hat{m}(x)$)

假设检验: $H_0:\beta_1=c$ (常为$c=0$) φ: $|\hat{\beta}-c|\le\hat{\sigma}S_x t_{n-2}(\frac{\alpha}{2})$ √. oth. ×

(一元回归中β_1的检验、回归显著性检验(F)、相关系数检验三者一致.)

使用问题: 1.解释回归方程(变化区间、人为控制) 2.外推(线性的把握)
3.不可逆转使用 $x=\hat{c}+\hat{d}y$, $y=\hat{a}+\hat{b}x$ (除非$\rho\approx0$时) 4. X随机 (仍可使用$\hat{\beta}$,但$V(\hat{\beta})$不同,$\hat{\sigma}^2$应视为固定X下Y方差)

• 多元线性回归 $Y_i=\beta_0+\beta_1X_{1i}+\cdots+\beta_kX_{ki}+e_i$ $(i=1,\cdots,n)$

$\hat{\beta}=(X^TX)^{-1}X^TY$ $\hat{\sigma}^2=\frac{\sum_i\hat{\delta}_i^2}{n-k-1}$. $\sum_i\hat{\delta}_i^2=\sum_i(Y_i-\bar{Y})^2-\hat{\beta}_1\sum_iX_{1i}Y_i-\cdots-\hat{\beta}_k\sum_iX_{ki}Y_i\sim\sigma^2\chi^2_{n-k-1}$

$V(\hat{\beta})=\sigma^2(X^TX)^{-1}$ (元素c_{ij}) (详细见6.4) $\frac{\hat{\beta}_j-\beta_j}{\hat{\sigma}\sqrt{c_{jj}}}\sim t_{n-k-1}$, $\frac{\hat{m}(x)-m(x)}{\hat{\sigma}\sqrt{\frac{1}{n}+\sum_{i,j}(x_i-\bar{x}_i)(x_j-\bar{x}_j)c_{ij}}}\sim t_{n-k-1}$

1.单个回归系数 $H_0:\beta_j=c$. 点 $y_0\in\hat{m}(x_0)\pm\hat{\sigma}\sqrt{1+\lambda^2(x_0)}\,t_{n-k-1}(\frac{\alpha}{2})$ ↓ $\lambda(x)$ ($V(\Sigma X)=\sum_{i,j}(\Sigma)_{ij}$)

2.全体回归系数(回归显著性) $H_0:\beta_1=\cdots=\beta_k=0$ $R_1=\sum_i\hat{\delta}_i^2$, $R_2=\sum_i(Y_i-\bar{Y})^2$ (只剩$\hat{\beta}_0$)
(F检验) $\frac{R_2-R_1}{\sigma^2}\sim\chi^2_k$, $\frac{R_2-R_1}{k\hat{\sigma}^2}\sim F_{k,n-k-1}$

3.部分回归系数 $H_0:\beta_1=\cdots=\beta_r=0$

$R_3=\sum_i(Y_i-\alpha_0-X_{r+1,i}\alpha_1-\cdots-X_{k,i}\alpha_{k-r})^2$ $\frac{R_3-R_1}{r\hat{\sigma}^2}\sim F_{r,n-k-1}$ (此重新拟合可通过X^TX=▦分块消元得出) $R_3-R_1=\hat{\beta}^T(X^TX)\hat{\beta}$.

使用问题: 1.X交互作用(X随机时不应压住其他X单改一个) 2.删去自变量 复共线性→稳定性 $|X^TX|\approx0$时 (高度相关)

可转为线性回归: 如多项式回归 $Y=b_0+b_1X+\cdots+b_pX^p+e$

$Y=b_0+b_1e^{tX}$. $Y=b_0e^{b_1X^2}$ (此处$\ln Y$是否有$e\sim N$待讨论.)

6.2 • **相关分析** 样本相关系数 $r = \dfrac{\sum_i (X_{1i}-\bar{X}_1)(X_{2i}-\bar{X}_2)}{\left(\sum_i (X_{1i}-\bar{X}_1)^2 \sum_i (X_{2i}-\bar{X}_2)^2\right)^{\frac{1}{2}}}$

$H_0: \rho=0$ $\varphi: |r| \leq C$ ✓ o.th.x. C 有 $\dfrac{\sqrt{n-2}\,r}{\sqrt{1-r^2}} \sim t_{n-2}$ ($\rho=0$时)，$C = \dfrac{t_{n-2}(\frac{\alpha}{2})}{\sqrt{n-2+t^2_{n-2}(\frac{\alpha}{2})}}$.

