EXERCISE 1 : It's Prob-time

STUDENT'S SCORE X IN ML : $0 < X < 1$

STUDENT'S SCORE Z IN STATISTICS : $0 < Z < 1$

SCORES ARE DISTRIBUTED ACCORDING TO THE FOLLOWING JOINT pdf:

$$f_{X,Z}(x,z) = \begin{cases} 8 \cdot (x \cdot z) & \text{for } 0 < z < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

1) (a) CHECK THAT $f_{X,Z}(x,z)$ IS A LEGIT JOINT pdf.

TO BE A LEGIT JOINT pdf,

(i) THE FUNCTION MUST BE GREATER THAN 0 (POSITIVE)

(ii) IT INTEGRATES TO 1.

<u>Check</u>:

(i) $f_{X,Z}(x,z) > 0 \Rightarrow 8 \cdot (x \cdot z) > 0$

$x > 0$

$z > 0$

(TRUE BECAUSE OF THE DOMAIN)

(ii) $\iint_A f_{X,Z}(x,z)\,dx\,dz = 1 \qquad A = \{(x,z) \in \mathbb{R}^2 : 0 < z < x < 1\}$

Design the domain

$0 < z < x < 1$

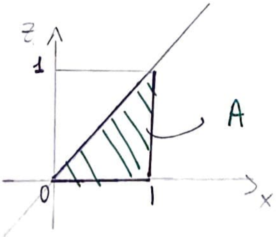
$x = z$

$x = 1$

Inequalities

$x > z$

$x < 1$



$$\int_0^1 z\,dz \int_z^1 8x\,dx = \int_0^1 z\,dz \left(\overset{4}{\cancel{8}} \frac{x^2}{\cancel{2}} \Big|_z^1 \right) = \int_0^1 z \cdot \left[4 - 4z^2\right] dz =$$

$$= \int_0^1 (4z - 4z^3)\,dz = \overset{2}{\cancel{4}} \frac{z^2}{\cancel{2}} \Big|_0^1 - \cancel{4} \frac{z^4}{\cancel{4}} = 2 - 1 = 1$$

→ THE PLOT IS IN THE ATTACHED JUPYTER NOTEBOOK.

(b) FROM THE JOINT pdf, WE CAN RECOVER THE MARGINALS BY INTEGRATING OVER ONE OF THE VARIABLES

$$f_z(z) = \int_z^1 8xz\,dx = \overset{4}{\cancel{8}} z \frac{x^2}{\cancel{2}} \Big|_z^1 = 4z(1-z^2) = 4z - 4z^3$$

THE PROPORTION OF STUDENTS THAT OBTAIN A SCORE GREATER THAN 0.5 IN STATISTICS IS EQUAL TO

By the fundamental theorem of calculus, $F_z'(z) = f_z(z)$

$$F_z(z) = \int f_z(z)\,dz$$

$$\mathbb{P}(z > 0.5) = 1 - \Pr(z \leq 0.5) = 1 - F_z(0.5) =$$

$$= 1 - \left[\int_0^{\frac{1}{2}} (4z - 4z^3)\,dz \right] =$$

$$= 1 - \left[\overset{2}{\cancel{4}} \frac{z^2}{\cancel{2}} \Big|_0^{\frac{1}{2}} - 4\frac{z^4}{4} \Big|_0^{\frac{1}{2}} \right] = 1 - \left[2z^2 \Big|_0^{\frac{1}{2}} - z^4 \Big|_0^{\frac{1}{2}} \right] = 1 - \left[\frac{1}{2} - \frac{1}{16} \right] =$$

$$= 1 - \frac{7}{16} = \frac{9}{16}$$

(c) THE PROBABILITY THAT A RANDOMLY SELECTED STUDENT WILL HAVE A MAT-SCORE EXACTLY EQUAL TO 0.5 IS 0.

$$\mathbb{P}(z = 0.5) = \int_{0.5}^{0.5} f_z(z)\,dz = 0$$

2) a) LET $W = \log(Z)$ BE THE LOG-STAT SCORE.

