approaches to statistical inference @ UTD of professor Wiorkowski: first session

Meisam Hejazinia

1/16/2013

Each week you will have assignment, and will be due next week. Assignments would be combination of theory and application. You need to do display and look at graphs, and you can use MATLAB or SAS or any other application.
Illustration will be done in the class in excel.
Mathematica is good, since it does symbolic manipulation could be done within.

Text is very good, and you can use it in future as well.
It is very non-intuitive in its derivation, but are not insightful.
Multivariate distribution is not included, but it is important.
We will spend some idea about logic and idea behind them, and use some math, but we will not put so much time in proving theorem.
Goal here is to cover many topics with enough depth.
Assignment means look at what you did and tell what it means. In other word it means what is practical implication.

Statistics will use math, but is different from math in the sense that math tries to come up with answer, and say answer is.

Here we give up on that here.

Real life straightlines are not really line, and are points around the line, and have variability.

Statistics are estimating real world.

In the real world we do not have to prove axioms as true or false, and people make decision under uncertainty.

There would be chance of saying things are percentage right or false.

Essense of working in statistics is a concept of random variable, which is an unknown quantity, which takes on various values (could be infinite) with fixed probabilities.

Discrete $x_1, x_2, , x_n$ so that $\exists x$ between x_i, x_j such that $p(x) = 0$.

$p(x_1), p(x_2), , p(x_n)$

$p(x_i) \geq 0$ and $\sum_i p(x_i) = 1$ and bionomial distribution or poisson distribution.

Continuous analog $f(x)$ associated with x is called probability density function $f(t)$ that $f(t) \geq 0$, and the interval will have $\int_{-\infty}^{\infty} f(x)dx = 1$

$p(x) = 0$

Gambling initiated probability principles.

Statistics came from biological sciences, as opposed to physics and chemistry. Physicians did not deal with uncertainty, and engineers as well, and that was why statistics did not stemed from them.

In bilogy you have organizations that behave by themselves and do not take care of other people concerns. They were independent and uncorrelated.

Normal distribution came from bilogy, since normally independent creature behaved.

In business and economics it is not really normal, since it is based on people and people care about what other people do.

In physics you had errors, and that could have been normal.

In the business word usually the curves are skewed.

Tails are heavier than normal distribution, and the economic problems are stemmed from this important matter.

Continuous and discrete are interchangable.

in discrete $p(x \leq a) \neq p(x < a)$ while in the continuous $p(x \leq a) = p(x < a)$

$F(a) = P(x \leq a) = \int_{-\infty}^{a} f(x)dx$

is called cumulative distribution function.

$p(a \leq x \leq b) = F(b) - F(a)$

Human mind bilogicially can not differentiate between S curves, and therefore they look at the probability density functions.

p.d.f. would be $\frac{d}{dx}F(x)$

parameters of the distribution

Location gives you the state of the parameter.

Then you have scale of distribution and one is wider and one is higher. It is like water baloon.

When it is pulled out on right or left is skewed.

The third one is kurtosis, which talkes about symmetry of the distribution.

Platocurtic is when you push the bull from up, they have heavier tails.

When it is higher than normal and push from two sides, it would become laptocurtic.

Bimodal is combination of two normal and has two picks, for man and woman intersection of curves.

Median is numerical value separating the higher half of a sample.

Expectation is the expected value $E(h(x)) = \int_{-\infty}^{\infty} h(x)f(x)d(x)$ and $\sum_{-\infty}^{\infty} h(x)p(x)$

Momement of distribution:

$\alpha_x = \int_{-\infty}^{\infty} x^{\alpha} f(x)dx = E(x^{\alpha})$

Momement at zero:

$E(x) = \alpha_1 = \int_{-\infty}^{\infty} xf(x)dx = \mu$
$E(x^2) = \alpha_2 = \int_{-\infty}^{\infty} x^2 f(x)dx = \sigma^2 + \mu^2$
$var(t) = \sigma^2 = E(x^2) - E^2(x)$

$\mu_n = \sum_{k=0}^{m}(-1)^k C(m,k)\alpha_1^k \alpha_{n-k}$

$\mu_2 = \alpha_2 - \alpha_1^2$
$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3$

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

Standard normal (0,1) $f(x) =$
$\mu = E(x) = \int_{-\infty}^{\infty} f(x)dx = -f(x) = 0 - 0 = 0$

$E(x^n) = \int_{-\infty}^{\infty} x^n f(x)dx = -x^{n-1}f(x) + \int_{-\infty}^{\infty}(n-1)x^{n-2}f(x)dx = 0$

This is key feature of normal distribution. Therefore:

$E(x^k) = (k-1)E(x^{k-2})$ for normal distribution

For standard normal distribution mean is zero and variance is 1.

Thefore normal distirbution will have:

$\alpha_k = (k-1)\alpha_{k-2}$

$\alpha_1 = 0$

$\alpha_2 = 1$

$\alpha_3 = 0$

$\alpha_4 = 3$

For normal measure of location is μ, our measure of scale is σ^2

our measure of skewness is:
Kurtosis would be $\frac{\mu_4}{\mu_2^2} = \gamma_2 = 3$

$e^{tx} = g(x,t)$

$\frac{\partial g}{\partial t} = xe^{tx}$

$E(e^{tx})$ = moment generating function $= 1 + \sum_{k=1}^{\infty} \alpha_k \frac{t^k}{k!}$

MGF is not necessarily unique.

if you can not define a neighborhood raound $t = 0$

$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-(\ln x)^2/2}$

$f_2(x) = f_1(x)(1 + sin(2\pi Ln(x)))$

Above two have the same moment generating function.

Suprisingly $E(e^{itx}) = \phi_f(t)$ $i = \sqrt{-1}$ is unique defined at $t = 0$

$\int_{-\infty}^{\infty} |\phi_f(t)|^2 dt < \infty)$

This is called Characteristic function

$\phi_f(t) = E_f(cos(tx)) + iE_f(sin(tx))$

$\alpha_k = \frac{\partial^n}{\partial t^n} \phi_f(t)/i^n$

Standard notation:

$E(e^{itx}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{itx - \frac{1}{2}x^2} dx$

$E(e^{itx}) = e^{-t^2/2}$

For the imaginery number the vertical axis would be i, and the horizontal axis would be x which is real number.

We are integrating with respect to x, and this helps us to derieve this.

$y = \mu + \sigma x$

$E(e^{ity}) = E(e^{t\mu + t\sigma x}) = e^{t\mu} E_x(e^{t\sigma x}) = e^{t\mu}\phi_x(\sigma t) = e^{t\mu - 1/2\sigma^2 t^2}$

We will take derivitive of different distribution and take a look at what their characteristic function are. Characteristic function uniquely determines the distribution function.

Discrete distribution function:

Probability Generating Function $G(t) = E(t^\alpha) = \sum_{x=0}^{\infty} p(x) z^x$
$E($
$G'(1) = \mu$
$G''(1) = E(x(x-1))$
$G^{(n)}(1) = E(\frac{x!}{(x-k)!})$

a) χ^2 distribution
Let $x_1, z_2, ... z_t$ be independent identically distribute N(0,1) random variable

they would be i.i.d. , $N(0,1)$

$y = \sum_{i=1}^{p} z_i^2$

let's start with normal distribution

Then y is distributed as the χ_p^2 distribution which means χ^2 with p level of freedom

$f(y) = \frac{1}{2^{p/2}\Gamma(p/2)} e^{y/2} y^{(p/2)-1}, y > 0$
$\phi_{x_p'}(t) = (1-2it)^{-\frac{p}{2}}$
$E(y) = p$
$var(y) = 2p$
$Skew(y) = \sqrt{\frac{8}{p}}$
$\gamma_1 = 2\sqrt{\frac{2}{p}}(2p)$
$=1\frac{}{p^2}$

We will have right skewd distribution.

As p starts to get large this coefficient will go to zero.

This will start to look like bell shape.

This means that the skewness will go away.

The idea is to increase the number of models you have, so that the data would fit them.

Let $z_1, z_2, .., z_k$ would be i.i.d $N(0,1)$.

$x \sim N(0,1)$

$x \sim N(0,1)$ would be i.i.d and would be independent.

$t = \frac{x\sqrt{n}}{\sqrt{\sum_{i=1}^{n} z_i^2}} \sim$ this is called t distirbution.
$f(t) = \frac{\Gamma((P+1)/2)}{\sqrt{p\pi}\Gamma(p/2)}(1+\frac{t^2}{p})^{-(p+1)/2}$

After 30 samples you can not distinguish between result and the degree of distribution.

$v \sim \chi_{p1}^2$ and $w \sim \chi_{p_2}^2$

$x = \frac{v/p_1}{w/p_2} = \frac{p_2}{p_1}\frac{v}{w}$

$x \sim F(p_1, p_2)$ Fisher distribution.

$p(x) = \frac{(p_1/p_2)^{p_1/2}}{B(p_1/2, p_2/2)} x^{(p_1/p_2)-1}(1+\frac{p_1}{p_2}x)^{\frac{p_1+p_2}{2}}$

$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$

$\mu_X(k) = \left(\frac{d_2}{d_1}\right)^k \frac{\Gamma(d_1 2+k)}{\Gamma(d_1 2)}\frac{\Gamma(d_2 2-k)}{\Gamma(d_2 2)}$

Wikipedia is great source for these distributions.

Bionomial distribution:

a) variable with 2 possible outcomes (two possible outcomes) which generically called success failure

b) conduct n-independent trails one trial is independent of another.

c) probability of success is constant (p) from trial to tiral

$x = \#$ of successes in trials.

$p(x) = \binom{n}{k} p^k (1-p)^{n-k} \quad \phi(x) = (1-p+pe^{it})^x$
$\mu_1 = np$
$\mu_2 = np(1-p)$
$\gamma_1 = \frac{1-2p}{\sqrt{np(1-p)}} \rightarrow^{x\to\infty} 0$
$\gamma_2 = 3 + \frac{1-6p(1-p)}{np(1-p)} \rightarrow^{x\to\infty} 3$

$n = 1000, p = \frac{1}{2}$
$p(x \leq 510) = p(x \leq 510.5)$
Continuous connotation:
$p(x < 510) = p(x \leq 509.5)$
$p(x = 510) = p(x \leq 510.5) - p(x \leq 509.5)$

Poisson distribution.

It means fish:

What happens for bionomial for large n

Suppose that we bound in np

Let p get small, and n increases, and have np constant

$p(x) = \frac{e^{-\lambda}\lambda^n}{n!}$ $x = 0, 1, 2, , \infty$
$\phi(t) = e^{x(it-1)}$
$\mu_1 = \lambda$
$\mu_2 = \lambda$ is not good since you need some flexibility
$\gamma_1 = \frac{1}{\sqrt{\lambda}}$ as $\lambda \to \infty$
$\lambda_2 = 3 + \frac{1}{\lambda}$ as $\lambda\infty$ and $\gamma_2 \to 0$

you can approximate poisson distribution with normal distribution.

Another way to define poisson is canonical definition:

Values occure or don't occured as 0, 1 in small intervals

assume probability of occurance is λdt in the small interval

Value occures or don't occured as 0, 1 in small intervals
$p_0(t) = (1 - \lambda dt)^{\frac{t}{dt}}$
We can control dt, and make it smaller
$p_0(t) = lim_{dt\to 0}(1 - \lambda dt)^{\frac{t}{dt}}$
$dt = \frac{t}{N}$
$\Rightarrow = lim_{N\to}(1 - \frac{\lambda t}{N})^N = e^{-\lambda t}$

$P_N(t + dt) = P_N(t)P_0(dt) + P_{N-1}P_1 dt = P_N(t)(1 - \lambda dt) + p_{N-1}(t)\lambda dt$
$\Rightarrow p_n(t + dt) = p(t) - \lambda p_n(t)dt + p_{n-c}(t)\lambda dt$
$p_n(t + \lambda t) - p(t) = \lambda p_n(t)dt + p_{n-1}(t)\lambda dt$
$\Rightarrow \frac{P_n(t+dt)-p(t)}{dt} = -\lambda p_n(t) + \lambda p_{n-1}(t) = -\lambda[p_{n-1}(t) - p_n(t)] = \frac{dp_n(t)}{t}$
$\Rightarrow$ Differential equation

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

① Cumulative function

$\ln(\phi(t/_i))$

$\frac{d^k}{dt^k}\ln(\phi(t/_i)) = K_k$

$K_1 = \mu$

$K_2 = \sigma^2$

$K_3 = \sigma^3\gamma_1$

$\gamma_2 = \frac{K_4}{\sigma^4} + 3$

Kurtosis

Karl Pearson:

$f(x)$ the normal density

$\frac{f'(x)}{f(x)} = \frac{(\mu - x)}{\sigma^2}$

Look at the differential Equation $\frac{f'(x)}{f(x)} = \frac{(m-x)}{a+bx+cx^2}$

$f(x) = C(a+bx+cx^2)^{-1/2c}$

$\exp\left[\frac{(b+2cm)\tan^{-1}\left[\frac{b+2cx}{\sqrt{4ac-b^2}}\right]}{c\sqrt{4ac-b^2}}\right]$

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

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

Gamma Distribution:

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$ $\alpha>0$

Gamma function

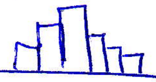
$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

if α is an integer k, then

$\Gamma(k) = (k-1)!$

$0! = 1$ is based on this rule

Sometimes used empirically



For low values gama & log-normal like eachother and as value increases the

$\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$

estimate

$\hat{\mu}_1 = \frac{\sum x_i}{n}$ $\hat{\mu}_2 = \frac{\sum(x_i-\bar{x})^2}{n}$ ..

② $f'(x) = \frac{d f(x)}{f(x)}$

$\frac{f'(x)}{f(x)} = \frac{(m-x)}{b_0+b_1x+b_2x^2}$

$m = -\frac{\mu_3(\mu_4+3\mu_2^2)}{A}$

$b_0 = -\frac{\mu_2(4\mu_2\mu_4 - 3\mu_3^2)}{A}$

$b_1 = m$

$b_2 = -\frac{(2\mu_2\mu_4 - 3\mu_3^2 - 6\mu_2^3)}{A}$

$A = 10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2$

- estimate in Equation gives estimate of $b_1, b_2, .., b_n$ & gives what distribution fits the data
- only ones in Pearson Family can be taken

Three parameter gamma

$f(x \mid k, \alpha, \lambda) = \frac{1}{\Gamma(\alpha)\lambda^\alpha} \times (x-a)^{\alpha-1} e^{-\left(\frac{x-a}{\lambda}\right)}$

$x > a$

a "start point of x"

λ: scale param.

α: Shape param

$E(x) = \mu_1 = \alpha\lambda$

$Var(x) = \mu_2 = \alpha\lambda^2$

$\gamma_1 = \frac{2}{\sqrt{\alpha}}$ Stand meas. skew

$\gamma_2 = 3 + 6/\alpha$ stan meas. kurtos.

- mode exist if $\alpha > 1$ occures at $\alpha - 1$

$\ln f(x) =$ Constant with respect to x $+$

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

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

$\frac{\partial}{\partial x} = \frac{(\alpha-1)}{x-a} - \frac{1}{\lambda} = 0$

③ $\frac{\partial^2}{\partial x^2} = -\frac{(\alpha-1)}{(x-a)^2}$

23nd Jan.

$a=0$
$\gamma=1$

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

$\gamma_1 \to 0$ $\alpha \to \infty \to$ Symmetry

$\gamma_2 \to 3$ $\alpha \to \infty \to$

$\Rightarrow$ Gama for large a can be estimated by normal

$\boxed{a=0, \lambda=2}$ $\alpha = \frac{df}{2}$

Gamma is χ^2 with df deg freedom

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

$\chi^2(df)$

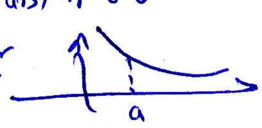
$\boxed{a=0 \quad \alpha=1} \Rightarrow$ Expon Dist

$f(x) = \frac{e^{-x/\lambda}}{\lambda}$ $f(x) = \theta \cdot e^{-\theta x}$

in software parametrization have difference

max anthr - on all distripution range $0-\infty$

measure of disorder



$-\int \ln f(x)\, f(x)\, dx$

has no memory

$\int_0^x f(z)dz = F(x) = 1 - e^{-x/\lambda}$

Assum $x > a$

$f(x \mid x>a) = \frac{f(x \text{ and } x>a)}{P(x \geq a)} =$

$\frac{\frac{1}{\lambda}e^{-x/\lambda}}{[1-(1-e^{-a/\lambda})]} = \frac{e^{-\left[\frac{x-a}{\lambda}\right]}}{\lambda}$ $x>a$

memoryless

- light bulbs are like that electron. Comp. have this dist.

onn Connection

x is Poisson with param λp,

then $p(x \geq m) = \sum_{k=m}^{\infty} \frac{e^{-\lambda p}\lambda p^k}{k!} =$

$\int_0^{\lambda p} \frac{e^{-z}z^{m-1}}{(m-1)!} = P\left(z < \lambda p \,\middle|\, \begin{matrix}\text{Gam} \\ a=0 \\ \alpha=m \\ \lambda=1\end{matrix}\right)$

m: int

unifying Concept

Beta Dist (also in Pearson family)

- in Bayes

④

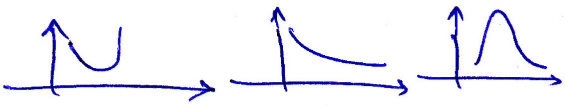
$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

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

---

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

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



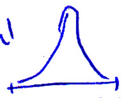
$E(x)=a+\frac{h\alpha}{\alpha+\beta}$

$Var(x)=\frac{h^2\alpha\beta}{(\alpha+\beta)^2}\cdot\frac{1}{(\alpha+\beta+1)}$

$\gamma_1=\frac{2(\beta-\alpha)}{(\alpha+\beta+2)}\sqrt{\frac{(\alpha+\beta+1)}{\alpha\beta}}$

$\gamma_2=\frac{6(\alpha-\beta)^2(\alpha+\beta+1)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}-\frac{6}{\alpha+\beta+3}$

it can become close to normal — but we are talking empirical



a ⟶ a+h

- Sym
- u-shape
- right/left skewed
- key: finite range

If x has an F distribution - two chi-sq div by each oth. & deg. freedom

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

ratio of chi-sq Related _ Beta dist

$\boxed{\text{oddity}}$ Let U binom rand varib param n,p

$P(U\ge m)=P_{beta}(x<p)$ $\quad \alpha=m$, $\beta=n-m+1$

- Connec Beta dist & Binom like wiered rel of Gama & Poisson
- there are dist. that don't fit

⑤ $\boxed{\text{Barr}}$ ~~Add~~

$F(x)=\int_{-\infty}^{x}f(z)dz$ → p.d.f

$dF=F(x)(1-F(x))g(x)dx$

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

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

$\Rightarrow \ln F(x)-\ln(1-F(x))=G(x)$

$\frac{dF}{F(x)(1-F(x))}=g(x)dx$

$\frac{F}{1-F(x)}=e^{G(x)}$

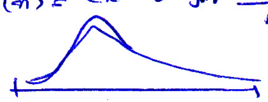
$F(x)=\frac{e^{G(x)}}{1+e^{G(x)}}=[1+e^{-G(x)}]^{-1}$

Uniform Dist:

$F(x)=(1-[1+x^{c}]^{-k})$

Household income in U.S.

$F(x)=\left(x-\sin\frac{2\pi x}{2\pi}\right)^{\alpha}$



close to bell shape

Johnson Family

let z be normal

$z=\gamma+\delta\ln x$ [log normal dist]

Normal z.

$z_B=\gamma+\delta\ln\left(\frac{x}{1-x}\right)$

Normal $\quad$ Johnson B Family curve

$z_u=\gamma+\delta\sinh^{-1}x$ $\boxed{\text{U Form}}$

$\sinh(x)=\frac{e^{x}-e^{-x}}{2}$

$\cosh(x)=\frac{e^{x}+e^{-x}}{2}$

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

p.df $\boxed{\text{B-form}}$

$f_B(x)=\frac{\delta}{\sqrt{2\pi}}\frac{1}{x(1-x)}e^{-\frac{(\gamma+\delta\log\frac{x}{1-x})^2}{2}}$

$0<x<1$

$-\infty<x<\infty$ $\quad f_u(x)=\frac{\delta}{\sqrt{2\pi}}\frac{1}{\sqrt{x^2+1}}e^{-\frac{(\gamma+\delta\ln(x+\sqrt{x^2+1}))^2}{2}}$

$-\infty<x<\infty$

---

x is log Normal

$\ln x$ is normal (μ,σ^2)

⑥

Let $g(x)$ be a monotone invertible continuous func g of x $\quad y=g(x)$ $\quad x=g^{-1}(y)$

Cont: no jumps

invert: one to one

if g is ↑ + x increasing

$F_Y(y)=F_X(g^{-1}(y))$

if g is ↓ decreasing $F_Y(y)=1-F_X(g^{-1}(y))$

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

Rand Variable x: $f(x)$

$y=e^{x}\Rightarrow x=\ln y$

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

$=\frac{1}{y}f_X(\ln y)$

$f_Y(y)=\frac{1}{y}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\ln y-\mu}{\sigma}\right)^2}$

$0<x<\infty$

p.d.f of Log normal distribution

---

x be random variable with c.d.f: $F(x)$

$y=F(x)$

what is the dist. of y?

$F_Y(y)=P(Y\le y)$ $\quad 0\le y\le 1$ dist func

$=P_X(F_X(x)\le y)=$

$P_X(x\le F^{-1}(y))=F_X(F_X^{-1}(y))=$

y: uniform on rang 0 & 1

Central limit theorem - mean

You can generate Random number

$F^{-1}(y)=x$

$F^{-1}(u_i)=x_i$

basis of simulation

7 23rd Jan

$f(f^{-1}(x)) = x = f^{-1}(f(x))$

$\frac{d f^{-1}(f(x))}{dx} = \frac{dx}{dx} = 1 \Rightarrow \frac{d f^{-1}(f(x))}{d f(x)}$

* $\frac{d f(x)}{dx} = 1$

$y = f(x) \Rightarrow \frac{d f^{-1}(y)}{dy} \cdot \frac{d f(x)}{dx} = 1$

$\frac{d f^{-1}(y)}{dy} = \frac{1}{\frac{d f(x)}{dx}} \Big|_{x = f^{-1}(y)}$

x : CDF $F(x)$

$y = F(x)$, $x = F^{-1}(y)$, p.d.f (y)

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

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

all simulation in Bayesian is based on this result

Richards Functions

$\frac{f(x)}{F(x)} = -\alpha \frac{(F(x)^{\alpha} - 1)}{\alpha} - \beta$

$F(x) = [1 + \gamma \lambda e^{-\alpha x}]^{-1/\lambda}$ $\quad \gamma > 0, \alpha > 0, \lambda > 0$

$\lim_{\lambda \to 0} F(x) = e^{-\gamma e^{-\alpha x}}$

Gompertz

$\lambda = -1$ exponential dist

$\lambda = 1$ Logistic Distribution

Extreme value distribution

- CLT: independ ident. dist. no matter what you start
- what flood rain look like
- Record of how fast people run
- if cont no matter where start, smooth relatively, no matter dist., max will go

① $f_1(x) = e^{-x^{-\alpha}}$ $\quad x > 0, \alpha > 0$

② $f_2(x) = \begin{cases} e^{-(-x)^{\alpha}} & x < 0 \quad \alpha > 0 \\ & \text{if } x > 0 \end{cases}$

③ $f_3(x) = e^{-e^{-x}}$ $\quad -\infty < x < \infty$

8

Goal is to get an approximation to "Real" data

Finance: Normal dist.

