3. A statistical decision problem.

Consider:

$i \in [\![1;n]\!]$: individual i in a population of size n.

$y = (y_i)_{i \in [\![1;n]\!]} \in \mathbb{R}^n$: the net income vector of the population.

$\tau = (\tau_i)_{i \in [\![1;n]\!]} \in \mathbb{R}^n$: the net income adjustment vector of the population

$L(\tau, y) = \sum_{i=1}^{n} (z - y_i - \tau_i)^2$: the government's income transfer policy performance-evaluating function.

$\hat{y} = (\hat{y}_i)_{i \in [\![1;n]\!]}$: the net income noisy estimate vector observed by the government.

Assume $\hat{y}_i \underset{iid}{\sim} \mathcal{N}(y_i, 1)$.

(a) The key elements of the statistical decision problem:

- The state space (where y lives): $\Theta = \mathbb{R}^n$
- The observations (where random variable $\hat{y}$ lives): $\mathcal{Y} = \mathbb{R}^n$, with draws distributed along $\mathcal{N}(y, I_n)$
- The action space (where τ lives): $\mathcal{A} = \mathbb{R}^n$
- The decision function (the function corresponding to the government mapping observation to action):

$$\delta : \mathcal{Y} \longrightarrow \mathcal{A}$$

$$\hat{y} \longmapsto \underset{\tau}{\operatorname{argmin}} \left\{ \mathcal{L}(\tau, \hat{y}) \right\}$$

$$= \underset{\tau}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} (z - \hat{y}_i - \tau_i)^2 \right\}$$

- The loss function (the loss that the government has to "pay" in reference to the true state of nature when actions are taken):

$$\mathcal{L} : \mathcal{A} \times \Theta \longrightarrow \mathbb{R}$$

$$(\tau, y) \longmapsto \mathcal{L}(\tau, y)$$

$$= \sum_{i=1}^{n} (z - y_i - \tau_i)^2$$

• The risk (the average loss that the government has to "pay" in reference to the true state of nature when a decision function is chosen):

$$\mathbb{R} : \mathcal{F}(\mathcal{Y}, \mathcal{A}) \times \Theta \longrightarrow \mathbb{R}$$
$$(\delta \times y) \longmapsto \mathbb{E}_{\hat{y}}\left[L(\delta(\hat{y}), y) \,\middle|\, y \right]$$
$$= \int_{\mathcal{Y}} L(\delta(t), y)\, dP(\hat{y} = t \mid y)$$

(b) Assume the decision rule is:

$$\tau_i^{(c)}(\hat{y}) = c\,(z - \hat{y}_i), \quad \text{with } c \in \mathbb{R}.$$

The risk then is:

$$R(\tau^{(c)}, y) = \mathbb{E}_{\hat{y}}\left(L\left[\left(\tau_i^{(c)}(\hat{y})\right)_{i \in [\![1;n]\!]}, y \right] \,\middle|\, y \right)$$

$$= \mathbb{E}_{\hat{y}}\left[\sum_{i=1}^{n} \left(z - y_i - \tau_i^{(c)}(\hat{y})\right)^2 \,\middle|\, y \right]$$

$$= E_{\hat{y}}\left[\sum_{i=1}^{n}\left(z - y_i - c\left(z - \hat{y}_i\right)\right)^2 \Big| \, y\right]$$

$$= \sum_{i=1}^{n} E_{\hat{y}}\left[\left(z - y_i - c\left(z - \hat{y}_i\right)\right)^2 \Big| \, y\right]$$

$$= \sum_{i=1}^{n} E_{\hat{y}}\left[\left(z - y_i - cz + c\hat{y}_i\right)^2 \Big| \, y\right]$$

$$= \sum_{i=1}^{n} E_{\hat{y}}\left[\left(z - y_i - cz\right)^2 + \left(c\hat{y}_i\right)^2 + 2\left(z - y_i - cz\right)c\hat{y}_i \,\Big|\, y\right]$$

$$= \sum_{i=1}^{n}\left[\left(z - y_i - cz\right)^2 + c^2 E_{\hat{y}}\left(\hat{y}_i^2 | y\right) + 2\left(z - y_i - cz\right)c E_{\hat{y}}\left(\hat{y}_i | y\right)\right]$$