($n=100$, $C\approx 0.2$，较微弱的相关需较大样本量)

(二维正态，Y对X回归可证：$Y_i = b+\beta_1 X_i + e_i$. $\dfrac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{\sum(X-\bar{X})^2}} \sim t_{n-2}$，$\hat{\beta}_1 \leftarrow r\sqrt{\dfrac{\sum(Y-\bar{Y})^2}{\sum(X-\bar{X})^2}}$，$\hat{\sigma} \leftarrow \sqrt{1-r^2}\sqrt{\dfrac{\sum(Y-\bar{Y})^2}{n-2}}$)

偏相关 $X_2' = X_2 - L(X_3,\cdots,X_p)$
$X_1' = X_1 - L(X_3,\cdots,X_p)$ 消去其他变量的影响 $\rho_{12\cdot(3\cdots p)} = \mathrm{Cov}(X_1', X_2')$

相关阵 $P = \begin{pmatrix} 1 & \rho_{12} & \vdots \\ \rho_{12} & 1 & \vdots \\ \cdot & \cdot & \cdot \end{pmatrix}$，$\rho_{12\cdot(3\cdots p)} = \dfrac{P_{12}}{\sqrt{P_{11}P_{22}}}$. (三元时 $\rho_{12\cdot 3} = \dfrac{\rho_{12}-\rho_{13}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}$.)

(P_{ij}为ij元余子式)

(Y对X及$X_3\cdots X_p$同时回归可证) $\dfrac{\sqrt{n-p}\;r_{12\cdot(3\cdots p)}}{\sqrt{1-r^2_{12\cdot(3\cdots p)}}} \sim t_{n-p}$

复相关 其他变量全体对一方的影响 $\rho_{1(2\cdots p)} = \sqrt{1-\dfrac{P}{P_{11}}}$.

$r^2_{1(2\cdots p)}$ 为 $Beta(\frac{p-1}{2}, \frac{n-p}{2})$ 分布，或 $\dfrac{n-p}{p-1}\dfrac{r^2_{1(2\cdots p)}}{1-r^2_{1(2\cdots p)}} \sim F_{p-1,n-p}$.

6.3 • **方差分析** Fisher试验设计三原则：重复、随机化(完全/部分)、分区组

因素、水平 单因素完全随机化实验：$Y_{ij} = a_i + e_{ij}$ ($j=1,\cdots,n_i$; $i=1,\cdots,k$)
$E(e_{ij})=0$, $V(e_{ij})=\sigma^2$, iid. (Normal)

$H_0: a_1=\cdots=a_k$. (平方和分解) $SS = SS_A + SS_e$ $SS = \sum_i\sum_j (Y_{ij}-\bar{Y})^2$.

$\varphi: \dfrac{MS_A}{MS_e} \leq F^{(\alpha)}_{k-1,n-k}$ ✓. $SS_e = \sum_i\sum_j (Y_{ij}-\bar{Y}_i)^2$. $SS_A = \sum_i n_i(\bar{Y}_i-\bar{Y})^2$, $MS_A = \dfrac{SS_A}{k-1}$.

(该因素的显著性：$\bar{Y}_i$与$\bar{Y}_j$的差异 $\leftrightarrow$ 随机误差 $\sqrt{MS_e}$ $MS_e = \dfrac{SS_e}{n-k} = \hat{\sigma}^2$.
显著不代表a中没有相同的，
就其中一对而言 $a_u - a_v \in \bar{Y}_u - \bar{Y}_v \pm \sqrt{\dfrac{n_u+n_v}{n_u n_v}}\,\hat{\sigma}\, t_{n-k}(\frac{\alpha}{2})$ 若包含0则达不到显著.
区间同时成立的概率更小，
可采用多重比较法)

双因素完全~实验：$Y_{ij} = \mu + a_i + b_j + e_{ij}$. ($i=1,\cdots,k$; $j=1,\cdots,l$)
$\sum_i a_i = 0$, $\sum_j b_j = 0$.

$Y_{\cdot\cdot} = \dfrac{\sum_i\sum_j Y_{ij}}{kl}$, $Y_{i\cdot} = \dfrac{\sum_j Y_{ij}}{l}$, $Y_{\cdot j} = \dfrac{\sum_i Y_{ij}}{k}$, $\hat{a}_i = Y_{i\cdot} - Y_{\cdot\cdot}$, $\hat{b}_j = Y_{\cdot j} - Y_{\cdot\cdot}$

$SS = SS_A + SS_B + SS_e$. $SS = \sum_i\sum_j (Y_{ij}-Y_{\cdot\cdot})^2$ $SS_e = \sum_i\sum_j (Y_{ij}-Y_{i\cdot}-Y_{\cdot j}+Y_{\cdot\cdot})^2$

$SS_A = \sum_i l(Y_{i\cdot}-Y_{\cdot\cdot})^2$ (注意不是 $Y_{ij} - \dfrac{Y_{i\cdot}Y_{\cdot j}}{Y_{\cdot\cdot}}$!)