THE PROBABILITY DENSITY FUNCTION f_W OF W IS CALCULATED AS FOLLOWS

$\propto e^W < 1 \Rightarrow \begin{cases} e^W > 0 \ \forall W \\ e^W < 1 \Rightarrow W < 0 \end{cases}$

$\Downarrow$

$W < 0$

$$F_W(w) = \Pr(\log Z \le w) = \Pr(Z \le e^w) = F_Z(e^w)$$

$$F_W(W) = F_Z(e^W)$$

DIFFERENTIATING BOTH SIDES, WE FIND

$$f_W(w) = F'_W(w) = \frac{\partial}{\partial w}\left(F_Z(e^w)\right) = f_Z(e^w) \cdot \frac{\partial}{\partial w}(e^w) =$$

$$= f_Z(e^w)\, e^w = \left[4e^w(1-e^{2w})\right] \cdot e^w = \left(4e^{2w} - 4e^{4w}\right) \mathbb{1}_{(-\infty,0)}(w)$$

→ THE PLOT IS IN THE ATTACHED JUPYTER NOTEBOOK.

b) WE WANT TO FIND THE PREDICTED VALUE OF W THAT HAS THE SMALLEST MEAN SQUARED ERROR.

→ THIS EXERCISE IS ABOUT VARIATIONAL CHARACTERIZATION OF THE EXPECTATION.

GIVEN THE RANDOM VARIABLE $W \sim f_W(w)$,

(IT IS AN UNCERTAIN MEASUREMENT)

WE WANT TO FIND THE "BEST" PREDICTION FOR W.

TO FORMALIZE THE IDEA OF BEST, WE INTRODUCE THE CONCEPT OF LOSS.

LET d BE OUR DETERMINISTIC PREDICTION,

THE LOSS FUNCTION IS DEFINED AS $L(W, d)$: IT REPRESENTS HOW MUCH WE LOSE USING d TO PREDICT W.

WE CHOOSE THE LOSS AS THE SQUARED LOSS

$$L(W, d) = (W - d)^2$$

BUT IT IS A RANDOM QUANTITY

TAKING THE EXPECTATION OF THE LOSS (WRT TO THE VARIABILITY W W), WE WILL OBTAIN THE RISK.

$$E_W L(W, d) = R_W(d) \quad \text{risk}$$

↓

TO OUR DECISION

$$d^* = \underset{\text{ALL } d}{\operatorname{argmin}} R(d)$$

by linearity of expectation

$$\rightarrow L(W, d) = (W - d)^2$$

$$E_W L(W, d) = E_W[(W - d)^2] = E_W[W^2 + d^2 - 2Wd] \overset{\downarrow}{=}$$

$$= E_W(W^2) + E_W(d^2) - 2dE_W(W) = E_W(W^2) + d^2 - 2d\,E_W(W) = R_W(d)$$

$$\Rightarrow R_W(d) = E_W(W^2) - 2d\,E_W(W) + d^2$$

WE CAN FIND THE MINIMIZING VALUE OF d BY DIFFERENTIATION (w.r.t. d)

$$R'_W(d) = -2E_W(W) + 2d$$

THEREFORE, WE CAN CONCLUDE THAT THE PREDICTED VALUE OF W THAT HAS THE SMALLEST RISK IS

$$d = E_W(W) = \int_{-\infty}^{+\infty} w\, f_W(w)\, dw = \int_{-\infty}^{0} w(4e^{2w} - 4e^{4w})\, dw =$$

$$= \int_{-\infty}^{0} 4we^{2w}\, dw - \int_{-\infty}^{0} 4we^{4w}\, dw = \lim_{L \to -\infty} \int_{L}^{0} 4w e^{2w}\, dw - \lim_{L \to -\infty} \int_{L}^{0} 4w e^{4w}\, dw =$$

$$\lim_{L\to-\infty} \underbrace{\int_L^0 4we^{2w}\,dw}_{①} - \lim_{L\to-\infty} \underbrace{\int_L^0 4we^{4w}\,dw}_{②}$$

① $\int 4we^{2w}\,dw = 4\int we^{2w}\,dw =$

①
| $f(w) = w$ $\quad f'(w) = 1$
| $g'(w) = e^{2w}$ $\quad g(w) = \frac{1}{2}\int 2e^{2w} =$
| $\quad = \frac{1}{2}e^{2w}$

$= 4\left[\frac{1}{2}we^{2w} - \frac{1}{4}\int 4\cdot\frac{1}{2}e^{2w}\right] = 2we^{2w} - e^{2w}$

② $\int 4we^{4w}\,dw = 4\int we^{4w}\,dw =$

②
| $f(w) = w$ $\quad f'(w) = 1$
| $g'(w) = e^{4w}$ $\quad g(w) = \frac{1}{4}\int 4e^{4w} =$
| $\quad = \frac{1}{4}e^{4w}$

