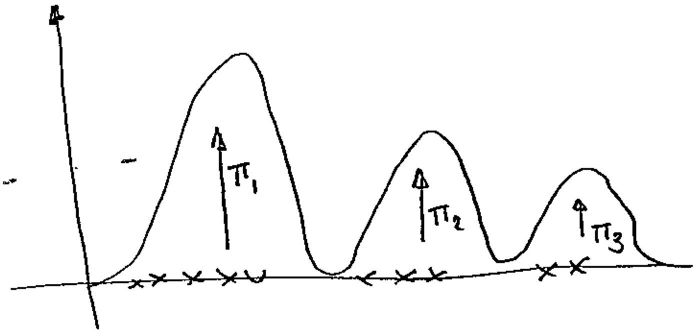
Mixtures of Gaussians

Start with params describing each cluster: μ_c σ_c 'size' π_c

Prob. distribution $p(x) = \sum_c \pi_c \mathcal{N}(x; \mu_c, \sigma_c)$ — ①

Equivalent 'latent variable' form

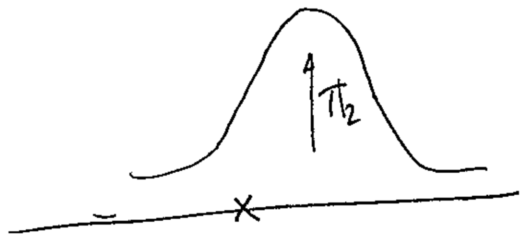
$p(Z=c) = \pi_c$ — ②



To draw a sample from $p(x)$ first select one of components with discrete prob π (large $\pi \Rightarrow$ selected more often

$p(x|z=c) = \mathcal{N}(x; \mu_c \sigma_c)$ — ③

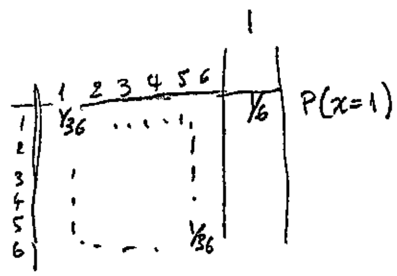
Then, given compont assignment $Z=C$ we can draw a value of x from the corresponding gaussian.



These 2 distributions form a joint ~~of~~ model over x and z,

Discarding value of z gives a sample from the marginal $P(x)$ (①)

~~defined~~



Models like this are called latent variable models. X modelled ~~with~~ jointly with additional variable Z that is hidden

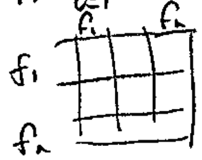
Z helps explains patterns in the values of X (eg groups)

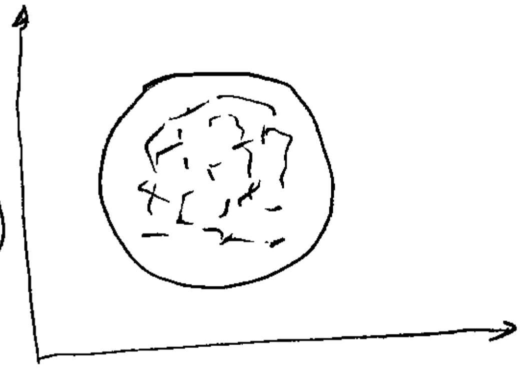
Multivariate Gaussian Models

$$N(\underline{x};\underline{\mu},\Sigma) = \frac{1}{(2\pi)^{d/2}}\Sigma^{-1/2} e^{\left\{-\frac{1}{2}(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})\right\}}$$

Features are multivariate, even high dimensional
Typically use multivariate Gaussian
- Vector mean μ, size number of features (n)
- Σ covariance matrix, ($n \times n$ size)

$$Cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})$$





If we are given data from a multivariate gaussian the MLE for the model params were $\hat{\mu}$ (mean of data) and $\hat{\Sigma}$ (covariance estimate)

$$\hat{\mu} = \frac{1}{m}\sum_i x^{(i)}$$

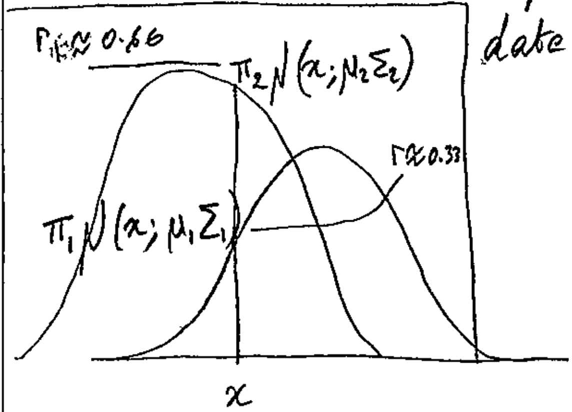
$$\hat{\Sigma} = \frac{1}{m}\sum_i (x^{(i)}-\hat{\mu})^T(x^{(i)}-\hat{\mu})$$

E-M (E Step)

- Start with clusters : Mean μ_c, Covariance Σ_c, "size" π_c
- For each Data point i and each cluster c

