WARM-UP PROBLEM

(1) cloud 1: there are clearly two subpopulations, but there is no clear upward or downward trend, so the correlation might be close to 0.

cloud 2: Clearly ρ is close to 1.

cloud 3: X and Y are clearly not independent, but most likely $\rho \approx 0$.

cloud 4: $\rho \gtrsim 0$.

) Since X and Y are defined through some transformation, they are clearly not independent. In fact, $\mathbb{P}(Y>0, X>c) = \mathbb{P}(X>c) \neq \mathbb{P}(X>c)\mathbb{P}(Y>0) = \frac{1}{2}\mathbb{P}(X>c)$,

since
$$\begin{aligned}\mathbb{P}(Y>0) &= \mathbb{P}(Y>0, |X|>c) + \mathbb{P}(Y>0, |X|\leq c)\\ &= \mathbb{P}(X>0, |X|>c) + \mathbb{P}(-X>0, |X|\leq c)\\ &= \mathbb{P}(X>c) + \mathbb{P}(X\in[-c,0])\\ &= \mathbb{P}(X>c) + \mathbb{P}(X\in[0,c]) \text{ by symmetry}\\ &= \mathbb{P}(X>0) = 1/2 .\end{aligned}$$

Now, $\forall y < -c$, $\mathbb{P}(Y\leq y) = \mathbb{P}(X\leq y) = \Phi(y)$

- For $y\in[-c,c]$, $\mathbb{P}(Y<-c) + \mathbb{P}(Y\in[-c,y]) = \mathbb{P}(Y\leq y) = \Phi(-c) + \mathbb{P}(-X\in[-c,y])$
$$\begin{aligned}&= \Phi(-c) + \Phi(y) - \Phi(-c)\\ &= \Phi(y).\end{aligned}$$
- For $y>c$, $\mathbb{P}(Y<c) + \mathbb{P}(Y\in[c,y]) = \Phi(c) + \mathbb{P}(X\in[c,y]) = \Phi(y)$
$\mathbb{P}(Y\leq y) =$

So that Y is Gaussian $\mathcal{N}(0,1)$.

PROBLEM : Flipping Cochran

(i) The orthogonal projection of X onto $V = \text{Span}\{(1, \dots, 1)^T\}$ is

$$\Pi_V[X] = V(V^T V)^{-1} V^T X = \overline{X}_n \begin{pmatrix} 1 \\ | \\ 1 \end{pmatrix}.$$

Cochran's Thm ensures that $\Pi_V[X]$ is independent of $X - \Pi_V[X]$,

so that $\overline{X}_n \perp s_n^2 = \frac{1}{n}\|X - \Pi_V[X]\|^2$.

Since $\sqrt{n}\,\overline{X}_n \sim N(0,1)$, then Cochran's Thm implies that

$$\|X - \Pi_V[X]\|^2 = n s_n^2 \sim \chi^2_{n-1}.$$

(ii) a/ Write $\widetilde{X}_i := X_i - \mu$.

$$n s_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2 = \sum (\widetilde{X}_i - \overline{\widetilde{X}})^2 = \sum \widetilde{X}_i^2 - n \overline{\widetilde{X}}_n^2,$$

so that

$$\mathbb{E}[n s_n^2] = \sum \mathbb{E}[\widetilde{X}_i^2] - n\,\mathbb{E}\left[\left(\frac{1}{n}\sum \widetilde{X}_i\right)^2\right] = n\sigma^2 - n\mathbb{V}\left[\frac{1}{n}\sum \widetilde{X}_i\right] = (n-1)\sigma^2$$

Since $\overline{X}_n$ and $n s_n^2$ are independent, and (X_i) are iid, then, $\forall \xi \in \mathbb{R}$,

$$\mathbb{E}\left[s_n^2 e^{i\xi n \overline{X}_n}\right] = \mathbb{E}[s_n^2]\,\mathbb{E}\left[e^{i\xi \sum X_i}\right] = \mathbb{E}[s_n^2]\;\mathbb{E}\left[e^{i\xi X_1}\right]^n = \Phi^n(\xi)\mathbb{E}[s_n^2]$$

b/ Note that $\Phi'(\xi) = i\,\mathbb{E}[X_1 e^{i\xi X_1}]$, $\Phi''(\xi) = -\mathbb{E}[X_1^2 e^{i\xi X_1}]$, $\Phi'(0) = i\mu$