$= 4\left[\frac{1}{4}we^{4w} - \frac{1}{16}\int 16\cdot\frac{1}{4}e^{4w}\,dw\right] =$

$= we^{4w} - \frac{1}{4}e^{4w}$

$$\lim_{L\to-\infty}\left[2we^{2w} - e^{2w}\right]_L^0 - \lim_{L\to-\infty}\left[we^{4w} - \frac{1}{4}e^{4w}\right]_L^0 =$$

$$= -1 - \left(-\frac{1}{4}\right) = \frac{-3}{4}$$

ⓒ FIND THE MEDIAN LOG-STAT SCORE.

THE MEDIAN OF A CONTINUOUS RANDOM VARIABLE IS THE "MIDDLE VALUE", THAT IS THE VALUE OF w THAT SPLITS THE AREA ENCLOSED BY THE CURVE $y = f_W(w)$ AND THE X-AXIS INTO TWO EQUAL AREAS, BOTH EQUAL TO 0,5. $\int_{-\infty}^{x} f_W(w)\,dw = \frac{1}{2}$

$$\int_{-\infty}^{x} (4e^{2w} - 4e^{4w}) \mathbb{1}_{(-\infty,0)} dw = \int_{-\infty}^{x} 4e^{2w} dw - \int_{-\infty}^{x} 4e^{4w} dw =$$

$$= 2\int_{-\infty}^{x} 2e^{2w} dw - \int_{-\infty}^{x} 4e^{4w} dw = 2e^{2w}\Big|_{-\infty}^{x} - e^{4w}\Big|_{-\infty}^{x} = 2e^{2x} - e^{4x} = \frac{1}{2}$$

$$\Rightarrow \quad -e^{4x} + 2e^{2x} = \frac{1}{2} \Rightarrow e^{4x} - 2e^{2x} + \frac{1}{2} = 0$$

$$e^{4x} - 2e^{2x} + \frac{1}{2} = 0 \qquad t = e^{2x}$$

$$t^2 - 2t + \frac{1}{2} = 0 \Rightarrow 2t^2 - 4t + 1 = 0$$

$$t = \frac{4 \pm \sqrt{8}}{4} = \frac{4 \pm 2\sqrt{2}}{4} = \begin{cases} t_1 = \frac{(2+\sqrt{2})}{2} \\ t_2 = \frac{2-\sqrt{2}}{2} \end{cases}$$

$$t_1 = e^{2x} \Rightarrow \frac{2+\sqrt{2}}{2} = e^{2x} \Rightarrow \ln\left(\frac{2+\sqrt{2}}{2}\right) = 2x \Rightarrow \cancel{x = \frac{1}{2}\ln\left(\frac{2+\sqrt{2}}{2}\right)} \Rightarrow$$

THIS VALUE $\notin (-\infty, 0)$

$$t_2 = e^{2x} \Rightarrow \frac{2-\sqrt{2}}{2} = e^{2x} \Rightarrow \ln\left(\frac{2-\sqrt{2}}{2}\right) = 2x \Rightarrow x = \frac{1}{2}\ln\left(\frac{2-\sqrt{2}}{2}\right)$$

$$\Downarrow$$

$x = \frac{1}{2}\ln\left(\frac{2-\sqrt{2}}{2}\right)$ IS THE MEDIAN LOG-SMAT SCORE.

3) ASSUMING A STUDENT GOT 0.8 IN ML : OBSERVED EVENT $X = 0.8$,
FIND ANALYTICALLY THE BEST MSE PREDICTOR FOR HER SMAT-SCORE.

THE REASONING IS THE SAME AS IN POINT 2(b) $\Rightarrow$
IT IS ABOUT THE VARIATIONAL CHARACTERIZATION OF THE EXPECTATION,
<u>BUT</u>, IN THIS CASE, <u>CONDITIONED</u> ON $X = x$: $x = 0.8$

$$E_{y,x}\left[(z-d)^2 \mid x=0.8\right] = E_{y,x}\left[z^2 \mid x=0.8\right] - 2d\,E_{y,x}\left[z \mid x=0.8\right] + d^2 = R(d)$$

BY DIFFERENTIATION WE OBTAIN THE MINIMUM MEAN SQUARED ERROR ESTIMATE OF z GIVEN $x=0.8$.