$H_{0A}: a_1=\cdots=a_k=0$, $H_{0B}: b_1=\cdots=b_l=0$ $SS_B = \sum_j k(Y_{\cdot j}-Y_{\cdot\cdot})^2$

$MS_e = \dfrac{SS_e}{(k-1)(l-1)}$. $\varphi: \dfrac{MS_A}{MS_e}; \dfrac{MS_B}{MS_e} \sim F$ 检验同上. 交互作用 SS_{AB} (S_e可能偏大)

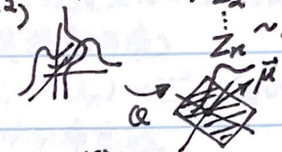
单因素随机区组实验：b_j 区组效应 SS_B 区组平方和 区组大小等于水平数.
划分不当(区组不显著时)，甚至$MS_B < MS_e$，加之以自由度损失，发现A效应更难(S_e更大)

正交表 部分实施水平组合保证[1] 仍可对 a_i, b_j, c_u, d_v 等估计[2] 总平方和SS仍可分解.
($SS_A = \dfrac{n(\hat{a}_1^2+\cdots+\hat{a}_k^2)}{k}$) $L_8(2^4 \times 4)$ n 因素/区组 (为n因数)

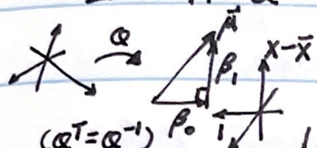
(两区组/两水平 F检验与双样本t检验一致)
('多总体(联合)均值t检验')

6.4 ◦ 矩阵变换视角 $Y \sim N(\beta_0+\beta_1(X-\bar{X}), \sigma^2 I)$

▷另一个例子：$Y \sim N(\mu, \sigma^2 I)$
$Z = QY$.
$Q=\begin{pmatrix} \frac{1}{\sqrt{n}} \cdots \frac{1}{\sqrt{n}} \\ \hline \end{pmatrix}$ $Z_1 \sim N(\sqrt{n}\mu, \sigma^2)$, $Z_2, \dots, Z_n \sim N(0,\sigma^2)$



$Z = QY$. $Q = \begin{pmatrix} \frac{1}{\sqrt{n}} \cdots \frac{1}{\sqrt{n}} \\ \frac{X_i-\bar{X}}{\sqrt{\sum(X-\bar{X})^2}} \\ \hline \end{pmatrix}$



$(Q^T=Q^{-1})$ 1与$X-\bar{X}$正交.

$Z_1 \sim N(\sqrt{n}\beta_0, \sigma^2)$
$Z_2 \sim N(\beta_1\sqrt{\sum(X-\bar{X})^2}, \sigma^2)$
$Z_3, \dots, Z_n \sim N(0,\sigma^2)$

为消除未知(待估)参数，
将坐标架转至该方向，剩余均值为0，
为真正(残差的)自由度.

$\frac{\sum Y}{\sqrt{n}} = Z_1 \to \sqrt{n}\beta_0$, $\hat{\beta}_0 = \frac{\sum Y}{n}$.

$\frac{\sum(X-\bar{X})Y}{\sqrt{\sum(X-\bar{X})^2}} = Z_2 \to \beta_1\sqrt{\sum(X-\bar{X})^2}$, $\hat{\beta}_1 = \frac{\sum(X-\bar{X})Y}{\sum(X-\bar{X})^2}$

自然$\hat{\beta}_0, \hat{\beta}_1$独立分布.

多元时 $Y \sim N(X\beta, \sigma^2 I)$
$\quad \llcorner \beta_0 + \beta_1(X-\bar{X}) + \beta_2(G-\bar{G}) + \cdots$

注意$X-\bar{X}$与$G-\bar{G}$一般不正交(相关，$\frac{\sum(X-\bar{X})(G-\bar{G})}{\sqrt{\sum(X-\bar{X})^2\sum(G-\bar{G})^2}} \neq 0$). 需正交化(如Schimt法).

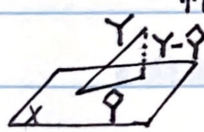
$EY = \beta_0 \cdot 1 + (\beta_1,\beta_2)(X-\bar{X}) + (\beta_1,\beta_2)(G'-\bar{G}') + \cdots$ (一元的"中心化"即正交化)

$(G'-\bar{G}') = \frac{Ⓖ-(Ⓖ\cdot Ⓧ)Ⓧ}{\sqrt{1-(Ⓖ\cdot Ⓧ)^2}}$ ($\nearrow^{X} \nearrow G$, G') 同理可由 $\sum G'Y = (\beta_1,\beta_2)$ 解得$\hat{\beta}_1, \hat{\beta}_2$.
$\sum XY = (\beta_1,\beta_2)$ (与∂偏导一致)

残差方和 $\sum_i e_i^2 = \sum_i (Y_i-\hat{Y}_i)^2 = \|Y-\hat{Y}\|^2 = \|Q(Y-\hat{Y})\|^2 = \|Z-Q\hat{Y}\|^2$

注意 $Q\hat{Y} = \begin{pmatrix} Z_1 \\ Z_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $\hat{Y}=X\hat{\beta}$, 就是用Z_1, Z_2估计出的$\hat{\beta}$
$\sim \chi^2_{n-2}$ (χ^2已达到$\sum_i \varepsilon_i^2{}_{min}$了)

◦ 帽子矩阵 $X\hat{\beta} \simeq Y \sim N(EY, \sigma^2 I)$, $EY = X\beta$ 其中设计矩阵 $X=(1\ X_1\ X_2 \cdots)$.