Data wasn't normal

arsenal - set of func. to find nice simple solution

Exponential Family

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

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

Gamma $f(y) = \beta^{\alpha} y^{\alpha-1} e^{-\beta y}$

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

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

Weibull Distrib

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

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

$\mu = \alpha + \gamma\Gamma(1 + 1/k)$

$\sigma^2 = \gamma^2\Gamma(1 + 2/k) - \mu^2$

$\gamma_1 = \frac{\Gamma(1+3/k)\gamma^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3}$

$\gamma_2 = \frac{\gamma^4\Gamma(1+4/k) - 3\sigma^4 - 4\Gamma(1+1/k)\sigma^2\mu - 6\mu^2\sigma^2 - \mu^4}{\sigma^4}$

- Quality Contrl
- Reliab Stud
- life time

$f(x \mid x \geq x_0) = \frac{f(x \text{ and } x \geq x_0)}{p(x \geq x_0)} =$

$f(x)$

9

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

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

- hazard function power at the rate
- like generalization of exponential curve

Logistic Distribution:

$F(x) = \frac{1}{1 + e^{-\frac{(x-\mu)}{\sigma}}}$ $\quad -\infty < x < \infty$ symmetric

$= \frac{e^{\frac{(x-\mu)}{\sigma}}}{1 + e^{\frac{(x-\mu)}{\sigma}}}$

Read documentation of package before work

$f(x) = \frac{e^{-\frac{(x-\mu)}{\sigma}}}{\sigma\left[1 + e^{-\frac{(x-\mu)}{\sigma}}\right]^2}$

$E(x) = \mu$

$V(x) = \frac{\pi^2\sigma^2}{3}$

$\gamma_1 = 0$

$\gamma_2 = 4.2$ Fatter tail than normal although Bell Shape

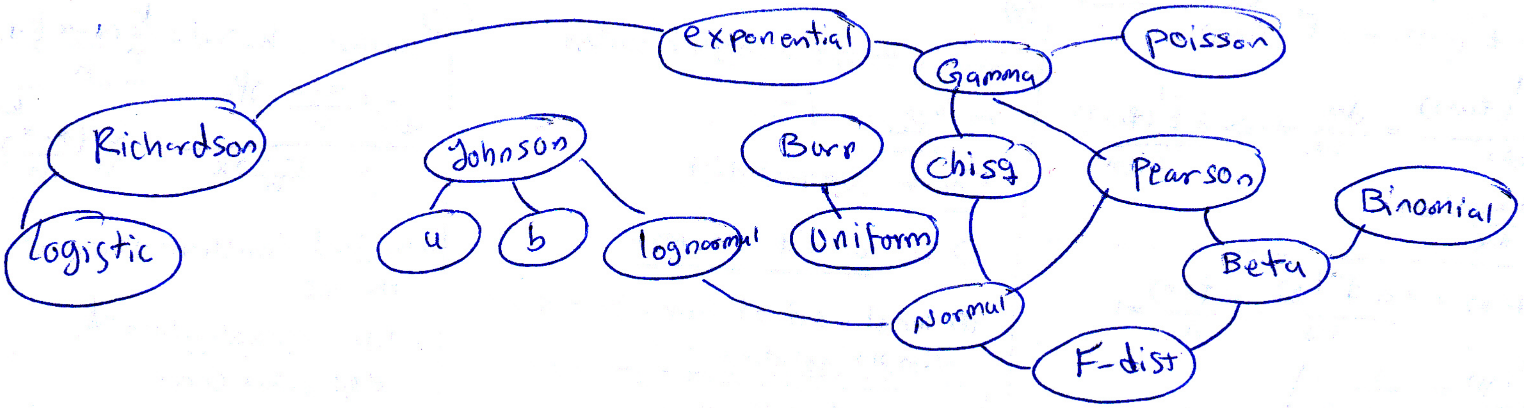
Regression

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$

Continuous Continuous

dichotomous → whether buy or not

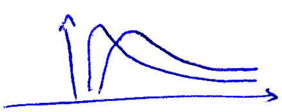
$p(x) = \frac{e^{\beta_0 + \sum_i \beta_i x_i}}{1 + e^{\beta_0 + \sum_i \beta_i x_i}}$


exponential
Gamma
poisson
Richardson
Johnson
Burr
chisq
Pearson
Binomial
Logistic
u
b
lognormal
Uniform
Beta
Normal
F-dist

30 Jan

Compare log normal and Gamma same mean

$\mu = 7$ SD $= 1$

 ... 

You should put all the steps - plot over to compare

always simplify

- skewness of normal is zero - when you standardize you would be able to compare better

- revist characteristic function

- Combine variables - what thousands of people bought
- Recall that characteristic function $\Phi_f(t) = E_f(e^{itx}) =$

$$\int_{R_x} e^{itx} f(x)\,dx = E_f(\cos tx) + i E_f(\sin tx)$$

if $\int_{-\infty}^{\infty} |\Phi_f(t)|\,dt < \infty$

then the characteristic func. uniquly determines the dist. of x.

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\,\Phi_f(t)\,dt$$

Also Recall $y = \alpha + \beta x$

$$\Phi_y(t) = e^{i\alpha t}\,\Phi_x(\beta t)$$

look $\sum_{i=1}^{n} x_i$

If $x_1, x_2, \dots, x_n$ are indpndnt r.v with

C.F (Char. func) $\Phi_1(t), \Phi_2(t), \dots, \Phi_n(t)$

then C.F. $\sum_{i=1}^{n} x_i$ is $\prod_{i=1}^{n} \Phi_i(t)$

Char.f. of sum $\longrightarrow$ mult of char. func

if $x_1, \dots, x_n$ are i.i.d (exact same dist). r.v's

$$\Phi_y(t) = (\Phi_x(t))^n$$

Past Curves

$N(0,1)$ $\Phi(t) = e^{-t^2/2}$

$N(\mu, \sigma^2)$ $\Phi(t) = e^{[it\mu - \frac{\sigma^2 t^2}{2}]}$

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

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

χ^2_p d.f. (deg fr) $\Phi(t) = (1-2it)^{-p/2}$

$y = \sum_{i=1}^{n} x_i$ $\Phi_y(t) = (1-2it)^{-np/2}$

$\Rightarrow$ y is χ^2_{np}

Binomial dist. x_i is $B(n,p)$ $\Phi(t) = (1-p+pe^{it})^n$

x_i is $B(n_i, p)$

p const

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

Poisson $\Phi(t) = e^{\lambda(e^{it}-1)}$

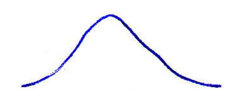
$y = \sum_i x_i$ $\Phi_y(t) = e^{n\lambda(e^{it}-1)} \Rightarrow$ poiss$(n\lambda)$

Gamma $\Phi(t) = \frac{1}{(1-it)^{\alpha}}$ x_i is $\Gamma(\alpha)$

$y = \sum_i x_i \Rightarrow \Phi_y(t) = \frac{1}{(1-it)^{n\alpha}}$

$\Rightarrow \sum_i x_i$ is $\Gamma(n\alpha)$

Chauchy Distribution

$f(x) = \frac{1}{\pi(1+x^2)}$ $-\infty < x < \infty$ 

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

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

$\Rightarrow$ doesn't converge mean Cauchy does not have mean

$\lim_{N\to\infty} \ln(1+x^2)\Big|_{-N}^{N} = 0$ you can write in diffrnt ways and that does not converge

it has median and mode $= 0$, but no momet

$\Phi_{Cauchy}(t) = e^{-|t|}$

x is Cauchy $f(x) = \frac{1}{\pi(1+x^2)}$

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

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

$\frac{1}{\pi(1+y^2)}$ $\Rightarrow$ the same.

The point is heavy tail

any power of x is overwhelmed with tails desirable for Economic — it is invariant to whether look at x or $\frac{1}{x}$.

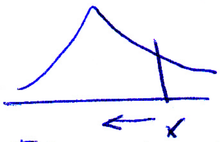
$\boxed{\text{Three parameter cauchy}}$

$f(x) = \frac{1}{\gamma\pi} \times \frac{1}{[1+(\frac{y-\alpha}{\gamma})^2]}$ α is like location Parameter.

it does thing mean does yet it is not

γ is scale parameter

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



$F^{-1}(p) = \alpha + \gamma\tan[\pi(p-\frac{1}{2})]$ $\Phi(t) = e^{i\alpha t - \gamma|t|}$

$x_1, x_2, \dots, x_n$ are iid Cauchy with (α, γ)

$\sum_i y_i$ $\Phi_y(t) = e^{i\alpha n t - n\gamma t}$

$\sum_i y_i$ is Cauchy $(n\alpha, n\gamma)$

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

$n\bar{y}$

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

$d\bar{y} = \frac{d(\sum_i y_i)}{n}$

$f(\bar{y}) = \frac{1}{\pi\gamma}\frac{1}{[1+(\frac{\bar{y}-\alpha}{\gamma})^2]}$ → which is same as what we start with

wild wiered ←

⇒ No Central theorem for Cauchy dist.

No moment.   Sum exact dist of Component

→ witch of Agnesi   More adding does not reduce variability

---

most dist is self perpetuate excep Beta since range changes, yet on all others work

---

$X_1, X_2, \dots, X_n$ indep r.v

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

$\Phi_j(t) = e^{it\mu_j - \frac{\sigma_j^2 t^2}{2}}$

$y = \sum x_i$ $\Phi_y(t) = e^{it\sum_j \mu_j - \sum_j \sigma_j^2 t^2/2}$

$\Rightarrow y \sim N(\sum_j \mu_j, \sum_j \sigma_j^2)$

---

x_j is $\chi^2(p_j) \Rightarrow \Phi_{x_j} = (1-2it)^{-p_j/2}$

$y = \sum_j x_j$ $\Phi_y(t) = (1-2it)^{-\sum_j p_j/2}$

$\Rightarrow y$ is $\chi^2(\sum_i p_i)$

still stay in chisq family

---

x_j is poisson $\Phi_{x_j}(t) = e^{\lambda_i(e^{it}-1)}$

$y = \sum_i x_i \Rightarrow \Phi_y(t) = e^{\sum\lambda_i(e^{it}-1)}$

$y \sim poiss(\sum_j \lambda_j)$

Binomial X_j is $B(n_j, p_j)$

$\Phi_{x_j}(t) = (1-p_j + p_j e^{it})^{n_j}$

$y = \sum_i x_i \Rightarrow \Phi_y(t) = \prod(1-p_j + p_j e^{it})^{n_j}$

≠ Binomial

not Binomial But very close

Gama x_j is $\Gamma(\alpha_j)$

$\Phi_{x_j}(t) = \frac{1}{(1-it)^{\alpha_j}}$ $y = \sum_j x_j$

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

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

$y = \sum_i x_i$ $\Phi_y(t) = e^{(\sum_j \alpha_j)it - (\sum_j \gamma_j)|t|}$

$C(\sum_j \alpha_j, \sum_j \gamma_j)$

The only dist that did not stay in family when adding up is Binomial

---

$\sum_j a_j x_j$ Cauchy is identical to t-dist with one degree of freedom

→ Dow Jones Industrial average
ai as weights

$\sum_i x_i$ iid

$\sum_j x_j$ indep ⇒ now look at $\sum_j a_j x_j$
x_j are indep r.v

---

x_j is $N(\mu_j, \sigma_j^2)$

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

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

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

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

Jan 30

(5)

χ^2 $\quad x_j$ is $\chi^2(p_j)$

c.f $(1-a_j it)^{-p_j t}$

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

$y = \sum_j a_j x_j = \prod_{i=1}^{n} (1-2a_j t i)^{-p_j/2} \neq \chi^2$ ⊗

sum of chi-sq is not theoritically χ^2

$H_0: \mu_1 = \mu_2$

$H_A: \mu_1 \neq \mu_2$

$x_1, \dots, x_{n_1}$

$x_2, \dots, x_{n_2}$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ is not t ⊗

Assume: $\sigma_1^2 = \sigma_2^2$ $\quad$ Pooled t-test with

saj the bate

Aithens

Hazarbazar

$\Rightarrow$ approximate by t although it is not t dist

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

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

$y = \sum_j a_j x_j$ $\quad \prod_{i=1}^{n} \frac{1}{(1-ia_j t)^{\alpha_j}} \neq \Gamma$ ⊗

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

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

$y = \sum_j a_j x_j$ $\quad e^{i(\sum_j a_j \alpha_j) t - (\sum_j a_j \gamma_j)|t|}$

|Cauchy| like Normal

Class of Distributions:

Class of Stable Distributions

r.v. y_1, y_2

$y_1 = \alpha_1 + \gamma_1 x_1$ $\quad e^{i\alpha_1 t} \Phi_{x_1}(\gamma_1 t)$

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

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

if there exist values c and d such that

$e^{i(\alpha_1+\alpha_2)t} \Phi_x(\gamma_1 t) \Phi_x(\gamma_2 t) = e^{idt} \Phi(ct)$

Then x is said to follow Stable distribution

There are an infinit num. of stable dist. of which only three have easy to write prob dens func:

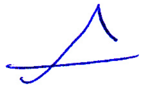
1) Normal: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $\quad -\infty < x < \infty$

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

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

Index		
	$\alpha = 2$	Normal
	$\alpha = 0$	Cauchy
	$\alpha = -1/2$	levy dist
	0	

Normal is one end, and every in the group will converge to Cauchy



$\Phi_N(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$ $\quad$ Normal

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

$\Phi_L(t) = e^{i\alpha t - \sqrt{-2i\alpha t}}$ $\quad$ levy dist

Normal only one has approp. moment

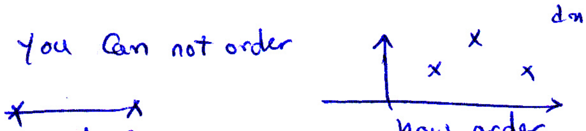
Multivariat: $\quad \underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$

$f(\underline{x})$

$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_p} f(x)\, d\underline{x}$

$d\underline{x} = dx_1 dx_2 \cdots$

You can not order



order $\quad$ how order

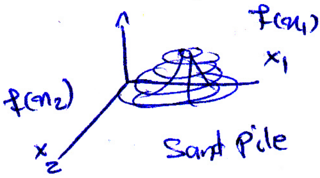
$\Rightarrow$ tendency to get index

$f(x_1, x_2)$

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

Marginal dist

Conditional dist $\quad x_2 | x_1$



Cutting sandpile is this dist

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

MVN : multivariate normal

start with $x_1, x_2, \cdots, x_p$ x_j is $N(\xi_j, \tau_j^2)$ indep

sie tau

p.d.f $(x) = \prod_{j=1}^{m} f_j(x_j) = \frac{1}{(2\pi)^{p/2} \prod_{j=1} \tau_j} e^{-\frac{1}{2}\sum_i \left(\frac{x_j - \xi_j}{\tau_j}\right)^2}$

$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_t \end{pmatrix}$ $\underset{p\times p}{D_{\tau^2}} = \begin{pmatrix} \tau_1^2 & & 0 \\ & \tau_2^2 & \\ 0 & & \ddots & \tau_p^2 \end{pmatrix}$ $\tau_j^2 > 0$

Diagonal matrix

$D_1 = I = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$ $D^{-1}_{\tau^2} = D^{-1}_{\tau} \cdot D_{\tau^2} = D_{\tau^2} D^{-1}_{\tau^2} = I$

$D^{-1}_{\tau^2} = \begin{pmatrix} 1/\tau_1^2 & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\tau_p^2} \end{pmatrix}$

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

$\sqrt{D_\lambda}\sqrt{D_\lambda} = D_\lambda$

$\Rightarrow$ p.df. $(x) = \frac{1}{(2\pi)^{p/2} |D_{\tau^2}|^{1/2}} e^{-\frac{1}{2}(\underline{x} - \xi)' D^{-1}_{\tau^2} (\underline{x} - \xi)}$

Quadratic Form

$x' = (x_1, x_2, \cdots, x_p)$ $\underline{x}' A \underline{x} = \sum_i \sum_j x_i a_{ij} x_j$

$A \begin{pmatrix} a_{11} & a_{12} \cdots a_{1p} \\ \vdots & \ddots \\ a_{p1} & a_{p2} \cdots a_{pp} \end{pmatrix}$

p.d.f $(x) = \frac{1}{(2\pi)^{p/2} |D_{\tau^2}|^{1/2}} e^{-\frac{1}{2}(\underline{x} - \xi)' D^{-1}_{\tau^2} (\underline{x} - \xi)}$

still all uncorrelated

$y_i = \sum_{j=1}^{p} \gamma_{ij} x_j$ vector vector

$i = 1, 2, \cdots,$

new matrix Γ

Special Γ that $\Gamma\Gamma' = \Gamma'\Gamma = I$ orthogonal matrices

- they Rotate things & mirror images, two three dimension

$\underline{y} = \Gamma \underline{x} \Rightarrow \underline{x} = \Gamma' \underline{y}$

$E(\underline{y}) = \Gamma E(\underline{x}) = \Gamma \underline{\xi} = \underline{\mu}$

$Cov(\underline{y}) = \Gamma D_{\tau^2} \Gamma' = \Sigma$ $\Sigma^{-1} = \Gamma D_{1/\tau^2} \Gamma'$

$\Sigma^{-1}\Sigma = \Gamma D_{1/\tau^2} \underbrace{\underbrace{\underbrace{\Gamma'\Gamma}_{I} D_{\tau^2}}_{I} \Gamma'}_{I}$

inverse maker

$\xi = \Gamma' \mu$ $\xi = \Gamma' \mu$

$f(\underline{y}) = \frac{1}{(2\pi)^{p/2}} \frac{1}{|\Gamma D_{\tau^2} \Gamma'|^{1/2}} e^{-\frac{1}{2}(\Gamma'\underline{y} - \Gamma'\underline{\mu})' D^{-1}_{\tau^2} (\Gamma'\underline{y} - \Gamma'\underline{\mu})}$

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

if Determinant, you would chang orthog Σ

$\frac{dx}{dy} = ||\Gamma'|| = 1$

$f(\underline{y}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\underline{y} - \underline{\mu})' \Sigma^{-1} (\underline{y} - \underline{\mu})}$

factor analy: factors Corr & extract fact. uncorr.

if y is $MVN(\mu, \Sigma)$

there exist p Var $x_1, x_2, \cdots, x_p$

which are uncorrelated

there is always factor.

Question is whether it makes sense.

full rank: Γ factors - if ~~least~~ fewer facto

Rotate in space, since it does not change anything verimax

There is always something there

property of multivar:

$Cov(y_i, y_j) = \sigma_{ij} = \rho_{ij} \sigma_i \sigma_j$

$Cov(\underset{k\times p}{A} \underset{p\times 1}{\underline{y}}, \underset{k\times}{B} \underline{y}) = A \Sigma B'$

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

Dist of $\underset{k\times p}{A} \underset{p\times 1}{\underline{y}}$ is $MVN(A\mu, A\Sigma A')$

Multinomial Dist

$x_1, x_2, \cdots, x_k$ CAT_1 CAT_2 .. CAT_k

p_1 p_2 p_k $\sum_{i=1}^{n} p_i = 1$

R.S. obsrv $n_1, n_2 \cdots n_k$

$\sum_i x_i = n$

$$p(x_1,\dots,x_k) = \frac{n!}{\prod_{i=1}^{k} x_i!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

$E(x_i) = np_i$

$Var(x_i) = np_i(1-p_i)$

$Cov(x_i, x_j) = -np_ip_j$

Drichlet Dist: Generaliz. of Beta dist

obsrv $x_1, x_2, \dots, x_n$ $\quad x_i \geq 0$ $\quad \sum_{i=1}^{n} x_i = 1$

$$p(x_1, x_2, \dots, x_k) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} x_i^{\alpha_i - 1}$$

Bayesian - Binomial - prior $\quad i = 1, 2, \dots n$

Drichlet is used

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

$$\alpha_+ = \sum_{i=1}^{n} \alpha_i$$

$$Cov(x_i, x_j) = \frac{-\alpha_i \alpha_j}{(\alpha_+)^2(\alpha_+ + 1)}$$

$$x_{md} = \left\{ \frac{\alpha_i - 1}{\alpha_+ - k} \right\} \quad \alpha_i > 1$$

mode

If $Y_1, Y_2, \dots, Y_k$ are $\Gamma(\alpha_i, 1)$

then $\frac{Y_1}{\sum_i Y_i}, \frac{Y_2}{\sum_i Y_2}, \dots, \frac{Y_i}{\sum Y_i}$ is Drichelet

---

- Next time Convergence & Probabilities
- on Defense you only need to memorize everything

statistics | Feb06

Convergence of random variable pdf $F_1, F_2, \dots, F_n$

Let $S_1, S_2, \dots, S_n$ be sequence of R.V.s

S_i be the average of i variable gamma distribution

Let S be a r.v. with cdf =

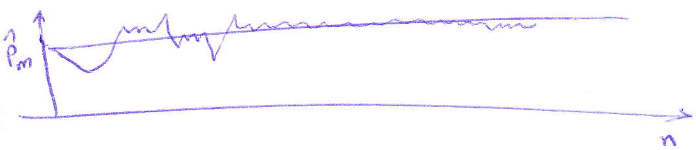
one says, that

S_n Converges in probability to S

$(S_n \xrightarrow{P} S)$

if for every $\varepsilon > 0$ $P(|S_n - S| > \varepsilon) \xrightarrow{0} 0$ as $n \to \infty$

Think of binomial $\hat{p}_m = x/m$



p as random variable - deviation will eventually Converge

Stronger form "Measure Theory"

Point is zero dimension

infinite sequence - Newtonian

Almost sure Convergance

all possible sequences $S_1, S_2, \dots, S_n$

if $S_m \longrightarrow S$ almost surely if

$P(\{S_n\}$ don't Converge$) = 0$

- Consider set of infinite sequences, and all will Converge

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

if that is true it implies in Converges to Probability

$\boxed{WLLN}$ let $x_1, x_2, \dots, x_n$ be i.i.d. r.v's so that $E(x_i)$ exist

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

If δ is a Constant, and $S_n \xrightarrow{P} \delta$

and $h(x)$ is a Continuous func

$h(S_n) \xrightarrow{P} h(S)$

I_j = r.v $I_j = 1$ pr π

$I_j = 0$ $1 - \pi$

$E(I_j) = \pi$ $V(I_j) = \pi(1-\pi)$

$\hat{\pi}_n \longrightarrow \sum_{j=1}^{n} I_j/n$

$\hat{\pi}_m \xrightarrow{P} \pi$

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

$\widehat{\frac{\pi(1-\pi)}{m}}\Big|_m = \frac{\hat{\pi}_m(1-\hat{\pi}_m)}{n}$ since $(1-x)x$ is Continuous

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

If S_n is an estimate of a parameter θ, it is said to be "Consistent", if $S_n \xrightarrow{P} \theta$ as $n \to \infty$ $\forall \theta$

no to saying unbiased

unbiased $E(S_n | \theta)$

Random Sample $x_1, x_2, \dots, x_n$ $E(x_i) = \mu$
$Var(x_i) = \sigma^2$

we want to estimate μ^2

$\bar{x}$

$\hat{\mu}^2 = \bar{x}^2$

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

$E(\bar{x}^2) = Var(\bar{x}) + E^2(\bar{x})$

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

it is biased

if big enough sample if n increases would be zero

Asymptotically unbiased

$S_1, S_2, \dots, S_n$ Converge in distribution to S

$S_n \xrightarrow{D} S$ or $S_n \longrightarrow S$

$\Rightarrow P(S_n \le t) \longrightarrow P(S \le t)$ as $n \xrightarrow{\text{goes}} \infty$

for every t where $P(S \le t)$ is Continuous

= nice smooth function

distribution is Converging

If $E(|S_n - E(S_n)|^2) < \infty$ then the speed of Convergence is $\frac{1}{\sqrt{n}}$

CLT (Central limit Theorem)

Limit theorem - asymptote

it is central, since it you haven't it statistics would be different, and you had to take sample on and on.

Let $x_1, x_2, \dots, x_n$ be i.i.d with mean μ and finite variance σ^2 (not work for Cauchy) then

$\bar{x} = \sum_{i=1}^{n} x_i/n$

$\bar{x} \xrightarrow{D} N(\mu, \sigma^2/n)$

and if $E(|x_i - \mu|)^2 < \infty$ then the convergence has a speed of $\frac{1}{\sqrt{n}}$, and it is uniform. ($f_{\bar{x}}(x) \longrightarrow N(x)$ not ... on x)

Heuristic proof of CLT: recall that if $\Phi_x(x)$ is c.f. of x, and $x_1, x_2, \dots, x_n$ be i.i.d $\Rightarrow$

c.f of $\sum x_i$ is $[\Phi_x(t)]^n$

recall that if cf has $\Phi_x(t)$

$y = \alpha + \beta x$

Then y has $e^{\alpha} \Phi_x(\beta t)$

Let x be $\Rightarrow$ $E(x) = 0$ $\quad Var(x) = 1$

$\Phi_x(t) = \sum_{j=1}^{\infty} \frac{\alpha_j (t)}{j!\, i^j} = 1 + \frac{0 \cdot t}{1 i} + \frac{1 t^2}{2 i^2} + o(t^2)$

$i = \sqrt{-1}$

Notation: $f(t)$ is $o(g(t))$ if $\lim_{t \to 0} \frac{f(t)}{g(t)} = 0$

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

$\Phi_x(t) = [1 - t^2/2 + o(t^2)]$

look at $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}}{\sigma} \sum_{i=1}^{n} \frac{1}{n}(x_i - \mu) = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \left(\frac{x_i - \mu}{\sigma}\right)$

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

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

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

$t/\sqrt{n} \to t$

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

$\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)$

$\lim_{n \to \infty} [1 - t^2/2n + o(t^2/n)]^n$

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

Convergence in mean: look at $|S_n - S|^r$

$\lim_{n \to \infty} E(|S_n - S|^r) = 0$

$S_n \xrightarrow{L^r} S$ $\quad r \geq 1$

if $s < r$ then $S_n \xrightarrow{L^r} S \Rightarrow S_n \xrightarrow{L^s} S$

$S_n \xrightarrow{a.s} S$ $\quad S_n \xrightarrow{P} S$

$\rightarrow S_n \xrightarrow{D} S$

The practical is: People want to know how convergence in probability is.

it is the easiest way

what if $E(x_i) = \mu_i$ and $V(x_i) = \sigma_i^2$

$x_1, x_2, \dots, x_n$ independent

Let $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$

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

$\lim_{n \to} \frac{r_n}{s_n} \longrightarrow 0$ then $\sum_i x_i \xrightarrow{L} X \sim N(\sum_i \mu_i, \sum_i \sigma_i^2)$

Let's assume that $E(x_i) = \mu_i = \pi_i$

$\quad B(1, \pi_i)$

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

$x_i = \begin{cases} 1 & \pi_i \\ 0 & 1 - \pi_i \end{cases}$

$\sum_i x_i$

$E(|x_i - \mu_i|^2) = (1 - \pi_i)^3 \pi_i + |(0 - \pi_i)^3|(1 - \pi_i)$

$= (1 - \pi_i)^3 \cdot \pi_i + \pi_i^3 (1 - \pi_i) =$

$\underbrace{\pi_i(1 - \pi_i)}_{x}\left((1 - \pi_i)^2 + \pi_i^2\right) \leq \pi_i(1 - \pi_i)$

$0 < x < 1$

$x^2 + (1 - x)^2$

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

$\leq \left[\sum_i \pi_i(1 - \pi_i)\right]^{1/3}$

$s_n = \sqrt{\sum_i \pi_i(1 - \pi_i)}$

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

$\frac{1}{\left[\sum_i \pi_i(1 - \pi_i)\right]^{1/6}} \longrightarrow 0$

statistics | Feb06

5

application - Take sample $x_1, \dots, x_n$

$I_j(y) = \begin{cases} 1 & \text{if } x_j \le y \\ 0 & \text{otherwise} \end{cases}$

Sample c.f.

