VI REVISION

Example: The Pareto distribution (or 'power law' distribution) has been used by economists to model income distributions and by astronomers to model size (eg diameter) distributions of 'small' objects in space eg comets/asteroids. The pdf has 2 parameters $\underset{\sim}{\theta} = (\phi, y_0)$ and is given by

$$f_Y(y;\underset{\sim}{\theta}) = \phi y_0^{\phi} y^{-(\phi+1)}, \quad y > y_0 > 0, \text{ zero elsewhere } (\phi > 0)$$

Exercise for Prob 2 not Stat. Theory:

Show it integrates to 1 and

(i) $E[Y] \equiv \mu = \dfrac{\phi y_0}{\phi - 1}$ (provided $\phi > 1$)

(ii) $E[Y^2] = \dfrac{\phi y_0^2}{\phi - 2}$ (..... $\phi > 2$)

& hence (iii) $Var[Y] \equiv \sigma^2 = \dfrac{\phi y_0^2}{(\phi-1)^2(\phi-2)}$ (........)

Moreover the cdf is

$$F(y) = P[Y \leq y] = 1 - (y_0/y)^{\phi} \quad y > y_0$$

(and zero for $y \leq y_0$)

so the q^{th} percentile, defined by $F(y_q) = q/100$

is given by $y_q = y_0\left(1 - q/100\right)^{-1/\phi}$

or $\log y_q = \log y_0 - \frac{1}{\phi}\log\left(1 - \frac{q}{100}\right)$

Also show (using 1, 1 transformation) that

$U = \log(Y/y_0)$ has pdf $\phi e^{-\phi u}$ $u > 0$, zero elsewhere

ie exponential with mean ϕ^{-1}

Q1: Find the MOMEs when (a) y_0 known (b) y_0 unknown

(a) $p = 1$ so equate (only) the first sample & pop. moments to solve for ϕ

$$\bar{y} = \frac{\phi y_0}{\phi - 1} \; (\equiv \mu) \longrightarrow \phi(\bar{y} - y_0) = \bar{y}$$

$(\phi > 1)$

$$\tilde{\phi} = \frac{\bar{y}}{\bar{y} - y_0}$$

(b) (equate the first two sample & pop moments, solving for ϕ & y_0)

$$\bar{y} = \frac{\phi y_0}{\phi - 1} \qquad \frac{1}{n}\sum y_i^2 = \frac{\phi y_0}{\phi - 2}$$

Eliminate y_0:

$$\frac{\sum y_i^2}{n\bar{y}^2} = \frac{\phi \cancel{y_0^2}}{\phi-2}\frac{(\phi-1)^2}{\phi^2 \cancel{y_0^2}} = 1 + \frac{1}{\phi(\phi-2)}$$

so $\phi(\phi-2) = 1\Big/\left(\frac{\sum y_i^2}{n\bar{y}^2} - 1\right) = \frac{n\bar{y}^2}{\sum(y_i-\bar{y})^2} > 0$ so quadratic has 2 solutions

and $$\tilde{\phi} = \frac{2 + \sqrt{4 + 4\frac{n\bar{y}^2}{\sum(y_i-\bar{y})^2}}}{2} = 1 + \sqrt{1 + \frac{n\bar{y}^2}{\sum(y_i-\bar{y})^2}}$$
$$= 1 + \sqrt{\frac{\sum y_i^2}{\sum(y_i-\bar{y})^2}}$$

and therefore $$\tilde{y}_0 = \frac{\tilde{\phi}-1}{\tilde{\phi}}\bar{y} = \frac{\bar{y}}{1 + \sqrt{\frac{\sum(y_i-\bar{y})^2}{\sum y_i^2}}}$$

Q2: Find the MLEs in the 2 cases

(a) y_0 known

$$L(\phi; y) = \phi^n y_0^{n\phi}\left(\prod y_i\right)^{-(\phi+1)} \quad ; y_i > y_0 \; i=1,\dots,n$$
(see (b))

so $$\ell(\phi; y) = n\log\phi + n\phi\log y_0 - (\phi+1)\sum\log y_i$$

$$\frac{\partial \ell}{\partial \phi} = 0 \Rightarrow \frac{n}{\phi} + n\log y_0 - \sum\log y_i = 0$$

$$\longrightarrow \hat{\phi} = \frac{n}{\left(\sum\log y_i - n\log y_0\right)} = \frac{n}{\sum\log(y_i/y_0)}$$

Remark Note that this is the same MLE as when we transform the data to $u_i = \log(y_i/y_0)$ $i=1,\dots,n$ obtaining (as $U \sim Exp(\phi)$) a random sample from $Exp(\phi)$ with MLE $\hat{\phi} = n/\sum u_i \left(\text{or } \frac{1}{\bar{u}}\right)$