$X^TX\hat{\beta} = X^TY$. $\hat{\beta} = (X^TX)^{-1}X^TY \sim N(\beta, \sigma^2(X^TX)^{-1})$

$X^TX = \begin{pmatrix} n & 0 \\ 0 & \sum X_i X_j \cdots \end{pmatrix}$

(二元时例:
$\begin{pmatrix} \sum X^2 & \sum XG \\ \sum XG & \sum G^2 \end{pmatrix}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \sum XY \\ \sum GY \end{pmatrix}$)

则 $E\hat{\beta} = \beta$, $V\hat{\beta} = \sigma^2 \cdot (X^TX)^{-1}$

$(X^TX)^{-1} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \end{pmatrix}$

$\hat{Y} = HY$. $\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}\sqrt{c_{ii}}} \sim t_{n-k-1}$.

(H一元时例: $h_{ij} = \frac{1}{n} + \frac{(X_i-\bar{X})(X_j-\bar{X})}{\sum(X-\bar{X})^2}$.)

$(\frac{1}{n} \le h_{ii} \le 1)$ $H = X(X^TX)^{-1}X^T$. 对称、幂等、半正定. $HX = X$, $H1 = 1$.
(投影阵) $(I-H)X = 0$.

$\sum_i h_{ii} = \mathrm{tr}\,H = r\,H = k+1$
$\sum_i\sum_j h_{ij} = n$. (X^TX满秩)

$e = Y - \hat{Y} \sim N((I-H)EY, \sigma^2(I-H))$ 注意残差不独立，
$\qquad\qquad\qquad\qquad \overset{\|}{0}$
但与估计值独立且正交.

$V(e_i) = E(e_i^2) = \sigma^2(1-h_{ii})$

$Cov(\hat{Y}, e) = H^T CovY(I-H)$
$= \sigma^2 H^T(I-H) = 0$.

$\hat{\sigma}^2(n-k-1) = \sum_i e_i^2 = e^Te = (X\beta+\varepsilon)^T(I-H)(X\beta+\varepsilon)$

$(=\beta^TX^TX\beta = n\bar{Y}^2 + \sum_k \hat{\beta}\sum XY$ 或 $= \|Q\hat{Y}\|^2 = Z_1^2 + \cdots + Z_{k+1}^2)$ $= \varepsilon^T(I-H)\varepsilon \longrightarrow \chi^2_{n-k-1}$ (该二次型经过旋转可以化为$\Lambda = \begin{pmatrix} 0 & \\ & I \end{pmatrix}$).

$\|Y\|^2 = \|\hat{Y}\|^2 + \|Y-\hat{Y}\|^2$ (ε为误差项, $V(\varepsilon)=\sigma^2$)

— $\sigma^2 H$, 由于$H1=1$对角化后有1个1属于β_0 ($EY=\beta_0 1$, Q第一行$\frac{1}{\sqrt{n}}$ 应删去.

$\downarrow \qquad \downarrow$
$n\bar{Y}^2 + \chi^2_k + \chi^2_{n-k-1}$ (解释: 1. Q正交化 2. 用H直接写出$\hat{Y}$, $Y-\hat{Y}$分布)

$\downarrow$
$n\bar{Y}^2 + \chi^2_{k-r} + \underbrace{\chi^2_r + \chi^2_{n-k-1}}_{R_3}$

▷ 残差图诊断：线性、方差齐性、不相关性、正态性；
异常点、强影响点/高杠杆点(h_{ij}).

学生化残差 $r_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$. ($|r_i| \ge 2$.)

标准化残差 $\frac{e_i}{\hat{\sigma}}$.

○ 伽马、贝塔、指数、泊松分布相关：

$Ga: f(x) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} \left(= \frac{(\lambda x)^{\alpha-1}}{(\alpha-1)!} e^{-\lambda x} \cdot \lambda\right)$　$\sum_n$ i.d. 指数分布 $(\lambda) = Ga(n, \lambda)$

$\int_t^{+\infty} Ga(n,\lambda)\,dx = \sum_{i=0}^{n-1} Po_i(\lambda t)$　与泊松分布"互逆"，为发生 n 次的时间.

（泊松过程，每一段都是指数分布）

（由此可解释 $2\lambda(X_1+\cdots+X_n) \sim \chi^2_{2n} = Ga(n, \frac{1}{2})$）

Beta: $Beta(\alpha,\beta) \times Bin(n,k) \longrightarrow Beta(\alpha+k, \beta+n-k)$　二项分布先验

$f(x) = \frac{1}{B(\alpha,\beta)} x^\alpha (1-x)^\beta$　且 $X \sim Beta(\alpha,\beta)$ 时，$\frac{\beta}{\alpha}\frac{X}{1-X} \sim F_{2\alpha, 2\beta}$.