$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^{m} I_i(y)$ $\quad \pi = F(y)$

$I_j(y)$ is Bernouli $\pi = F(y)$

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

Generalized CLT

let $x_1, x_2, \dots, x_n$ be i.i.d. r.v. then $\exists$ a sequence a_n, b_n and a non-degenerate r.v. Z, such that

$a_n(\sum_i x_i) - b_n \xrightarrow{D} Z \Leftrightarrow Z$ is α-stable $0 < \alpha \le 2$

Bond price, and stock will converge to something

if you take Cauchi and add them up, it converges to something.

it will not blow up

Characteristic func. of stable dist

$\Phi_Z = e^{it\mu - |ct|^{\alpha}(1 - i\beta \, \mathrm{sgn}(t) \Phi)}$

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

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

Scale $= c$
Shape $= \alpha$
Shift $= \mu$
Skew $= \beta$

has tails of magnitude $\frac{1}{|x|^{\alpha+1}}$

Suppose you have a statistics, x_n

$\frac{(x_n - \mu)}{\sqrt{Var(x_n)}} \longrightarrow Z$ and $n Var(x_n) \xrightarrow{P} \tau^2 > 0$

Let $h(x)$ be a function that has is continous and at zero derivative ($h'(\mu)$) is not zero, then $h(x_n) \xrightarrow{D} N(h(\mu), Var(x_n)[h'(\mu)]^2)$

$\bar{y} \sim N(\mu, \sigma^2/n)$ $\quad h(\mu) = e^{\mu}$

$h'(\mu) = e^{\mu} > 0$

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

Log Normal

$e^{\mu + \sigma^2/2n}$ $\quad Var(e^{\sigma^2/n} - 1)e^{2\mu + \sigma^2/n}$

6

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

$\underset{p\times 1}{\bar{X}}$ is $MVN(\underset{p\times 1}{\mu}, \underset{p\times p}{\Sigma})$ $\quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$

Let $h(x)$ be a continous function of x, such that $\left.\frac{\partial h(x)}{\partial x_i}\right|_{\mu} \ne 0$ $\quad i = 1, 2, \dots, p$

then $h(x) \xrightarrow{D} N(h(\mu), \underset{1\times p}{[h'(\mu)]} \underset{1\times p}{\Sigma} \underset{p\times 1}{h(\mu)})$

$\underset{p\times 1}{h'(\mu)} = [\partial h(\mu)/\partial \mu_i]$

Propagation of Error

$\binom{x}{y}$ $\quad E(x) = \mu_1$ $\quad E(y) = \mu_2$

$V(x) = \sigma_1^2$ $\quad Var(y) = \sigma_2^2$

$Corr(x, y) = \rho$

μ_2/μ_1 $\quad h(\mu) = \mu_2/\mu_1$

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

$\bar{x}, \bar{y}$ $\quad \widehat{\frac{\mu_2}{\mu_1}} = \frac{\hat{\mu}_1}{\hat{\mu}_2} = \bar{y}/\bar{x}$ $\quad Cov\binom{x}{y} = \frac{1}{n}\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

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

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

$\widehat{Var}(\bar{y}/\bar{x}) = \frac{1}{n\bar{x}^2}[(\frac{\bar{y}}{\bar{x}})^2 s_1^2 - 2\hat{\rho} s_1 s_2 (\frac{\bar{y}}{\bar{x}}) + s_2^2]$

Randomization and simulation

Keep it simple stupid (kiss) principle

The only way to solve is through simulation

1. standerdize
2. mixing

- Computer does what you tell them to do, and not what you want to
- Computer generates Psuedo random every observation and pair has same probability
- proportion that are non patten are very big

Psuedo random numbers generate two ways

- Congruential Random number generator

$\begin{array}{r} 34.12 \\ \times\ 54.86 \\ \hline 1871.8232 \end{array}$

you through two last decimal digit out, since there is uncertainty in measure

$X_{n+1} = aX_n + b$ and m

you want to generate long stream of Random numbers

$m=90$ $a=5$ $b=7$

seed a random number: milisecond on Computer

$= mod(a \times b, 90)$

after some time the number Repeats

- $9821x + 218327$ → Random number . Excel up to

- There are tests for random Variables: how often Repeated, how often digits repeated

mathematica "mersenne Twister" generates super long series, with good statistical properties

- this - white noise - mixture of all different colors Convert to numbers.

→ make sure you use this method to make sure your random numbers are random indeed.

$x(n+1) = 1650\ x_n\ \ln(2^{31}-1)$

$F(y) \sim U(0,1)$

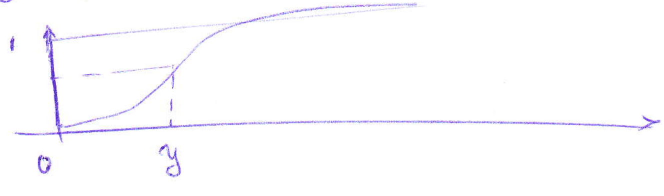
Continuous Curve $F(y)$

generate r.n u $u \sim (0,1)$ and evaluate $F^{-1}(u)$

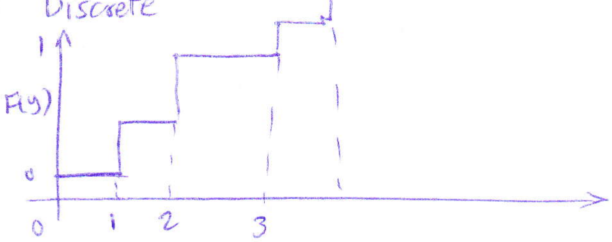
① random sample of normal in Excel =normsinv(Rand(A1))

what about discrete?

② binomdist (a, b, c, False)



Discrete



=lookup (Rand ( ))

(*) without replacement taking random sample from class you need to round up

(*) ordering random numbers that is generated for population is also another way to generate random number

---

P-P plot

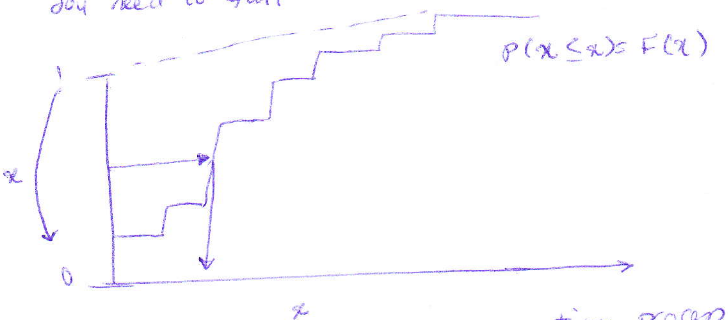
Statistics

Feb 13

Simulation: poisson with mu=8

=poisson(x, mu, true); in Excel

(*) when you simulate from discrete, after sometime you need to quit.



(*) pseudo random number generation process

generate random number given series of x and then do the lookup.

(*) it is very easy when you have invertible

$x \sim U(0,1)$ $\quad x \sim U(0,1)$

$x = F^{-1}(x)$

F( ) which is probability function

e.g =Betainv(probability, alpha, Beta)

=CritBinom for inverse binomial

→ Critbinom function

- FDist
- FINV
- normDist
- normInv
- weibull

Exponential Distribution $\quad u \sim Unif(0,1)$

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

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

$x = -\lambda \ln(1-x)$

Cauchy distribution

$F(x) = \frac{r}{\pi(r^2 + (x-\alpha)^2)}$

$F(x) = \frac{1}{\pi}\tan^{-1}\left(\frac{x-\alpha}{r}\right) + \frac{1}{2} = u$

$\Rightarrow X = r \cdot \tan[\pi(u - 1/2)] + \alpha$

Weibull $F(x) = 1 - e^{-\left(\frac{x-\alpha}{r}\right)^k}$

$x = \alpha + r(-\ln u)^{1/k}$

easy to do if you have cumulative dist func.

$\Phi(x) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x} dx$ $\quad$ Close form does not exist

Box-Muller

Generates 2 $u_1, u_2 \sim U(0,1)$ + independent

Pseudo random little dependency exists

$x_1 = \mu + \sigma\sqrt{-2\ln u_1}\cos(2\pi u_2)$

$x_2 = \mu + \sigma\sqrt{-2\ln u_1}\sin(2\pi u_2)$

Normal $\quad z_1, z_2, \dots, z_n \sim N(0,1)$

Simulate χ^2_n $\quad \sum_{i=1}^{n} z_i^2$

Levy – $X = \frac{r}{z^2} + \alpha$ $\quad Z \sim N(0,1)$

Gamma k degree of freedom

$Gamma_\lambda = \sum_{i=1}^{k}$ indep expon $\quad u_1, u_2, \dots, u_k$

$= \sum_{i=1}^{k} \lambda \ln(u_i)$ $\quad U(0,1)$ independent

Can't get $F^{-1}(u)$ or canonical definition

Acceptance-Rejection method:

a) Suppose you can easily simulate from some function $g(x)$

Suppose it has cumulative dist func $G(x)$

but you want to simulate from $f(x)$, which is hard or impossible

b) Try to find value $c > 1$, such that $f(y) \le c\,g(y)$ $\forall y$.

Method: ① generate a random value y from $g(y)$

② generate another random number u

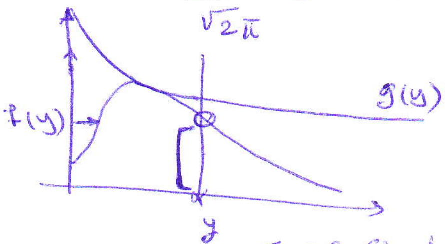
③ if $u \cdot c \cdot g(y) \le f(y)$, set $x = y$ output

if not skip and go back to step1 + don't output.

Suppose we want to simulate $|Z|$ where

$Z \sim N(0,1)$

p.d.f: $\frac{2e^{-Z^2/2}}{\sqrt{2\pi}} = f(y)$



$\longleftrightarrow$ g simulates this $\underline{u}$

Given y $\quad P[u\, c\cdot g(y)) < f(y)] = \frac{f(y)}{c \cdot g(y)}$

probability of accepting a value $= \int_{\text{Range}(y)} \frac{f(y)}{c\, g(y)} g(y)\,dy$

$= \frac{1}{c}$

You want c to be close to 1, since most of numbers you generate will be accepted

$P(Y \le x \text{ \& accept}) = \int_{-\infty}^{x} \frac{f(y)}{c\,g(y)} g(y)\,dy = \frac{\int_{-\infty}^{x} f(y)\,dy}{c}$

$P(Y \le x \mid \text{accepted}) = \int_{-\infty}^{x} \frac{f(y)\,dy}{c} \Big/ \frac{1}{c} = \int_{-\infty}^{x} f(y)\,dy = F(x)$

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

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

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

$\Rightarrow \quad x = 1$

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

$= \frac{2e^{x - x^2/2}}{\sqrt{2\pi}} \left[\cancel{-1} + \cancel{1} - 2x + x^2 \right] \Big|_{x=} < 0$

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

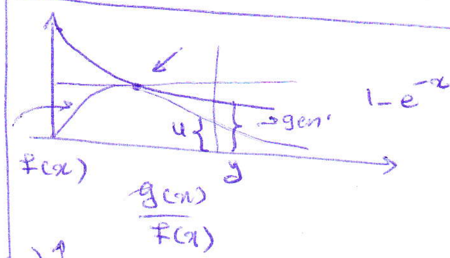
1.3 $\qquad 1.315489 = c$

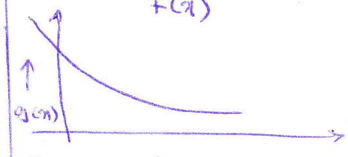
$\frac{1}{c} = 1/1.315489 = .760173$

~~Ready~~

$1.315489 e^{-x}$ ~~$\geq$~~ $f(x)$

Generate values in multidimensional space





⊛ Question what is smallest value of c to make them tangent and make them efficient

function that is easy to evaluate

Multivariate

$MVN(\mu, \Sigma)$

$p \times 1 \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$

$\Sigma > 0$ means $x' \Sigma x > 0 \quad \forall\, \underline{x} \neq 0$

positive definite

Then there exist an orthogonal matrix Γ

and Diagonal matrix D_λ

$\Sigma = \Gamma D_\lambda \Gamma'$

since Σ is always symmetric

and diagonal of D_λ are the eignvalues of Σ $\quad (\lambda_i > 0)$ + Columns of Γ are eigen vectors of Σ

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

Column of Γ

to generate $MVN(\mu, \Sigma)$

a) generate $Z_1, Z_2, \dots, Z_p$ indep $N(0,1)$

$\underline{x} = \underline{\mu} + \Gamma' D \sqrt{\lambda}\, \mathbf{Z} \qquad D_{\sqrt{\lambda}} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_p} \end{pmatrix}$

Statistics

Feb 13 ⑤

x_1 $\mu_1=\mu_2=0$ $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\Sigma=\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$
x_2 $\sigma_1=\sigma_2=1$

$\rho = Cor(x_1, x_2)$

Eigen values are $1+\rho$ $1-\rho$

$= \beta$ $\begin{matrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{matrix}$

$$\begin{matrix} x_1 \\ x_2 \end{matrix} = \begin{pmatrix} \sqrt{\frac{1+\rho}{2}}\, z_1 + \sqrt{\frac{1-\rho}{2}}\, z_2 \\ \sqrt{\frac{1+\rho}{2}}\, z_1 - \sqrt{\frac{1-\rho}{2}}\, z_2 \end{pmatrix}$$

Gibbs (physisist)

Gibbs Sampler — Finance, Econometrics

$x_1, x_2, \dots, x_p$ is hard

but x_i | others is easier

do one at a time

* Start with arbitrary then you get sequence of vectors

x is $N(20, 10)$ (σ) $y \sim N(15, 5)$

$\rho = Corr(x, y) = .6$ MVN

$y|x \sim N(\alpha_{y|x} + \beta_{y|x} \cdot x$ $\sqrt{1-\rho^2}$
9 .3 4 $\sqrt{.64} = .8$

$y|x = N(9 + .3x, 4)$

$x|y$ is $N(2 + 1.2\,y, 8)$

$\beta_{y|x} = \rho \frac{\sigma_y}{\sigma_x} = .6 \frac{(5)}{10} = .3$

$\alpha_{y|x} = \mu_y - \beta_{y|x} \cdot \mu_x = 15 - 0.3(x) = 9$

Excel

x_1 = norminv(rand(), 20, 10)
y_1 = norminv(rand(), 9 + .3 * I4, 4)
x_2 = norminv(rand(), 2 + 1.2 * y_1, 8)

no matrix inversion and eigen value here

They don't have to be exact approximate is enough — a lot of numerical integration is done with this

⑥

- it is recursive process

---

p-p plot | q-q plot

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

$E(x_{(i)})$ as $n \to \infty$ $F^{-1}(i/n+1)$

in a q-q plot, you plot $x_{(i)}$ vs. $F^{-1}(i/n+1)$

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

$x_1, x_2, \dots, x_n$

$\mu + \sigma F^{-1}(i/n+1)$

You must look at the tails -

* You can do beta, gamma probability plot and you can see whether fits normal

β: miss something is going on but you can't see

Power: something going on and you see it, probability

- most tests do not have power

$r = [xp]$ Sample size

r, p the desired quantile

as $n \to \infty$ $\frac{\sqrt{n}(x(p)\ F^{-1}(p))}{\sqrt{p(1-p)}} \cdot F(F^{-1}(p))$

$\xrightarrow{D} Z$ $N(0,1)$

F Normal $p = .5$

$r = [x \cdot .5]$

$F^{-1}(.5) = \mu$ symmetric $\frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}\sigma^2}$

$F^{-1}(.5) = \mu$

$F(F^{-1}(.5)) = F(\mu)$ median $x_{md} \sim N(\mu_{md}, \tau)$
as $x \to \infty$

as $n \to \infty$ $\tau^2 = \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot 2\pi\sigma^2}{n}$

asymptotic dist $= \pi \frac{\sigma^2}{2n}$

of order statistic n, r $\tau = \sqrt{\frac{\pi}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 1.25 \frac{\sigma}{\sqrt{n}}$

quantiles

$\int_{-\infty}^{\infty} f(x)\,dx$ Calculus

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

no close form / solution

Numerical Quadratics

simple formula that works well on polynomial

SIMSON Rule

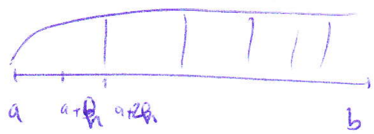
$$\int_a^b F(x)\,dx = \frac{(b-a)}{6}\left[F(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right]$$

read documantation of your package and see what it uses

$\int_0^1 F(x)\,dx$

$F(x)=1$	$\int_0^1 1\,dx = 1$	$\frac{1}{6}(1+4+1] = 1$
$f(x)=x$	$\int_0^1 x\,dx = -\frac{1}{2}x^2 = 1/2$	$1/6\,[0+4*\frac{1}{2}+1] = 3/6 = .5$
$f(x)=x^2$	$\int_0^1 x^2dx = \frac{1}{3}x^2 = 1/3$	$1/6\,[0+4\cdot\frac{1}{4}+1] = 2/6 = 1/3$
$f(x)=x^3$	$\int_0^1 x^3dx \; 1/4\,x^3 = 1/4$	$\frac{1}{6}[0+4\times\frac{1}{8}+1] = \frac{1.5}{6} = \frac{1}{4}$
$f(x)=x^4$	$\int_0^1 x^4dx = \frac{1}{5}x^4 = 1/5$	$\frac{1}{6}[0+4\times\frac{4}{16}+1] = 1.25$

does not work ←
but is pretty close



divide into n intervals with $h = \left(\frac{b-a}{n}\right)$

$\frac{2h}{6}$ integral $\frac{h}{3}\left(f(0)+4f(a+h)+f(a+2h)\right)$

add up

$\frac{h}{3}\left[g(a)+g(b)\right.$

$+ 2\left(g(a+2h)+g(a+4h)+\dots\right)$

$\left.+ 4\left[g(a+h)+g(a+3h)+\dots\right]\right]$

Excel

Beta distribution alpha = beta = 2

density func

mean 0.5 $\quad 6x(1-x)*x$

Var $1/20$

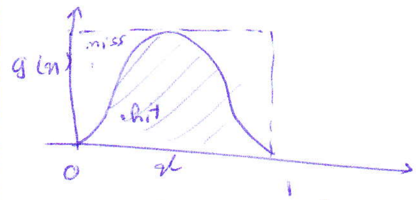
$h = .5$

x	$f(x)$	
0	$=6\cdot x(1-x)$	$=(.5/3)\times(x_0+4x_1+x_2)$
0.5	:	
1	:	

Monte carlo integration

Suppose you want to find

$\int_0^1 g(x)\,dx$ $\quad$ $\max g(x) = 1$, $0<x<1$; $\min g(x) = 0$, $0<x<1$



Post

generate two uniform rand variable and evaluate $g(x')$

Count a hit, if $y' < g(x')$

Count a miss otherwise

Do n times $x = \#$ hit

$\int_0^1 g(x)\,dx = x/n$ $\quad E\left(\frac{x}{n}\right) \longrightarrow p$ area $\quad$ binomial

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

Want to know integral with an accuracy of δ with probability $1-\alpha$ $\quad n \geq \frac{z^2_{\alpha/2}}{4\delta^2}$

Alternatively

let $\theta = \int_0^1 g(x)\,dx$ $\quad x \sim U(0,1)$

$= E(g(x))$

generate $u_1, u_2, \dots, u_n$

Compute $\hat{\theta} = \sum_{i=1}^{n} \frac{g(u_i)}{n}$

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

$g(x) = 1-x^2$

$\int_0^1 (1-x^2)\left[x-\frac{x^3}{3}\right]_0^1 = 1-\frac{1}{3} = 2/3$

Excel

$=1-\text{rand}()^2 \longrightarrow$ 100 times & take average

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

Generate $x_1, x_2, \dots, x_n$ from $F(x)$

then compute $g(x_1), g(x_2), \dots, g(x_n)$

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

next time we will talk about estimation

Statistics Feb 20 (1)

- data of random sample of employees
- histogram
- in between
- You are asked to fit this data, and cut down parameter
- Beta could not be, since beta has α, β
- Gamma and lognormal could fit
- which one is right?
- You can always do q-q ordered value vs. inverse func
- you can fit chisq
- Gamma is one possible choice

$$P(x|\theta, r) = \frac{\theta^r x^{r-1} e^{-\theta x}}{\Gamma(r)} \quad x \geq 0$$

density func that we want to fit?

Estimation

Given DATA (Random Sample), estimate the parameters (r, θ) of the model.

OLDEST Method : method of moments

easy to use, and has little draw back

Find moments of your model

estimate moments from your Data

then set them Equal and solve

(use as many sample moments as parameters)

$$\bar{x} = \sum_i x_i / n$$

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

$$\hat{\mu}_3 = \sum_i \frac{(x_i - \bar{x})^3}{n} \quad \text{3rd moment}$$

$$\hat{\mu}_4 = \sum_i \frac{(x_i - \bar{x})^4}{n}$$

Gamma

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

$$E(x) = \frac{\Gamma(r+1)}{\Gamma(r)\theta} = r/\theta$$

(2)

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

$$Var(x) = r^2/\theta^2 + r/\theta^2 - \left(\frac{r}{\theta}\right)^2 = r/\theta^2$$

two unknown param, and two equations

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

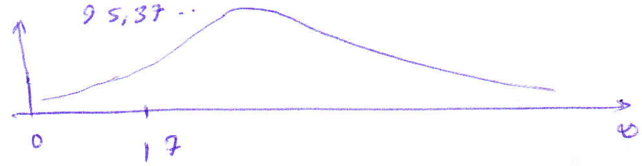
$$\boxed{\frac{s^2}{\bar{x}} = \frac{1}{\hat{\theta}} \Rightarrow \hat{\theta} = \frac{\bar{x}}{s^2} \qquad \bar{x} = \hat{r}/\hat{\theta} \Rightarrow \hat{r} = \bar{x} \cdot \hat{\theta} = \frac{\bar{x}^2}{s^2}}$$

for this data: $\bar{x} = 34.62$

$s^2 = 95.934949$

$$\hat{\theta} = \frac{34.62}{95.93\ldots} = .3609$$

$$\hat{r} = \frac{(34.62)^2}{95.37\ldots} = 12.49$$



the other possibility

look at $x - 17$

$$\hat{\theta} = \frac{34.62 - 17}{95.934949} = .1837$$

$$\hat{r} = \frac{(34.62 - 17)^2}{95.934948} = 3.2361971$$

Could be Correlated !

which way?

Can do histogram and Q-Q plot to see which one is better.

* are you happy?

Method of moment gives point estimate?

Confidence?

* Question $\pm$?

are they Correlated?

Quick and easy to find point estimate, but there are still Questions.

* only in normal dist we have no ~~also~~ dependence of μ, σ^2.

usually

Good size n: 100, 1000 would be fine, all are asymptotic, so sample of 10 not good

$\hat{\theta}, \hat{r}$ are functions of $(\bar{x}, s^2)$

$f(x,y) \simeq f(\mu_1,\mu_2) + \frac{\partial f}{\partial x}\Big|_{\xi_1 \text{ b/w } x,\mu_1}(x-\mu_1) + \frac{\partial f}{\partial y}\Big|_{\xi_2 \text{ b/w } y,\mu_2}(y-\mu_2)$

for x close to μ_1, y close to μ_2

this implies that $E(f(x,y)) \approx f(\mu_1,\mu_2)$

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

$E(\bar{x}^2) = \mu^2$

with bigger sample we will have this

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

$Var(y) + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}Cov(x,y)$

- it is correlation of error terms, and we can generalize it

$\bar{x}$, s^2 our parameters

$\hat{\theta} = \frac{\bar{x}}{s^2}$, $Var(\bar{x}) = \sigma^2/n = (r/\theta^2)/n = \frac{r}{\theta^2 n}$ μ_4

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

at least 50 sample size.

$Cov(\bar{x}, s^2) = ?$

for normal it is 0, but you can derive it.