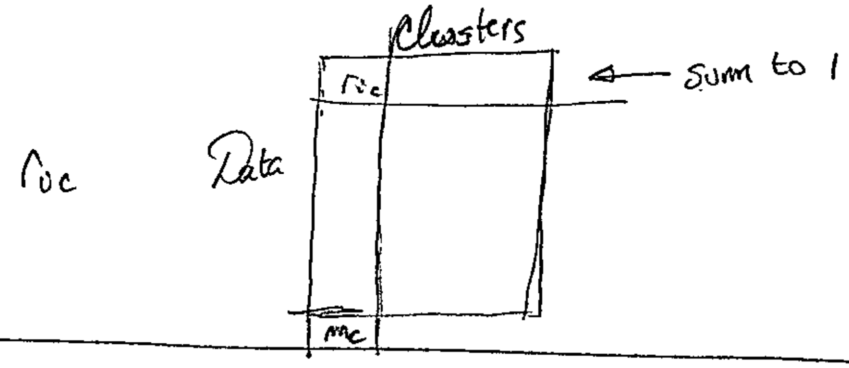
Compute $$r_{ic} = \frac{\pi_c N(x_i ; \mu_c, \Sigma_c)}{\sum_{c'} \pi_{c'} N(x_i ; \mu_{c'}, \Sigma_{c'})}$$

r_{ic} : responsibility value, relative prob that data pt i belongs to c



if c_i is not a good explanation for x it will result in small r_{ic}

if best explanation for x, r_{ic} will approach 1



E-M (M Step)

From r_{ic} update μ_c, Σ_c, π_c

For each cluster Gaussian $z = c$

Update its param using weighted data pts $\quad m_c = \sum_i r_{ic}$

cluster c has the sum of these soft memberships

$$\pi_c = \frac{m_c}{m}$$ fraction of data pts assigned to c

$$\mu_c = \frac{1}{m_c} \sum_i r_{ic} x^{(i)}$$ weighted average.

r_{ic} small – no influence on average

r_{ic} large – large infl.

$$\Sigma_c = \frac{1}{m_c} \sum_i r_{ic} \left(x^{(i)} - \mu_c\right)^T \left(x^{(i)} - \mu\right).$$

weighted by r_{ic}.

$$\sigma_c = \sqrt{\frac{1}{m_c} \sum_i r_{ic} \left(x_i - \mu_c\right)^2}$$

each step increases log likelihood

$$\log p(\underline{X}) = \sum_i \log \left[\sum_c \pi_c \mathcal{N}(x_i ; \mu_c , \Sigma_c) \right]$$

log prob of data points under mixture model

<u>Iterate until convergence</u>

0.2.

From our model equation

STEP 1 : INIT

$$T_i = \beta_0 + \beta_1 V_i + \epsilon_i$$

V_i is transformed temp

SOS $$S(\beta_0, \beta_1) = \sum_{i=1}^{M} (T_i - \beta_0 - \beta_1 V_i)^2$$

Max. $$\frac{\partial S}{\partial \beta_0} = \sum_{i=1}^{M} 2(T_i - \beta_0 - \beta_1 V_i) = 0$$

$$\sum_{i=1}^{M} T_i - \beta_0 - \beta_1 V_i = 0 \qquad \text{—— (1)}$$

Max $$\frac{\partial S}{\partial \beta_1} = \sum_{i=1}^{M} 2(T_i - \beta_0 - \beta_1 V_i)\, V_i = 0$$

$$\sum_{i=1}^{M} V_i (T_i - \beta_0 - \beta_1 V_i) = 0 \qquad \text{—— (2)}$$

FROM (1) $$\sum T_i - M\beta_0 - \beta_1 \sum V_i = 0 \qquad \text{—— (3)}$$

FROM (2) $$\sum V_i T_i - \sum V_i \beta_0 - \beta_1 \sum V_i^2 = 0 \qquad \text{—— (4)}$$

FROM (3) $$M\beta_0 + \beta_1 \sum V_i = \sum T_i \qquad \text{—— (5)}$$

FROM (4) $$\sum V_i \beta_0 + \beta_1 \sum V_i^2 = \sum V_i T_i \qquad \text{—— (6)}$$

$$\begin{bmatrix} M & \sum V_i \\ \sum V_i & \sum V_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum T_i \\ \sum V_i T_i \end{bmatrix} \qquad \text{—— (7)}$$

~~For~~. BY DEFN OF VAR

$$\sigma^2 = \mathbb{E}[\epsilon_i^2] - (\mathbb{E}[\epsilon_i])^2$$

since $$\mathbb{E}[\epsilon_i] = 0$$

$$\sigma^2 = \mathbb{E}[\epsilon_i^2] = \mathbb{E}\left[(T_i - \beta_0 - \beta_1 V_1)^2\right]$$

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^{M} (T_i - \hat{\beta}_0 - \hat{\beta}_1 V_1)^2 \qquad \text{—— (8)}$$

(7) + (8) are initial parameters.

For censored data.

$$\mu_i' = \beta_0' + \beta_1' V_i \quad \text{——— (9)}$$

Standardize input

$$x_i = \frac{t_i - \mu_i}{\sigma'} \quad \text{——— (10)}$$

$$H(x) = \frac{\phi(x)}{1 - \Phi(x)} \qquad \begin{array}{l}\phi(x) = \text{pdf} \\ \Phi(x) = \text{cdf}\end{array} \quad \text{——— (11)}$$

First order expectation, mean correction.

$$E[Z_i \mid Z_i > t_i^*, \theta_i'] = \mu_i' + \sigma' H(x) \quad \text{——— (12)}$$

2nd Order contribution, variance correction

$$E[Z_i^2 > t_i^*, \theta'] = (\mu_i')^2 + (\sigma')^2 + \sigma'(t_i^* + \mu_i') H(x) \quad \text{——— (13)}$$

STEP 3: MAXIMIZATION

Revised eqns using Expected values.