T_1 (k_1)　(k_2) T_2

$Ga(k_1,\lambda)$　$Ga(k_2,\lambda)$

比例的分布：则 $\frac{T_1}{T} \sim Beta(k_1,k_2)$.　$\left(\frac{\chi^2_{2\alpha}}{2\alpha} \Big/ \frac{\chi^2_{2\beta}}{2\beta}\right)$，$\chi^2_{2\alpha} = Ga(\alpha, \frac{1}{2})$，$\chi^2_{2\beta} = Ga(\beta, \frac{1}{2})$

$T = T_1 + T_2 \sim Ga(k_1+k_2, \lambda)$　（甚至只需 i.i.d. 即可，可视作 p_1 $p_2 = 1-p_1$）

○ $\hat\beta$ 为 BLUE 证明：1. $Cov(\hat\beta) = \sigma^2 (X^TX)^{-1}$　若另一估计量 $\beta' = A^T Y$（A与X同形状）

应满足 $E(\beta') = \beta \longrightarrow A^T X\beta = \beta$, $\forall\beta$! $A^TX = I$. 则 $V(\beta') = \sigma^2 A^TA \geqslant \sigma^2 (X^TX)^{-1}$

故 $V(c\beta') = c^T V(\beta') c$ 始终 $\geqslant V(c\hat\beta)$,　因 $A^TA - (X^TX)^{-1} = A^T(I-H)A \geqslant 0$

$\forall c_i\beta_i$ 组合均有 $\hat\beta$ 最小方差.　└ 半正定.

2.（假设 $e \sim Normal$）$Y \sim N(X\beta, \sigma^2 I)$，$I(\beta) = \int \frac{\left(\frac{\partial \vec f}{\partial \beta}\right)^2}{\vec f} d\vec y$　其中

故 CR 下界 $V_{min}(\beta) = \frac{1}{I(\beta)} = \sigma^2 (X^TX)^{-1}$.　$\frac{\partial \vec f}{\partial \beta} = \vec f \cdot \frac{1}{\sigma^2} \cdot X^T(Y - X\beta)$，（平方应视作 xx^T.）（向量求导）

$\left(\vec f = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{(y-X\beta)^T(y-X\beta)}{2\sigma^2}}\right)$　$I(\beta) = \int \frac{1}{\sigma^2} \cdot \frac{X^T(Y-X\beta)(Y-X\beta)^T X \vec f}{\sigma^2} d\vec y = X^T \frac{I}{\sigma^2} X$.

○ 组间方差分布证明：$\bar Y_i \sim N(a_i, \frac{\sigma^2}{n_i})$，$\sqrt{n_i}\bar Y_i \sim N(\sqrt{n_i}a_i, \sigma^2 I)$（将 $X^T \bigcirc X$ 提出归一）（参考线代）

$\frac{n_i \bar Y_i}{n} = \bar Y$. 则 Q 第一行取 $\left(\frac{\sqrt{n_i}}{\sqrt n}\right)^T \longrightarrow \sqrt n \bar Y$. 得证.

○ 全书小结：

概率论
- 概率（几何）与分布　古典概率、贝叶斯、随机变量、分布（三大）、E.V...
- 极限理论　大数、中心极限
- 随机过程与 Markov 过程
- 鞅论等

数理统计
- 参数估计与假设检验（统计推断）　奈曼、贝叶斯、大样本、极大似然
- 非参数统计　符号检验、χ^2 拟合优度
- 相关与回归分析　H. r...
- 统计决策
- 试验设计与分析、抽样理论等　ANOVA.
- 时间序列分析

〈补〉概率导论

1 样本空间与概率

条件独立　　计算概率的方法: 1.计数法(排列/组合/分割)(古典概型)
2.序贯树形图 3.全概率公式

· 邦费罗尼不等式 $P(\bigcap_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$　　条件概率的全概率公式
$P(A|B) = \sum_i P(C_i|B) P(A|BC_i)$

· 博雷尔-坎泰利引理 $\{P_i\}$ $i=1,\cdots,+\infty$. N为序列事件均未发生, I为有无穷多次发生.
若 $\sum_{i=1}^{\infty} p_i < \infty$ 则 $P(I)=0$.