Alternatively plot L against ϕ (for fixed y_0)

$$\phi^n \exp[-\phi \underbrace{\sum\log(y_i/y_0)}_{\sum u_i}]$$

with maximum at $\hat{\phi} = \frac{n}{\sum u_i}$

(b) For y_0 unknown

$L(\underset{\sim}{\theta}; \underset{\sim}{y}) = \phi^n y_0^{n\phi} \left(\prod y_i\right)^{-(\phi+1)}$

$\Rightarrow l(\underset{\sim}{\theta}; \underset{\sim}{y}) = n\log\phi + n\phi\log y_0 - (\phi+1)\sum\log y_i$

$y_i > y_0 \quad i = 1, \ldots, n$, Zero otherwise

is an increasing function of y_0 (for fixed $\phi > 0$)

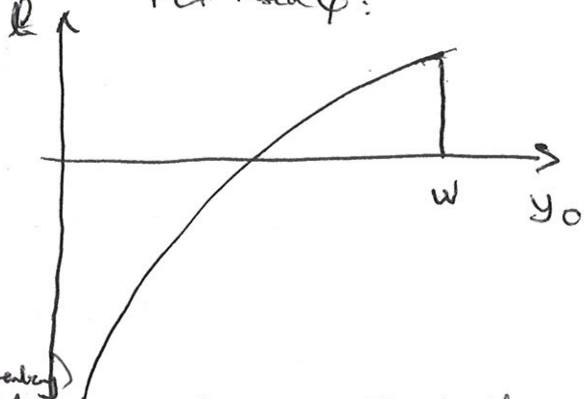
so it is maximised at the largest possible value of y_0 (for fixed $\underset{\sim}{y}$)

Now $y_0 < y_i \quad i = 1, \ldots, n$ so $y_0 < \min_i(y_i) = w$ (say)

ie $\hat{y}_0 = \min_i y_i$

and by (a) $\hat{\phi} = \dfrac{n}{\underbrace{\sum_{i=1}^{n}\log(y_i/\hat{y}_0)}_{>0}}$

Alternatively sketch l (or L) for fixed ϕ:



Q2 (Supplementary) What is the MLE of the q^{th} percentile $y_q = y_0(1-q/100)^{-1/\phi}$?

(a) $y_0(1-q/100)^{\sum\log(y_i/y_0)/n}$ (b) $\hat{y}_0(1-q/100)^{\sum\log(y_i/\hat{y}_0)/n}$

<u>Q3</u> Is the Pareto in the exponential family in (a) or (b)?

$f_Y(y;\underset{\sim}{\theta}) = \exp\left[\log\phi + \phi\log y_0 - (\phi+1)\log y\right]$

$= \exp\left[\underset{b(\phi)}{-\phi}\,\underset{a(y)}{\log y} + \underbrace{\phi\log y_0 + \log\phi}_{c(\phi)} \underbrace{-\log y}_{d(y)}\right]$

(a) <u>Yes</u> (see above)

(b) although $f_Y(y;\underset{\sim}{\theta})$ <u>is</u> of the correct form

the range depends on $\underset{\sim}{\theta}$ so it is <u>not</u> in the E.F.

~~What further property does S possess that enables the search for a MVUE?~~

~~Given that E[~~

Q4 (a) Show that $S = \sum \log Y_i$ is sufficient for ϕ when y_0 is known and mention another property of S by quoting (but ~~not~~ proving) a Theorem.

Given that $U_i = \log(Y_i/y_0) \sim \text{Exp}(\phi)$ and that $X \sim \text{Gamma}(n;\phi)$ has moments

$$E[X^r] = \frac{\Gamma(n+r)}{\Gamma(n)\phi^r} \qquad r > -n,$$

Find an unbiased estimator of ϕ which is a function of S (and $\log y_0$), (and is therefore a MVUE)

[HINT: start with the MLE $\hat{\phi} = \frac{n}{S - n\log y_0} = \frac{n}{\sum \log(y_i/y_0)}$]

From Q3, since $t(y_i) = \log y_i$ the Exponential Family Theorem shows that S is both sufficient and complete.

Now $S - n\log y_0 = \sum U_i \sim \text{Gamma}(n;\phi)$ as $U_i \sim \text{Exp}(\phi)$ $i=1,\dots,n$.

Hence $$E\left[\frac{n-1}{S - n\log y_0}\right] = \frac{\Gamma(n-1)\, n-1}{\Gamma(n)\,\phi^{-1}} = \phi$$

gives an unbiased estimator which is the MVUE in case (a).