$1 \to k$ obs
$k \to 2$ hidden

$$\sum_{i=1}^{k} T_i + \sum_{i=k}^{m} E[Z_i] = m\beta_0 + \beta_1 \sum_{i=1}^{m} V_i \quad \text{——— (14)}$$

$$\sum_{i=1}^{k} V_i T_i + \sum_{i=k}^{m} E[Z_i^2] = \beta_0 \sum_{i=1}^{m} V_i + \beta_1 \sum_{i=1}^{m} V_i^2 \quad \text{——— (15)}$$

$$\begin{bmatrix} m & \sum V_i \\ \sum V_i & \sum V_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum T_i + \sum E[Z_i] \\ \sum V_i T_i + \sum V_i E[Z_i^2] \end{bmatrix} \quad \text{——— (16)}$$

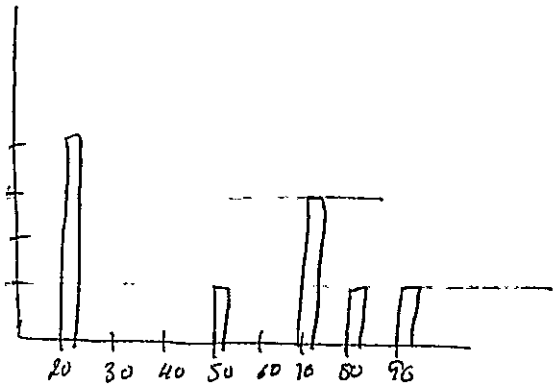
$$\sigma^2 = \frac{1}{m} \left[\sum_{i=1}^{k} (T_i - \beta_0 - \beta_1 V_i)^2 + \sum_{i=k}^{m} E[Z_i^2 \mid Z_i > t_i^*, \theta] - 2 \sum_{i=k}^{m} E[Z_i \mid Z_i > t_i^*, \theta'] (\beta_0 + \beta_1 V_i) \right] \quad \text{——— (17)}$$

$$2 \sum E[Z_i](\beta_0 + \beta_1 V_i) = 2\beta_0 \sum E[Z_i] + 2\beta_1 \sum E[Z_i] V_i$$

REPEAT UNTIL CONVERGE.

eg Jack knife.

f	
4	20
3	70
1	50
1	90
1	80
10	



$$\bar{x} = \frac{80+210+50+90+80}{10}$$

$$= 51$$

$$s^2 = \sqrt{\frac{(4\times 961)+(3\times 361)+1+1521+841}{10}}$$

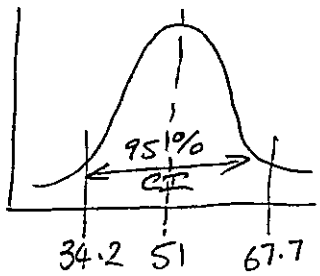
$$= \sqrt{\frac{7920}{10}} = 27$$

$$\sigma_{\bar{x}} = \frac{27}{\sqrt{10}} = 8.538$$

95% confidence interval $CI = 51 \pm 1.96\times 8.538$

$$= 51 \pm 16.734$$

$$= [34.265, \quad 67.734]$$



SAMPLE 1

$$E[X^2] = \int_0^\infty x^2 f(x)\,dx.$$

$$= \int_0^\infty x^2 \lambda e^{-\lambda x}\,dx = \lambda \int_0^\infty x^2 e^{-\lambda x}\,dx$$

$$\boxed{\begin{array}{lll} \text{Let} & u = x^2 & du = 2x\,dx. \\ & v = -\dfrac{e^{-\lambda x}}{\lambda} & dv = e^{-\lambda x}\,dx \end{array}}$$

$$E[X^2] = \cancel{\lambda}\left(\left[-\frac{x^2 e^{-\lambda x}}{\cancel{\lambda}}\right]_0^\infty - \int_0^\infty -\frac{2x e^{-\lambda x}}{\cancel{\lambda}}\right)$$

$$\left[-x^2 e^{-\lambda x}\right]_0^\infty$$

$x^2 e^{-\lambda x} = 0$, $x = 0$

$$\lim_{x \to \infty} x^2 e^{-\lambda x} = 0$$

$$\therefore \quad E[X^2] = 2\int_0^\infty x e^{-\lambda x}.$$

$$\boxed{\begin{array}{lll} \text{Let} & u = x & du = dx. \\ v = & -\dfrac{e^{-\lambda x}}{\lambda} & dv = e^{-\lambda x}\,dx \end{array}}$$

$$E[X^2] = 2\left(\left[-\frac{x e^{-\lambda x}}{\lambda}\right]_0^\infty - \int_0^\infty -\frac{e^{-\lambda x}}{\lambda}\right)dx$$

Consider $\left[-\dfrac{x e^{-\lambda x}}{\lambda}\right]_0^\infty$

$= 0$, $x = 0$

$$\lim_{x \to \infty} -\frac{x e^{-\lambda x}}{\lambda} = 0$$

$$E[X^2] = 2 \int_0^\infty \frac{e^{-\lambda x}}{\lambda}\,dx$$

$$= \left[\frac{-2\, e^{-\lambda x}}{\lambda^2}\right]_0^\infty$$

$$\lim_{x \to \infty} -\frac{2e^{-\lambda x}}{\lambda^2} = 0$$

$$E[X^2] = 0 - \left(-\frac{2}{\lambda^2}\right) = \frac{2}{\lambda^2}$$

$$V_{AR}(X) = E[X^2] - (E[X])^2.$$

$$E[X^2] = \int_0^\infty x^2 f(x)$$

$$= \int_0^\infty x^2 \lambda e^{-\lambda x} \, dx.$$

$$= \lambda \int_0^\infty x^2 e^{-\lambda x} \, dx.$$

Let $u = x^2$ $\quad du = 2x$

$v = \dfrac{-e^{-\lambda x}}{\lambda}.$ $\quad dv = e^{-\lambda x}.$

$$\int u \, dv = uv - \int v \, du.$$

$$E[X^2] = \lambda \left(\left[-\frac{x^2 e^{-\lambda x}}{\lambda} \right]_0^\infty - \int_0^\infty -\frac{e^{-\lambda x}}{\lambda} \cdot 2x \, dx \right)$$