Now, $$n s_n^2 = \sum X_i^2 - n(\overline{X}_n)^2 = \sum_{i=1}^n X_i^2 - \frac{1}{n}\sum_{k,j=1}^n X_k X_j$$

$$= \left(1 - \frac{1}{n}\right)\sum_{i=1}^n X_i^2 - \frac{1}{n}\sum_{j \neq k} X_j X_k$$

We can thus write

$$\mathbb{E}\left[n s_n^2 e^{i\xi n \overline{X}_n}\right] = \left(1-\frac{1}{n}\right)\sum_{i=1}^{n} \mathbb{E}\left[X_i^2 \prod_{h=1}^{n} e^{i\xi X_h}\right] - \frac{1}{n}\sum_{j\neq k} \mathbb{E}\left[X_j X_k \prod_{h=1}^{n} e^{i\xi X_h}\right]$$

$$= \left(1-\frac{1}{n}\right)\sum_{i=1}^{n} \mathbb{E}\left[X_i^2 e^{i\xi X_i} \prod_{h\neq i} e^{i\xi X_h}\right] - \frac{1}{n}\sum_{j\neq k} \mathbb{E}\left[X_j X_k e^{i\xi X_j} e^{i\xi X_k} \prod_{\substack{h\neq j \\ h \neq k}} e^{i\xi X_h}\right]$$

$$= \left(1-\frac{1}{n}\right)\sum_{i=1}^{n} \left\{\mathbb{E}\left[X_i^2 e^{i\xi X_i}\right] \mathbb{E}\left[\prod_{h\neq i} e^{i\xi X_h}\right]\right\}$$

$$- \frac{1}{n}\sum_{j\neq k} \left\{\mathbb{E}\left[X_j e^{i\xi X_j}\right] \mathbb{E}\left[X_k e^{i\xi X_k}\right] \mathbb{E}\left[\prod_{\substack{h\neq j \\ h \neq k}} e^{i\xi X_h}\right]\right\}$$

by independence.

Since the sample is iid, we can rewrite this as

$$\mathbb{E}\left[n s_n^2 e^{i\xi n \overline{X}_n}\right] = \frac{n-1}{n}\sum_{i=1}^{n} \left(-\phi''(\xi)\right)\phi(\xi)^{n-1} - \frac{1}{n}\sum_{j\neq k}\left(-\phi'(\xi)^2\right)\phi(\xi)^{n-2}$$

$$= (1-n)\,\phi''(\xi)\,\phi(\xi)^{n-1} + (n-1)\,\phi''(\xi)^2\,\phi(\xi)^{n-2}$$

$$\overset{\text{Using } (a/)}{=} \phi(\xi)^n (n-1)\sigma^2,$$ and the ODE follows

c/ Since $\displaystyle (\log \phi)''(\xi) = \frac{\phi''(\xi)}{\phi(\xi)} - \left[\frac{\phi'(\xi)}{\phi(\xi)}\right]^2$, then the ODE from b/

reads $(\log \phi)''(\xi) = -\sigma^2$, so that $\log \phi(\xi) = -\frac{\sigma^2}{2}\xi^2 + a\xi + b$.

Plugging in the boundary conditions yield

$$\phi(\xi) = \exp\left\{-\frac{\sigma^2}{2}\xi^2 + i\mu\xi\right\} \qquad \square$$

PROBLEM: Gender (in)Equalities & Pay Gap

(i) α represents the average male salary.