· 生日问题

2 离散随机变量

条件期望　全期望定理 $E(X) = \sum_y P_Y(y) E(X|Y=y)$ (或积分)

· 巴拿赫火柴问题　· 圣彼得堡悖论　　容斥恒等式(示性函数证明)

· 熵与信息量 $H(X) = -\sum_i p_i \log p_i$. $H(X) \leq -\sum_i p_i \log q_i$ 交叉熵 $(\sum_i q_i = 1)$

互信息 $I(X,Y) = H(X) - H(X|Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \geq 0$
$= H(X) + H(Y) - H(X,Y)$

3 一般随机变量

· 布丰投针 $\frac{2l}{\pi d}$　联合累积分布函数　· 混合随机变量 $F_X(x) = pF_Y(x) + (1-p)F_Z(x)$

随机个随机独立变量和 $T = X_1 + \cdots + X_N$ (X_i. N i.d.)　$E(T)$. $V(T) = V(X) \cdot E(N) + E^2(X) \cdot V(N)$
$= E(X) \cdot E(N)$

4 随机变量高级主题

导出分布　　· 拉普拉斯分布　　和的方差 $V(\sum_i X_i) = \sum_i V(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$

重期望法则 $E(E(X|Y)) = E(X)$　　用条件期望作为估计量: $\hat{X} = E(X|Y)$, $\tilde{X} = \hat{X} - X$.
(外面的E为E_Y)　　(如Bayes估计) 则 $E(\tilde{X}) = 0 = E(\tilde{X}|Y)$,　　(含X和Y)

条件方差　全方差定理 $V(X) = E(V(X|Y)) + V(E(X|Y))$.　$Cov(\hat{X}, \tilde{X}) = 0$. $V(X) = V(\hat{X}) + V(\tilde{X})$.
(代入 $E(\tilde{X}^2) = E(V(X|Y))$ 得全方差式)

矩母函数 $M_Y(s) = M_N(\ln M_X(s))$.　　可得 $V(XY) = E^2(X)V(Y) + E^2(Y)V(X) + V(X)V(Y)$
$M_X(s) = \sum_x e^{sx} p(x)$　($Y = X_1 + \cdots + X_N$)　　(X. Y i.d.)
$= E(e^{sX})$ (常见分布的矩母函数, 由形式判断分布)

5 极限理论

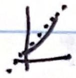
$\lim_{n\to\infty} P(|Y_n-a|\geq\varepsilon)=0, \forall\varepsilon>0.$ $P(\lim_{n\to\infty} Y_n=a)=1$

弱大数定律 依概率收敛 强大数定律 以概率1收敛（几乎处处收敛）

· 切诺夫界 利用马尔可夫不等式。

$P(X\geq a)\leq \frac{M_X(s)}{e^{sa}}$. 令 $\sum_i X_i = X$. 说先

紧缩矩母：针对 $X_i\in[0,1]$（如伯努利），e^{sX} 为凸函数.



$E(e^{sX})\leq(e^s-1)E(X)+1\leq e^{E(X)(e^s-1)}$ （可Y取等为泊松分布. 0/1. $p\to 0$）

$M_X(s)=E(e^{sX})\leq\prod_i e^{E(X_i)(e^s-1)}=e^{\mu(e^s-1)}$

则 $\forall s\geq 0$ 有 $P(X\geq a)\leq e^{\mu(e^s-1)-sa}$ $(\mu=E(X))$

取 $a=(1+\xi)\mu$, $s=\ln(1+\xi)$ 得 $=e^{\mu(\xi-(1+\xi)\ln(1+\xi))}\leq e^{-\frac{\mu\xi^2}{2}}$ （适当放）

同理 $s\leq 0$ 由 $P(X\leq a)\leq e^{-sa}M_X(s)$ 得 $P(X\leq(1-\xi)\mu)\leq e^{-\mu(\xi+(1-\xi)\ln(1-\xi))}\leq e^{-\frac{\mu\xi^2}{3}}$.

中心极限定理的证明：$Z_n=\frac{X_1+\cdots+X_n}{\sigma\sqrt{n}}$.

$M_{Z_n}(s)=\left(M_X\left(\frac{s}{\sigma\sqrt{n}}\right)\right)^n$

利用了 $M_{Z_n}\to M_Z$ 则 $F_{Z_n}\to F_Z$.

对0展开 $M_X(s)=1+0\cdot s+\frac{\sigma^2}{2}\cdot s^2+o(s^2)$

得到 $\lim_{n\to\infty} M_{Z_n}(s)=e^{\frac{s^2}{2}}$. 标准正态.

· 琴生不等式（期望的） f 凸 $(\frac{\partial^2 f}{\partial x^2}>0)$：$f(E(X))\leq E(f(X))$.
（代入 $f(a)+(x-a)\frac{df}{dx}(a)\leq f(x)$）

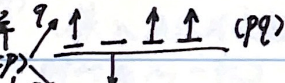
· 依均方收敛 $\lim_{n\to\infty}E(X_n-a)^2=0$
$\subset$ 依概率收敛（切比雪夫不等式）（强于）