$$= \left[-x^2 e^{-\lambda x} \right]_0^\infty + \int_0^\infty e^{-\lambda x} 2x \, dx.$$

$$E[X] = \lambda \left(\left[-\frac{x e^{-\lambda x}}{\lambda} \right]_0^\infty + \int_0^\infty \frac{e^{-\lambda x}}{\lambda} (1)\, dx \right)$$

$$= \left[-x e^{-\lambda x} \right]_0^\infty + \int e^{-\lambda x}\, dx$$

$$-x e^{-\lambda x} = 0 \text{ when } x = 0$$

$$\lim_{x \to \infty} -x e^{-\lambda x} = 0$$

$$E[X] = \int_0^\infty e^{-\lambda x}\, dx.$$

$$= \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty$$

$$0 - \left(-\frac{1}{\lambda} \right) = \frac{1}{\lambda}.$$

$$\text{VAR} = E[X^2] - E[X]^2$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$$L(x,\theta) = \prod_{i=1}^{M} f(x_i, \theta)$$

$$= \prod_{i=1}^{M} \lambda e^{-\lambda x_i}$$

$$= \lambda^n e^{-\sum_{i=1}^{M} \lambda x_i}$$

log likelihood.

$$l(x,\theta) = n \log \lambda - \lambda \sum_{i=1}^{M} x_i$$

~~Max~~ $\frac{\partial l(x,\theta)}{\partial \lambda} = \frac{M}{\lambda} - \sum_{i=1}^{M} x_i = 0$ at max

$$\frac{M}{\lambda} = \sum_{i=1}^{M} x_i$$

$$\hat{\lambda} = \frac{M}{\sum_{i=1}^{M} x_i} = \frac{1}{\bar{x}}$$

Regression line

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$S = \sum_{i=1}^{m} \varepsilon_i^2 = \sum_{i=1}^{m} (y_i - \beta_0 - \beta_1 x_i)^2$$

~~on thrash~~

find $\underset{\beta_0 \beta_1}{\arg\min}\ S$

$$\max \frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^{m} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\sum_{i=1}^{m} y_i - \beta_0 - \beta_1 x_i = 0$$

$$\sum_{i=1}^{m} y_i - m\beta_0 - \beta_1 \sum_{i=1}^{m} x_i = 0 \quad \text{--- } (1)$$

$$\max \frac{\partial S}{\partial \beta_1} = 2 \sum_{i=1}^{m} (y_i - \beta_0 - \beta_1 x_i)(-x_i) = 0$$

$$= \sum_{i=1}^{m} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$= \sum_{i=1}^{m} x_i y_i - x_i \beta_0 - \beta_1 x_i^2 = 0$$

$$= \sum_{i=1}^{m} x_i y_i - \beta_0 \sum_{i=1}^{m} x_i - \beta_1 \sum_{i=1}^{m} x_i^2 = 0 \quad \text{--- } (2)$$

$$\sum_{i=1}^{m} y_i \sum_{i=1}^{n} x_i - m\beta_0 \sum_{i=1}^{m} x_i - \beta_1 \left(\sum_{i=1}^{n} x_i\right)^2 = 0$$

$$m\sum x_i y_i - m\beta_0 \sum_{i=1}^{m} x_i - m\beta_1 \sum_{i=1}^{m} x_i^2 = 0$$

$$\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i - \beta_1\left(\sum_{i=1}^{m} x_i\right)^2 = m\sum_{i=1}^{m} x_i y_i - m\beta_1 \sum_{i=1}^{m} x_i^2$$

$$\beta_1\left(m\sum_{i=1}^{m} x_i^2 - \left(\sum_{i=1}^{m} x_i\right)^2\right) = m\sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i$$

$$\hat{\beta}_1 = \frac{m\sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i}{m\sum_{i=1}^{m} x_i^2 - \left(\sum_{i=1}^{m} x_i\right)^2}$$

$$= \frac{m\bar{x}\bar{y} - m^2\bar{x}\bar{y}}{m\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$\beta_1 \cancel{\sum x_i} = \frac{m\bar{x}\bar{y} - m^2\bar{x}\bar{y}}{m\sum x_i^2 - \left(\sum x_i\right)^2} \sum x_i$$

$$\beta_0 = \frac{\cancel{m}\bar{x}\bar{y} - m^{\cancel{2}}\bar{y}\bar{x}}{m\sum x_i^2 - \left(\sum x_i\right)^2} \frac{\sum x_i}{\cancel{m}} - \frac{\sum y_i}{m}$$

$$= \frac{\left(\bar{x}\bar{y} - m\bar{x}\bar{y}\right)}{m\sum x_i^2 - m^2\bar{x}^2}(m\bar{x}) \quad - \bar{y}$$

$$= \frac{m\bar{x}^2\bar{y} \cancel{- m^2\bar{x}^2\bar{y}} - m\sum x_i^2\bar{y} + \cancel{m^2\bar{x}^2\bar{y}}}{m\sum x_i^2 - m^2\bar{x}^2}$$

$$= \frac{\cancel{m}\,\bar{x}^2\bar{y} - \cancel{m}\sum x_i^2\bar{y}}{\cancel{m}\sum x_i^2 - \cancel{m}\bar{x}^2}$$

$$-\bar{y}\,\frac{\cancel{\left(\bar{x}^2 - \sum x_i^2\right)}}{\cancel{\left(\bar{x}^2 - \sum x_i^2\right)}}$$

$$-\bar{y}$$

Q2

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}^3 + \epsilon_i$$

$$S = \sum_{i=1}^{M} \epsilon_i^2 = \sum_{i=1}^{M} \left(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}^3\right)^2$$