β is the average salary spread: if $\beta = 0$: no gap.

if $\beta < 0$: women earn less than men.

ii) The model is clearly a Gaussian linear model, and hence

$$Y \sim N(\alpha \mathbb{1} + \beta X, \sigma^2 I_n)$$

the log likelihood function reads (with $\theta := (\alpha, \beta)$)

$$\ell_n(\theta) := -\frac{1}{n} \log L_n(\theta) = -\frac{1}{n} \log \left[\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2} (Y_i - \alpha - \beta x_i)^2 \right\} \right]$$

$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{2\sigma^2 n} \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2$$

$$= \log(\sigma\sqrt{2\pi}) + \frac{1}{2\sigma^2 n} \|Y - \theta \cdot \tilde{X}\|^2, \text{ with } \tilde{X} = \begin{pmatrix} 1 & x_1 \\ | & | \\ 1 & x_n \end{pmatrix},$$

so that minimising the log-likelihood is equivalent to least-square estimators, which exist if and only if the matrix $\tilde{X}$ has full rank $(=2)$, e.g. iff $\sum_{i=1}^{n} x_i \neq 0$ and $\sum_{i=1}^{n} x_i \neq n$

Then, the max. likelihood estimator is given by

$$\boxed{\hat{\theta}_n = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T Y .}$$

Now, $\tilde{X}^T \tilde{X} = \begin{pmatrix} 1 & x_1 \\ | & | \\ 1 & x_n \end{pmatrix}^T \begin{pmatrix} 1 & x_1 \\ | & | \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} = n \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x} \end{pmatrix},$

with $\overline{x} := \frac{1}{n} \sum x_i$, using the fact that $x_i^2 = x_i$ for any i,

Hence $(\tilde{X}^T \tilde{X})^{-1} = \frac{1}{n\overline{x}(1-\overline{x})} \begin{pmatrix} \overline{x} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$, and $\tilde{X}^T Y = \begin{pmatrix} n\overline{Y} \\ \sum x_i Y_i \end{pmatrix}$,

so that $\hat{\alpha} = \dfrac{n\overline{Y} - \sum x_i Y_i}{n(1-\overline{x})}$ and $\hat{\beta} = \dfrac{\sum x_i Y_i}{n\overline{x}} - \dfrac{n\overline{Y} - \sum x_i Y_i}{n(1-\overline{x})}$

(iii) Since the model is Gaussian Linear, we can write

$$\widehat{\theta}_n \sim \mathcal{N}\left(\vartheta, \sigma^2 (\tilde{X}^\top \tilde{X})^{-1}\right), \text{ and hence}$$

$$\sqrt{n}\sqrt{(1-\overline{x})\overline{x}}\ \frac{\widehat{\beta}_n - \beta}{\sigma} \sim \mathcal{N}(0,1) \text{ as } n \text{ tends to infinity,}$$

and therefore $1-\gamma = \mathbb{P}\left(-q_{1-\gamma/2} \leq \frac{\widehat{\beta}_n - \beta}{\tau} \leq q_{1-\gamma/2}\right)$

$$= \mathbb{P}\left(\widehat{\beta}_n - q_{1-\gamma/2}\tau \leq \beta \leq \widehat{\beta}_n + q_{1-\gamma/2}\tau\right),$$

with $q_{1-\gamma/2} := \Phi^{-1}\left(1-\frac{\gamma}{2}\right)$ and $\tau := \sqrt{\frac{\sigma^2}{(1-\overline{x})n\overline{x}}}$.

iv) We want to test $H_0 : \{\beta_0 = 0\}$ vs $H_1 : \{\beta \neq 0\}$.

We reject H_0 if 0 is not in the confidence interval above, which gives the rejection region $|\widehat{\beta}_n| > q_{1-\gamma/2}\tau$.

The power of the test reads, $\forall \beta \neq 0$:

$$\mathbb{P}_\beta\left(|\widehat{\beta}| > q_{1-\gamma/2}\tau\right) = \mathbb{P}_\beta\left(\widehat{\beta} > q_{1-\gamma/2}\tau\right) + \mathbb{P}\left(\widehat{\beta} < -q_{1-\gamma/2}\tau\right)$$

$$= 1 - \Phi\left(q_{1-\gamma/2} - \beta/\tau\right) + \Phi\left(-q_{1-\gamma/2} - \beta/\tau\right)$$

Clearly the rejection region increases with γ.