6 伯努利过程和泊松过程

伯努利过程 无记忆性与独立性

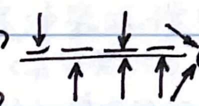
1. 相邻到达的间隔时间：$p(1-p)^{t-1}$ 几何分布.
2. k 次到达时间：$T_i \overset{i.i.d.}{\sim} Geo(p)$. $T=\sum_{i=1}^k T_i\sim k$ 阶帕斯卡分布.
3. t 时间内的到达次数：$\binom{t}{k}p^k(1-p)^{t-k}$ $\binom{t-1}{k-1}p^k(1-p)^{t-k}$ 布.
二项分布. （k 次到达前的失败次数 ~ 负二项分布. 可代入 $t'=t-k$）
（△注意2.1.3.的'互逆'关系）

4. 分裂与合并



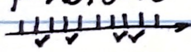
(pq)

（下面均可如此类推）



$(p+q-pq)$ $(=1-(1-p)(1-q))$

$p=\lambda\delta, \delta\to 0$



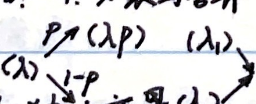
泊松过程（点过程） 到达率 λ 时间同质性与独立性（无记忆性） 小区间概率

$P(0,\tau)=1-\lambda\tau+o(\tau)$
$P(1,\tau)=\lambda\tau+o(\tau)$
$P(k,\tau)=o(\tau)$ $(k\geq 2)$

1. 相邻到达的间隔时间：$\lambda e^{-\lambda t}$ 指数分布.
2. k 次到达时间 $T_i\overset{i.i.d.}{\sim}Exp(\lambda)$, $T=\sum_{i=1}^k T_i\sim k$ 阶埃尔朗分布. $\frac{\lambda^k t^{k-1}e^{-\lambda t}}{(k-1)!}$ $(Ga(k,\lambda))$ $(\sum_{k=1}^{\infty}Ga(t)=\lambda)$
3. t 时间内的到达次数：$\frac{(\lambda t)^k e^{-\lambda t}}{(k-1)!}$

△求概率分布：
① （分析事理）分析该点/事件发生概率 $(f(x)dx)$
② 累积分布函数求导（$P(X\leq x)$）. 矩母函数等.

4. 分裂与合并



$X_1\sim Exp(\lambda_1)$. $X_2\sim Exp(\lambda_2)$. i.d.
则 $\min\{X_1,X_2\}\sim(\lambda_1+\lambda_2)$（串联系统）$E(\min)=\frac{1}{\lambda_1+\lambda_2}$.
但 $\max\{X_1,X_2\}$ 不是 Exp.! $E(\max)=\frac{1}{\lambda_1+\lambda_2}+\frac{1}{\lambda_2}\cdot\frac{\lambda_1}{\lambda_1+\lambda_2}+\frac{1}{\lambda_1}\cdot\frac{\lambda_2}{\lambda_1+\lambda_2}$

· 随机个随机独立变量和 — 随机过程

· 对称性：倒流的伯努利/泊松过程没有区别.

随机插入

无记忆性导致应为两段 i.d. 加和.（二阶帕斯卡/埃尔朗）（注意调查的固定/参考点.）

· 无限服务队列 $\pi(\lambda E(X))$

7 马尔可夫过程

马尔可夫链模型　转移概率矩阵　n步转移概率. K-C方程

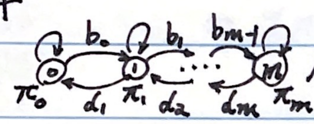
状态：可达的 $i\to j\in A(i)$　常返/非常返的 $\forall j\in A(i),\ i\in A(j)$.　（$A$存在性：$\forall i$，一定$\exists j\in A(i)$且常返）

（性质）常返类R　周期的 $i\in S_k$ 且 $p_{ij}>0$ 则 $j\in S_{k \bmod T+1}$.　$\left(\begin{matrix}\cup S=R\\ \cap S=\phi\end{matrix}\right)$　/非周期的 $\exists n,\ s.t.\ \forall i,j\in R,\ p_{ij}(n)>0$.

除多类链（看初态：可能有进入过程）、周期类（$p_{ij}(n)$摆动）外，$\exists$稳态概率 $\pi_j=\lim\limits_{n\to\infty}P(X_n=j)$
（非无穷链）　　　　　　　　　　　　（与初始X_0无关）

平衡分布. ～方程组 $\begin{cases}\pi_j=\sum\limits_i \pi_i p_{ij}\\ 1=\sum\limits_i \pi_i\end{cases}$　长期行为的频率解释（大数Th.）

·门口没有伞

·生灭过程  局部平衡方程 $\pi_i b_i=\pi_{i+1}d_{i+1}$.

吸收概率 $a_i=P(X_n=s|X_0=i)$，s为吸收的.　～期望时间方程组 $\begin{cases}\mu_i=0 & i\text{常返}\\ \mu_i=1+\sum\limits_j p_{ij}\mu_j & \text{非～}\end{cases}$

（期望）首访时间与回访时间（n不够大）　$P_{ss}=1$　$(i\to s)\begin{cases}t_s=0\\ t_i=1+\sum\limits_j p_{ij}t_j\end{cases}$，$t_s^*=1+\sum\limits_j p_{sj}t_j=\dfrac{1}{\pi_s}$.