Q5. Show that (S, W), where $W = \min_i Y_i = \hat{y}_0$, is (jointly) sufficient for $\underset{\sim}{\theta} = (\phi, y_0)^T$.

$$L(\underset{\sim}{\theta}; y) = \phi^n y_0^{n\phi} \left(\prod y_i\right)^{-(\phi+1)} \qquad y_i > y_0 \quad i=1,\dots,n$$

(zero otherwise)

so we can write

$$L(\underset{\sim}{\theta}; y) = \phi^n y_0^{n\phi} \exp\left[-\phi \sum \log y_i\right] I(w > y_0) \cdot \frac{1}{\prod_{i=1}^n y_i}$$

as $w = \min_i y_i > y_0 \iff y_i > y_0$ $i=1,\dots,n$.

where $I(\cdot)$ is the indicator function (1 if true, zero if False)

By the (Neyman) Factorization theorem (S, W) is sufficient as we have written L in the form

$$g(s, w; \underset{\sim}{\theta})\, h(\underset{\sim}{y}) \qquad \text{where } h(\underset{\sim}{y}) = \frac{1}{\prod_{i=1}^n y_i}$$

70

Q6 From the cdf $F(y) = 1-(y_0/y)^\phi$ of the Pareto distribution show that W also has a Pareto distribution but with parameters $n\phi$ and y_0. Hence show that W is approximately unbiased for y_0 if n is large, and correct W for bias (in the case when ϕ is known).

W is a minimum so look at

$$\Pr[W > w] = \Pr[Y_1 > w, \dots, Y_n > w]$$
$$= \prod_{i=1}^{n} \Pr[Y_i > w] = [1 - F(w)]^n = (y_0/w)^{n\phi}$$

Hence the cdf of W is $F_W(w) = 1 - (y_0/w)^{n\phi}$

which gives the required Pareto distribution

Now by the properties of the Pareto

$$E[W] = \frac{n\phi y_0}{n\phi - 1} = y_0\left(1 - \frac{1}{n\phi}\right)^{-1}$$
$$\doteqdot y_0\left(1 + \frac{1}{n\phi} + \dots\right)$$

so W is approx. unbiased

and $\frac{n\phi - 1}{n\phi} W$ is exactly unbiased for y_0 for all n.

Q7. By using $E[\log Y] = \log y_0 + 1/\phi$ when $Y \sim \text{Pareto}(\phi, y_0)$ find an unbiased estimate of $\log y_0$ that is a function of S and $\log W$

From Q4, $E[S] = n \log y_0 + n/\phi$
and from Q6, $E[\log W] = \log y_0 + \frac{1}{n\phi}$

Hence $E\left[\log W - S/n^2\right] = \log y_0 (1 - 1/n)$

so that $\frac{n}{n-1}\left[\log W - \frac{S}{n^2}\right]$ is a UE of $\log y_0$.

[which is the MVUE if the family is complete]

Q8. In case (a) when y_0 is known, Find the approximate sampling distribution of $\hat{\phi}$ when n is large and give an asymptotic pivot for ϕ and an approximate 95% C.I. for ϕ

$$\hat{\phi} \sim N(\phi\,;\ CRLB(\phi))$$

where
$$l(\phi\,;y) = n\log\phi + n\phi\log y_0 - (\phi+1)\sum\log y_i$$
$$\rightarrow \frac{dl}{d\phi} = \frac{n}{\phi} + n\log y_0 - \sum\log y_i \quad \text{as before}$$
$$\rightarrow \frac{d^2l}{d\phi^2} = -\frac{n}{\phi^2}$$

so that the Fisher information is $E\left[-\frac{d^2l}{d\phi^2}\right] = \frac{n}{\phi^2}$

and $CRLB(\phi) = 1/\frac{n}{\phi^2} = \phi^2/n$

$$\frac{\hat{\phi}-\phi}{\sqrt{\widehat{CRLB}(\phi)}} = \frac{\hat{\phi}-\phi}{\hat{\phi}/\sqrt{n}} = \sqrt{n}\left(1-\phi/\hat{\phi}\right)$$

$\sim N(0,1)$ approx is an asymptotic pivot for ϕ

so that $\hat{\phi} \pm 1.96\,\hat{\phi}/\sqrt{n}$ gives an approximate 95% C.I.

Q9: By stating (but not proving) a well known result, Find a most powerful test of size α of

$$H_0: \phi = \phi_0$$
versus $H_1: \phi = \phi_1$ where $\phi_1 > \phi_0$

in the case when y_0 is known

The Neyman Pearson Lemma shows that the MP test of size α H_0 versus H_1 is to reject H_0 if and only if

$$\lambda(y) = \frac{L(\phi_0\,;y)}{L(\phi_1\,;y)} \leq a_\alpha$$

where a_α is chosen to give size α.

