Worksheet 2: Probability for Prediction

Name: Due: September 23, 2024

For questions 1-3, you will be working with probability measure $\mathbb{P}_{\mathcal{X}\times\mathcal{Y}}$ on $\mathcal{X}\times\mathcal{Y} = \mathbb{R}\times\mathbb{R}$ defined by density function $f_{\mathcal{X}\times\mathcal{Y}}$:

$$f(x,y) := x \cdot y^{-1/2}\mathbb{1}_{x\geq 0, y>0, x+y\leq 1},$$

where $\mathbb{1}_{\mathtt{cond}(t)} := \begin{cases} 1 & \text{if } \mathtt{cond}(t) \\ 0 & \text{else} \end{cases}$ denotes the *indicator* function for condition $\mathtt{cond}(t)$.*

1. Does $f_{\mathcal{X}\times\mathcal{Y}}$ define a valid pdf? If so, articulate why, else explain why not, and if not, modify f to make it a pdf. Also sketch (shading the appropriate region in $\mathbb{R}^2$) the *support* of f, i.e. the (closure of) points (x,y) for which $f(x,y) \neq 0$.

2. Define the marginal density $f_{\mathcal{Y}}(y)$, express as a function of y, sketch a plot, and compute the marginal probability $\mathbb{P}_{\mathcal{Y}}(y=0)$. Also report $\lim\limits_{y_0\to 0^+} f_{\mathcal{Y}}(y_0)$ and $\lim\limits_{y_0\to 0^+} \mathbb{P}_{\mathcal{Y}}([0,y_0))$.

3. Components x, y define random variables $\mathbb{R}^2 \xrightarrow{\pi_\cdot} \mathbb{R}$ by projection, $\pi_x : (x,y) \mapsto x$ and $\pi_y(x,y) \mapsto y$. Are these random variables independent? Justify your answer.

4. Consider dataset on $\mathcal{X}\times\mathcal{Y} = \{0,1\}^2$ in the table below. The right most column denotes the output of *model* $\overline{y} : \mathcal{X} \to \mathcal{Y}$. Formally define the covariance $\mathrm{Cov}(x,y)$ as an expectation and estimate this value from the table. Also estimate accuracy $\mathbb{P}_{\mathcal{X}\times\mathcal{Y}}(\overline{y} = y)$ and true positive rate $\mathbb{P}_{\mathcal{X}|\mathcal{Y}}(\overline{y} = 1 | y = 1)$.

x	y	$\overline{y}$
0	1	0
1	0	1
1	1	1
0	0	0
0	0	0
1	1	1
1	0	1

*While $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, we distinguish components with notation to e.g. distinguish the marginals $f_{\mathcal{X}}$ and $f_{\mathcal{Y}}$.

5. Consider, now, probability space $\mathcal{X} := \prod_{j=1}^{n} \mathcal{X}_j$ with joint measure $\mathbb{P}_{\mathcal{X}}$, and suppose that for any proper subset (projection) $\{i_1, \ldots, i_k\} \subsetneq \{1, \ldots, n\}$ setting $\mathcal{X}' := \prod_{j=1}^{k} \mathcal{X}_{i_j}$, we have equality measures $\mathbb{P}_{\mathcal{X}'} = \prod_{j=1}^{k} \mathbb{P}_{\mathcal{X}_{i_j}}$. Are projections $\pi_1, \ldots, \pi_n$ then independent? Why or why not? (Rationale for your answer may take the form of proof, example/counterexample, "just the definition," picture, or something else.)

6. Imagine that you are tasked with building a model to interpret data from some device or machine which TSA will use for security screening. Reason through how you will evaluate performance and discuss countervailing interests which should factor into your design (e.g. choice of threshold).

7. Consider joint probability space $\left(\mathcal{X} \times \mathcal{Y} = [0,1] \times \{0,1\}, \mathbb{P}_{\mathcal{X}\times\mathcal{Y}}\right)$ with measure defined by

$$\mathbb{P}_{\mathcal{X}\times\mathcal{Y}}([a,b] \times \{j\}) := p_j \int_a^b f_j(t)\,dt$$

for $0 \leq a \leq b \leq 1$ and $f_j(t) := \begin{cases} 2t & \text{if } j = 1 \\ 2(1-t) & \text{else.} \end{cases}$ Determine, analytically, what $p_j(x) = \mathbb{P}_{\mathcal{Y}|\mathcal{X}}(y = 1|x)$ must be in order for marginal $\mathbb{P}_{\mathcal{X}}$ to be uniform ($\mathbb{P}_{\mathcal{X}}([a,b]) = b - a$).