Lecture 17, Nov 3, 2023

Properties of Conditional PDFs

- Conditional PDFs can be treated just like regular PDFs; the condition parameterizes the PDF
  - Marginalization: $f_{x|z}(x|z) = \sum_{y \in \mathcal{Y}} f_{xy|z}(x, y|z)$
  - Conditioning: $f_{x|yz}(x|y, z) = \dfrac{f_{xy|z}(x, y|z)}{f_{y|z}(y|z)}$
  - Notice that if we remove z we just get the regular marginalization and conditioning laws
    * The parameter z can be e.g. a mean or standard deviation for a normal distribution, etc
- Note that everything to the left of the bar is a random variable and everything to the right is a fixed conditioning variable (the bar has lower "precedence" than the comma)
- $f_{xyz}(x, y, z) = f_{xy|z}(x, y|z)f_z(z) \implies f_{xz}(x, z) = \sum_{y \in \mathcal{Y}} f_{xyz}(x, y, z) = \sum_{y \in \mathcal{Y}} f_{x,y|z}(x, y|z)f_z(z) \implies$ $f_{x|z}(x|z) = \dfrac{f_{xz}(x, z)}{f_z(z)} = \sum_{y \in \mathcal{Y}} f_{x,y|z}(x, y|z)$
- $f_{x|yz}(x|y, z) = \dfrac{f_{xyz}(x, y, z)}{f_{yz}(y, z)} = \dfrac{f_{xy|z}(x, y|z)f_z(z)}{f_{y|z}(y|z)f_z(z)} = \dfrac{f_{xy|z}(x, y|z)}{f_{y|z}(y|z)} \implies f_{xy|z}(x, y|z) = f_{x|yz}(x|y, z)f_{y|z}(y|z)$
  - $f_{x|yz}(x|y, z)$ is conditioned on both y and z; so we can think of the above as multiplying by a distribution of y moves y to the left of the bar
  - Since both distributions are conditioned on z, the resulting distribution will also be conditioned on z
- Random variables x and y are *independent* if $f(x|y) = f(x)$, i.e. knowing y does not give us additional information about x
  - $f(x, y) = f(x|y)f(y) = f(x)f(y)$ is an equivalent definition
  - This also implies that $f(y|x) = f(y)$
- Random variables x and y are *conditionally independent* on z if $f(x|y, z) = f(x|z)$, i.e. knowing z makes x and y independent (y has no information on x if we already know z)
  - This is equivalent to $f(x, y|z) = f(x|z)f(y|z)$ similar to above
- Independence greatly simplifies algorithms and allows us to decouple information and processes
  - Suppose $y_i = g_i(x, w_i)$ where y_i are measurements, x is the state and w_i are noise
  - $f(y_1, \cdots, y_N|x) = \prod_{i=1}^{N} f(y_i|x)$ if we have independence, which greatly simplifies the problem

Bayes' Theorem

- $f(x|y)f(y) = f(y|x)f(x) \implies f(x|y) = \dfrac{f(y|x)f(x)}{f(y)}$ is *Bayes' theorem* (or rule)
- This applies to both continuous and discrete and mixed PDFs
- Example: disease diagnosis
  - Doctors were asked to estimate the probability that a woman with no symptoms between 40 and 50 years old with a positive mammogram actually has breast cancer
  - Given: 7% of mammograms give false positives, 10% of mammograms gives false negatives, actual incidence of breast cancer in this age group is 0.8%
  - Let x be whether the patient has cancer (0) or not (1), and let y be whether the test is negative (1) or positive (0)
  - We want $f_{x|y}(0, 0)$, i.e. the patient has cancer given a positive test
  - The 7% false positives is $f_{y|x}(0|1)$, the 10% false negatives is $f_{y|x}(1|0)$, the 0.8% is $f_x(0)$
  - $f_{x|y}(0|0) = \dfrac{f_{y|x}(0|0)f_x(0)}{f_y(0)}$
    * $f_{y|x}(1|0) = 0.10 \implies 1 - f_{y|x}(1|0) = f_{y|x}(0|0) = 0.90$
    * $f_x(0) = 0.008 \implies 1 - f_x(0) = f_x(1) = 0.992$

        * $f_{y|x}(0|1) = 0.007 \implies f_{y|x}(0|1)f_x(1) + f_{y|x}(0|0)f_x(0) = f_y(0) = 0.07664$
    - Using these numbers we get $f_{x|y}(0,0) = 9.4\%$
    - Since the actual probability of cancer is very low, very few positives will be due to true positives
- Example: girls named Lulu
    - A family has two children; given that one is a girl, what is the probability that both are girls?
        * Let x be 1 if there are no boys in the family and 0 if there are boys in the family
        * Let y be 1 if there are no girls in the family and 0 if there are girls in the family
        * $f_y(0) = \frac{3}{4}, f_y(1) = \frac{1}{4}$ because there are 4 cases, 3 of which have at least 1 girl
        * $f_x(1) = \frac{1}{4}, f_x(0) = \frac{3}{4}$ similarly
        * $f_{y|x}(0|1) = 1$ since if there are no boys, there must be girls
        * We want to know $f_{x|y}(1|0) = \frac{f_{y|x}(0|1)f_x(1)}{f_y(0)} = \frac{1}{3}$
    - If we are given that one of them is a girl named Lulu (given that this is an uncommon name), what is the probability that both are girls?
        * Let x be 1 if there are no boys in the family and 0 if there are boys in the family
        * Let y be 1 if there are no girls named Lulu in the family and 0 if there are
        * $f_{x|y}(1|0) = \frac{f_{y|x}(0|1)f_x(1)}{f_y(0)}$
            - $f_{y|x}(0|1)$ is the probability that given 2 girls, at least one is named Lulu
                - If $p \ll 1$ is the probability of a girl being named Lulu
                - We can enumerate all possible outcomes:
                    * Both children not named Lulu: $(1-p)(1-p)$
                    * First child not Lulu, second child Lulu: $(1-p)p$
                    * First child Lulu, second child not Lulu: p
                    * Both children named Lulu: 0
                - Therefore $f_{y|x}(0|1) = 2p - p^2 \approx 2p$
            - $f_y(0) = f_{y|x}(0|0)f_x(0) + f_{y|x}(0|1)f_x(1) = p - \frac{1}{4}p^2$
                - $f_{y|x}(0|0)$ has at least one boy in the family, so the probability of having a girl is $\frac{2}{3}$ so having a girl named Lulu is $\frac{2}{3}p$
                - $f_x(0) = \frac{3}{4}, f_x(1) = \frac{1}{4}$
        * This gives us $f_{x|y}(1|0) \approx \frac{1}{2}$