Our aim is to find $\underset{\beta_0 \beta_1 \beta_2 \beta_3}{\arg\min}\, S$, ~~which is achieved.~~ that is

$$\frac{\partial S}{\partial \beta_0} = \frac{\partial S}{\partial \beta_1} = \frac{\partial S}{\partial \beta_2} = \frac{\partial S}{\partial \beta_3} = 0$$

$$\frac{\partial S}{\partial \beta_0} = 2 \sum_{i=1}^{M} y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}^3 = 0$$

$$\frac{\partial S}{\partial \beta_1} = 2 \sum_{i=1}^{M} \left(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}^3\right)(-x_{i1}) = 0$$

$$\frac{\partial S}{\partial \beta_2} = 2 \sum_{i=1}^{M} \left(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}^3\right)(-x_{i2}) = 0$$

$$\frac{\partial S}{\partial \beta_3} = 2 \sum \left(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3}^3\right)(-x_{i3}^3) = 0$$

$$\sum y_i = M\beta_0 + \beta_1 \sum x_{i1} + \beta_2 \sum x_{i2} + \beta_3 \sum x_{i3}^3$$

$$\sum x_{i1} y_i = \beta_0 \sum x_{i1} + \beta_1 \sum x_{i1}^2 + \beta_2 \sum x_{i2} x_{i1} + \beta_3 \sum x_{i3}^3 x_{i1}$$

$$\sum x_{i2} y_i = \beta_0 \sum x_{i2} + \beta_1 \sum x_{i1} x_{i2} + \beta_2 \sum x_{i2}^2 + \beta_3 \sum x_{i3}^3 x_{i2}$$

$$\sum x_{i3}^3 y_i = \beta_0 \sum x_{i3}^3 + \beta_1 \sum x_{i1} x_{i3}^3 + \beta_2 \sum x_{i2} x_{i3}^3 + \beta_3 \sum x_{i3}^6$$

KNN

Let x_i be a data point

c_j be centroids $x \; j \in \{1, 2, \dots, k\}$.

$$j^{*} = \underset{j \in \{1, \dots, k\}}{\operatorname{argmin}} \; |x_i - c_j|$$

j^{*} index of closest centroid

$$c_{j^*} = c_{j^*} \cup \{x_i\}$$

Calculate means

$$c_{j_{new}} = \overline{c_{j^*}}$$

$$\Delta = |c_{j_{new}} - c_j|$$

$$\hat{\Delta} \leq M \quad \text{exit.}$$

$\sigma^2 =$ c_j i

$$\sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

$$\begin{bmatrix} \sum y_i \\ \sum x_{i1} y_i \\ \sum x_{i2} y_i \\ \sum x_{i3}^3 y_i \end{bmatrix} = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \sum x_{i3}^3 \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1} x_{i2} & \sum x_{i1} x_{i3}^3 \\ \sum x_{i2} & \sum x_{i1} x_{i2} & \sum x_{i2}^2 & \sum x_{i2} x_{i3}^3 \\ \sum x_{i3}^3 & \sum x_{i1} x_{i3}^3 & \sum x_{i2} x_{i3}^3 & \sum x_{i3}^6 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

Setting X as the design matrix

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13}^3 \\ \vdots & & & \\ 1 & x_{n1} & x_{n2} & x_{n3}^3 \end{bmatrix}$$

we notice that

$$X^T X = \left[\text{This matrix} \right]$$

also $X^T y = \left[\text{This matrix} \right]$

$$L(\underline{x}, \underline{\theta}) = \prod_{i=1}^{n} (f_i, \underline{\theta}) \qquad x \geq x_m \quad 3$$

Pareto $\dfrac{\alpha x_m^{\alpha}}{x^{\alpha+1}}$

$$L(\underline{x}, \underline{\theta}) = \prod_{i=1}^{n} \frac{\alpha x_m^{\alpha}}{x_i^{\alpha+1}}$$

$$l(\underline{x}, \underline{\theta}) = \sum_{i=1}^{n} \log \left(\frac{\alpha x_m^{\alpha}}{x_i^{\alpha+1}} \right) \quad 2$$

$$= \sum \left(\log \alpha x_m^{\alpha} - \log x_i^{\alpha+1} \right)$$

$$= \sum_{i=1}^{n} \log \alpha x_m^{\alpha} - \sum_{i=1}^{n} \log x_i^{\alpha+1}$$

$$= \sum \log \alpha + \sum \log x_m^{\alpha} - \sum \log x_i^{\alpha+1}$$

4

$$= n\log\alpha + n\log x_m^{\alpha} - \sum_{i=1}^{n}(\alpha+1)\log x_i$$

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

$$\frac{\partial l(\underline{x};\underline{\theta})}{\partial\alpha} = \frac{n}{\alpha} + n\log x_m - \sum_{i=1}^{n}\log x_i = 0 \quad \text{— ①}$$

$$\frac{\partial l(\underline{x};\underline{\theta})}{\partial x_m} = \frac{n\alpha}{x_m} \qquad x_m \le \underline{x}$$

$\dfrac{\partial l(\underline{x};\underline{\theta})}{\partial x_m}$ is valid for $x_m > 0$

So the minimum x_m possible
which means min l
is min $(\underline{x})$

$$\therefore \hat{x}_m = \min(x)$$

For ①

$$\alpha = \frac{n}{\sum\log x_i - n\log x_m}$$

$$= \frac{n}{\sum\log x_i - \sum\log x_m}$$

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log \frac{x_i}{x_m}}$$

where $x_i \geq x_m$.

and $\hat{x}_m = \min(\underline{x})$

ie $$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log \frac{x_i}{\min(\underline{x})}}$$