American statistician 2007 pp. 159-160

Covariance$(\bar{x}, s^2) = \mu_3/n = E(x-\mu)^3 \to$ Gamma

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

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

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

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

$E(x) = r/\theta$

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

$\frac{r}{\theta^3}[\cancel{r^2} + 3r + 2 - \cancel{3r^2} - 3r + \cancel{2r^2}]$

$= 2r/\theta^2$

$Cov(\bar{x}, s^2) = \frac{2r}{\theta^2 n}$

$E(x-\mu)^4 = \frac{6r}{\theta^4}$

$Var(\bar{x}) = \frac{r}{n\theta^2}$

$E(x) = r/\theta$

$\Rightarrow \hat{\theta} = \bar{x}/s^2$

$\frac{\partial\hat{\theta}}{\partial\bar{x}} = 1/s^2$ $\quad \frac{\partial\hat{\theta}}{\partial s^2} = \frac{-\bar{x}}{s^4}$

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

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

$\Rightarrow Var(\hat{\theta}) =$

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

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

$\Rightarrow Var(\hat{\theta}) = ($

$Var(f(x,y)) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)\begin{pmatrix} V(x) & Cov(x,y) \\ Cov(x,y) & V(y)\end{pmatrix}\begin{pmatrix}\frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y}\end{pmatrix}$

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

$\times \begin{pmatrix}\frac{\hat{\theta}^2}{\hat{r}} \\ \frac{-\hat{\theta}^3}{\hat{r}}\end{pmatrix} = \frac{3\hat{\theta}^2}{n\hat{r}} + \frac{2\hat{\theta}^2}{(n-1)}$

When you want to get variability, usually you need to do lot of work

in real life you just have to multiply numbers, tedious part is to put formula and simplify it

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

$\times\begin{pmatrix}\hat{\theta}^2 \\ -\hat{\theta}^3\end{pmatrix} = \frac{2\hat{r}}{n}\left[1 + \frac{n\hat{r}}{n+1}\right] \simeq \frac{2\hat{r}(\hat{r}+1)}{n}$

$\hat{\beta}_1 = f_1(x_1, x_2)$ $\quad \hat{\beta}_2 = f_2(x_1, x_2)$

$\hat{\theta}$: $\bar{x}$, s^2, $\hat{\theta}$

$\hat{\beta}_1 = \frac{\partial f_1}{\partial x_1}(x_1 - \mu_1) + \frac{\partial f_1}{\partial x_2}(x_2 - \mu_2) + f_1(\mu_1, \mu_2)$

$\hat{\beta}_2 \simeq \frac{\partial f_2}{\partial x_1}(x_1 - \mu_1) + \frac{\partial f_2}{\partial x_2}(x_2 - \mu_2) + f_2(\mu_1, \mu_2)$

Statistics $\qquad$ Feb 20 $\qquad$ 5

$$\mathrm{Cov}(\hat\beta_1,\hat\beta_2) = \left(\frac{\partial f_1}{\partial x_1}\cdot\frac{\partial f_2}{\partial x_1}\right)\mathrm{Var}(x_1) + \left(\frac{\partial f_1}{\partial x_2}\cdot\frac{\partial f_2}{\partial x_2}\right)\cdot$$

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

$\mathrm{Cov}(\hat r,\hat\theta)$

- many often you have to solve - iterate - kth estimate, until converge -
  appl: starting moment, start somewhere
  start up with the moment estimator

- it could be a right model - Q-Q it fits empirically - on the set of data - ~~just~~ b/c gamma fit, not say it is gamma - could have been approximated
- human being is sensitive to linearity, so q-q is better than histogram

intercept → location } for q-q
slope → scale

you have to do this once, and say I did it

(*) Statistics never say it is certain, there is always chance

* Coefficient of variation

$$\boxed{c_v = \sigma/\mu}$$

- measure in fore cast - Error as func of mean
  tricky measure when $\mu=0$
- Dow Jones industrial $\frac{3000}{400{,}000} = 0.07 \Rightarrow$ not bad
- if the data is normal $(1+\frac{1}{4n})\frac{s}{\bar x}$ is unbiased estimate of $\frac{\sigma}{\mu}$
- Assume DATA is Normal $\mathrm{Cov}(\bar x, s^2)=0$

$$\hat C_v = \frac{\sqrt{s^2}}{\bar x}$$

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

$$\hat V(\hat C_v) = (-s/\bar x^2)^2\,\mathrm{Var}(\bar x) + \left(\frac{1}{2\bar x s}\right)^2 \mathrm{Var}(s^2) =$$

$$\frac{s^2}{\bar x^4}\cdot\frac{\sigma^2}{n} + \frac{1}{4\bar x^2 s^2}\,\frac{2\sigma^4}{n} = \sigma^2/n\left[\frac{s^2}{\bar x^4} + \frac{1}{2\bar x^2}\,\sigma^2/s^2\right]$$

6

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

$$\boxed{\hat\sigma = s}$$

$$= \frac{1}{n}\left(\frac{s}{\bar x}\right)^2\left[\left(\frac{s}{\bar x}\right)^2 + \frac{1}{2}\right] = \frac{1}{n}(\hat C_v)^2\cdot [\hat C_v)^2 + \tfrac12]$$

very useful formula

application | CRIME

① lot of people don't report
don't know how many are
probability reporting
$x_1, \dots, x_n$

k ↓ crime $\qquad p$ ↘ probability of reporting

x_i is $Bin(k, p)$ $\quad$ $E(x_i) = kp$ $\quad$ $SD(x_i) = kp(1-p)$

Goal is to estimate k + p

$$\bar x = \Sigma_i x_i/n \qquad E(\bar x) = \sum_{i=1}^n \frac{E(x_i)}{n} = \frac{nkp}{n} = kp$$

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

$E(s^2)$: Need $E(x_i^2)$ $\quad E(\bar x^2)$

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

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

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

$$\frac{1}{(n-1)}\left[nkp(1-p) + nk^2p^2 - n\left[\frac{kp(1-p)}{n} + k^2p^2\right]\right]$$

$$= \frac{1}{n-1}\left[nkp(1-p) + \cancel{nk^2p^2} - kp(1-p) - \cancel{nk^2p^2}\right]$$

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

$$\bar x = kp \qquad s^2 = kp(1-p)$$

$$\Rightarrow \frac{s^2}{\bar x} = 1-p \Rightarrow \hat p = 1 - s^2/\bar x \qquad 0<\hat p<1$$

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

hard to do with max likelihood & Bays and play to find what stat properties are
Very Strong Result

t-test

One sample t-test

$H_N = \mu = \mu_0$

$H_A = \mu \neq \mu_0$ $\quad x_1, x_2, \ldots, x_n$

$t_{obs} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim$ t-dist with n-1 deg. frdm

originally thought normal

two samples

$x_{11} \quad x_{21}$

$\vdots \quad \vdots$

$x_{1n_1} \quad x_{2n_2}$

$H_0 : \mu_1 = \mu_2$

$H_N : \mu_1 \neq \mu_2$

t-dist depends on degree of freedom

1- deg freedom $\rightarrow$ Cauchy

25 deg freedom $\rightarrow$ indisting

$\downarrow$

observation

- For small sample size t-dist is used

$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$ Not A t-dist

Behrens Fisher problem

assumption for long time $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$\hat{\sigma} = \frac{(n_1-1)\, s_1^2 + (n_2-1)\, s_2^2}{n_1 + n_2 - 2}$

$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2/n_1 + s^2/n_2}} \sim t(n_1 + n_2 - 2)$

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

$H_a : \sigma_1^2 \neq \sigma_2^2$

Saither width Welch Huzurbazar : Let's empirically study the sample dist $t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$

approximately t: df ?

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

it is $\longleftarrow$

two sample t-test with unequal variance

$\min(n_1-1, n_2-1) \le df \le (n_1 + n_2 - 2) \leftarrow s_1^2 = s_2^2$

degree of freedom come from method of moment

Let $Y_1, Y_2, \ldots, Y_N$ be iid

$Y_i \sim \chi^2(f_i)$

$\sum_i a_i Y_i$ Not χ^2

but assume it is close to $c\chi^2(\nu)$

Goal : find c & v approx

$E(\sum_i a_i Y_i) = \sum_i a_i E(y_i) = \sum_i a_i f_i = c\nu$ $\leftarrow$ MM method of moment

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

$$\boxed{\hat{c} = \frac{\sum_i a_i^2 f_i}{\sum_i a_i f_i} = \frac{\hat{c}^2 \hat{\nu}}{\hat{c}\hat{\nu}} \qquad \hat{\nu} = \frac{(\sum_i a_i f_i)^2}{\sum_i a_i^2 f_i}}$$

method of moment gives us approximation

we are looking at $s_1^2 = \sum_i \frac{(x_{1i} - \bar{x}_1)^2}{(n-1)} = \sum_i \frac{(x_{1i} - \bar{x}_1)^2}{f_1}$

$\frac{\sum_i (x_{1i} - \bar{x}_1)^2}{\sigma_1^2} = \sum \left(\frac{x_{1i} - \bar{x}_1}{\sigma_1}\right)^2 \rightarrow \chi^2_{f_1}(n_1 - 1)$

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

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

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

$a_1 = \frac{\sigma_1^2}{n_1 f_1} \qquad a_2 = \frac{\sigma_2^2}{n_2 f_2}$

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

- method of moment has application in applied Statistical method
- people became tired of them, so went over maximum likelihood

Statistics

Feb 27

- the hw should be redone on numericals as well

Maximum Likelihood

- Suppose we have $x_1, x_2, \dots, x_n$
  density $p(x, \underline{\theta}) \leftarrow$ key step: choosing model
  with $prob(x_1, x_2, \dots, x_n | \underline{\theta}) = \prod_{i=1}^{n} p(x_i, \underline{\theta})$
  then we will check whether model fits data
  entire method dependent upon model choice.

assums (philosophical)
- Nature is indifferent to you.
  no bias in sample - set of parameters pop up
  the one that is most likely to happen
  no systematic bias exists

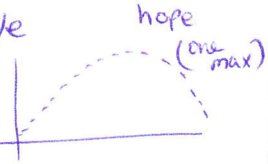
Fisher
$\mathcal{L}(\theta) = \prod_{i=1}^{n} p(x_t | \theta)$

value param are things that make this happen
maximum likelihood estimator $\hat{\underline{\theta}}$, as that value of $\underline{\theta}$, which maximizes $\mathcal{L}(\underline{\theta})$

- we assume implicitly model reasonable
  value most likely happen will happen

2 Cases - θ is unrestrained $(-\infty < \theta < \infty)$
whole range is possible

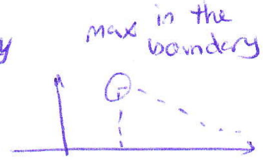
hope (one max)



if θ is variance $0 - \infty$
if θ is proportion $0 - 1$
based on characteristic of the parameter

② it is constraint in some way

max in the boundary



we may have

- worry about relative max as opposed to absolute max.
  check whether you keep converging
- three or four time from different point $\rightarrow$ make sure absolute max

ESP: Extra Sensory perception

$\bigcirc \ \triangle \ \square \ \int\int \ \#$ 20 each

No ESP on averg $\frac{1}{5} 100 = 20$
$\frac{1}{5}$th right

ESPI measure of your success
0 IF No ESP
1 IF perfect

$\theta = \frac{p - 1/5}{4/5}$ $\boxed{p = \text{proportion of the successes}}$

- if you estimate monotone func. of param you get the same est. param
e.g. max likelihood (var) = sqrt (std)

Max likelihood estimate of θ

real score π
p: binomial
- Beauty of max likelihood: you can write way you want

observe random variable: $S_i = \begin{cases} 1 & \pi \\ 0 & 1-\pi \end{cases}$

$\theta = \frac{\pi - 1/5}{4/5}$

$\pi = \frac{1}{5} + \frac{4}{5}\theta$

$1 - \pi = \frac{4(1-\theta)}{5}$

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

index that make sense in Economic sense

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

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

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

statistics is additive
world is multiplicative, but log is used to convert additive

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

$\downarrow$

$\ln(\frac{1}{5} \cdot (1 + 4\theta))$

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

$+ \sum_i (1-S_i) \cdot \ln(1-\theta) + \sum_i (1-S_i) \ln 4$

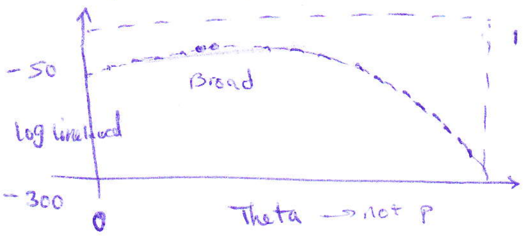
$- \sum_i (1-S_i) \ln 5$

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

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

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

$x=30 \quad n=100$



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

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

$$+4\hat\theta x = 0 \Rightarrow 4x - n + x + n \cdot 4\hat\theta$$

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

$$\Rightarrow \begin{matrix} n=100 \\ x=30 \end{matrix} \quad \hat\theta = \frac{5}{4}(.3) - .25 = (1.25)(.3) - 0.25 = 0.375 - 0.25 = 0.125$$

how get mean and Variance?

Suppose you have a "density", θ be an r-dimensional vector of parameters

Define ("score function)

$$S_i(x, \theta) = \frac{\partial}{\partial \theta_i} \ln(x, \theta)$$

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

Define $p(x, \theta)$ as "regular" w.r.t its first derivatif

if $E_x(S_i(x,\theta)) = 0 \quad i = 1, 2, \ldots, r$

is regular with respect to second derivative

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

then if $x_1, x_2, \ldots, x_n$ is a sample with p.d.f $p(x,\theta)$ is regular with respect to first and 2nd derivative, then (*)

{all density func do not have this feature, and you need to check these two conditions}

(*) ① a maximum likelihood Estimator exists.

② maximum likelihood Estimator satisfies the

Equation $\sum_{j=1}^{n} S_i(x_i, \hat\theta_{ML}) = 0$

③ If $\hat\theta_{ML}$ is unique for $n >$ some n_0 then $\hat\theta_{ML}$ is asymptotically multivariate normal with mean vector θ and Covariance matrix $B^{-1}(\theta, \theta)/n$ where $B_{ij} = E(S_i(x,\theta), S_j(x,\theta))$ $= -E(S_{ij}(x,\theta))$

product of first partial

mixed partial

Some people when things become complicated, in real life

Estimate $E(S_{ij}(B,\theta)) \simeq \frac{1}{n}\sum_{k=1}^{n} S_{ij}(x_k, \hat\theta_{ML})$

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

$$\hat B^{-1} = \hat B^{-1}$$

$$\hat\theta \sim^{MVN} \left(\theta, \left(-\sum_{k=1}^{n} S_{ij}(x_k, \hat\theta_{ML})\right)^{-1}\right)$$

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

program puts little h and evaluates, they compute this. Estimate on Estimate

- have big samples when you use these Programs

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

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

$1 - \pi = \frac{4(1-\theta)}{5}$

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

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

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

$$(1+4\theta)^{S_i} 4^{1-S_i} (1-S_i)^{1-S_i}/5$$

Probab dens. func. Simple Guess

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

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

include this when plot

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

Statistics

Feb 27

6

check whether regular

$$E\left(\frac{\partial L(\theta)}{\partial\theta}\right) = \frac{4\,E(S_i)}{1+4\theta} - \frac{E(1-S_i)}{(1-\theta)} = \frac{4(1+4\theta)}{5(1+4\theta)} - \frac{4(1-\theta)}{5(1-\theta)} = \frac{4}{5}-\frac{4}{5}=0$$

regular w.r.t 1st derivative

check reg. with res. to 2nd deriv.

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

$S_i = \{^0_1$

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

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

$$\frac{4}{5}\left[\frac{4-4\theta+1+4\theta}{(1+4\theta)(1-\theta)}\right] = \frac{4}{(1+4\theta)(1-\theta)} \rightarrow var(\hat\theta)$$

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

$$\boxed{-E(S_{ij}(x,\theta)) = E(S_i(x,\theta)\,S_j(x,\theta))}$$

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

- unless well defined func : Computer Calculates but You're not sure the variance exists

---

$$f(x) = \frac{\theta^r e^{-\theta x} x^{r-1}}{\Gamma(r)} \qquad \theta = \binom{r}{\theta}$$

$$\ln f(x) = r\ln\theta - \theta x + (r-1)\ln x - \ln\Gamma(r)$$

$x_1, x_2, \dots, x_n$

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

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

$$\overline{\ln x} = \sum_{i=1}^n \frac{\ln x_i}{n}$$

Gamma func is regular

DiGamma function

1) $$\frac{\partial L(r,\theta)}{\partial r} = n\left[\ln\theta + \overline{\ln x} - \frac{\Gamma'(r)}{\Gamma(r)}\right] = 0$$

2) $$\frac{\partial L(r,\theta)}{\partial\theta} = n\left[\frac{r}{\theta} - \bar x\right] = 0 \quad \text{II}$$

Trigamma func

2nd Derivative $$\frac{\partial^2 L(r,\theta)}{\partial r^2} = -n\left[\frac{\Gamma(r)\Gamma''(r) - [\Gamma'(r)]^2}{\Gamma(r)^2}\right]$$

6

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

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

II $$\frac{\hat r}{\hat\theta} - \bar x = 0 \Rightarrow \hat\theta = \frac{\hat r}{\bar x}$$

$$L(r,\theta) = {}^{at}\,L\left(r,\frac{r}{\bar x}\right) = n\left[r\ln\left(\frac{r}{\bar x}\right) - \frac{r}{\bar x}\cdot\bar x + (r-1)\right.$$

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

$$B(r,\theta) = \begin{pmatrix} .078622 & -2,62023 \\ -2,62023 & 90,71748 \end{pmatrix}$$

$$\frac{1}{100}\hat\theta^{-1} = \begin{pmatrix} 3,405589 & 0,098371 \\ 0,098 & 0,002952 \end{pmatrix}$$

$sd(\hat r) = 1.8454$ Sqrt

$11,25 \le \hat r < 1613$

$sd(\hat\theta) = 0.05435$

$275 \le \hat\theta \le 1488$

Sqrt

$\hat\theta = r/\bar x$

1st derivative

$$n\left[\ln\left(\frac{r}{\bar x}\right) + \overline{\ln x} - \frac{\Gamma'(r)}{\Gamma(r)}\right] = 0$$

$$n\left[\log r - \ln\bar x + \overline{\ln x} - \frac{\Gamma'(r)}{\Gamma(r)}\right] = 0$$

Newton raveson method

$x = \sqrt{K}$

$f(x) = x^2 - K = 0$

$f(x) = f(\mu) + f'(\xi)(x-\mu)$

$0 = f(\mu) + f'(\xi)(x-\mu)$ neighborhood of solution

Guess a solution x_k not sol. but close

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

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

$\rightarrow$ stay reasonable

$$x_{k+1} = x_k - \frac{(x_k^2 - K)}{2x_k} = x_k + \frac{K - x_k}{2x_k} = x_k - \frac{K}{2x_k} - \frac{x_k}{2}$$

$$= \frac{1}{2}\left(x_k + \frac{K}{x_k}\right)$$

for -68 you will see that it is not Converging

you have to be wory about Convergence

in this Case probably you need to start other values

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_k(x) \end{pmatrix} \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}$$

(7)

$$f_i(\underline{x}) = f_i(\underline{\mu}) \sum_{j=1}^{n} \frac{\partial f_i(x)}{\partial x_j}(x_j - \mu_j)$$

$g(x)$ max

$$f(x) = \begin{pmatrix} \frac{\partial g_1(\underline{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial g(\underline{\theta})}{\partial \theta_2} \end{pmatrix} = \underline{0} \qquad \underline{0} = f(\underline{\mu}) + \left\{ \frac{\partial^2 g(\underline{\theta})}{\partial \theta_i \partial \theta_j} \right\} (\underline{\theta} - \underline{\mu})$$

$k \times k$ matrix

$$\underline{\mu} = \hat{\underline{\theta}}_k$$

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

Stat inference   March 06

- Maximum likelihood estimation of transformation

Family of transformations:

$-\infty < \lambda < \infty$

$Z_i(\lambda) = \dfrac{x_i^{\lambda} - 1}{\lambda}$ $\left[\begin{array}{l}\lambda = 0 \\ Z_i(\lambda) = \ln x_i\end{array}\right.$

transform your data to something simple

- want to see if there is λ such that $Z_i(\lambda)$ is normally distributed

the idea is to estimate transformation

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

whatever you did then you will check with p-p or q-q

random sample $x_1, x_2, \ldots, x_n$

$Z_1(\lambda), Z_2(\lambda), \ldots, Z_n(\lambda)$

with joint density

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma(\lambda)} e^{-\frac{1}{2}\left(\frac{Z_i(\lambda) - \mu(\lambda)}{\sigma(\lambda)}\right)^2} dZ_1(\lambda) dZ_2(\lambda) \ldots dZ_n(\lambda)$$

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

$\dfrac{dZ_i(\lambda)}{dx_i} = \dfrac{\lambda x_i^{\lambda-1}}{\lambda} = x_i^{\lambda-1}$ $[\lambda = 0 \quad 1/x_i]$

$f(x_1, x_2, \ldots, x_n \mid \lambda, \mu(\lambda), \sigma(\lambda))$
↳ pdf

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma(\lambda)} e^{-\frac{1}{2}\left[\frac{\frac{x_i^{\lambda}-1}{\lambda} - \mu(\lambda)}{\sigma(\lambda)}\right]^2} x_1^{\lambda-1} x_2^{\lambda-1} \ldots x_n^{\lambda-1} dx_1 dx_2 \ldots dx_n$$

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

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

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

Q: How to estimate $\mu(\lambda), \sigma^2(\lambda), \lambda$?

Given λ, $\hat{\mu}(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i^{\lambda}-1}{\lambda}\right) = \bar{Z}(\lambda)$

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

$= \frac{1}{n}\sum_{i=1}^{n}(Z_i(\lambda) - \bar{Z}(\lambda))^2$

① $L(\lambda, \hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_i \frac{(Z_i(\lambda) - \bar{Z}(\lambda))^2}{\hat{\sigma}^2(\lambda)}$ ②

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

$= -\frac{n}{2}\ln 2\pi - \frac{n}{2} \underbrace{-\frac{n}{2}\hat{\sigma}^2(\lambda) + (\lambda - 1)\sum_i \ln x_i}_{\text{only part depend on } \lambda}$

Numerically

Steps: ① pick λ

② Compute for each x_i $\dfrac{x_i^{\lambda} - 1}{\lambda} = Z_i(\lambda)$

③ Compute $\overline{Z(\lambda)}$ and $\hat{\sigma}^2(\lambda)$

④ Compute $-\frac{n}{2}\hat{\sigma}^2(\lambda) + (\lambda - 1)\sum_i \ln x_i$ *

⑤ plot * against λ

⑥ go back to 1 after incrementing λ

- Start from large range and then narrow down

$\hat{\lambda} = -0.07$

* There is many ways to do thing and fit data but at last you need to check by p-p q-q

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

$x_1, x_2, \ldots, x_n \quad f(x) = 1/\theta$

$x_{(n)} = \max(x_1, x_2, \ldots, x_n) \leq \theta$

$\prod_{i=1}^{n} f(x_i) = \frac{1}{\theta^n}$

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

$L(\theta \mid x)$ is monotone decreasing $\frac{1}{\theta^n}$ is maximized $\hat{\theta} = x_{max}$

$\Rightarrow \hat{\theta}_{max} = x_{(n)} = \max(x_1, x_2, \ldots, x_n)$

$\text{Prob}(x_1, x_2, \ldots, x_n \leq z) = (z/\theta)^n$

$\Rightarrow$ pdf $x_{(n)} = \dfrac{n z^{n-1}}{\theta^n} \quad 0 \leq z \leq \theta$

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

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

Bias of an Estimator $\hat{\theta}$ of θ

as $|E(\hat{\theta}) - \theta| = \text{Bias}(\hat{\theta})$

Bias of $\hat{\theta}_{ML} = \left|\frac{n}{n+1}\theta - \theta\right| = \theta\left|\frac{n - n - 1}{n+1}\right| = \frac{\theta}{n+1}$

asymptotically unbiased: for big sample no problem

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

$$= \frac{n\theta^3}{n+2}$$

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

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

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

---

what we try to do is θ

Variance checks against mean

so define concept

$$\boxed{\text{Mean squared Error } (\hat{\theta})}$$

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \quad \text{not variance but related}$$

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

$$E(\hat{\theta} - E(\hat{\theta}))^2 + \underbrace{E(E(\hat{\theta}) - \theta)^2}_{\text{we called it bias}}$$

$$+ 2E(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)$$

$$= Var(\hat{\theta}) + Bias^2(\theta) + 0$$

$$\Rightarrow \boxed{MSE(\hat{\theta}) = V(\hat{\theta}) + Bias^2(\hat{\theta})}$$

only Equal when unbiased Estimate

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

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

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

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

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

$$Bias(\hat{\theta}_u) = 0$$

$$Var(\hat{\theta}_u) = Var\left(\frac{(n+1)}{n} x_n\right) = \frac{(n+1)^2}{n^2} Var(x_n)$$

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

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

$$\frac{MSE(\hat{\theta}_{ML})}{MSE(\hat{\theta}_u)} = \frac{\frac{2\theta^2}{(n+2)(n+1)}}{\frac{\theta^2}{n(n+2)}} = \frac{2n}{n+1} \geq 1$$

$$MSE(\hat{\theta}_{ML}) \geq MSE(\hat{\theta}_u)$$

Can we do better?

$$\hat{\theta}_c = c\, x_m$$

$$Var(\hat{\theta}_c) = c^2 Var(x_n)$$

$$E(\hat{\theta}_c) = c\, E(x_n)$$

☒ Blue estimate: Best linear unbiased

☒ MVUE minimum = Variance unbiased estimate

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

Best c that minimizes MSE

$$\frac{d}{dc} MSE(\hat{\theta}_n) = 2c\,E(x_n^2) - 2E(x_n)\cdot\theta$$

$$\frac{d^2}{dc^2} MSE(\hat{\theta}_c) = 2E(x_n^2) > 0$$

$$\Rightarrow c = \frac{\theta E(x_m)}{E(x_n^2)} = \frac{\frac{\theta \cdot n}{n+1}\theta}{\frac{n\theta^2}{n+2}} = \frac{\frac{n\theta^2}{n+1}}{\frac{n\theta^2}{(n+2)}} = \frac{(n+2)}{(n+1)}$$

$$\hat{\theta}_c = \frac{n+2}{n+1} x_m \qquad \hat{\theta}_u = \frac{n+1}{n} x_n \qquad \hat{\theta}_{ML} = x_n$$

for big n all equal numerically

$$Bias(\hat{\theta}_c) = |c\,E(x_n) - \theta| = \left|\frac{n+2}{n+1}\frac{n}{n+1}\theta - \theta\right|$$

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

$$\left. - 2n - 1\right) = \frac{\theta}{(n+1)(n+1)}$$

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

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

$$\frac{\theta^2}{(n+1)^4}\left[n^2+2n+1\right] = \frac{\theta^2(n+1)^2}{(n+1)^4} = \frac{\theta^2}{(n+1)^2}$$

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

dist. non Regular → max likelihood

need → unbiased then use $\hat{\theta}_u$

MSE → use $\hat{\theta}_c$

Bays Estimation

based on philosophical assumption:

unknown is not unknow you may know

you have two random x, θ

θ must also have prob. dist

Let $\underline{\theta}$ be a vector of parameters.