$$= \sum_{i=1}^{n}\left[\left(z - y_i - cz\right)^2 + c^2\left(V_{\hat{y}}\left(\hat{y}_i | y\right) + \left[E_{\hat{y}}\left(\hat{y}_i | y\right)\right]^2\right) + 2\left(z - y_i - cz\right)c\, y_i\right]$$

$$= \sum_{i=1}^{n}\left[\left(z - y_i - cz\right)^2 + c^2\left(1 + y_i^2\right) + 2\left(z - y_i - cz\right)c\, y_i\right]$$

<u>ie</u> $R(\tau^{(c)}, y)$

$$= \sum_{i=1}^{n} \left[(z - y_i - cz)^2 + (z - y_i - cz)\, 2c y_i + c^2 (1 + y_i^2) \right]$$

Since we focus on a single individual i, we focus on one term of this sum only and call it the risk r from know on.

We have:

$$r : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

$$(c, y_i) \longmapsto (z - y_i - cz)^2 + (z - y_i - cz)\, 2c y_i + c^2 (1 + y_i^2)$$

(c) $r_1'(c, y_i) = -z2(z - y_i - cz) + 2c(1 + y_i^2) + 2z y_i - 2y_i^2 - 4z y_i c$

$$= -2z^2 - 2y_i^2 + 4z y_i + 2c\left[z^2 + 1 + y_i^2 - 2z y_i\right]$$

$$= -2\left[z^2 + y_i^2 - 2zy\right] + 2c\left[1 + z^2 + y_i^2 - 2z y_i\right]$$

$$= -2(z - y_i)^2 + 2c(z - y_i)^2 + 2c$$

ie $r_1'(c, y_i) = 2(z - y_i)^2 \left[c - 1\right] + 2c$

Therefore $\forall y_i \in \mathbb{R}, \forall c > 1, \; r_1'(c, y_i) > 0.$

This implies that the risk is strictly increasing in c on $]1; +\infty[$. And since the risk is continuous in c on the interval:

$$\forall y_i \in \mathbb{R}, \forall c \in]1; +\infty[, \quad \exists \delta > 0 : r(c - \delta, y_i) < r(c, y_i)$$

Therefore, for $c > 1$, the decision rule $z_i^{(c)}$ is inadmissible.

(d) Note that:

$$\bullet\ r_1'(c, y_i) = 0 \iff (z-y_i)^2 = c\left[(z-y_i)^2+1\right]$$

$$\iff c = 1 - \frac{1}{(z-y_i)^2+1}$$

$$\bullet\ r_1'(c, y_i) > 0 \iff c > 1 - \frac{1}{(z-y_i)^2+1}$$

By continuity, we infer the following variations for the risk:

$\forall\, y_i \in \mathbb{R}$,

c	$-\infty$	$1-\frac{1}{(z-y_i)^2+1}$	$+\infty$
$r_1'(c, y_i)$	$-$	0	$+$
$r(c, y_i)$	$\searrow$	$r\left(1-\frac{1}{(z-y_i)^2+1}, y_i\right)$	$\nearrow$

Now the worst case risk is equal to

$$\max_{y_i \in \mathbb{R}} \left\{ r(c, y_i) \right\}$$

Recall that:

$$r(c, y_i) = (z-y_i-cz)^2 + (z-y_i-cz)\, 2c y_i + c^2(1+y_i^2)$$

Hence:

$$r_2'(c, y_i) = -2(z - y_i - cz) + 2cz - 2c^2 z - 4cy_i + 2c^2 y_i$$

$$= -2z + 2cz + 2cz - 2c^2 z + 2y_i - 4cy_i + 2c^2 y_i$$

$$= -2z\left[1 - 2c + c^2\right] + 2y_i\left[1 - 2c + c^2\right]$$

Note that:

- If $c \in \mathbb{R} \setminus \{1\}$:
  - $r_2'(c, y_i) = 0 \Leftrightarrow 2y_i\left[1 - 2c + c^2\right] = 2z\left[1 - 2c + c^2\right]$