Also, $\lim_{\beta \downarrow -\infty} \mathbb{P}_\beta(-) = 1 = \lim_{\beta \uparrow +\infty} \mathbb{P}_\beta(-)$

The higher the $|\beta|$, the easier it gets to discriminate H_0 vs H_1.

(v) Test : $\mathcal{H}_0 : \{\beta \le 0\}$ vs $\{\beta > 0\}$.

We reject $\hat{\beta}_n > c$ for c such that $\sup_{\beta \le 0} \mathbb{P}_\theta(\hat{\beta}_n > c) = \gamma$, i.e.

$$\mathbb{P}_\theta(\hat{\beta}_n > c) = \mathbb{P}_\theta\left(\frac{\hat{\beta}_n - \beta}{\tau} > \frac{c-\beta}{\tau}\right) = 1 - \Phi\left(\frac{c-\beta}{\tau}\right),$$

and $\sup_{\beta \le 0} \mathbb{P}_\theta(\hat{\beta}_n > c) = 1 - \Phi(c/\tau)$, so that we choose $c = \tau q_{1-\gamma}$,

and we reject $\mathcal{H}_0$ if $\hat{\beta}_n > \tau q_{1-\gamma}$.

PROBLEM: Hellinger Distance

(i) Since f and g belong to $\mathcal{D}$, then $\int f(x)\,dx = \int g(x)\,dx = 1$, so that

$$\mathbb{H}(f,g) := \frac{1}{2}\int\left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx = \frac{1}{2}\int\left(f(x)+g(x) - 2\sqrt{f(x)\,g(x)}\right)dx$$

$$= 1 - \int\sqrt{f(x)\,g(x)}\,dx, \text{ so that } \mathbb{H}(f,g) \le 1.$$

Now, $2\,\mathbb{H}(f,g) = \|\sqrt{f} - \sqrt{g}\|_2^2$, so that $\mathbb{H}$ is positive, symmetric and satisfies the triangle inequality, and $\mathbb{H}(f,g)=0$ iff $f=g$ almost everywhere.

(ii) Let $\vartheta \le \vartheta'$. Then $\mathbb{H}(f,g) = 1 - \int_0^{\vartheta}\frac{dx}{\sqrt{\vartheta\vartheta'}} = 1 - \sqrt{\frac{\vartheta}{\vartheta'}}$.

(iii) Clearly, the likelihood function reads $L_n(\vartheta) = \frac{1}{\vartheta^n}\mathbb{1}_{\vartheta \ge X_{(n)}}$,

so that $\widehat{\vartheta}_n = X_{(n)}$, and $\mathbb{P}(\widehat{\vartheta}_n \le \tau) = \mathbb{P}(X_1 \le \tau)^n = (\tau/\vartheta)^n$, for $\tau \in [0,\tau]$.

and $\mathbb{P}(\widehat{\vartheta}_n \le 0) = 0$, $\mathbb{P}(\widehat{\vartheta}_n \ge 1) = 1$.

The density then reads $f(\tau) = \frac{n\tau^{n-1}}{\vartheta^n}\mathbb{1}_{[0,\vartheta]}(\tau)$, and

$$\mathbb{E}_\vartheta\left[\widehat{\vartheta}_n^{1/2}\right] = \int_0^{\vartheta} f(\tau)\,d\tau) = \frac{n\sqrt{\vartheta}}{n+1/2}.$$

(iv) We can write, by the tower property,

$$\mathbb{E}\left[\mathbb{H}(f_{\widehat{\vartheta}_n}, f_\vartheta)\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{H}(f_{\widehat{\vartheta}_n}, f_\vartheta)\,\middle|\,\widehat{\vartheta}_n\right]\right] = \mathbb{E}\left[1 - \sqrt{\frac{\widehat{\vartheta}_n}{\vartheta}}\right], \text{ and hence}$$

$$= \frac{1}{2n+1}.$$

Direct computations yield $\mathbb{E}_\vartheta\left[(\widehat{\vartheta}_n - \vartheta)^2\right] = \frac{2\vartheta^2}{(n+1)(n+2)} = \mathcal{O}(1/n^2)$ for large n, whereas the Hellinger distance is $\mathcal{O}(1/n)$.