ML✓ $p(x|\underline{\theta})$ be the p.d.f of $x|\underline{\theta}$

(1) - all assumption about pick model holds

(2) let $g(\theta)$ be p.df on θ so that

$$\int_{R\underline{\theta}} g(\underline{\theta})d\theta = 1$$

prior dist. on param $\underline{\theta}$

- Someone in the field say that I think it should be between x, y, an bell shap - experts know agaist non experts

- if you say I don't know, Just uniform

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 $L(\underline{\theta}, \hat{\underline{\theta}})$ — function to be minimized by Estimator

(a) $L(\underline{\theta}, \hat{\underline{\theta}}) = \min_{\hat{\theta}} E(\hat{\underline{\theta}} - \underline{\theta})^2$ Quadratic Loss

$$E_\theta(\theta - m)^2 = E_\theta(\theta - E(\theta) + E(\theta) - m)^2 =$$

$$E(\theta - E(\theta))^2 + (E(\theta) - m)^2 \qquad M = E(\hat{\theta})$$

(b) if you pick $L(\theta, \hat{\theta}) = E|\hat{\theta} - \theta|$

$\hat{\theta}$ = median of posterior dist.

$$\text{if } L(\theta, \hat{\theta}) = \begin{cases} 0 & \text{if } \hat{\theta} = \theta \\ k & \text{if } \hat{\theta} \neq \theta \end{cases}$$

$\hat{\theta}$ = mode of posterior dist

Example: $\delta_i = \begin{cases} 1 & \pi \\ 0 & 1-\pi \end{cases}$

$\delta_1, \delta_2, \cdots, \delta_n$

Estimate π — from $\boxed{\frac{\partial LL}{\partial \pi} = 0}$

$$\hat{\pi} = \sum_i \frac{\delta_i}{n} = \frac{X}{n} \leftarrow \text{MVUE}$$ minimum variance unbiased Estimate

$$E(\hat{\pi}) = \pi$$

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

→ Now doing Bayesian

Natural prior:

pick π uniform on $(0,1)$

$1 = g(\pi) \qquad 0 \le \pi \le 1$

$$p(x|\pi)g(\pi) = \frac{n}{x!(n-x)!}\pi^x(1-\pi)^{n-x} - 1$$

$$h(x) = \int_0^1 \cdots = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}\int_0^1 \pi^x(1-\pi)^{n-x}d\pi$$

$$= \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}\int_0^1 \pi^{(x+1)-1}(1-\pi)^{(n-x-1)-1}d\pi$$

$$= \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)} B(x+1, n-x+1)$$

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

$Beta(x+1, n-x+1)$

pick Quadratic Loss

$$\frac{\alpha}{\alpha+\beta} = \frac{x+1}{(x+1)+(n-x+1)} = \frac{x+1}{n+2}$$

$$\hat{\pi}_b = \frac{x+1}{n+2} \qquad \hat{\pi}_{ML} = \frac{x}{n}$$

$$\hat{\pi}_b = \frac{n\hat{\pi}_{ML}+1}{n+2}$$

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

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

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

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

$n \to \infty$

$.5 \pm .5\sqrt{.5}$

$c, .1464 \le \pi \le$

Challanges

- The idea of reproducability hinders science acceptance of Bayesian
  it is about belief system
- max likelihood θ^2 gives you, but on bayesian it has different estimation

Conjugate priors

Jeffrys prior or uniform $0 - \infty$

Bays estimate of θ

$x \sim U(0, \theta)$

$f(x|\theta) = 1/\theta \quad 0 \le x \le \theta$

$f(\underline{x}|\theta) = \frac{1}{\theta^n} \quad 0 \le x \le \theta$

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

$x_m \le \theta \le N$

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

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

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

$\left[\frac{1}{x_n^{n-1}} - \frac{1}{N^{n-1}} \right]$

$\Rightarrow p(\theta|x) = \underbrace{\frac{n-1}{\left[\frac{1}{x_n^{n-1}} - \frac{1}{N^{n-1}}\right]}}_{c} \cdot \frac{1}{\theta^n} \quad x_n \le \theta \le N$

- use Quadratic loss

$\hat{\theta}_b$ - mean of posterior dist

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

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

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

$\frac{\left[1 - \left(\frac{x_m}{N}\right)^{n-2}\right]}{\left[1 - \left(\frac{x_m}{N}\right)^{n-1}\right]}$

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

degenerate prior as $N \to \infty$ will lead to zero and it will disapper

$\hat{\theta}_b = \frac{(n-1)}{(n-2)} x_m$

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

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

Aside $Bias(\hat{\theta}_{ML}) = \frac{\theta}{n+1}$ $\boxed{Bias(\hat{\theta}_{ML}) > Bias(\hat{\theta}_b)}$

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

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

$\hat{\theta}_c \quad \hat{\theta}_u \quad \hat{\theta}_{ML} \quad \hat{\theta}_b$

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

$\boxed{\underset{\hat{\theta}_u}{\frac{n+1}{n} x_m} \quad \underset{\hat{\theta}_{ML}}{x_m} \quad \underset{\hat{\theta}_b}{\frac{n-1}{n+2} x_m} \quad \underset{\hat{\theta}_c}{\frac{n+2}{n+1} x_m}}$

to be consistent you need to incorporate your belief ; Bayesian say

Let $g(\theta)$ pdf of θ

$p(x|\theta)$ let be pdf of (x) shared wit max likelihood

$P(\underline{\theta}|\underline{x}) = \frac{P(\underline{x}|\underline{\theta})g(\underline{\theta})}{h(x)}$

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

$\ln p(\underline{\theta}|x) = \ln p(x|\underline{\theta}) + \ln g(\underline{\theta}) - \ln h(x)$

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

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

$n \cdot E[\ln p(x_i|\underline{\theta})] + \ln g(\underline{\theta}) - \ln h(x)$

increasing in n while other not

- as sample becomes big enough
  the first part which is likelihood
  dominates

- the bigger the sample both converge to
  likelihood

① statistics 20 march

Bayes ③ (frequentist) ① paragmatist ②

Max likelihood [ML]

$f(x|\theta)$ $f(x|\theta)$ data should speak for it self

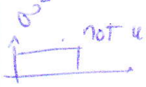
$g(\theta)$

Loss function : could be complicated & Bays handles that

Assume

$\widehat{g(\theta)}_{ML} = g(\hat{\theta}_{ML})$: means $\hat{\sigma}_{ML} \Downarrow \hat{\sigma}^2_{ML}$

σ



$\hat{\sigma}^2$ not uniform

- if x is large effect of prior bw

-- don't have to check for the regularity

- don't have to take analytical solution

paragmatist : if you have prior use it.

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

observe first x_1 & later x_2

① Frequentists: assume independent

$x_1 + x_2 \sim B(n_1+n_2, \pi)$

charac. func

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

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

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

from indep

$$+ n_2\pi(1-\pi)) = \frac{\pi(1-\pi)}{n_1+n_2}$$

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

$$\hat{P}_i = \frac{x_i}{n_i}$$

③ Bayesian

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

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

$g(\pi) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}$ Quadratic Loss ②

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

Beta dist prop since: integral any not π cancel out

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

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

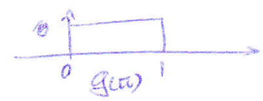
when we use Quadratic loss we use $E(\pi|posterior)$ as estimator (*)

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

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

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

$\alpha = 1, \beta = 1$



$$\hat{\pi}_3 = \frac{x_1+x_2+1}{n_1+n_2+2}$$

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

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

- difference in estimator

② Pragmatist empirical Bayes

first get result of method 1

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

- do some observations, and get some observatio[ns]

- Assume a beta prior and pick α, β so that the first two moment match estimated value of mean & variance

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

$$\begin{cases} \alpha = \hat{P}_1(n_1-1) \\ \beta = (1-\hat{P}_1)(n_1-1) \end{cases}$$

posterior prior $\propto \pi^{\hat{\alpha}-1}(1-\pi)^{\hat{\beta}-1}$

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

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

③

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

$\hat{p}_i = \frac{x_i}{n_i}$

$$= \frac{(n_1 - 1)\hat{p}_1}{(n_1 + n_2 - 1)} + \frac{n_2 \hat{p}_2}{(n_1 + n_2 - 1)}$$

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

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

Class of estimators $\hat{\pi} = w\hat{p}_1 + (1-w)\hat{p}_2$ $w \geq 0$, $0 \leq w \leq 1$

independence

$Var(\hat{\pi}) = w^2 Var(\hat{p}_1) + (1-w)^2 Var(\hat{p}_2)$

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

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

$\Rightarrow w\, Var(\hat{p}_1) - (1-w)Var(\hat{p}_2) = 0$ is the best w

$\Rightarrow w(Var(\hat{p}_1) + Var(\hat{p}_2)) = Var(p_2)$

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

$\Rightarrow$ weights $\frac{n_1}{n_1 + n_2}$ $\frac{n_2}{n_1 + n_2}$

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 assumption of knowing something (gamma, beta, ...)

good for hypoth. testing, but not estimation

"Jack knife" "Boot Strap"

: QUENUILLE:

- suppose observe iid $x_1, x_2, \dots, x_n$ from some F (cdf)

$\hat{\theta} = (x_1, x_2, \dots, x_n)$

$BIAS(\hat{\theta}) = E_F(\hat{\theta} - \theta)$

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

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

④

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

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

Correct $(\hat{\theta})$ by subtracting the bias:

Corrected estimator $\hat{\theta} - \widehat{Bias} = \hat{\theta} - (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta})$

$= n\hat{\theta} - (n-1)\hat{\theta}_{(\cdot)}$

Estimate μ^2 $\quad Var(\bar{x}) = E(\bar{x}^2) - E^2(x)$

$x_1, x_2, \dots, x_n$ $\quad \bar{x}$ $\quad \frac{\sigma^2}{n} = E(\bar{x}^2) - \mu^2$

$\hat{\mu}^2 = \bar{x}^2$

$E(\hat{\mu}^2) = E(\bar{x}^2) = \mu^2 + \frac{\sigma^2}{n}$ ⊛ so it is biased

---

Excel

Generat x, x^2

Condition on removing observations

$100\,\bar{x}^2 - 99\,\bar{x}_{(\cdot)}^2 \simeq \frac{\sigma^2}{n}$

$\theta = \mu$ $\quad \hat{\theta} = \bar{x}$

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

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

$$= \frac{1}{n}\frac{\sum_{i=1}^{n}[n\bar{x} - x_i]}{(n-1)} = \frac{\sum_{i=1}^{n}\bar{x}}{n-1} - \frac{\sum_i x_i}{n(n-1)}$$

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

Bias $= 0$

$\Rightarrow$ Correction $= 0$

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

$= \hat{\theta}$

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

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

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

$n\sum_{j=1}^{n} x_j^2 - \sum_{j=1}^{n} x_j^2 = (n-1)\sum_i x_i^2$

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

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

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

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

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

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

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

$$\hat\theta = \frac{\sum_i x_i^2}{n} - \bar x^2$$

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

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

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

$$\hat\theta = \hat\theta - \widehat{Bias} = \frac1n\sum_i(x_i-\bar x)^2 + \frac{1}{n(n-1)}\sum_i(x_i-\bar x)^2$$

$$= \frac1n\sum_i(x_i-\bar x)^2\left[1+\frac{1}{n-1}\right] = \frac{\sum_i(x_i-\bar x)^2}{(n-1)}$$

For a lot of estimators the expected value of $\hat\theta$: $E(\hat\theta) = \theta + \frac{-a_1}{n} + \frac{a_2}{n^2} + \dots$

$\mu, \bar x^2$ — a_i dont depend on θ

$E(\bar x)^2 = \mu^2 + \frac{\sigma^2}{n} + 0 \dots$

asymptotically unbiased

$$E(\hat\theta_{(i)}) = \theta + \frac{a_1}{n-1} + \frac{a_2}{(n-1)^2}$$

$$n\hat\theta - (n-1)\hat\theta_{(\cdot)} = n\theta + \cancel{a_1} + \frac{a_2}{n} + \dots - (n-1)\theta - \cancel{a_1} - \frac{a_2}{(n-1)} + \dots$$

$$= \theta + \frac{a_2}{n} - \frac{a_2}{n-1} + \dots$$

$$= \theta - \frac{a_2}{n(n-1)} \leftarrow O\left(\frac{1}{n^2}\right)$$

You can apply to all situations and it works very well

For any Quadratic functional

$\widehat{Bias} = (n-1)(\hat\theta_{(\cdot)} - \hat\theta)$ is unbiased

1958 (Tukey)

Noticed $Var(\bar x) = \frac{\sigma^2}{n}$

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

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

$$\bar x_{(\cdot)} = \frac1n\sum_i \bar x_{(i)} = \frac1n \dots = \bar x$$

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

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

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

for any statistic $\hat\theta$ estimate θ

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

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

robust

estimate $\theta = \log\frac{M_x}{M_y}$ $\quad y \sim U(0,1)$, $x \sim$ expon

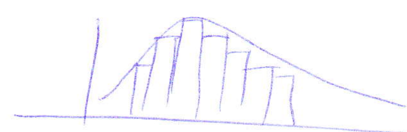
Only problem: not worked median
reason clear, for positional measure
due to data removal doesn't work

Brealey Efron - Boot strap

(Estimating of Variance or Covariance

① Since the empirical c.d.f ($\frac1n$ x_i) is the max likelihood of actual F

① Draw sample of size n from sample density function (with replacement)



② Calculate based on pseudosample

③ repeat step 2, N time (10,000 e.g)

④ Estimate $\hat\theta_{\cdot} = \sum_{b=1}^N \frac{\hat\theta_b}{N}$

$$\widehat{SD} = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^N(\hat\theta_n - \hat\theta_\cdot)^2}$$

For Regression:

$g_i(x_i,\beta) = e^{x_i'\beta}$

$$D = \sum_i |y_i - g_i(x_i,\beta)|$$

First compute LS estimator of β

$\hat e_i = y_i - g_i(x_i,\hat\beta) \Rightarrow e_i^*$ $\quad i = 1,2,\dots,n$
draw from error

- $y_i^* = g_i(x_i, \beta) + e_i^*$
- Construct new estimator $\hat{\beta}_n$

$$\hat{Cov}(\hat{\beta}_b) = \frac{1}{B-1} \sum_{i=1}^{B} (\hat{\beta}_b^* - \hat{\beta}^*)(\hat{\beta}_b^* - \hat{\beta}^*)$$

$$\hat{\beta}^* = \frac{1}{B} \sum_{i=1}^{B} \hat{\beta}_b^*$$

easier than doing bais. almost non parametric
can be used for complex situation
no integration needed

- after estimation { classical, Bayesian, non parametric } that give parameter that are important for hypothesis testing
- hypothesis testing is used for freq, unlike Bayesian that get posterior and interpret

---

Likelihood Ratio test (LRT) $\underset{\sim}{\theta}$ in space Ω

$\omega \epsilon \Omega$

Need μ, σ^2 $\underset{\sim}{\theta} = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$

$\omega \subset \Omega$

$H_N: \mu = \mu_0$ $\begin{pmatrix} \mu_0 \\ \sigma^2 \end{pmatrix}$ = set of all is ω

$H_A: \mu \neq \mu_0$ dim $\omega \leq$ dim ω

$H_N: \underset{\sim}{\theta} \in \omega$ if $H_N: \mu = \mu_0$ & $\sigma^2 = \sigma_0^2$

$H_A: \underset{\sim}{\theta} \in \Omega - \omega$ $\omega = \begin{pmatrix} \mu_0 \\ \sigma_0^2 \end{pmatrix}$ zero dim point

Compute $L(\hat{\theta}_{ML})$ likelihood in Ω

$L(\hat{\theta}_{ML})$ max likelihood in ω

$$LRT = \frac{L(\hat{\underset{\sim}{\theta}}_\omega)}{L(\hat{\underset{\sim}{\theta}}_\Omega)}$$

Reject H_0 if this ratio is too small

$-2[L(\hat{\theta}_\omega) - L(\hat{\theta}_\Omega)]$ Reject when this is too big

↓

logarithm of L

Big result (theorem)

Let $x_1, x_2, \ldots, x_n$ be random sample from some density $P(x, \underset{\sim}{\theta})$, where $P(x, \underset{\sim}{\theta})$ is regular with respect to first and second derivative

a) if $\underset{\sim}{\theta} \in \omega$ (hypothesis is true)

$$\lim_{n \to \infty} -2[L(\hat{\theta}_\omega) - L(\hat{\theta}_\Omega)] \sim \chi^2(k - k')$$

dif

k = dim of Ω
k' = dim of ω

⚠ wrong result when the distribution is not regular

b) if $\underset{\sim}{\theta} \notin \omega$ (hypothesis is false)

$$\forall \alpha \quad \lim_{n \to \infty} P[-2[L(\hat{\theta}_\omega) - L(\hat{\theta}_\Omega)] > \chi^2_{1-\alpha}(k-k')] = 1$$

⊛ Basis for almost every statistical test that is used: mean, proportion

Statistical Inference — March 27 — ①

- sufficient statistic

Assume $x_1, x_2, \dots, x_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$p.d.f.(x) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2}\sum_i \left(\frac{x_i-\mu}{\sigma}\right)^2}$$

notice $\sum_i \left(\frac{x_i-\mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_i (x_i - \bar{x} + \bar{x} - \mu)^2 =$

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

$$p.d.f.(x) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\bar{x})^2 - \frac{n}{2\sigma^2}(\bar{x}-\mu)^2}$$

$\bar{x} \sim N(\mu, \sigma^2/n)$

$$p(\bar{x}) = \frac{1}{\sqrt{2\pi}\,\sigma/\sqrt{n}} e^{-\frac{1}{2}\frac{(\bar{x}-\mu)^2}{\sigma^2/n}}$$

- since $\bar{x}$ is sufficient for μ; $\sum(x_i-\bar{x})^2$ is sufficient for σ^2, $\hat{\mu} = \bar{x}$

$$f(x|\bar{x}) = \frac{f(x \text{ and } \bar{x})}{f(\bar{x})} = \frac{f(x)}{f(\bar{x})} = \frac{\frac{1}{(2\pi)^{n/2}}\sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\bar{x})^2}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(\bar{x}-\mu)^2}}$$

$- \frac{n}{2\sigma^2}(\bar{x}-\mu)^2$

$$f(x|\bar{x}) = \frac{1}{(2\pi)^{\frac{(n-1)}{2}}\sigma^{n-1}} e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\bar{x})^2}$$

does not depend on μ

- $\bar{x}$ has captured all information about μ

Definition: A statistic $T(x)$ is a sufficient (anything derived from data) statistic for θ, if the conditional distribution of the sample given $T(x)$ does not depend on θ

Factorization theorem

Let $f(x|\theta)$ denote p.d.f. of x, the $T(x)$ is sufficient statistic for θ, if there exist a function $g(t|\theta)$ + $h(x)$ such that

$$f(x|\theta) = g(T(x)|\theta)h(x) \quad \forall x, \theta$$

Exponential family of distributions — func of $T(x)$ + θ

$$f(x|\theta) = h(x) \cdot c(\theta) \exp\left[\sum_i w_i(\theta) t_i(x)\right]$$

$\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_p)'$

$$T(x) = \left(\sum_{j=1}^{n} t_1(x_j), \sum_{j=1}^{n} t_2(x_j), \dots, \sum_{j=1}^{n} t_p(x_j)\right)$$

② Example

Let w be an unbiased estimator of θ.

Let T be the sufficient statistic of θ.

$\phi(T) = E(w|T)$

then $\phi(T)$ is also an unbiased estimate of θ

$Var(\phi(T)) \le var(\theta) \quad \forall \theta$

practically means it is carrying information about θ

Big result that we calculated next session χ^2 the regularity condition should hold & is derived from this

---

$x_1, x_2, \dots, x_n \overset{iid}{} x_i \sim N(\mu, \sigma^2)$

$H_1: \mu = \mu_0$

$H_2: \mu \ne \mu_0$

$$L(\mu, \sigma|x) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\frac{1}{2}\sum_i \left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$\Rightarrow L(\mu,\sigma|x) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_i (x_i-\bar{x})^2 - \frac{n}{2\sigma^2}(\bar{x}-\mu)^2$$

$H_0: \mu = \mu_0 \iff \omega \begin{pmatrix}\mu_0 \\ \sigma^2\end{pmatrix}$

$\mu \ne \mu_0 \iff \Omega \begin{pmatrix}\mu \\ \sigma^2\end{pmatrix}$

In Ω: $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \frac{\sum_i (x_i-\bar{x})^2}{n}$

ω: $\hat{\mu} = \mu_0$, $\hat{\sigma}_\omega^2 = \frac{\sum_i (x_i-\mu_0)^2}{n}$

---

max likelihood does not give unbiased estimator; they are asymptotically unbiased

---

$$L_\Omega = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\hat{\sigma}^2_\Omega - \frac{1}{2\hat{\sigma}^2_\Omega}\sum_i (x_i-\bar{x})^2$$

$$= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\hat{\sigma}^2_\Omega - \frac{n}{2}$$

$$L_\omega = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\hat{\sigma}^2_\omega - \frac{n}{2\sigma^2_\omega}\sigma^2_\omega$$

$$\Rightarrow L_\omega - L_\Omega = -\frac{n}{2}\left[\ln\hat{\sigma}^2_\omega - \ln\hat{\sigma}^2_\Omega\right]$$

$$-2\left[L_\omega - L_\Omega\right] = n \ln\frac{\hat{\sigma}^2_\omega}{\hat{\sigma}^2_\Omega}$$

- Reject when its too big

$\Rightarrow$ reject when $\frac{\hat{\sigma}^2_\omega}{\hat{\sigma}^2_\Omega}$ is too big

Reject when $\frac{\frac{1}{n}\left[\sum_i (x_i-\bar{x})^2 + n(\bar{x}-\mu_0)^2\right]}{\frac{1}{n}\sum_i (x_i-\bar{x})^2}$ is too big

Reject when $1 + \frac{n(\bar{x}-\mu_0)^2}{\hat{\sigma}^2_\Omega}$ is too big $\Rightarrow$ Reject when $\frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2_\Omega/n}$ too big means you reject when $\left(\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right)^2$ is too big

t-statistics

* $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ is t with n-1 deg of freedom

* asymptotically it is chisq

Appendix

① assume in sub space → estimate param
② find likelihood
③ $LR = \frac{L(\hat{\theta}_\omega)}{L(\hat{\theta}_\Omega)}$ reject when too small
or take log

- it was two sided test since it was square
in practical life every thing is two sided

$x \sim Bin(n, \pi)$

$H_0: \pi = \pi_0$ $\quad P(x|n,\pi) = \binom{n}{x}\pi^{x}(1-\pi)^{n-x}$

$H_A: \pi \neq \pi_0$ $\quad L(\pi|x,n) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$

$\pi \quad \omega \quad \pi = \pi_0$ zero dimensional

$\Omega \quad 0 \leq \pi \leq 1$

$\hat{\pi}_\Omega \quad L'(\pi) = \frac{x}{\pi} - \frac{(n-x)}{1-\pi}$ $\quad (n-x)\ln(1-\pi)$

$L''(\pi) = \frac{x}{\pi^2} - \frac{(n-x)}{(1-\pi)^2} < 0$

$\Rightarrow \hat{\pi} = x/n$

$L_\Omega = L(\hat{\pi}_\Omega) = n\hat{\pi}\ln\hat{\pi}_A + n(1-\hat{\pi})\ln(1-\hat{\pi}_A) + \ln\binom{n}{x}$

$L_\omega = n\hat{\pi}\ln\pi_N + n(1-\hat{\pi})\ln(1-\pi_N) + \ln\binom{n}{x}$

$L_\omega - L_\Omega = n\hat{\pi}\ln\frac{\pi_0}{\hat{\pi}} + n(1-\hat{\pi})\ln\left[\frac{1-\pi_0}{(1-\hat{\pi})}\right]$