$$R'(d) = -2\,E_{y,x}\left[z \mid x=0.8\right] + 2d$$

THEREFORE,

$$d = E_{y,x}\left[z \mid x=0.8\right] = \int z \cdot f_{z|x}(z|x)\,dz = \int z \cdot \frac{f_{x,z}(x,z)}{f_x(x)}\,dz =$$

$$= \int_0^1 z\,\frac{\cancel{8}^{2}\cancel{x}z}{\cancel{4}x^{\cancel{3}2}}\,dz = \int_0^1 2\,\frac{z^2}{x^2}\,dz =$$

$$= \frac{2}{x^2}\,\frac{z^3}{3}\Big|_0^1 = \frac{2}{3x^2} = \frac{2}{3}\cdot\left(\frac{8}{10}\right)^{-2} = \frac{25}{24}$$

$$f_x(x) = \int_0^x f_{x,z}(x,z)\,dz =$$

$$= \int_0^x 8xz\,dz = 8x\,\frac{z^2}{2}\Big|_0^x =$$

$$= \cancel{8}^{4}x\,\frac{x^2}{\cancel{2}} = 4x^3$$

EXERCISE 2 : Stat | 1st contact

1) THE FIRST THING THAT WE ARE REQUIRED TO DO IS TO CHECK THAT THE GIVEN FUNCTION $f_X(x|\alpha) = \frac{1}{2\pi}(1+\alpha \cdot \cos x)$ IS A VALID PROBABILITY DENSITY FUNCTION. TO DO SO, WE HAVE TO CHECK THAT:

① THE FUNCTION IS GREATER THAN 0 (POSITIVE)

② IT INTEGRATES TO 1

CHECK :

① $\frac{1+\alpha \cdot \cos x}{2\pi} > 0 \rightarrow 1+\alpha\cdot\cos x > 0 \rightarrow \alpha \cdot \cos x > -1 \rightarrow \begin{cases} \text{ALWAYS TRUE SINCE} \\ \alpha \in [-1/3, 1/3] \end{cases}$

② $\int_0^{2\pi} \frac{1}{2\pi}(1+\alpha\cdot\cos(x))\,dx = \frac{1}{2\pi}\int_0^{2\pi} 1+\alpha\cdot\cos(x)\,dx = \frac{1}{2\pi}\left[\int_0^{2\pi} dx + \int_0^{2\pi}\alpha\cdot\cos x\,dx\right] =$

$= \frac{1}{2\pi}\left[x\Big|_0^{2\pi} + \alpha\sin x\Big|_0^{2\pi}\right] = \frac{1}{2\pi}\left[2\pi + 0\right] = 1$

2) PARAMETER FAMILY'S VISUALIZATION IN THE ATTACHED JUPYTER NOTEBOOK

3) WE'RE NOW COMPUTING THE METHOD OF MOMENTS ESTIMATOR FOR α BASED ON m INDEPENDENT AND IDENTICALLY DISTRIBUTED MEASUREMENTS $\{X_1 \ldots X_N\}$ FROM $f_X(x|\alpha)$. TO OBTAIN OUR RESULT WE HAVE TO SOLVE THE FOLLOWING SYSTEM OF EQUATIONS:

$$S_m(\alpha|X_m) = \begin{cases} \text{POPULATION MOMENT}_1 = \text{EMPIRICAL MOMENT}_1 \\ \text{POPULATION MOMENT}_2 = \text{EMPIRICAL MOMENT}_2 \end{cases}$$

WHERE
- WE DEFINE THE j-th POPULATION MOMENT AS:

$$\mu_j = \mu_j(\alpha) = \mathbb{E}_\alpha(X^j) = \int x^j \, dF(X|\alpha)$$

- WE DEFINE THE j-th EMPIRICAL MOMENT AS:

$$\hat{m}_j = \hat{m}_j(X_m) = \frac{1}{m}\sum_{i=1}^{m} X_i^j$$

SO:

$$S_m(\alpha|X_m) = \begin{cases} \mu_1(\alpha) = \hat{m}_1(X_m) \\ \mu_2(\alpha) = \hat{m}_2(X_m) \end{cases} = \begin{cases} \int x \, dF(X|\alpha) = \frac{1}{m}\sum^{m}(X_i) \\ \int x^2 dF(X|\alpha) = \frac{1}{m}\sum^{m}(X_i)^2 \end{cases}$$