    $\Leftrightarrow y_i = z$

  - $r_2'(c, y_i) > 0 \Leftrightarrow y_i > z_i$

- If $c = 1$:
  - $\forall y_i \in \mathbb{R},\ r_2'(c, y_i) = 0$
  - $r(c, y_i) = (z - y_i - z)^2 + (z - y_i - z)2y_i + 1 + y_i^2$

    $= (z - y_i - z + y_i)^2 + 1$

    $= 1$

By continuity, we infer the following variations for the risk:

$\forall\, c \in \mathbb{R} \setminus \{1\}$:

y_i	$-\infty$		z		$+\infty$
$r_2'(c, y_i)$		$-$	0	$+$	
$r(c, y_i)$		$\searrow$	$r(c, z)$	$\nearrow$	

If $c = 1$:

y_i	$-\infty$		$+\infty$
$r_2'(c, y_i)$		0	
$r(c, y_i)$		$\xrightarrow{\quad 1 \quad}$	

Therefor, the worst case risk is:

$$\max_{y_i \in \mathbb{R}} \left\{ r(c, y_i) \right\} = \begin{cases} +\infty \text{ (unbounded) if } c \in \mathbb{R} \setminus \{1\} \\ 1 \quad \text{if } c = 1. \end{cases}$$

Therefore the value of c that minimizes the worst-case risk is:

$$\underset{c \in \mathbb{R}}{\operatorname{argmin}} \left\{ \max_{y_i \in \mathbb{R}} \left\{ r(c, y_i) \right\} \right\} = 1.$$

(e) Call π the prior distribution for y_i over $\mathbb{R}$, such that $y_i \sim \mathcal{N}(0, \gamma^2)$.

The Bayes risk is:

$$r_B(\pi, c) = \int_{\mathbb{R}} r(c, y_i)\, d\pi(y_i)$$

$$= E_{y_i}\left[(z - y_i - cz)^2 + (z - y_i - cz)\, 2c y_i + c^2(1 + y_i^2) \,\middle|\, c\right]$$

$$= E_{y_i}\left[(z - cz)^2 + y_i^2 - 2(z-cz)y_i + 2czy_i - 2cy_i^2 - 2c^2 z y_i + c^2 + c^2 y_i^2 \,\middle|\, c\right]$$

$$= (z - cz)^2 + E_{y_i}(y_i^2 | c) - 2(z - cz)\, E_{y_i}(y_i | c) + 2cz\, E_{y_i}(y_i | c)$$

$$- 2c\, E_{y_i}(y_i^2 | c) - 2c^2 z\, E_{y_i}(y_i | c) + c^2 + c^2 E_{y_i}(y_i^2 | c)$$

$$= (z - cz)^2 + \gamma^2 - 2c\gamma^2 + c^2 + c^2\gamma^2$$

$$= (z - cz)^2 + c^2 + \gamma^2\left[1 - 2c + c^2\right]$$

$$\boxed{\underline{\text{ie}}\ r_B(\pi, c) = z^2(1-c)^2 + c^2 + \gamma^2\left[1 - 2c + c^2\right]}$$

We can now find c that minimizes the Bayes risk. The Bayes risk is a polynomial of degree two in c with a positive coefficient for its second-degree monomial. Therefore its minimum is reached when its first order derivative in c is null and this point is unique. Therefore:

$$c_{Bayes} = \underset{c \in \mathbb{R}}{\arg\min} \left\{ r_B(\pi, c) \right\}$$

$$= \left\{ c \in \mathbb{R} : r_{B_2}'(\pi, c) = 0 \right\}$$

$$= \left\{ c \in \mathbb{R} : -2z^2(1-c) + 2c - 2\gamma^2 + 2\gamma^2 c = 0 \right\}$$

$$= \left\{ c \in \mathbb{R} : -z^2 - \gamma^2 + z^2 c + c + \gamma^2 c = 0 \right\}$$

$$= \left\{ c \in \mathbb{R} : c\left[1 + z^2 + \gamma^2\right] = z^2 + \gamma^2 \right\}$$

$$= \frac{z^2 + \gamma^2}{1 + z^2 + \gamma^2}$$

$$= 1 - \frac{1}{1 + z^2 + \gamma^2}$$

Therefore Bayes rule corresponds to

$$c_{Bayes} = 1 - \frac{1}{1+z^2+\sigma^2}.$$

Now we switch back to the original setting.