$L_\omega - L_\Omega$

$f(z) = \ln(1+z)$ $\quad f(0) = 0$

$f'(z) = \frac{1}{1+z}$ $\quad f'(0) = 1$

$f''(z) = -\frac{1}{(1+z)^2}$ $\quad f''(0) = -1$

$f'''(z) = \frac{2}{(1+z)^3}$ $\quad f'''(0) = 2$

$\Rightarrow \ln(1+z) = z - z^2/2 + z^3/3 + \dots$

$\ln\left(\frac{\pi_0}{\hat{\pi}}\right) = \ln\left[1 + \underbrace{\frac{(\pi_0-\hat{\pi})}{\hat{\pi}}}_{z}\right]$

$\ln\left(\frac{\pi_0}{\hat{\pi}}\right) \approx \frac{\pi_0-\hat{\pi}}{\hat{\pi}} - \frac{(\pi_0-\hat{\pi})^2}{2\hat{\pi}^2}$

$n\hat{\pi}\ln\left(\frac{\pi_0}{\hat{\pi}}\right) \simeq n(\pi_0-\hat{\pi}) - n\frac{(\pi_0-\hat{\pi})^2}{2\hat{\pi}}$

$\ln\left[\frac{1-\pi_0}{1-\hat{\pi}}\right] = \ln\left(1 + \frac{\hat{\pi}-\pi_0}{1-\hat{\pi}}\right) \simeq \frac{(\hat{\pi}-\pi_0)}{1-\hat{\pi}} - \frac{(\hat{\pi}-\pi_0)^2}{2(1-\hat{\pi})^2}$

$n(1-\hat{\pi})\ln\left(\frac{1-\pi_0}{1-\hat{\pi}}\right) \simeq n(\hat{\pi}-\pi_0) - \frac{n(\hat{\pi}-\pi_0)^2}{2(1-\hat{\pi})}$

$L_\omega - L_\Omega = -n\frac{(\pi_0-\hat{\pi})^2}{2\hat{\pi}} - n\frac{(\hat{\pi}-\pi_0)^2}{2(1-\hat{\pi})}$

$-2[L_\omega - L_\Omega] = n\frac{(\pi_0-\hat{\pi})^2}{\hat{\pi}} + n\frac{(\hat{\pi}-\pi_0)^2}{1-\hat{\pi}}$

$= n(\pi_0-\hat{\pi})^2\left[\frac{1}{\hat{\pi}} + \frac{1}{1-\hat{\pi}}\right] = \frac{n(\pi_0-\hat{\pi})^2}{\hat{\pi}(1-\hat{\pi})}$

$= \left(\frac{\hat{\pi}-\pi_0}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}}\right)^2$ Natural statistics
not one sided, but
two sided due to square

- Every thing you have seen in basic stat course cames from this
- this is powerful theory
- asymptotic theory to find out what stat. use

two independent samples

$x_{11}, \dots, x_{n1}$ $\quad x_{21}, \dots, x_{n2}$

μ_1 $\quad \mu_2$ $\quad$ Normal

σ_1^2 $\quad \sigma_2^2$

$H_0: \mu_1 = \mu_2 = \mu$

$H_A: \mu_1 \neq \mu_2$

$\underset{\sim}{\theta}_\Omega = \begin{pmatrix}\mu_1\\ \mu_2\\ \sigma_1^2\\ \sigma_2^2\end{pmatrix}$ $\quad \underset{\sim}{\theta}_\omega \begin{pmatrix}\mu\\ \mu\\ \sigma_1^2\\ \sigma_2^2\end{pmatrix}$ $\quad x_{1.} = \sum_i \frac{x_{1i}}{n_i}$

dim 4 $\quad$ dim 3

$L(\underset{\sim}{\theta}) = -\frac{n_1}{2}\ln 2\pi - \frac{n_1}{2}\ln\sigma_1^2 - \frac{1}{2\sigma_1^2}\sum_j(x_{1j}-\bar{x}_{1.})^2$

$-\frac{n_1}{2\sigma_1^2}(\bar{x}_{1.}-\mu_1)^2 - \frac{n_2}{2}\ln 2\pi - \frac{n_2}{2}\ln\sigma_2^2 - \frac{1}{2\sigma_2^2}\sum_j$

$(x_{2j}-\bar{x}_{2.})^2 - \frac{n_2}{2\sigma_2^2}(\bar{x}_2-\mu_2)^2$

$\Omega \quad \hat{\mu}_1 = \bar{x}_{1.} \quad \hat{\mu}_2 = \bar{x}_2$

$\hat{\sigma}_1^2 = \frac{\sum_j(x_{1j}-\bar{x}_{1.})^2}{n_1}$ $\quad \hat{\sigma}_2^2 = \frac{\sum_j(x_{2j}-\bar{x}_{2.})^2}{n_2}$

$\omega \quad \mu_1 = \mu_2 = \mu$

$\frac{\partial L(\theta)}{\partial\mu} = \frac{2n_1}{2\sigma_1^2}(\bar{x}_{1.}-\hat{\mu}) + \frac{2n_2}{2\sigma_2^2}(\bar{x}_{2.}-\hat{\mu}) = 0$

Statistical inference march 27 (5)

$$\hat{\mu} = \frac{\frac{n_1 \bar{x}_1}{\hat{\sigma}_1^2} + \frac{n_2 \bar{x}_2}{\hat{\sigma}_2^2}}{\frac{n_1}{\hat{\sigma}_1^2} + \frac{n_2}{\hat{\sigma}_2^2}}$$

$$\hat{\sigma}_{1W} = \frac{1}{n_1} \sum_j (x_{1j} - \hat{\mu})^2$$

Assume You know $\cancel{\sigma_1^2}$ and $\cancel{\sigma_2^2}$

$$\left(\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)^2$$ Standard test

Behner Fisher

Assume $\sigma_1^2 = \sigma_2^2$ GRE different question but same AVG & variation

LR $$\left(\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_{\mu}^2}{n_1} + \frac{\sigma_{\mu}^2}{n_2}}} \right)^2$$

$$\hat{\sigma}_{,\mu m} = \frac{\sum_j (x_{1j} - \bar{x}_j)^2 + \sum_j (x_1 - \bar{x}_1)^2}{(n_1 + n_2 - 2)}$$

t_{α} $(n_1 + n_2 - 2)$

- hypothesis test $\boxed{\text{Appendix}}$

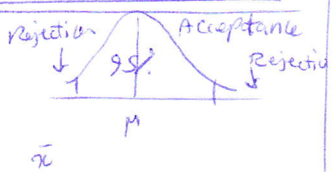
if null hypothisis is true 5% you say it or not

theory : absolutely equal

range of possible value & uncertainty involve

not taken into account

Confidence interval



Suppose you have hypothesis θ_0

$H_0 : \theta \in \omega$

$\theta \in \Omega$

Region $\Rightarrow A(\underline{\theta}_0)$ acceptance region

Have statistics $T(\underline{x})$

if $T(\underline{x}) \in A(\underline{\theta}_0)$ we accept H_0

observed

$C(\underline{X})$ = Confidence interval $= \{ \underline{\theta}_0 : T(x) \in A(\underline{\theta}_0) \}$

- is data Consistent with hyp $\boxed{\text{Appendix}}$ or not
- any other value in Confidendence interval also accept
- Certain product tolerant, reject, how big is the mean

(6)

$$\left| \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} \right| \le t_{\alpha/2}$$

$1 - \alpha$

$\Rightarrow \bar{x} \pm t_{\alpha/2} \hat{\sigma}/\sqrt{n}$ Confid interval μ

Have a random sample $x_1, x_2, \dots, x_n$

we assume there exists a λ such that $z = \frac{x^{\lambda} - 1}{\lambda}$

is Normal (μ, σ^2)

we get a Confidence interval on λ

$H_0 : \lambda = \lambda_0$

$H_a : \lambda \ne \lambda_0$

$$L(\lambda_0) = -\frac{n}{2} \ln \hat{\sigma}^2(\lambda_0) - \sum_i \frac{(z_i(\lambda_0) - \hat{\mu}(\lambda_0))^2}{\hat{\sigma}^2(\lambda_0)}$$

$+ (\lambda_0 - 1) \sum \ln x_i$

$z_i(\lambda_0) = \frac{x_i^{\lambda_0} - 1}{\lambda_0}$ $\hat{\mu}(\lambda_0) = \sum \frac{z_i(\lambda_0)}{n}$

$\hat{\sigma}(\lambda_0) = \sum \frac{z_i(\lambda_0) - \mu(\lambda_0)^2}{n}$

Since dimension is low (1)

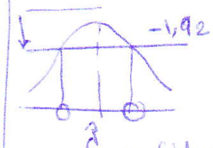
$-2[L(\lambda_0) - L(\hat{\lambda})] \sim X^2(3-2) = X^2_{1\,df}$

Pick $\alpha = .05$ $X^2_{1\,df} = 3.84$

Accept when $-2[L(\lambda_0) - L(\hat{\lambda})] \le 3.84$

$\Rightarrow L(\lambda_0) - L(\hat{\lambda}) \ge -1.92$

$\Rightarrow \{ \lambda_0 \mid L(\lambda_0) \ge L(\hat{\lambda}) - 1.92 \}$ Confidence interval for λ_0



$\ln x = 0$ $\lambda = 0$

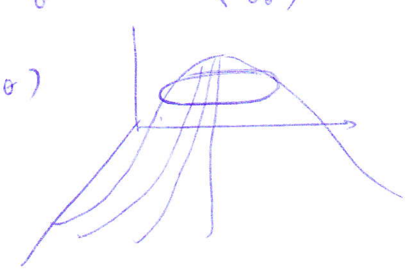
$\ln 1/x = -1$ $\lambda = -1$

-.7 Confidence interval

Gamma dist

Ω $\hat{r}$ $\hat{\theta}$ ω $\binom{r_0}{\theta_0}$

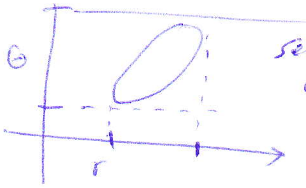
$L(r, \theta)$



$-2[L(r_0, \theta_0) - L(\hat{r}, \hat{\theta})] < X^2_{.95}($ 5.99

set of all $\binom{r_0}{\theta_0}$

$L(r_0, \theta_0) > L(\hat{r}, \hat{\theta}) - 3$



sets oval

Bayesian says, I don't know Confidence interval
I have whole distribution

(.95)
(1-α) $g(\theta \mid \underline{x})$ on bayesian posterior prediction

$H_N: \mu_1 = \mu_2 = \mu_3 = \mu \iff \mu_1 = \mu_2$
$\mu_1 = \mu_3$
$\mu_2 = \mu_3$

$H_N: \mu_1 = \mu_2 = \cdots = \mu_n \Rightarrow \frac{n(n-1)}{2}$
$(\mu_i - \mu_j)$

- Confidence interval a to b when (a, b) does not include '0', so reject excel

---

	No difference	H_0 is true
Test 1	$1-\alpha$	α
Test 2	$1-\alpha$	α

what is the chance that we got both true?
$(1-\alpha)^2$
$\Rightarrow$ chance of at least one positive $1-(1-\alpha)^2$

For three test $1-(1-\alpha)^3$

3 → 0,420
6 → 0,265
10 → 0,40
→ lot's of significant result when nothing is going on

α_E experiment wide Error rate
α_I $\alpha_E = 1-(1-\alpha_I)^k$

$\Rightarrow (1-\alpha_I)^3 = 0.95 \Rightarrow 1-\alpha_I = \sqrt[3]{0,95}$
$\Rightarrow \alpha_I = 1-{,95}^{1/3} \Rightarrow \boxed{\alpha = 0,0169}$

- This method gives you that even b/w slow & fast music d

- Not right due to assumption of anovance

Bonfroni:

Recall $P(\bigcap A_i^c) = P((\bigcup A_i)^c) = 1 - P(\bigcup A_i)$
$\Rightarrow P(\bigcup A_i) = 1 - P(\bigcap A_i^c)$
know $P(\bigcup A_i) \leq \sum_i P(A_i)$
$\Rightarrow 1 - P(\bigcap A_i^c) \leq \sum_i P(A_i)$

Let A_i = {reject null hypothesis when true for the i'th Variable}
$\Rightarrow A_i^c$ = {Correctly accepting true null hypothesis
$P(\bigcap A_i^c)$ = {Correctly accepting all the null hypoth}
$1 - P(\bigcap A_i^c) = P_r$(rejecting at least one true null hypothesis)
α_E — test A_i at α_i
$\alpha_E \leq \sum_i P(A_i)$
$\alpha \leq \sum \alpha_i$ if you do p test each at $\alpha_i = \alpha$
$\alpha_E \leq p\,\alpha_I$

Bonfroni approach: take $\alpha_I = \frac{\alpha_E}{p}$ $\boxed{0.5/3 = 0,167}$
p = # of hypothesis you will test

- Next week on ANOVA

Statistics April 03 ①

Bonfroni approach

- $\frac{\alpha}{n}$ replaced with α where n is number of Comparison.
- it cuts down number of results many result is positive ⇒ false positive
- at the Price of missing significant result when it is there

Holm - Bonfroni

- the same α : chance of making any mistake
- β prob missing sth when it is there

k tests

- Compute the p-value for each test
  p-value = Prob of the observed result or anything more extreme
- max likelih two sided not one sided

P_i

1) ordered the $P_i \to P_{(i)}$ smallest to largest
2) check if $P_{(\cdot)} \le \frac{\alpha}{k}$
   (reject hyp. of Bonfronie)
   I (don't reject) accept all hypothesis
   otherwise reject (i) → (i)+1
3) check if $P_{(2)} \le \frac{\alpha}{k-1}$
   if (not) accept hypothesis (2) ... k)
   otherwise reject
4) check if $p(3) \le \frac{\alpha}{k-2}$ etc.

Excel slow vs. fast vs. none
3 p-values, put them in order from smallest to largest
① check smallest $\frac{.05}{3}$ reject null
② check second smallest $\frac{.05}{2} = 0.025$ reject
③ check largest $0.05 \Rightarrow$ reject due to lower p-val

Bonfronie checkes all against 0.01666 and so does not reject the ③ since greater than this amount

k Groups $1, 2, \dots, k = i$

$x_{ij} = \mu_i + e_{ij}$ $\quad$ j applicant

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

②

key assump in ANOVA (analysis of variance)
all group same variance

$$\mathrm{Cov}(e_{ij}, e_{i'j'}) = \begin{cases} \sigma^2 & \text{if } i=i',\ j=j' \\ 0 & \text{otherwise} \end{cases} \quad \text{means indep}$$

$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_n \end{pmatrix}$

$$L(\mu, \sigma^2) = \frac{-n_+}{2} \ln 2\pi - \frac{n_+}{2} \ln \sigma^2 \cancel{- \frac{1}{2\sigma^2} \sum_i \sum_j (x_{ij} - \mu_i)^2}$$

$$\boxed{n_+ = \sum_i n_i \qquad x_{i.} = \sum_j \frac{x_{ij}}{n_i}}$$

$\to \frac{n_+}{2}$

$$-\frac{1}{2\sigma^2} \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 - \frac{1}{2\sigma^2} \sum_i n_i (\bar{x}_{i.} - \mu_i)^2$$

Ω $\quad \hat{\mu}_i = \bar{x}_{i.} \quad \hat{\sigma}^2_\Omega = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2}{n_+}$

$$L_\Omega = \frac{-n_+}{2} \ln 2\pi - \frac{n_+}{2} \ln \hat{\sigma}^2_\Omega - \frac{n_+}{2}$$

$H_0 : \mu_1 = \mu_2 = \dots = \mu_n = \mu$
H_A : at least one per μ_i, μ_j different

Let $c_1, c_2, \dots, c_k$ be values such that $\sum_i c_i = 0$ Contrast Vector

1	1	1	0	$\frac{1}{2}$
2	-1	0	1	$\frac{1}{2}$
3	0	-1	-1	$-\frac{1}{2}$

→ infinit Contrast Vector

$\mu_1 - \mu_2$ $\quad$ $\mu_1 - \mu_3$ $\quad$ $\mu_2 - \mu_3$ $\quad$ $(\frac{\mu_1 + \mu_2}{2}) - \mu_3$

$\sum_i c_i \mu_i$ $\quad$ Contrast of Group means

H_A : $\exists$ A Contract $c_1, c_2, \dots c_n$ which is different from zero
some combination of mean which is not zero

ω : $\mu_1 = \mu_2 = \dots = \mu_k = \mu$

$$L(\mu, \sigma^2) = -\frac{n_+}{2} \ln 2\pi - \frac{n_+}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2$$
$$- \frac{1}{2\sigma^2} \sum_i n_i (\bar{x}_{i.} - \mu)^2$$

$$\hat{\sigma}^2_\omega = \frac{\sum_i \sum_j (x_{ij} - \hat{\mu})^2}{n_+} \qquad \hat{\mu} = \frac{\sum_i n_i \bar{x}_{i.}}{n_+} = \bar{\bar{x}}$$

$$L_\omega = -\frac{n_+}{2} \ln 2\pi - \frac{n_+}{2} \ln \hat{\sigma}^2_\omega - \frac{1}{2\hat{\sigma}^2_\omega} \sum_i \sum_j (x_{ij} - \hat{\mu})^2$$
$$= -\frac{n_+}{2} \ln 2\pi - \frac{n_+}{2} \ln \hat{\sigma}^2_\omega - \frac{n_+}{2}$$

$$-2[L_\omega - L_\Omega] = -[\cancel{n_+} \ln 2\pi - n_+ \ln \hat{\sigma}^2_\omega + \frac{n_+}{2}]$$
$$+ n_+ \ln 2\pi + n_+ \ln \hat{\sigma}^2_\Omega + \frac{n_+}{2}$$
$$= n_+ \ln \left(\frac{\hat{\sigma}^2_\omega}{\hat{\sigma}^2_\Omega}\right)$$

reject if this is big

$\Rightarrow$ Reject $\frac{\hat{\sigma}^2_\omega}{\hat{\sigma}^2_\Omega}$ is big

$$\Rightarrow \frac{\sum_i (x_{ij} - \bar{x}_{ij})^2 + \sum n_i (\bar{x}_{i.} - \bar{\bar{x}})^2}{\sum n_i (x_{ij} - \bar{x}_{ij})^2} \text{ is big}$$

Reject when

$$\frac{\sum_i n_i(\bar{x}_{i.}-\bar{\bar{x}})^2}{\sum_i\sum_j (x_{ij}-\bar{x}_{i.})^2} \text{ is large}$$

Fisher for the moment assume $n_i = n \ \forall i$)

$$\begin{array}{cccc} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & & x_{kn} \\ \hline \bar{x}_{1.} & \bar{x}_{2.} & \cdots & \bar{x}_{k.} \\ s_1^2 & s_2^2 & \cdots & s_n^2 \end{array} \quad \hat{\sigma}^2 \quad \text{assume common variance}$$

$$s_i^2 = \frac{\sum_i (x_{ij}-\bar{x}_{i.})^2}{n_i - 1}$$

will (I) $\hat{\sigma}_w^2 = \frac{\sum_i (n_i-1) s_i^2}{(n_+ - k)}$

CLT $\bar{x} \sim N(\mu, \sigma^2/n)$ $\quad \bar{\bar{x}}$

$$\frac{\sum_i (\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{\hat{\sigma}^2}{n} \Rightarrow n\frac{\sum_i(\bar{x}_i - \bar{\bar{x}})}{K-1} = \hat{\sigma}_B^2)$$

Between estim

(II) $\hat{\sigma}_B^2 = \frac{\sum_i n (\bar{x}_{i.} - \bar{\bar{x}})^2}{k-1}$

two estimator at same type

Fisher	H_0 is true	H_0 is false
$E(\hat{\sigma}_w^2)$ (I)	σ^2	σ^2
$E(\hat{\sigma}_B^2)$ (II)	σ^2	$\sigma^2 + \frac{\sum_i n_i(\mu_i - \bar{\bar{\mu}})^2}{k-1}$ $\bar{\bar{\mu}} = \frac{\sum_i n_i x_i}{n_+}$

$F_{ratio} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_w^2}$

↳ after Fisher

⇒ if F_{ratio} is large reject it

we are looking at same ratio of likelihood of large rejection

F - distribution

Table

Src	df	SS	MS	P-value
Between	$k-1$	$\sum_i n_i(\bar{x}_{i.}-\bar{\bar{x}})^2$	$SS_B/df = \hat{\sigma}_B^2$	
within	$\sum_i n_i - k$	$\sum_i\sum_j(x_{ij}-\bar{x}_{i.})^2$	$SS_w/d. = \hat{\sigma}_w^2$	
Total	$n_+ - 1$	$\sum_i\sum_j (x_{ij} - \bar{\bar{x}})^2$		

③

④ Factor music → 3 level (slow-fast, no music)

2 factor → week day afternoon night

replication → observation per cell
there equal n obs = 12

	df
b/w group	2
within G	12

ANOVA - if accept nothing going on
if reject means something going on

it is α_E and not indiv comparison
Check with Fisher

in our example

lots of method to find what is going on

- Fisher modified LSD test has best power by simulation

Least significant Difference method

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

$df = \sum_i n_i - k$

Small probability that none is different

	diff b/w mean	+/- figure
Fast vs none	2474.6	$t_{\alpha/2}\sqrt{2MS}$
Fast vs slow	Fast-slow	
slow vs fast	1477	

Put at same level as F test → α_E
.05 e.g.

= tinv(.05, 12) = 2.178813

low high
+ −

0 A result in Confidence interval ⟺ (+/-) < diff

This was analysis of variance

$y_{ij} = \mu + \delta_i + e_{ij}$ $\quad e_{ij} \sim N(0, \sigma_e^2)$

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

decompose $\sum_i\sum_j (y_{ij} - \mu - \delta_i)^2$

$= \sum_i\sum_j (y_{ij} - \bar{y}_{ij})^2 + \sum_i n_i(\bar{y}_{i.} - \mu - \delta_i)^2$

$E(y_{ij} - \bar{y}_{i.})^2 = E(\mu + \delta_i + e_{ij} - \mu - \delta_i - \bar{e}_{ij})^2$

$= E(e_{ij} - \bar{e}_{i.})^2$ $\quad \bar{e}_{i.} = \sum_j \frac{e_{ij}}{n_i}$

$= Var(e_{ij}) + Var(\bar{e}_{ij}) - 2Cov(e_{ij}, \bar{e}_{i.})$

$= \sigma^2 + \frac{\sigma^2}{n_i} - \frac{2\sigma^2}{n_i} = \sigma^2 - \frac{\sigma^2}{n_i}$

Statistics ⑤ April 03

$\sum_j E(y_{ij} - \bar{y}_{i.})^2 = n_i\sigma^2 - \sigma^2 = \sigma^2(n_i - 1)$

$$\boxed{\sum_i \sum_j E(y_{ij} - \bar{y}_{i.})^2 = \sigma^2 \sum_i (n_i - 1)}$$

our nex

- Suff stat $y_{i.} - \mu - \delta_i$ means $\longrightarrow$ $\underline{\underline{k}}$ means

$y_{ij} = \mu + \alpha_i + e_{ij}$ $\quad \boxed{\sum \alpha_i = 0}$ ← Some restriction to identity ↓ $\boxed{\sum_i n_i \alpha_i = 0}$

$\Rightarrow \hat{\mu} = \bar{\bar{y}} \quad \hat{\alpha}_i = \bar{y}_{i.} - \bar{\bar{y}}_{..}$

$E(\sum_i n_i(\bar{y}_{i.} - \bar{y}_{..})^2)$

$E(\bar{y}_{i.} - \bar{y}_{..})^2 = E(\mu + \alpha_i + \bar{e}_{i.} - \mu - \bar{e}_{..})$ // $\bar{e}_{..}$

$\bar{y}_{..} = \frac{\sum_i n_i \bar{y}_{i.}}{n_+} = \sum_i \frac{n_i(\mu + \alpha_i + \bar{e}_{i.})}{n_+} = \mu + 0 + \frac{\sum n_i \bar{e}_{i.}}{n_+}$

$= E(\alpha_i + \bar{e}_{i.} - \bar{e}_{..})^2 = \alpha_i^2 + E(\bar{e}_{i.} - \bar{e}_{..})^2 = \alpha_i^2$

$+ Var(\bar{e}_{i.}) + Var(\bar{e}_{..}) - 2Cov(\bar{e}_{i.}, \bar{e}_{..})$

$= \cancel{\alpha_i^2} + \frac{\sigma^2}{n_i} + \frac{\sigma^2 n_i^2}{(n_+)^2 \cdot n_i} - \frac{2n_i\sigma^2}{(n_+)^2}$

$Var(\sum_i \frac{n_i \bar{e}_{i.}}{n_+}) = \sum_i \sum_j$

$\boxed{\alpha_i^2 + \frac{\sigma^2}{n_i} - \frac{\sigma^2}{n_+}}$ ⟶ should be result

$E(n_i(\bar{y}_{i.} - \bar{y}_{..})^2) = n_i\alpha_i^2 + \sigma_e^2 - n_i\sigma_e^2/n_+$

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

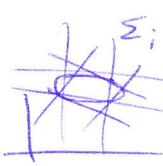
Source	df	SS	MS	E(MS)
within	$n_+ - K$	$\sum_i\sum_j(y_{ij} - \bar{y}_{..})^2$	SS/df	$\frac{\sigma^2(n_+ - k)}{(n_+ - k)} = \sigma^2$
Betw	$n_+ - k$	$\sum_i n_i(\bar{y}_{i.} - \bar{y}_{..})^2$		$\sigma^2 + \frac{1}{n-1}\sum_i n_i\alpha_i^2$