We want to find $E[\hat{\alpha}]$

where $\hat{\alpha} = \dfrac{n}{\sum_{i=1}^{n} \ln\left(\frac{X_i}{x_m}\right)}$ $\quad x_m > 0$

$$X_i \sim \text{Pareto}(\alpha, x_m)$$

We focus on the denominator,

$$S = \sum_{i=1}^{n} \ln\left(\frac{X_i}{x_m}\right)$$

to understand its distribution, so that we can arrive to the estimate $E[S]$

So let us start by understanding the distribution of $\ln\left(\frac{X_i}{x_m}\right)$

If we define $Y_i = \ln\left(\frac{X_i}{x_m}\right)$

$$\Rightarrow X_i = x_m e^{Y_i}$$

The pdf can be found by differentiating the CDF

$$F_Y(y) = P(Y \leq y) \quad \forall y$$

$$= P\left(\ln\left(\frac{X}{x_m}\right) \leq y\right)$$

$$= P\left(\frac{X}{x_m} \leq e^{y}\right)$$

$$= P\left(X \leq x_m e^{y}\right)$$

the CDF of Pareto (α, x_m)

$$F_X(x) = P(X \leq x) = 1 - \left(\frac{x_m}{x}\right)^{\alpha} \quad x \geq x_m$$

so $$F_Y(y) = P\left(X \leq x_m e^{y}\right)$$

$$= 1 - \left(\frac{x_m}{x_m e^{y}}\right)^{\alpha}$$

$$= 1 - e^{-y\alpha} \quad y \geq 0$$

and hence

$$Y = \ln\left(\frac{X}{x_m}\right) \sim \text{Exponential}(\alpha)$$

we now move to $S = \sum_{i=1}^{n} \ln\left(\frac{X_i}{x_m}\right)$

where all X_i are iid and exponential with rate α, so S follows a Gamma distribution

$$S \sim \text{Gamma}(n, \alpha)$$

the pdf of Gamma distr.

$$f_S(s) = \frac{\alpha^{n}}{\Gamma(n)} s^{n-1} e^{-\alpha s}$$

~~where $E[S] = \frac{n}{\alpha}$~~

To find $E[\hat{\alpha}]$

since $\hat{\alpha} = \frac{n}{S}$ $\quad E[\hat{\alpha}] = n\,E\left[\frac{1}{S}\right]$

Consider ~~a Gamma~~

$$S \sim \text{Gamma}(k, \lambda)$$

pdf $\quad f_S(s) = \frac{\lambda^k}{\Gamma(k)} s^{k-1} e^{-\lambda s} \qquad s > 0$

$$E\left[\frac{1}{S}\right] = \int_0^\infty \frac{1}{s} f(s)\,ds$$

$$= \int_0^\infty \frac{1}{s} \cdot \frac{\lambda^k}{\Gamma(k)} s^{k-1} e^{-\lambda s}\,ds$$

$$= \frac{\lambda^k}{\Gamma(k)} \int_0^\infty \frac{s^{k-1}}{s} e^{-\lambda s}\,ds$$

$$= \frac{\lambda^k}{\Gamma(k)} \int_0^\infty s^{k-2} e^{-\lambda s}\,ds$$

$*$ $\boxed{\text{since } \Gamma(k) = \int_0^\infty s^{k-1} e^{-s}\,ds}$

$\cancel{\int_0^\infty s^{k-2} e^{-\lambda s}\,ds}$

$$\cancel{\int_0^\infty s^{(k-2)} e^{-(k-1)} e^{(k+1)-\lambda s}}$$

Let $u = \lambda s$, $s = \frac{u}{\lambda}$ $ds = \frac{du}{\lambda}$

$$\int_0^\infty s^{k-2} e^{-\lambda s} \, ds$$

$$= \int_0^\infty \left(\frac{u}{\lambda}\right)^{k-2} e^{-u} \frac{1}{\lambda} \cdot du$$

$$= \frac{1}{\lambda^{k-1}} \int_0^\infty u^{k-2} e^{-u} \, du$$

$$= \frac{1}{\lambda^{k-1}} \Gamma(k-1)$$

$$\therefore E\left[\frac{1}{S}\right] = \frac{\lambda^k}{\Gamma(k)} \cdot \frac{1}{\lambda^{k-1}} \Gamma(k-1)$$

$$= \lambda \frac{\Gamma(k-1)}{\Gamma(k)} = \lambda \frac{\cancel{(k-1)}\,\cancel{\Gamma(k-1)}}{(k-1)\,\Gamma\cancel{(k-1)}}$$

$$= \frac{\lambda}{k-1}$$

(ii)

$$E\left[\frac{1}{\delta^2}\right] = \int_0^\infty \frac{1}{s^2} f(s)\, ds$$

$$= \frac{\lambda^k}{\Gamma(k)} \int_0^\infty \frac{s^{k-1}}{s^2} e^{-\lambda s}\, ds$$

$$= \frac{\lambda^k}{\Gamma(k)} \int_0^\infty s^{k-3} e^{-\lambda s}\, ds$$

let $u = \lambda s$, $s = \frac{u}{\lambda}$ $ds = \frac{1}{\lambda} du$

$$\int_0^\infty s^{k-3} e^{-\lambda s}\, ds$$

$$= \int_0^\infty \left(\frac{u}{\lambda}\right)^{k-3} e^{-u} \frac{1}{\lambda}\, du$$

$$\frac{1}{\lambda^{k-2}} \int_0^\infty u^{k-3} e^{-u}\, du = \frac{1}{\lambda^{k-2}} \Gamma(k-2)$$

$$E\left[\frac{1}{\delta^2}\right] = \frac{\cancel{\lambda^k}}{\Gamma(k)} \cdot \frac{1}{\cancel{\lambda^{k-2}}} \Gamma(k-2) = \lambda^2 \frac{\Gamma(k-2)}{\Gamma(k)}$$

$$= \frac{\lambda^2 \, \cancel{\Gamma(k-2)}}{\cancel{\Gamma(k-2)}(k-2)(k-1)} = \frac{\lambda^2}{(k-2)(k-1)}$$