IN ORDER TO SOLVE THE SYSTEM WE HAVE TO COMPUTE $d\,F(X|\alpha)$. BY THE FUNDAMENTAL THEOREM OF INTEGRAL CALCULUS WE HAVE THAT:

$$F'(X|\alpha) = f_X(x|\alpha)$$

THE NEXT STEP IS USING THIS FORMULA TO COMPUTE μ_1, μ_2

$$\mu_1 = \int_0^{2\pi} \frac{1}{2\pi} x \cdot (1+\alpha\cos x)\,dx = \int_0^{2\pi} \frac{1}{2\pi} x\,dx + \int_0^{2\pi} \frac{1}{2\pi} x\,\alpha \cdot \cos x\,dx = \frac{1}{2\pi} \frac{x^2}{2}\Big|_0^{2\pi} + \frac{1}{2\pi}\alpha \int_0^{2\pi} x\cos x\,dx \Rightarrow$$

INTEGRATING $\int_0^{2\pi} x\cos x\,dx$

$$\left[\int_0^{2\pi} x\cos x\,dx = x\cdot \sin x\Big|_0^{2\pi} - \int_0^{2\pi} \sin x = x\sin x\Big|_0^{2\pi} - (-\cos x)\Big|_0^{2\pi} = (x\cdot\sin x + \cos x)\Big|_0^{2\pi} \right]$$

$$\Rightarrow \frac{1}{4\pi}\cdot 4\pi^2 + \frac{1}{2\pi}\alpha(x\cdot\sin x + \cos x)\Big|_0^{2\pi} = \pi + \frac{1}{2\pi}\alpha\cdot[0] = \pi$$

$$\mu_2 = \int_0^{2\pi} \frac{1}{2\pi} \cdot x^2(1+\alpha\cos(x))\,dx = \frac{\alpha}{2\pi}\int_0^{2\pi} x^2\cos x + \frac{1}{2\pi}\int_0^{2\pi} x^2\,dx = \left[\frac{\alpha((x^2-2)\sin x + 2\cdot x\cdot\cos(x))}{2\pi} + \frac{x^3}{6\pi}\right]_0^{2\pi} =$$

$$= \frac{6\alpha + 4\pi^2}{3} = 2\alpha + \frac{4}{3}\pi$$

NOTICE THAT

$\mu_j = \mathbb{E}(x^j)$ SO IN OUR CASE $\mathbb{E}(X) = \pi$, $\mathbb{E}(X^2) = 2\alpha + \frac{4}{3}\pi^2$

SO:

$$\alpha = \left(\mathbb{E}(X^2) - \frac{4}{3}\pi^2\right)/2$$

4) WE CAN NOW FIND α'S VALUES STARTING FROM THE FOLLOWING RELATIONS:

$$\mathbb{V}ar(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \quad \text{AND} \quad \mathbb{V}ar(X) = \frac{1}{m}\sum^m (x_i - \mu)^2$$

WE SUBSTITUTE THE PREVIOUSLY OBTAINED VALUES INSIDE THE EQUATIONS AND OBTAIN:

$$\mathbb{E}(X^2) = \mathbb{V}ar(X) + \mathbb{E}(X)^2 \rightarrow 2\alpha + \frac{4}{3}\pi^2 = \frac{1}{20}\sum (x_i - \bar{x}_m)^2 + \pi^2$$

$$\Rightarrow \alpha = \frac{\frac{1}{20}\sum^{10}(x_i - \bar{x}_m)^2 + \pi^2 - \frac{4}{3}\pi^2}{2} = 0.30477$$

Note

$\bar{X}_m = \frac{1}{20}\sum^m X_i$

5) WHEN WE HAVE IID RANDOM VARIABLES, THE LOG-LIKELIHOOD FUNCTION IS DEFINED AS THE SUM OVER m = NUMBER OF SAMPLES, OF THE LOGARITHM OF THE PDF, IN OUR CASE:

$$\ell(\alpha) \overset{IID}{=} \sum_i^m \log(f_x(x|\alpha)) = \sum_i^m \log\left(\frac{1}{2\pi} + \frac{\alpha \cdot \cos}{2\pi}\right) = \sum_i^m \log(1 + \alpha\cos x) + \underbrace{\sum_i^m \log\left(\frac{1}{2\pi}\right)}_{\boxed{\text{THIS FACTOR CAN BE REMOVED SINCE IT DOESN'T DEPEND ON } \alpha}} =$$

$$= \sum_i^m \log(1 + \alpha\cos x) = m \log(1 + \alpha\cos(\bar{x}_m))$$