连续时间马尔可夫链 $q_{ij}=\nu_i p_{ij}$　转移速率矩阵　同理有稳态收敛定理

$X(t)\ \underbrace{\xrightarrow{\ i\ \quad j\ }}_{Exp.(\nu_i)}$ $(q_{ij}\delta+o(\delta))$　（默认忽略 $p_{ii}=q_{ii}=0$）　$\forall i,\ \lim\limits_{t\to\infty}P(X(t)=j|X(0)=i)=\pi_j$

嵌入马尔可夫链 X_n　π为 $\begin{cases}\pi_j\sum\limits_{k\neq j}q_{jk}=\sum\limits_{k\neq j}\pi_k q_{kj}\\ 1=\sum\limits_k \pi_k\end{cases}$ 的解.

（当$\nu_i\equiv\nu$时两者π_i一致）

·（排队论）缓冲器存信息（丢包概率）　局部～（时逆性）$\pi_j q_{ji}=\pi_i q_{ij}$.　其中所有非常返态$j$有$\pi_j=0$.

抽样马尔可夫链 $Y_n=X_{ln}$. $r_{ij}(l)$. π_i一致.

8 贝叶斯统计推断

最大后验概率(MAP)　最小均方(LMS)　线性最小均方(LLMS)

$E(\Theta|X)$　　$E((\Theta-aX-b)^2)_{min}$

·匹配的滤波器 $X_i=a_i+W_i$　$\|a\|^2$. $\sum\limits_i a_iX_i$.　（正态噪声模型下，LMS即LLMS.）$\longrightarrow \hat\Theta=E(\Theta)+\rho\dfrac{\sigma_\Theta}{\sigma_X}(X-E(X))$　（其中$E(X)=E(E(\Theta|X))$.）

此时方差为$(1-\rho^2)\sigma_\Theta^2$. $\rho=\dfrac{Cov(\Theta,X)}{\sigma_\Theta\sigma_X}$

（多元$\sum\limits_i a_iX_i$时求偏导得 $\hat\Theta=\dfrac{\frac{\mu}{\sigma_0^2}+\sum\limits_i\frac{X_i}{\sigma_i^2}}{\sum\limits_i\frac{1}{\sigma_i^2}+\frac{1}{\sigma_0^2}}$）

9 经典统计推断

贝叶斯线性回归（最大后验概率，$\dfrac{\theta_0^2}{2\sigma_0^2}+\dfrac{\theta_1^2}{2\sigma_1^2}+\sum\limits_i\dfrac{(y_i-\theta_0-\theta_1x_i)^2}{2\sigma^2}\longrightarrow min$.　在$\sigma_0^2\sigma_1^2\to\infty$不提供先验信息时与经典回归一致）

y LLMS.　（为最小二乘法提供了合理性）

回归方法的考虑：1.异方差性（加权～）
2.非线性 3.多重共线性 4.过度拟合 5.因果关系

（f为对参数取max，此处为两点式）

似然比检验 $L(x)=\dfrac{f(x;H_1)}{f(x;H_0)}$ 临界值ξ s.t. $P(L(x)>\xi;H_0)=\alpha$，

当观察x有$L(x)>\xi$时拒绝H_0.

奈-皮引理：$P(L(x)>\xi;H_0)=\alpha$

$P(L(x)\leq\xi;H_1)=\beta$ 若∃其他检验拒绝域为R，

$P(X\in R;H_0)\leq\alpha$，则$P(X\in R;H_1)\geq\beta$.

（证明：$\xi=\dfrac{p_\Theta(\theta_0)}{p_\Theta(\theta_1)}$ 此时MAP其犯错概率 $\leq\dfrac{\xi}{1+\xi}(<\alpha)+\dfrac{1}{1+\xi}(\geq\beta)$） （$H_0$. H_1试作两点） （应能计算$L(x)$及其分布）

广义似然比检验 最大似然$p_X(x;\theta)$得$\hat\theta$，计算$\dfrac{p_X(x;\hat\theta)}{p_X(x;\theta_0)}$根据$\alpha$判断$H_0$.

χ^2检验推导：$p_X(x;\theta)=c\theta_1^{n_1}\cdots\theta_m^{n_m}\longrightarrow\hat\theta_i=\dfrac{n_i}{n}$ $(n=\sum_i n_i)$.

φ：$\prod_i\dfrac{(\frac{n_i}{n})^{n_i}}{(\theta_{0i})^{n_i}}>\xi$ ✓. 即$\sum_i n_i\ln\dfrac{n_i}{n\theta_{0i}}>\ln\xi$ 当$n\to\infty$时（大数）$\dfrac{n_i}{n}\to\theta_{0i}$（$H_0$下），

二阶展开得$\approx\dfrac{1}{2}\cdot\sum_i\dfrac{(n_i-n\theta_{0i})^2}{n\theta_{0i}}$.

（分布证明见前）.