5

since $\hat{\alpha} = \frac{n}{S}$

and $\hat{\alpha}^2 = \frac{n^2}{S^2}$

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

$$E(\hat{\alpha}^2) = \frac{n^2 \alpha^2}{(n-1)(n-2)}$$

$$Var(\hat{\alpha}) = E(\hat{\alpha}^2) - \left(E[\hat{\alpha}]\right)^2$$

$$= n^2 E\left[\frac{1}{S^2}\right] - n^2 \left(E\left[\frac{1}{S}\right]\right)^2$$

$$= \left(\frac{n^2\alpha^2}{(n-1)(n-2)} - \frac{n^2\alpha^2}{(n-1)^2}\right)$$

$$= n^2\alpha^2 \left(\frac{(n-1) - (n-2)}{(n-1)^2 (n-2)}\right)$$

$$= \frac{n\alpha^2}{(n-1)^2 (n-2)}$$

this variance is greater than the CRLB

$$\frac{n\alpha^2}{(n-1)^2(n-2)^2} > \frac{\alpha^2}{n} \quad \forall n > 2$$

6

as $n \to \infty$, the variance approaches $\frac{\alpha^2}{n}$, hence $\hat{\alpha}$ becomes asymptotically unbiased

---

Let us now consider x_m, where we have found that

$$\hat{x}_m = \min(x_1, \ldots, x_m)$$

where $X_i \sim \text{Pareto}(\alpha, x_m)$

If $Y = \min(x_1, \ldots, x_m)$

CDF $F_Y(y) = 1 - \left(\frac{x_m}{y}\right)^{n\alpha}$

PdF $f_Y(y) = \frac{d}{dy}\left(1 - \frac{x_m^{n\alpha}}{y^{n\alpha}}\right), \quad y \geq x_m$

$$= -\left(-n\alpha \frac{x_m^{n\alpha}}{y^{n\alpha+1}}\right)$$

$$= \frac{n\alpha\, x_m^{n\alpha}}{y^{n\alpha+1}} \qquad y \geq x_m$$

$$E[\hat{x}_m] = E[y]$$

$$= \int_{x_m}^{\infty} \cancel{y}\; n\alpha \, \frac{x_m^{n\alpha}}{y^{n\alpha+1}} \, dy$$

x^2

$\frac{x^3}{3}$

$$= n\alpha \, x_m^{n\alpha} \int_{x_m}^{\infty} \frac{1}{y^{n\alpha}}$$

$y^{-n\alpha}$

$$= n\alpha \, x_m^{n\alpha} \left[\frac{y^{-n\alpha+1}}{-n\alpha+1} \right]_{x_m}^{\infty}$$

$\lim_{y \to \infty} = 0$

$$= n\alpha \, x_m^{n\alpha} \left(- \frac{x_m^{-n\alpha+1}}{-n\alpha+1} \right)$$

$n\alpha > 1$

$$= n\alpha \, x_m^{n\alpha} \left(\frac{x_m^{1-n\alpha}}{n\alpha-1} \right)$$

$$= \frac{n\alpha \, x_m}{n\alpha-1}$$

$E[\hat{x}_m] \neq x_m$ that is $\hat{x}_m$ is biased!

$\hat{x}_m = \min(x_1 .. x_n)$ is a ~~biased est~~ + does not attain CRLB

8

Let us consider the sufficiency of $\hat{x}_m$

as $$f(x_i ; x_m) = \begin{cases} \dfrac{\alpha x_m^{\alpha}}{x_i^{\alpha+1}} & x_i \geq x_m \\ 0 & \text{otherwise} \end{cases}$$

$$f(x_1, \dots, x_n ; x_m) = \prod_{i=1}^{n} \frac{\alpha x_m^{\alpha}}{x_i^{\alpha+1}} \qquad x_m \leq x_i$$

$$= \alpha^n x_m^{\alpha n} \prod_{i=1}^{n} x_i^{-(\alpha+1)} \qquad x_m \leq x_i$$

we can write this using an indicator fn

$$= \alpha^n x_m^{\alpha n} \prod_{i=1}^{n} x_i^{-\alpha-1} \cdot 1_{\{x_m \leq \min x_i\}}$$

$$= \alpha^n x_m^{\alpha n} \cdot 1_{\{x_m \leq \min x_i\}} \prod x_i^{-\alpha-1}$$

Let $T(x) = \min(x_1, \dots, x_n)$

$$= \left(x_m^{\alpha n} \cdot 1_{\{x_m \leq T(x)\}} \right) \left(\alpha^n \prod_{i=1}^{n} x_i^{-\alpha-1} \right)$$

$$= g(T(x) ; x_m) \; h(x)$$

So by Neyman - Fisher
$T(x) = \min(x_1 \dots x_n)$ is a sufficient statistic of x_m