(f) Recall question (b). We had found:

$$R(\tau^{(c)}, y) = \sum_{i=1}^{n} r(c, y_i)$$

ie

$$R(\tau^{(c)}, y) = \sum_{i=1}^{n} \left[(z - y_i - cz)^2 + (z - y_i - cz)\, 2c\, y_i + c^2 (1 + y_i^2) \right]$$

(g) Recalling the computations from question (b) again, we immediately see that we have:

$$\frac{\partial R}{\partial c}(\tau^{(c^*)}, y) = 0 \iff \sum_{i=1}^{n} \left(2(z-y_i)^2 [c^* - 1] + 2c^* \right) = 0$$

$$\iff \left[c^* - 1\right] \sum_{i=1}^{n} (z - y_i)^2 + n c^* = 0$$

$$\iff c^* \left[n + \sum_{i=1}^{n} (z - y_i)^2 \right] = \sum_{i=1}^{n} (z - y_i)^2$$

$$\iff c^* = 1 - \frac{n}{\sum_{i=1}^{n} (z - y_i)^2 + n}$$

Therefoe the total risk across all individuals is minimized for :

$$c^* = 1 - \frac{n}{\sum_{i=1}^{n} (z - y_i)^2 + n} \neq 1 .$$

Therefore, minimizing the total risk across all individuals is not compatible with minimizing the individual worst case risk (unless all people already earn the target income level, a case for which the whole programme is pointless).

Recall question (d) where we showed that minimizing the individual risk yielded an optimum:

$$c^*_{\text{individual}} = 1 - \frac{1}{(z-y_i)^2+1}$$

Therefore, minimizing total risk and individual risk are not necessarily compatible.

Note that we have:

$$c^* = 1 - \frac{n}{\sum_{i=1}^{n}(z-y_i)^2 + n}$$

$$= 1 - \frac{1}{\frac{\sum_{i=1}^{n}(z-y_i)^2}{n} + 1}$$

Therefore, minimizing total risk minimizes individual risk only for those whose squared difference with the target income is equal to the average one in the whole population. There may be no individual for whom this is the case.

(h) I propose the following estimator:

$$\hat{c}^{*} = 1 - \frac{1}{\frac{\sum_{i=1}^{n} (z - \hat{y}_i)^2}{n} + 1}$$

Which gives:

$$\tau_i^{(\hat{c}^{*})}(\hat{y}) = \left[1 - \frac{1}{\frac{\sum_{i=1}^{n} (z - \hat{y}_i)^2}{n} + 1} \right] (z - \hat{y}_i)$$

Since $\hat{y}_i \sim \mathcal{N}(y_i, 1)$, $\hat{y}_{i\,JS} = \left[1 - \frac{n-2}{\sum_{i=1}^{n} \hat{y}_i^2} \right] \hat{y}_i$

Then:

$$\tau_i^{(\hat{c}^{*})}(\hat{y}_{JS}) = \left[1 - \frac{1}{\frac{\sum_{i=1}^{n} (z - \hat{y}_{i\,JS})^2}{n} + 1} \right] (z - \hat{y}_{i\,JS})$$

ie $\tau_i^{(\hat{c}^*)}(\hat{y}_{JS})$

$$= \left[1 - \frac{1}{\frac{\sum_{i=1}^{n}\left[z - \left[1 - \frac{n-2}{\sum_{i=1}^{n}\hat{y_i}^2}\right]\hat{y_i}\right]^2}{n} + 1}\right] z - \left[1 - \frac{n-2}{\sum_{i=1}^{n}\hat{y_i}^2}\right]\hat{y_i}$$

I'm not sure I fully understood what you meant here but it seems that the two estimate don't coincide. I don't understand what value you want me to estimate with Jones-Stein.