$\Rightarrow \alpha_i = 0$

$E(\bar{y}_{i.}) = \mu + \alpha_i = 0$

$H_0: \sum_i c_i\mu_i = 0 \qquad \sum_i c_i = 0$

$\sum_i c_i \bar{y}_{i.} \qquad Var(\sum_i c_i \bar{y}_{i.}) = \sum_i c_i^2 \frac{\sigma^2}{n_i}$



look at diff Contrast diffent parallell line

it do elepsis in space

$$\frac{\sum_i c_i \bar{y}_{i.}}{\sqrt{\sum_i c_i^2 \frac{\sigma^2}{n_i}}} \le \sqrt{(k-1)F_{1-\alpha}(k-1, n_+ - k)} \quad \forall \sum c_i$$

deg freedom — all Contrast

Shfey limit formula, but should not be used since lot of them insignificant

Contrast that maximize this ration ⑥

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

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

this is Contrast that maximizes F-stat

$\bar{x}_{non} = 12,564,8$

$\bar{x}_{slow} = 14 \qquad \bar{x}.\ 13,882$

$\bar{x}_{fast} = 15,47,4$

$c_1 \propto -1317.2 \qquad -1$

$c_2 \propto 1591 \qquad .87$

$c_3 \propto 1157,40 \qquad .123$

Max Contrast fast no music

$0.12\mu_2 + .87\mu_3 - \mu_1$ = max Contrast in music

slow

$\boxed{\text{Fixed effect}}$ ⟶ when groups are different

Variability { ① inherent mk product ② b/w workers

just how var

not interest in AVG, just if there is diff

$\Rightarrow$ train uniter if there is $\Rightarrow$ product les variable high quality

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

$e_{ij} \sim N(0, \sigma^2) \qquad Cov(e_{ij}, e_{i'j'}) = \begin{cases} 1 & i=i', j=j' \\ 0 & \text{oth} \end{cases}$

$a_i \sim N(0, \sigma_a^2) \qquad Cov(a_i, a_{i'}) = \begin{cases} \sigma_a^2 & \text{if } i=i', j=j' \\ 0 & \text{other} \end{cases}$

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

$H_N: \sigma_a^2 = 0$?

source	df	SS	MS	E(MS)
Between	Same as before			$\sigma_e^2 + \frac{\sigma_a^2}{(n-1)}\left[n_+ - \frac{\sum n_i^2}{n_+}\right]$
within				σ_e^2

if all $n_i = n$ $\qquad \left. \begin{matrix} \sigma_e^2 + n\sigma_a^2 \\ \sigma_e^2 \end{matrix} \right\}$

Random effect if null central F

$\Rightarrow$ different dist based on whether { Random / Fixed

mixed model ⟶ some Random, some Fixed

repeated measure design

Diet – befor after on same people

metabolism exercise

people walk more when feel better

→ 6 week befor – 12 week later

j = time β_j fixed (7)
i = person a_i → person, random

$y_{ij} = \mu + a_i + \beta_j + e_{ij}$ $\sum_j \beta_j = 0$

Some restriction to get estimability

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

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

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

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

- these observations no longer correlated

$Cov(y_{ij} - y_{i'j})$ $i \neq i'$ $= 0$

$Cov(y_{ij}, y_{ij'}) = Cov(\mu + a_i + \beta_j, \mu + a_i + \beta_{j'} + e_{ij'})$

$= \sigma_a^2$ $j = 1, 2, \ldots, J$, $i = 1, 2, \ldots, I$

Source	df	SS	ms
rows (random)	$(I-1)$	$J \sum (\bar{y}_{i.} - \bar{y}_{..})^2$	✓
Column (fixed)	$(J-1)$	$I \sum_{j=1}^{J} (\bar{y}_{.j} - \bar{y}_{..})^2$	✓
Error term	$(I-1)(J-1)$	$\sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	✓

	E(ms)
Rows	$\sigma_e^2 + J\sigma_a^2$
Column	$\sigma_e^2 + \frac{I}{J-1} \sum \beta_j^2$
Err. term	σ_e^2

Reject Column effect ⇒ differences in Column

$\bar{y}_{.j} - \bar{y}_{.j'} = \cancel{\mu} + \cancel{a} + \beta_j + \bar{e}_{.j} - \cancel{\mu} - \cancel{a} - \beta_{j'} - \bar{e}_{.j'}$

$y_{ij} = \mu + a_i + \beta_j + e_{ij}$ $= (\beta_j - \beta_{j'}) + \bar{e}_{.j} - \bar{e}_{.j'}$

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

$(\bar{y}_{.j} - \bar{y}_{.j'}) \pm t_{\alpha/2} \sqrt{\frac{2\sigma_e^2}{I}}$ d.f $(I-1)(J-1)$

Multivariate analysis

$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \underline{x}$ $f(\underline{x}) = \frac{1}{(2\pi)^{3/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})}$

$\underline{x} \sim MVN(\underline{\mu}, \Sigma)$

$L(\underline{\mu}, \Sigma) | x_1, x_2, \ldots, x_p$ $\hat{\underline{\mu}} = \sum_{i=1}^{n} \frac{\underline{x}_i}{n} = \bar{\underline{x}}$

$\hat{\Sigma} = S = \frac{\sum_{i=1}^{n} (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})'}{n-1}$

$H_0: \underline{\mu} = \underline{\mu}_0$

$-2(\ln \omega - \ln \Omega) = n(\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0) = T^2_{obs}$

multivariate t test

$p = 1$ $n(\bar{x} - \mu_0) \cdot \frac{1}{s^2} (\bar{x} - \mu_N) = \frac{(\bar{x} - \mu_N)^2}{s^2_{\bar{x}}}$ (8)

accep $H_N: \underline{\mu} = \underline{\mu}_0$

$T^2_{obs} \leq \frac{p(n-1)}{n-p} F_{\alpha, p, n-p}$

$H_N: \underline{a}'\underline{\mu}$

$T_{\alpha, p, n-p} = \sqrt{\frac{p(n-1)}{n-p} F_{\alpha, p, n-p}}$

$\underline{a}'\underline{\mu}$ $\sqrt{\frac{a'Sa}{N}}$

$\underline{a}'\bar{\underline{x}} - T_{\alpha, p, n-p} \leq \underline{a}'\underline{\mu} \leq \underline{a}'\bar{\underline{x}} + \sqrt{\frac{1}{n} a'Sa} \cdot T_{\alpha, p, n-p}$

$\forall a$ is always true

$\underline{x} \sim MVN(\underline{\mu}, \Sigma)$

$H_N: \underline{\mu} = \mu \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ means every same mean

$H_A: \underline{\mu} \neq \mu \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

$\underset{I-1 \times I}{C} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ 0 & & & & \end{pmatrix}$

$\underline{y} = C\underline{x} \sim MVN(C\underline{\mu}, C\Sigma C')$

$n(C\bar{x} - \underline{0})(CSC')^{-1}(C\bar{x}) \leq \frac{(I-1)(n-1)}{n-I+1} F_{\alpha, a-1, n-I+1}$

- Constant variability assumption
basic assumption of same variance
test but may be pattern in it

next time - interaction
- regression
- time series
- logistic
- non linear

Contrast vector

$\sum_{j=1}^{s} c_j \bar{y}_j \pm \sqrt{\frac{c'Sc}{n}} \sqrt{\frac{(p-1)(n-1)}{(n-p+1)} F_{1-\alpha}(p, n-p+1)}$

Statistical inference April 10th ①

- Couple of things on ANOVA

INTeraction

- Two way fixed ANOVA with multiple observations in each cell
- unlike situation in one way (don't care # of obs in cell equal) $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$; $k = 1, 2, \dots, K$

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

Rows, column, Cell

e_{ijk} i.i.d $\sim N(0, \sigma_e^2)$

unestimable model $\Rightarrow$ need restriction

$$\Sigma_i \alpha_i = \Sigma_j \beta_j = \Sigma_i (\alpha\beta)_{ij} = \Sigma_j (\alpha\beta)_{ij} = 0$$

idea: interaction

α_i : Row effect

α_j : Column effect

Compare $\bar{y}_{ij.}$ to $\bar{y}_{ij'.}$

all obs per cell

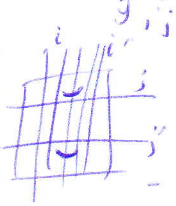
$$i: \bar{y}_{ij.} - \bar{y}_{ij'.} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \bar{e}_{ij.} - \mu - \alpha_i - \beta_j' - (\alpha\beta)_{ij'} - \bar{e}_{ij'.} = (\beta_j - \beta_{j'}) + [(\alpha\beta)_{ij} - (\alpha\beta)_{ij'}] + [\bar{e}_{ij.} - \bar{e}_{ij'.}]$$

difference b/w inter entry of cells

diff b/w interaction

if went to other row

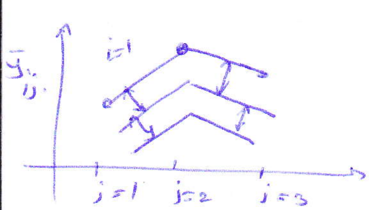
$$\bar{y}_{i'j} - \bar{y}_{i'j'} = (\beta_j - \beta_{j'}) + [(\alpha\beta)_{i'j} - (\alpha\beta)_{i'j'}] + \dots$$



not same diff due to interaction

if no ~~difference~~ interaction identical difference

- no interaction corresponding to the generalized parallelism



No interaction

- diff b/w row not depend on row, and row diff not depend on Column

if diff $\rightarrow$ interaction

		df
- Sum of square for rows	$KJ\Sigma_i (\bar{y}_{i..} - \bar{y}_{...})^2$	$(I-1)$
SS col	$KI \Sigma_j (\bar{y}_{.j.} - \bar{y}_{...})^2$	$(J-1)$
SS interaction	$= K \Sigma_i \Sigma_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{...})^2$	$(I-1)(J-1)$
SS error	$= \Sigma_i \Sigma_j \Sigma_k (y_{ijk} - \bar{y}_{ij.})^2$	$IJ(K-1)$

②

under the restrictions

$\hat{\mu} = \bar{y}_{...}$

$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...}$

$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$

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

non Central F

Source	E(MS)	
Rows	$\sigma_e^2 + \frac{KJ\Sigma_i \alpha_i^2}{I-1}$	$\frac{RMS}{\text{Err Mean Squr}} \sim F(I-1, IJ(K-1))$
Columns	$\sigma_e^2 + \frac{\Sigma_j \beta_j^2}{J-1}$	$\frac{CMS}{\text{Err mean sq}} = F(J, IJ(K-1))$
Interaction	$\sigma_e^2 + \frac{K \Sigma_i \Sigma_j (\alpha\beta)_{ij}^2}{(I-1)(J-1)}$	$\frac{\text{Intact MS}}{\text{Err}} \sim F((I-1)(J-1), IJ(K-1))$
Error	σ_e^2	

what iff err random

$$y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk}$$

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

n

	df		E(M)
Rows	$(I-1)$	2	$\sigma_e^2 + k\sigma_{ab}^2 + Jk\sigma_a^2$
Columns	$(J-1)$	3	$\sigma_e^2 + k\sigma_{ab}^2 + Ik\sigma_b^2$
Interaction	$(I-1)(J-1)$	6	$\sigma_e^2 + k\sigma_{ab}^2$ ①
Err	$IJ(K-1)$	48	σ_e^2

$I = 3$, $J = 4$, $k = 5$

Sensitivity increase when denom. df high $\Rightarrow$ for 2, 3 $\Rightarrow$ easy to miss

① $H_0: \sigma_{ab}^2 = 0$
$H_A: \sigma_{ab}^2 \neq 0$

$\Rightarrow$ $\boxed{\text{pooling will happen} \Rightarrow 6 + 48 = 54 = \text{add ssq/df} \quad \text{both} = \sigma_e^2}$

$\Rightarrow$ test hyp from bottom up: pool to increase degree of freedom up to find things

to test this you need to: $\frac{MS(INT)}{MS(Error)}$

$$MS(\text{rows}) \sim (\sigma_e^2 + k\sigma_{ab}^2 + Jk\sigma_a^2) \frac{\chi^2(I-1)}{I-1}$$

$$MS(\text{Column}) \sim (\sigma_e^2 + k\sigma_{ab}^2 + IK\sigma_b^2) \frac{\chi^2(J-1)}{(J-1)}$$

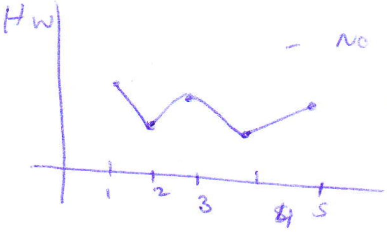
$$MS(\text{Interact}) \sim (\sigma_e^2 + k\sigma_{ab}^2) \frac{\chi^2(I-1)(J-1)}{(I-1)(J-1)}$$

$$MS(\text{Error}) \sim \sigma_e^2 \frac{\chi^2(IJ(k-1))}{IJ(k-1)}$$

~~non Central~~ chi sq $\rightarrow$ no non Centrality

- so You can not divide [Error] unless they are fixed
so you have to go back to chi-sq

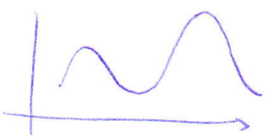
ANOVA very non parametric



- No assump about func. form
gives you some idea about response func

- model helps to understand things, so we want to use it
  - Father, son, central tendency / Converge, Regress
  - Random makes things stable

Regression



$\bar{x} = \frac{H+L}{2}$ $\quad 65 \le \bar{x} \le 70$ $\quad$ Degree day DD = 0

$\bar{x} \ge 70$ $\quad DD = \bar{x} - 70$

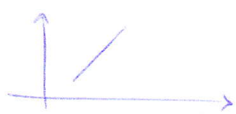
$\bar{x} \le 65$ $\quad DD = 65 - \bar{x}$

DD ~ KWH

- anything you try to predict is y

$y = \beta_0 + \beta_1 x + e$ $\quad e \sim N(0, \sigma^2)$

$E(y|x)$



$y = \beta_0 + \beta_1 e^{x^2/2}$

$y = \beta_0 + \beta_1 g(x) + e$

$h(y) = \beta_0 + \beta_1 g(x) + e$

- plot against log straight line
- transformation of y and x should be on line

$\frac{y^\lambda - 1}{\lambda}$ $\quad x \quad y$

$\sqrt{x} \quad \sqrt{y}$

$\ln x \quad \ln y$

$1/x \quad 1/y$

- linear should be relation of unknown to known

$y = \beta_0 + \beta_1 \ln x + \beta_2/x + e$ $\quad$ still linear

$y = \Sigma$ known · unknown

$y = \beta_0 + \beta_1 \cdot e^x + e$

next: fit model

$e \sim N(0, \sigma_e^2)$

Max likelihood $\quad \min_{\beta_0, \beta_1} \Sigma_i (y_i - \beta_0 - \beta_1 x_i)^2$

minimize vertical distance

- You can do minimize horizontal, why not?
due to need for prediction

$\hat{e}$ = act - pred $= y_i - \hat{y}_i$ $\Big|$ $\min \frac{\Sigma_i (y_i - \beta_0 - \beta_1 x_i)^2}{\beta_1^2}$ $\quad$ Horizontal

Closest distance

$\min_{\beta_0, \beta_1} \frac{\Sigma_i (y_i - \beta_0 - \beta_1 x_i)^2}{1 + \beta_1^2}$ $\quad$ vertical distance
min distance key

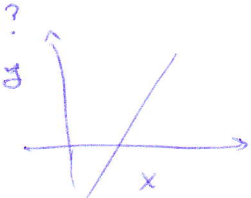
intercept: if not use electricity for heat, use it

total cost = Fixed cost + Variable cost · x

FC < 0 → somebody paying?

data not close to zero

usually data close to zero, sensible intercept



Ask: Does x help me to predict y?

Find the model using x

Residual Sum sq (using x) $= \Sigma_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x)^2$

then say suppose you don't use x.

- if just y → use mean

Don't use x → $\hat{y}_i = \bar{y}_R + 0x + e$

$TSS = \Sigma_i (y_i - \bar{y})^2$

$0 \le MRSS \le TSS$ (residual) $\quad 0 \le \frac{MRSS}{TSS} \le 1$

$0 \le 1 - \frac{RSS}{TSS} \le 1$

adjusted $R^2 = 1 - \frac{RSS/(n-2)}{TSS/(n-1)}$ ⟷ R^2 using degree of freedom

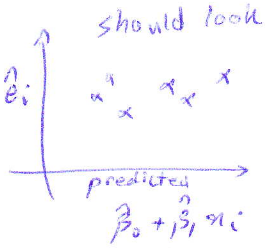
R^2 = proportion of variability in y eliminated by using x as a predictor

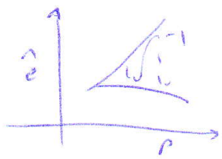
0.85 → .15 unexplained by response

- outlier → study and clean up if needed
Model apply to majority of people

Residual plot

should look Random skatter





lower

think percentage

look logarithm

fan in

Normal prob plot of

$\hat{e}_i$ vs predicted

- Normal probability plot
- Forecast and see how it works → summer 25%

error rate

multiple Regression

- Normalize data first so that Computer don't overflow

$y = x\beta + e$ $\quad y = \beta' x = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$

$C.S.\ E = (x'x)^{-1} x'y$ $\quad \begin{cases} x_1 \to \frac{x_1}{10} \\ \beta_1 \to 10\beta_1 \end{cases}$

- two dimensions

relationship that Can't see

- still plot to find out outliers e.g New York way higher rest of points

multi y : multi variate regression

- plot crime rate per independent just to find outliers

3 basic steps reg

1. backward Elimination
2. add Variable Forward selection
3. stepwise

Start everything & drop things that don't need

- if take intercept out dist R^2 will screw out

leave it in

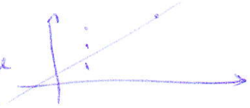
- don't look at Coefficients, you Can modify by divide
- p-values → Can't do that multicoll says you don't need if you have other
  - ↳ R-sq, F-stat says relation
- Collin → Can do, but not Come up with easy inter-Pretation
- P-val : you don't need one if there other

- What Can: Remove the highest p-value

Backward: generate p-value

Keep going either all left with lower .05 or elemin everything

- ask software to see steps.
- next step: residual plot
- when not random add square
- influential point one point determine slope - at least put half half



random residuals

- residual plot for every variable
- now hard part: have to explain it

model:

Crime rate $= 66.551 + 1.9986 \times$ Area $- 1.369 \times$ (90765)

↳ larger area, the harder to maintain crime (plausible)

No data close to 0

old people knows what going on operation crime watch

$+ 8.6791 \times \frac{\#\ doctor}{1000}$ → doctor go to places with more money

- statistics → relationship and not causality

$- 2.408 \times \frac{\#\ hospital}{1000}$ → politicians social service not seek to need money to rob

- in Regression don't through things out ; don't decide ahead on through things out

another parameter : $\hat{\sigma}_e$

$$\boxed{\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\frac{\alpha}{2}}(n-1)\, \hat{\sigma}_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}}$$

± 2

→ value trying to predict

- People usually skip plots but it is easy and so you need to try it

next week

- –
- logit
- log linear

multivar reg Could be done

MANOVA

multiv t-test

⋮

Counted data

- Regression continuous
- multiple categories last session
- now many observations

$\hat{p}_1, \hat{p}_2, \dots, \hat{p}_m$ proportions

$x_1, x_2, \dots, x_k$

$\hat{p}_i = \beta_0 + \Sigma_i \beta_i x_{ij} + e_{ij}$

y_i = # of successes at x_i

n_i = total # of cases

$p(x_i) = \frac{e^{x_i'}}{1 + e^{x_i'\beta}}$ $\qquad \hat{p}(x_i) = \frac{y_i}{n_i}$

$\ln\left[\frac{\hat{p}(x_i)}{1-\hat{p}(x_i)}\right] = x_i'\beta$

$f(p) = \ln p - \ln(1-p) = \ln\left(\frac{p}{1-p}\right)$

$f'(p) = \frac{1}{p} - \frac{1}{(1-p)} = \frac{1}{p(1-p)}$

$Var\left(\ln\frac{\hat{p}}{1-\hat{p}}\right) = \left[\frac{1}{\hat{p}(1-\hat{p})}\right]^2 \cdot \frac{\hat{p}(1-\hat{p})}{n_i}$

$= \frac{1}{n_i \hat{p}_i (1-\hat{p}_i)}$

Regression on the logit : weighted regression inversely proportional to var

$w_i = n_i \hat{p}_i (1-\hat{p}_i)$

---

sometimes n_i counted data as well

PICNIC
{ CRAB MEAT
{ POTATO SALAD (PS)

CRAB MEAT Yes

	PS YES	PS NO	
ILL	120	4	124
Not ill	80	31	111
	200	35	235

CRAB MEAT NO

	PS Y	PS NO		
	22	0	22	ILL
	24	23	47	NOT ILL
	46	23	69	

Generalize to five, six way tables (here three way table)

log linear models

two kind of analysis

- no row, col or slab is favored

motivate

$$\begin{array}{c|ccc} & & \text{Col} & \\ & 1 & 2 \cdots & J \\ \hline \text{Row} \; 1 & & & \\ \vdots & & x_{ij} & \\ I & & & \end{array}$$

$N = x_{++}$

$m_{ij} = E(x_{ij}) = N\pi_{ij}$

H_I: $\pi_{ij} = \pi_i \pi_j$ (indep)

H_A : Not

$\ln m_{ij} = \ln N + \ln \pi_{ij}$ $\qquad H_A$

$\ln m_{ij} = \ln N + \ln(\pi_i \pi_j)$ $\qquad H_0$

$= \ln N + \ln(\pi_i) + \ln(\pi_j)$ $\qquad$ additive model

$\Omega \quad m_{ij} = \ln(E(x_{ij})) = \ln N + \ln(\pi_{ij})$

$= \underbrace{\ln N + \ln \pi_i + \ln \pi_j}_{H_0} + [\ln \pi_{ij} - \ln \pi_i - \ln \pi_j]$

$= \mu + \alpha_i + \beta_j + \underbrace{(\alpha\beta)_{ij}}_{\text{interaction}}$

log of expected value

H_0: indep $\iff (\alpha\beta)_{ij} = 0$

Max likelihood

$\Omega \quad \hat{\pi}_{ij}\big|_\Omega = \frac{x_{ij}}{x_{++}}$

$\omega \quad \hat{\pi}_i = \frac{x_{i+}}{x_{++}} \quad \hat{\pi}_j = \frac{x_{+j}}{x_{++}}$

log likelihood ratio $\quad 2\Sigma_i \Sigma_j \underbrace{x_{ij}}_{\text{observe value}} \left[\ln \hat{\pi}_{ij} - \ln\left[\frac{x_{i+} x_{+j}}{(x_{++}^2)}\right]\right]$

$\ln\left(\frac{\hat{\pi}_{ij} N}{\hat{\pi}_i \hat{\pi}_j N}\right)$

$G = 2\Sigma_i \Sigma_j OBS_{ij} \ln\left[\frac{O_{ij}}{E_{ij}}\right]$

$G = 2\Sigma_i O_i \ln\left(\frac{O_i}{E_i}\right) = 2\Sigma_i O_i \ln\left[1 + \frac{O_i - E_i}{E_i}\right]$

$\delta_i = O_i - E_i$ $\qquad O_i = E_i + \delta_i$

$= 2\Sigma_i O_i \ln\left[1 + \frac{\delta_i}{E_i}\right]$

$= 2\Sigma_i (E_i + \delta_i)\ln\left(1 + \frac{\delta_i}{E_i}\right)$

$f(x) = \ln(1+x) \Rightarrow \ln(1+x) = 0 + x - \frac{1}{2}x^2$

Mc loran Series $\quad f'(x) = \frac{1}{1+x}$

$f''(x) = -\frac{1}{(1+x)^2}$

$f(x) = f(0) + f'(0)\cdot x + \frac{1}{2!} f''(0) x^2 + \dots O(x^3)$

$\ln\left[1 + \frac{\delta_i}{O_i}\right] = \frac{\delta_i}{E_i} - \frac{1}{2}\left(\frac{\delta_i}{E_i}\right)^2 + O\left(\frac{\delta_i}{E_i}^3\right)$

$\therefore G \approx 2\Sigma_i (E_i + \delta_i)\left(\frac{\delta_i}{E_i} - \frac{1}{2}\left(\frac{\delta_i}{E_i}\right)^2\right) + O\left(\frac{\delta_i}{E_i}^3\right)$

$G \approx 2\left[\underbrace{\Sigma_i \delta_i}_{=0} - \Sigma_i \frac{1}{2}\frac{\delta_i^2}{E_i} + \Sigma_i \frac{\delta_i^2}{E_i} + O\left(\left(\frac{\delta_i}{E_i}\right)^3\right)\right]$

$G \approx 2 \times \frac{1}{2}\Sigma_i \frac{\delta_i^2}{E_i} = \Sigma_i \frac{(O_i - E_i)^2}{E_i}$ Chi-square

approximation of log likelihood Estimation

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

$\{x_{ijk}\}$

Poisson $P(\{x_{ijk}\}) = \prod_i \prod_j \prod_k e^{-m_{ijk}} [m_{ijk})^{x_{ijk}}$