Sufficiency of $\hat{\alpha}$

$$f(x_i ; \alpha) = \frac{\alpha x_m^{\alpha}}{x_i^{\alpha+1}} \qquad x_i \geq x_m$$

$$f(x_1 \ldots x_n ; \alpha) = \prod_{i=1}^{n} \frac{\alpha x_m^{\alpha}}{x_i^{\alpha+1}}$$

$$= \alpha^n x_m^{\alpha n} \prod_{i=1}^{n} x_i^{-\alpha-1}$$

$$= \alpha^n \, e^{\ln x_m^{\alpha n}} \, e^{\ln \prod_{i=1}^{n} x_i^{-\alpha-1}}$$

$$= \alpha^n \, e^{\alpha n \ln x_m} \, e^{-\alpha-1 \sum_{i=1}^{n} \ln x_i}$$

$$= \alpha^n \, e^{(\alpha n \ln x_m) + (-\alpha-1) \sum \ln x_i}$$

let $T(x) = \sum \ln x_i$

$$= \left(\alpha_m \, e^{(\alpha n \ln x_n) + (-\alpha-1) T(x)} \right) \times 1.$$

$$= \quad g\big(T(x) \mid \alpha\big) \quad h(x)$$

so $\sum \ln x_i$ is a sufficient stat of α

but $\hat{\alpha} = \dfrac{n}{\sum \ln \frac{x_i}{x_m}} = \dfrac{n}{\sum (\ln x_i - \ln x_m)} = \dfrac{n}{(\sum \ln x_i) - (n \ln x_m)}$

= so $\hat{\alpha}$ is a function of $T(x)$

~~For $\hat{\alpha}$, deriving M.L.E.~~

$$\cancel{\frac{\partial l(x;\alpha,x_m)}{\partial \alpha} = 0}$$

$$\cancel{\hat{\alpha} = \log \bar{x} - \log x_m}$$

~~the~~

For α, let us derive the MLE

for n observations $(x_1 \ldots x_n)$ $\quad x_i \geq x_m$

$$L(\alpha) = \prod_{i=1}^{n} f(x_i;\alpha,x_m)$$

$$= \prod_{i=1}^{n} \frac{\alpha\, x_m^{\alpha}}{x_i^{\alpha+1}} \qquad \begin{array}{l} x \geq x_m \\ \alpha > 0 \\ x_m > 0 \end{array}$$

$$= \alpha^{n} x_m^{n\alpha} \prod_{i=1}^{n} \frac{1}{x_i^{\alpha+1}}$$

log likelihood.

$$l(\alpha) = \log L(\alpha) =$$

$$= n\log\alpha + \cancel{\alpha}\,n\alpha\log x_m + \sum_{i=1}^{n} \frac{1}{x_i^{\alpha+1}}$$

$$= n\log\alpha + n\alpha\log x_m - (\alpha+1)\sum_{i=1}^{n} \log x_i$$

$$\frac{\partial l(\alpha)}{\partial \alpha} = \frac{n}{\alpha} + n \log x_n - \sum_{i=1}^{n} \log x_i$$

for max likelihood

$$\frac{\partial l(\alpha)}{\partial \alpha} = 0$$

$$\frac{n}{\alpha} = \sum_{i=1}^{n} \log x_i - n \log x_n.$$

$$= \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \log x_n$$

$$= \sum_{i=1}^{n} \log\left(\frac{x_i}{x_n}\right)$$

thus

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log\left(\frac{x_i}{x_n}\right)}.$$

Consider the likelihood for a single observation

$$L(x; \alpha, x_m) = \frac{\alpha\, x_m^{\alpha}}{x^{\alpha+1}} \qquad \begin{array}{l} x \geqslant x_m \\ \alpha > 0 \\ x_m > 0 \end{array}$$

$$\ell(x; \alpha, x_m) = \log \alpha + \alpha \log x_m - (\alpha+1) \log x.$$

Score for x_m

$$\frac{\partial \ell(x; \alpha, x_m)}{\partial x_m} = \frac{\alpha}{x_m} \qquad \left| \begin{array}{l} \alpha > 0 \\ x_m > 0 \\ x \geqslant x_m. \end{array} \right.$$

$$\frac{\partial^2 \ell(x; \alpha, x_m)}{\partial x_m^2} = -\frac{\alpha}{x_m^2} \qquad \left| \begin{array}{l} \alpha > 0 \\ x_m > 0 \\ x \geqslant x_m \end{array} \right.$$

The expected value

$$\mathbb{E}\left[\frac{\partial^2 \ell(x; \alpha, x_m)}{\partial x_m^2}\right]_{\substack{\alpha > 0 \\ x_m > 0 \\ x \geqslant 0}} = \frac{-\alpha}{x_m^2} \qquad \begin{array}{l} \alpha > 0 \\ x_m > 0 \\ x \geqslant x_m. \end{array}$$

Fisher information for n samples. $I(x_m) = +\dfrac{n\alpha}{x_m^2}$ $\qquad \begin{array}{l} \alpha > 0 \\ x_m > 0 \end{array}$

So CRLB is $\boxed{\mathrm{Var}(\hat{x}_m) \geqslant \dfrac{x_m^2}{n\alpha}.}$ $\qquad \begin{array}{l} \alpha > 0 \\ x_m > 0 \end{array}$

Score function for α

$$\frac{\partial \ell(x; \alpha, x_m)}{\partial \alpha} = \frac{1}{\alpha} + \log x_m - \log x \qquad \begin{array}{l} x \geq x_m \\ \alpha > 0 \\ x_m > 0 \end{array}$$

$$\frac{\partial^2 \ell(x; \alpha, x_m)}{\partial \alpha} = -\frac{1}{\alpha^2} - \frac{1}{[illegible]} \qquad \begin{array}{l} x \geq x_m \\ \alpha > 0 \\ x_m > 0 \end{array}$$

$$= [illegible]$$

$$= -\frac{[illegible]}{\alpha^2};$$

Expected Value

$$\mathbb{E}\left[\frac{\partial \ell(x; \alpha, x_m)}{\partial \alpha}\right] = -\frac{1}{\alpha^2}$$

Fisher information for ~~m~~ ~~=~~ observations

$$I_m(\alpha) = \frac{m}{\alpha^2}.$$

CRLB

$$\boxed{Var(\hat{\alpha}) \geq \frac{\alpha^2}{m}.}$$