From Q2(a)

$$\lambda(y) = \left(\frac{\phi_0}{\phi_1}\right)^n y_0^{\,n(\phi_1-\phi_0)} \exp\left[(\phi_1-\phi_0)\sum\log Y_i\right] \leq a_\alpha$$

$$\Leftrightarrow \sum\log Y_i \leq c_\alpha \qquad \Leftrightarrow (\phi_1-\phi_0)\sum\log Y_i \leq b_\alpha$$

Now $S - n\log y_0 \sim \text{Gamma}(n;\phi)$

so under H_0

$$P[S \le c_\alpha] = F_n(c_\alpha - n\log y_0) = \alpha$$

where $F_n(\cdot)$ is the cdf of $\text{Gamma}(n;\phi_0)$

~~Hence $c_\alpha = n\log y_0 + F_n^{-1}(\alpha)$ 100 th percentile~~

For large n $F_n(\cdot)$ is approximately a normal cdf with the same mean (n/ϕ_0) and variance (n/ϕ_0^2)

so $$\alpha = \Phi\left(\frac{c_\alpha - n\log y_0 - n/\phi_0}{\sqrt{n/\phi_0^2}}\right)$$

ie $$c_\alpha = n\log y_0 - z_\alpha\sqrt{n}/y_0 + n/\phi_0$$

OMIT

Find the power of the test.

Under H_1, $F_n(\cdot)$ is approx. a normal cdf with same mean n/ϕ_1 & variance n/ϕ_0^2 so

$$\beta = P[S \le c_\alpha] = F_n[c_\alpha - n\log y_0] = \Phi\left(\frac{c_\alpha - n\log y_0 - n/\phi_1}{\sqrt{n/\phi_1^2}}\right)$$

$$= \Phi\left(\frac{n/\phi_0 - n/\phi_1 - z_\alpha\sqrt{n}/y_0}{\sqrt{n/\phi_1^2}}\right)$$

Q10. Does a UMP test (of size α) exists in each of these cases? If it does, give its critical region, If not, why not?

(a) $H_0: \phi = \phi_0 \quad H_1: \phi > \phi_0$: Yes, and same as in Q9

(b) $H_0: \phi = \phi_0$: $H_1: \phi < \phi_0$: Yes but now $S \ge c_\alpha$

where $c_\alpha = n\log y_0 + z_\alpha\sqrt{n}/y_0 + n/\phi_0$

(c) $H_0: \phi = \phi_0$: $H_1: \phi \ne \phi_0$

$\nexists$ UMP, as CR of MP test is different for (a) and (b).

Q11: Find the Wilks' statistic for testing

$$H_0: \phi = \phi_0 \quad \text{vs} \quad H_1: \phi = \phi_1$$

(a) For known y_0

(b) For unknown y_0

and state its large sample distribution

(a) $\hat{\phi} = \dfrac{n}{S - n\log y_0}$ and $\hat{\phi}_0 = \phi_0$

so from Q2(a), as $\ell(\phi; y) = n\log\phi - \phi(S - n\log y_0) - S$

$$-2\log \Lambda(\underset{\sim}{y}) = 2\left[\ell[\hat{\phi}; \underset{\sim}{y}] - \ell[\hat{\phi}_0; \underset{\sim}{y}]\right]$$

$$= 2\left[n\log n - n\log(S - n\log y_0) - n - n\log\phi_0 + \phi_0(S - n\log y_0)\right]$$

$p = 1$ $p_0 = 0$ so $-2\log\Lambda(\underset{\sim}{y}) \sim \chi^2_1$ under H_0 for large n.

(b) $\hat{\phi} = \dfrac{n}{S - n\log\hat{y}_0}$ and under H_0, we maximise

$$\ell(\phi_0, y_0; y) = n\log\phi_0 + n\phi_0\log y_0 - (\phi_0 + 1)S \quad , w > y_0$$

w.r.t. y_0

As before maximised at $\hat{y}_0 = w$ so

$$\hat{\underset{\sim}{\theta}}_0 = (\phi_0, \hat{y}_0)$$

$$-2\log\Lambda(\underset{\sim}{y}) = 2\left[\ell(\hat{\phi}, w; \underset{\sim}{y}) - \ell(\phi_0, w; \underset{\sim}{y})\right]$$

$$= 2\left[n\log n - n\log(S - n\log w) - n - n\log\phi_0 + \phi_0(S - n\log w)\right]$$

$p = 2$, $p_0 = 1$ so $-2\log\Lambda(\underset{\sim}{y}) \sim \chi^2_1$ under H_0.