$m_{ijk} = E(x_{ijk})$

Multinomial:

$$p(\{x_{ijk}\}) = \frac{x_{++}!}{\prod_i \prod_j \prod_k x_{ijk}!} \prod_i \prod_j \prod_k (\pi_{ijk})^{x_{ijk}}$$

log likelihood

$\ln p\{x_{ijk}\} = C + \sum_i \sum_j \sum_k x_{ijk} * \ln m_{ijk}$

doesn't depend on x_{ijk}

sufficient since in exponential family

model: $\ln m_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$
$+ (\alpha\beta\gamma)_{ijk} \to$ ANOVA

$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0$

$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = \sum_i (\alpha\beta)_{ik} = \sum (\alpha\gamma)_{ik} = \sum (\beta\gamma)_{jk}$
$= \sum (\beta\gamma)_{jk} = 0$

$\sum_i (\alpha\beta\gamma)_{ijk} = \sum_j (\alpha\beta\gamma)_{ik} = \sum_k (\alpha\beta\gamma)_{ijk} = 0$

$\sum_i \sum_j \sum_k x_{ijk} \ln m_{ijk} = \mu \cdot x_{+++} + \sum_i \alpha_i x_{i++}$
$+ \sum_j \beta_j x_{+j+} + \sum_k \gamma_k x_{++k} + \sum_i \sum_j (\alpha\beta)_{ij} x_{ij+}$
$\sum_i \sum_k (\alpha\gamma)_{ij} + \sum_j \sum_k (\beta\gamma) x_{+jk} + \sum_i \sum_j \sum_k (\alpha\beta\gamma)_{ijk}$
$\cdot x_{ijk}$

$\mu \quad x_{++} \quad \alpha_i \{x_{i++}\} \quad \beta_j \{x_{+j+}\} \quad \gamma_k \{x_{++k}\}$

$(\alpha\beta)_{ij} \{x_{ij+}\} \quad (\alpha\gamma) \{i+k\} \quad (\beta\gamma)_{jk} \{x_{+jk}\}$

$(\alpha\beta\gamma)_{ijk} \quad x_{ijk}$

Ω

ω

$2 \sum_i \sum_j \sum_k x_{ijk} \ln \left(\frac{\hat{m}_{ijk\Omega}}{\hat{m}_{ijk\omega}} \right)$

$\omega \subset \Omega$

α & β in interaction put in

$\chi^2 \quad df = \dim \Omega - \dim \omega$

$(\alpha\beta\gamma)_{ijk} = 0 \quad \forall i,j,k$

Iterative proportional fitting

l = # of cycles $\quad r$ = # substeps in the cycle

$\hat{m}_{ijk}[(l-1)m+r] = \hat{m}_{ijk}^{[(l-1)n+r-1)} \times \theta_r / \hat{m}_{\theta_r}^{(l-1)n+r-1)}$

θ_r = r'th minimal sufficient statistics

Three Way Log Linear Models
(Row = i , Column = j, Slab = k)

Interpretation	Parameterization	df	Estimate of m_{ijk}
Model 1	$(\alpha\beta\gamma)=0\ (\alpha\beta)=0\ (\alpha\gamma)=0\ (\beta\gamma)=0$	(IJK –I – J –K +2)	$x_{i++}x_{+j+}x_{++k} / x_{+++}^2$
Model 2	$(\alpha\beta\gamma)=0\ (\alpha\beta)=0\ (\alpha\gamma)=0$	(KJ -1) (I -1)	$x_{+jk}x_{i++} / x_{+++}$
Model 3	$(\alpha\beta\gamma)=0\ (\alpha\beta)=0\ (\beta\gamma)=0$	(IK -1) (J -1)	$x_{i+k}x_{+j+} / x_{+++}$
Model 4	$(\alpha\beta\gamma)=0\ (\alpha\gamma)=0\ (\beta\gamma)=0$	(IJ – 1) (K-1)	$x_{ij+}x_{++k} / x_{+++}$
Model 5	$(\alpha\beta\gamma)=0\ (\alpha\beta)=0$	(I-1) (J -1) K	$x_{i+k}x_{+jk} / x_{++k}$
Model 6	$(\alpha\beta\gamma)=0\ (\alpha\gamma)=0$	(I – 1) J (K -1)	$x_{ij+}x_{+jk} / x_{+j+}$
Model 7	$(\alpha\beta\gamma)=0\ (\beta\gamma)=0$	I (J – 1) (K -1)	$x_{ij+}x_{i+k} / x_{i++}$
Model 8	$(\alpha\beta\gamma)=0$	(I – 1) (J -1) (K -1)	no closed form Use Deming Proportional Algorithm

crab meat j   potato salad k

Model 1 = All Factors independent of each other   all interaction = 0 ; all factors indep of eachother

Model 2 = Rows independent of Columns and Slabs but Columns and Slabs related   keep $\beta\gamma$ j and k → neither pot salad & crab meat seekness but random

Model 3 = Columns independent of Rows and Slabs but Rows and Slabs related

Model 4 = Slabs independent of Rows and Columns but Rows and Columns related

Model 5 = Given a Slab, Rows and Columns independent (but independence pattern may differ from slab to slab} → indep pattern at slabs different
(Conditional independence or Rows and Columns given Slab)   different set of numbers per slab

Model 6 = Given a Column, Rows and Slabs independent (but independence pattern may differ from Column to Column) (Conditional independence of Rows and Slabs given Column) → may not be same indep proportions

Model 7 = Given a Row, Columns and Slabs independent (but independent pattern may differ from Row to Row) → indep pattern at diff rows
(Conditional independence of Columns and Slabs given Row)

Model 8= Rows Depend on Columns (and vice versa), Rows depend on Slabs (and vice versa), Columns depend On Slabs (and vice versa) but there is no three way dependence → no simple explanation

Original Table

		Crab k=1 PS j=1	Crab k=1 No PS j=2	No Crab k=2 PS j=1	No Crab k=2 No PS j=2
Ill	i=1	120	4	22	0
Not Ill	i=2	80	31	24	23

observe

Marginals

Ill	i=1	146
Not Ill	i=2	158

		Crab k=1	No Crab k=2
PS	j=1	200	46
No PS	j=2	35	23

Model: Main Effects Plus Interaction Between Potato Salad and Crab

Expected Table

$$\ln m_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk}$$ → default model

$$\hat{m}_{ijk} = \frac{x_{i++}\, x_{+jk}}{x_{+++}}$$

		Crab k=1 PS j=1	Crab k=1 No PS j=2	No Crab k=2 PS j=1	No Crab k=2 No PS j=2
Ill	i=1	96.05	16.81	22.09	11.05
Not Ill	i=2	103.95	18.19	23.91	11.95

PS No PS Crab No Crab

Marginals

Ill	i=1	146
Not Ill	i=2	158

		Crab k=1	No Crab k=2
PS	j=1	200	46
No PS	j=2	35	23

Chi-Square Contribution

Chi sq Contrib

		Crab k=1 PS j=1	Crab k=1 No PS j=2	No Crab k=2 PS j=1	No Crab k=2 No PS j=2
Ill	i=1	5.97044	9.76107	0.000384	11.04605
Not Ill	i=2	5.516989	9.019723	0.000355	10.20711

add up

Chisquare= 51.52212 on 3 df

pvalue= 3.79E-11

= illness related to something

⇒ REJECT

Original Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

i ill ness
j crab
k potato salad

ij

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

ik I

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

jk

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

ITERATION 1

Start Data

match

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	1	1	2
i = 2	not ill	1	1	2
		2	2	4

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	1	1	2
i = 2	not ill	1	1	2
		2	2	4

	j = 1	j = 2	
i = 1	2	2	4
i = 2	2	2	4
	4	4	8

	k =1	k =2	
i = 1	2	2	4
i = 2	2	2	4
	4	4	8

	k = 1	k = 2	
j = 1	2	2	4
j = 2	2	2	4
	4	4	8

a) i x j adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	71	2	73
i = 2	not ill	52	27	79
		123	29	152

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	71	2	73
i = 2	not ill	52	27	79
		123	29	152

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	73	73	146
i = 2	79	79	158
	152	152	304

	k = 1	k = 2	
j = 1	123	123	246
j = 2	29	29	58
	152	152	304

b) i x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120.6027	3.39726	124
i = 2	not ill	73.06329	37.93671	111
		193.666	41.33397	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	21.39726	0.60274	22
i = 2	not ill	30.93671	16.06329	47
		52.33397	16.66603	69

match I

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	193.666	52.33397	246
j = 2	41.33397	16.66603	58
	235	69	304

b) j x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	124.5471	2.876668	127.4238
i = 2	not ill	75.45287	32.12333	107.5762
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	18.80755	0.831813	19.63937
i = 2	not ill	27.19245	22.16819	49.36063
		46	23	69

No match I

	j = 1	j = 2	
i = 1	143.3547	3.708481	147.0632
i = 2	102.6453	54.29152	156.9368
	246	58	304

	k =1	k =2	
i = 1	127.4238	19.63937	147.0632
i = 2	107.5762	49.36063	156.9368
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Screws up

Original Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

ITERATION 2

Start Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	124.5471	2.876668	127.4238
i = 2	not ill	75.45287	32.12333	107.5762
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	18.80755	0.831813	19.63937
i = 2	not ill	27.19245	22.16819	49.36063
		46	23	69

	j = 1	j = 2	
i = 1	143.3547	3.708481	147.0632
i = 2	102.6453	54.29152	156.9368
	246	58	304

	k =1	k =2	
i = 1	127.4238	19.63937	147.0632
i = 2	107.5762	49.36063	156.9368
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

a) i x j adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	123.3702	3.1028	126.473
i = 2	not ill	76.44868	31.95085	108.3995
		199.8189	35.05364	234.8725

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	18.62983	0.8972	19.52703
i = 2	not ill	27.55132	22.04915	49.60048
		46.18115	22.94636	69.1275

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	126.473	19.52703	146
i = 2	108.3995	49.60048	158
	234.8725	69.1275	304

	k = 1	k = 2	
j = 1	199.8189	46.18115	246
j = 2	35.05364	22.94636	58
	234.8725	69.1275	304

b) i x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120.9579	3.042129	124
i = 2	not ill	78.28266	32.71734	111
		199.2405	35.75947	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.98917	1.010825	22
i = 2	not ill	26.10685	20.89315	47
		47.09602	21.90398	69

	j = 1	j = 2	
i = 1	141.947	4.052955	146
i = 2	104.3895	53.61049	158
	246.3366	57.66344	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	199.2405	47.09602	246.3366
j = 2	35.75947	21.90398	57.66344
	235	69	304

b) j x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.4189	2.97752	124.3965
i = 2	not ill	78.58106	32.02248	110.6035
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.50071	1.061404	21.56212
i = 2	not ill	25.49929	21.9386	47.43788
		46	23	69

	j = 1	j = 2	
i = 1	141.9197	4.038924	145.9586
i = 2	104.0803	53.96108	158.0414
	246	58	304

	k =1	k =2	
i = 1	124.3965	21.56212	145.9586
i = 2	110.6035	47.43788	158.0414
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Original Data

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

ITERATION 3

Start Data

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.4189	2.97752	124.3965
i = 2	not ill	78.58106	32.02248	110.6035
		200	35	235

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.50071	1.061404	21.56212
i = 2	not ill	25.49929	21.9386	47.43788
		46	23	69

	j = 1	j = 2	
i = 1	141.9197	4.038924	145.9586
i = 2	104.0803	53.96108	158.0414
	246	58	304

	k =1	k =2	
i = 1	124.3965	21.56212	145.9586
i = 2	110.6035	47.43788	158.0414
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

a) i x j adjustment

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.4877	2.948825	124.4365
i = 2	not ill	78.5204	32.04558	110.566
		200.0081	34.9944	235.0025

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.51232	1.051175	21.56349
i = 2	not ill	25.4796	21.95442	47.43402
		45.99192	23.0056	68.99752

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124.4365	21.56349	146
i = 2	110.566	47.43402	158
	235.0025	68.99752	304

	k = 1	k = 2	
j = 1	200.0081	45.99192	246
j = 2	34.9944	23.0056	58
	235.0025	68.99752	304

b) i x k adjustment

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0615	2.938481	124
i = 2	not ill	78.82863	32.17137	111
		199.8901	35.10985	235

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.92755	1.072454	22
i = 2	not ill	25.24646	21.75354	47
		46.17401	22.82599	69

	j = 1	j = 2	
i = 1	141.9891	4.010935	146
i = 2	104.0751	53.92491	158
	246.0642	57.93584	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	199.8901	46.17401	246.0642
j = 2	35.10985	22.82599	57.93584
	235	69	304

b) j x k adjustment

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.1281	2.929287	124.0573
i = 2	not ill	78.87195	32.07071	110.9427
		200	35	235

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.84868	1.08063	21.92931
i = 2	not ill	25.15132	21.91937	47.07069
		46	23	69

	j = 1	j = 2	
i = 1	141.9767	4.009916	145.9866
i = 2	104.0233	53.99008	158.0134
	246	58	304

	k =1	k =2	
i = 1	124.0573	21.92931	145.9866
i = 2	110.9427	47.07069	158.0134
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Original Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

ITERATION 4

Start Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.1281	2.929287	124.0573
i = 2	not ill	78.87195	32.07071	110.9427
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.84868	1.08063	21.92931
i = 2	not ill	25.15132	21.91937	47.07069
		46	23	69

	j = 1	j = 2	
i = 1	141.9767	4.009916	145.9866
i = 2	104.0233	53.99008	158.0134
	246	58	304

	k =1	k =2	
i = 1	124.0573	21.92931	145.9866
i = 2	110.9427	47.07069	158.0134
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

a) i x j adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.1479	2.922043	124.0699
i = 2	not ill	78.85431	32.0766	110.9309
		200.0022	34.99865	235.0009

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.8521	1.077957	21.93005
i = 2	not ill	25.14569	21.9234	47.06909
		45.99779	23.00135	68.99914

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124.0699	21.93005	146
i = 2	110.9309	47.06909	158
	235.0009	68.99914	304

	k = 1	k = 2	
j = 1	200.0022	45.99779	246
j = 2	34.99865	23.00135	58
	235.0009	68.99914	304

b) i x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0796	2.920395	124
i = 2	not ill	78.90342	32.09658	111
		199.983	35.01698	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.9186	1.081396	22
i = 2	not ill	25.10878	21.89122	47
		46.02739	22.97261	69

	j = 1	j = 2	
i = 1	141.9982	4.001791	146
i = 2	104.0122	53.9878	158
	246.0104	57.98959	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	199.983	46.02739	246.0104
j = 2	35.01698	22.97261	57.98959
	235	69	304

b) j x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0899	2.918979	124.0089
i = 2	not ill	78.91012	32.08102	110.9911
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.90616	1.082685	21.98884
i = 2	not ill	25.09384	21.91732	47.01116
		46	23	69

	j = 1	j = 2	
i = 1	141.996	4.001664	145.9977
i = 2	104.004	53.99834	158.0023
	246	58	304

	k =1	k =2	
i = 1	124.0089	21.98884	145.9977
i = 2	110.9911	47.01116	158.0023
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Original Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k = 1	k = 2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Start Data

Iteration 5

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0899	2.918979	124.0089
i = 2	not ill	78.91012	32.08102	110.9911
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.90616	1.082685	21.98884
i = 2	not ill	25.09384	21.91732	47.01116
		46	23	69

	j = 1	j = 2	
i = 1	141.996	4.001664	145.9977
i = 2	104.004	53.99834	158.0023
	246	58	304

	k = 1	k = 2	
i = 1	124.0089	21.98884	145.9977
i = 2	110.9911	47.01116	158.0023
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

a) i x j adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0933	2.917765	124.011
i = 2	not ill	78.90711	32.08201	110.9891
		200.0004	34.99977	235.0001

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.90674	1.082235	21.98897
i = 2	not ill	25.09289	21.91799	47.01088
		45.99963	23.00023	68.99985

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k = 1	k = 2	
i = 1	124.011	21.98897	146
i = 2	110.9891	47.01088	158
	235.0001	68.99985	304

	k = 1	k = 2	
j = 1	200.0004	45.99963	246
j = 2	34.99977	23.00023	58
	235.0001	68.99985	304

b) i x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0825	2.917506	124
i = 2	not ill	78.91485	32.08515	111
		199.9973	35.00266	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.91722	1.082777	22
i = 2	not ill	25.08708	21.91292	47
		46.0043	22.9957	69

	j = 1	j = 2	
i = 1	141.9997	4.000283	146
i = 2	104.0019	53.99807	158
	246.0016	57.99836	304

	k = 1	k = 2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	199.9973	46.0043	246.0016
j = 2	35.00266	22.9957	57.99836
	235	69	304

b) j x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0841	2.917284	124.0014
i = 2	not ill	78.9159	32.08272	110.9986
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.91527	1.08298	21.99825
i = 2	not ill	25.08473	21.91702	47.00175
		46	23	69

	j = 1	j = 2	
i = 1	141.9994	4.000264	145.9996
i = 2	104.0006	53.99974	158.0004
	246	58	304

	k = 1	k = 2	
i = 1	124.0014	21.99825	145.9996
i = 2	110.9986	47.00175	158.0004
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Original Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Iteration 6

Start Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0841	2.917284	124.0014
i = 2	not ill	78.9159	32.08272	110.9986
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.91527	1.08298	21.99825
i = 2	not ill	25.08473	21.91702	47.00175
		46	23	69

	j = 1	j = 2	
i = 1	141.9994	4.000264	145.9996
i = 2	104.0006	53.99974	158.0004
	246	58	304

	k =1	k =2	
i = 1	124.0014	21.99825	145.9996
i = 2	110.9986	47.00175	158.0004
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

a) i x j adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0846	2.917092	124.0017
i = 2	not ill	78.91542	32.08287	110.9983
		200.0001	34.99996	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.91536	1.082908	21.99827
i = 2	not ill	25.08458	21.91713	47.00171
		45.99994	23.00004	68.99998

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124.0017	21.99827	146
i = 2	110.9983	47.00171	158
	235	68.99998	304

	k = 1	k = 2	
j = 1	200.0001	45.99994	246
j = 2	34.99996	23.00004	58
	235	68.99998	304

b) i x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0829	2.917051	124
i = 2	not ill	78.91663	32.08337	111
		199.9996	35.00042	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.91701	1.082994	22
i = 2	not ill	25.08367	21.91633	47
		46.00068	22.99932	69

	j = 1	j = 2	
i = 1	142	4.000045	146
i = 2	104.0003	53.9997	158
	246.0003	57.99974	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	199.9996	46.00068	246.0003
j = 2	35.00042	22.99932	57.99974
	235	69	304

b) j x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0832	2.917016	124.0002
i = 2	not ill	78.9168	32.08298	110.9998
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.9167	1.083025	21.99972
i = 2	not ill	25.0833	21.91697	47.00028
		46	23	69

	j = 1	j = 2	
i = 1	141.9999	4.0000	145.9999
i = 2	104.0001	54.0000	158.0001
	246	58	304

	k =1	k =2	
i = 1	124.0002	21.9997	145.9999
i = 2	110.9998	47.0003	158.0001
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Original Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k = 1	k = 2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Iteration 7

Done not changing anymore

Start Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0832	2.917016	124.0002
i = 2	not ill	78.9168	32.08298	110.9998
		200	35	235

k = 2 no Crab

		j = 1	j = 2	
i = 1	ill	20.9167	1.083025	21.99972
i = 2	not ill	25.0833	21.91697	47.00028
		46	23	69

	j = 1	j = 2	
i = 1	141.9999	4.000042	145.9999
i = 2	104.0001	53.99996	158.0001
	246	58	304

	k = 1	k = 2	
i = 1	124.0002	21.99972	145.9999
i = 2	110.9998	47.00028	158.0001
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

a) i x j adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0833	2.916986	124.0003
i = 2	not ill	78.91672	32.08301	110.9997
		200	34.99999	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.91671	1.083014	21.99973
i = 2	not ill	25.08328	21.91699	47.00027
		45.99999	23.00001	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k = 1	k = 2	
i = 1	124.0003	21.99973	146
i = 2	110.9997	47.00027	158
	235	69	304

	k = 1	k = 2	
j = 1	200	45.99999	246
j = 2	34.99999	23.00001	58
	235	69	304

b) i x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.083	2.916979	124
i = 2	not ill	78.91691	32.08309	111
		199.9999	35.00007	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.91697	1.083028	22
i = 2	not ill	25.08313	21.91687	47
		46.00011	22.99989	69

	j = 1	j = 2	
i = 1	142	4.000007	146
i = 2	104	53.99995	158
	246	57.99996	304

	k = 1	k = 2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	199.9999	46.00011	246
j = 2	35.00007	22.99989	57.99996
	235	69	304

b) j x k adjustment

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0831	2.916974	124
i = 2	not ill	78.91694	32.08303	111
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.91692	1.083033	21.99996
i = 2	not ill	25.08308	21.91697	47.00004
		46	23	69

	j = 1	j = 2	
i = 1	142.0000	4.0000	146
i = 2	104.0000	54.0000	158
	246	58	304

	k = 1	k = 2	
i = 1	124.0000	22.0000	146
i = 2	111.0000	47.0000	158
	235	69	304

	k = 1	k = 2	
j = 1	200.0000	46.0000	246
j = 2	35.0000	23.0000	58
	235	69	304

Original Data

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Expectation under

Hypothesis: $(\alpha\beta\gamma)_{ijk} = 0$

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	121.0832	2.917016	124.0002
i = 2	not ill	78.9168	32.08298	110.9998
		200	35	235

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	20.9167	1.083025	21.99972
i = 2	not ill	25.0833	21.91697	47.00028
		46	23	69

Chi-sq → original & expected

Contribution

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	0.00969	0.402073	0.411764
i = 2	not ill	0.014868	0.036557	0.051425
		0.024558	0.43863	0.463188

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	0.056105	1.083025	1.139131
i = 2	not ill	0.046786	0.053518	0.100303
		0.102891	1.136543	1.239434

df 1
Chiobs 1.703
p-value = 0.191945 Accept

Conclusion: No three way interaction

Original Data

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Expectation under

Relationship b/w illness & Crabmeat

$$Hypothesis: \quad (\alpha\beta\gamma)_{ijk} = 0 \ and \ (\alpha\gamma)_{ik} = 0$$

$$\hat{m}_{ijk} = \frac{x_{ij+} \times x_{+jk}}{x_{+j+}}$$

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	115.4472	2.413793	117.8609
i = 2	not ill	84.55285	32.58621	117.1391
		200	35	235

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	26.55285	1.586207	28.13905
i = 2	not ill	19.44715	21.41379	40.86095
		46	23	69

Contribution

k = 1 Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	0.179549	1.042365	1.221913
i = 2	not ill	0.245153	0.077212	0.322365
		0.424702	1.119577	1.544279

k = 2 no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	0.780647	1.586207	2.366854
i = 2	not ill	1.065884	0.117497	1.18338
		1.846531	1.703704	3.550234

df	2		inc df	1	
Chiobs	5.095		inc Chi	3.392	
p-value =	0.078296	Accept	p-value	0.065518	Accept

Conclusion: There does not seem to be a relationship between Crabmeat and Illness

Original Data

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	120	4	124
i = 2	not ill	80	31	111
		200	35	235

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	22	0	22
i = 2	not ill	24	23	47
		46	23	69

	j = 1	j = 2	
i = 1	142	4	146
i = 2	104	54	158
	246	58	304

	k =1	k =2	
i = 1	124	22	146
i = 2	111	47	158
	235	69	304

	k = 1	k = 2	
j = 1	200	46	246
j = 2	35	23	58
	235	69	304

Expectation under

$$Hypothesis: \quad (\alpha\beta\gamma)_{ijk} = 0 \ and \ (\alpha\beta)_{ij} = 0$$

$$\hat{m}_{ijk} = \frac{x_{i+k} \times x_{+jk}}{x_{++k}}$$

ijk

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	105.5319	18.46809	124
i = 2	not ill	94.46809	16.53191	111
		200	35	235

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	14.66667	7.333333	22
i = 2	not ill	31.33333	15.66667	47
		46	23	69

Contribution

k = 1, Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	1.983528	11.33444	13.31797
i = 2	not ill	2.215833	12.6619	14.87773
		4.199361	23.99635	28.19571

k = 2, no Crab

		j = 1 PS	j = 2 No PS	
i = 1	ill	3.666667	7.333333	11
i = 2	not ill	1.716312	3.432624	5.148936
		5.382979	10.76596	16.14894

df	2		inc df	1	
Chiobs	44.345		inc Chi	42.642	
p-value =	2.35E-10	Reject	p-value	6.57E-11	Reject

Conclusion: There seems to be a relationship between Potato Salad and Illness

Meco 6315

For the following data, use Log Linear Model techniques to determine if political party affiliation is related to Sex and Ethnicity.

(*) Three way table

Male (k = 1)

		Ethnicity			
		Anglo (j = 1)	Hisp (j = 2)	Black (j = 3)	
	Dem (i =1)	40	56	87	183
Party	Ind (i = 2)	21	23	14	58
	Rep (i = 3)	62	41	38	141
		123	120	139	382

Female (k = 2)

		Ethnicity			
		Anglo (j = 1)	Hisp (j = 2)	Black (j = 3)	
	Dem (i =1)	51	66	98	215
Party	Ind (i = 2)	24	18	11	53
	Rep (i = 3)	39	27	21	87
		114	111	130	355