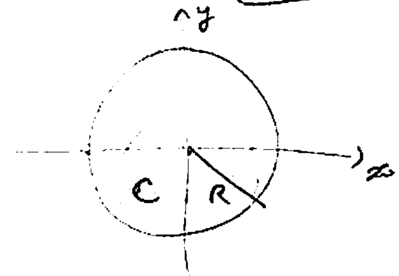
Lec. 30 03/23

Ex $(X,Y) \sim \text{Unif}(C)$ where $C = \{(x,y) : x^2+y^2 \le R\}$.



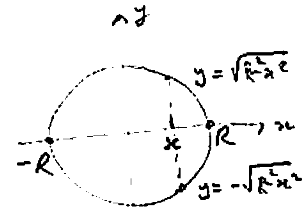
1) Marginal distr. $f_X(x)$, $f_Y(y)$?

2) Are X, Y independent ?

3) Conditional distr. $f_{X|Y}(x|y)$?

Joint pdf.

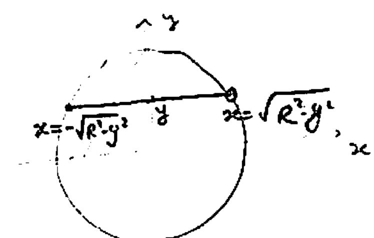
$$f(x,y) = \begin{cases} \dfrac{1}{\text{Area}(C)} = \dfrac{1}{\pi R^2}, & x^2+y^2 \le R \\ \text{elsewhere.} \end{cases}$$



1) $$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}\,dy = \begin{cases} \dfrac{2\sqrt{R^2-x^2}}{\pi R^2}, & -R \le x \le R \\ 0, & \text{elsewhere} \end{cases}$$

$$f_Y(y) = \begin{cases} \dfrac{2\sqrt{R^2-y^2}}{\pi R^2}, & -R \le y \le R \\ 0, & \text{elsewhere} \end{cases}$$

2) No, since $f(x,y) \ne f_X(x) \cdot f_Y(y)$



3) $$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{\pi R^2} \cdot \frac{\pi R^2}{2\sqrt{R^2-y^2}}$$

$$= \begin{cases} \dfrac{1}{2\sqrt{R^2-y^2}}, & -\sqrt{R^2-y^2} \le x \le \sqrt{R^2-y^2} \\ 0, & \text{elsewhere.} \end{cases} = \text{Unif}\left[-\sqrt{R^2-y^2}, \sqrt{R^2-y^2}\right]$$

$\Rightarrow X|Y \sim \text{Unif}\left(-\sqrt{R^2-y^2}, \sqrt{R^2-y^2}\right)$

Ex | Refer to Ex. from 03/21.

$Y \sim Unif(0,1)$, $X|Y \sim Unif(0,Y)$

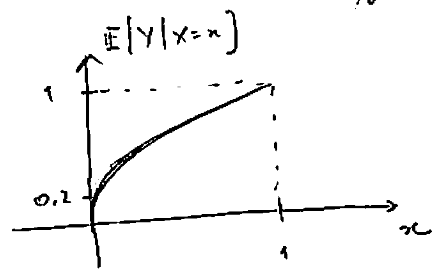
$Y|X \sim ?$ $\quad E[Y|X=x] = ?$ (Prediction of Y from $X=x$).

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}. \qquad f(x,y) = \begin{cases} 1/y, & 0 \le x \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

We already computed $\rightarrow$ $f_X(x) = \begin{cases} -\ln x, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$

$$\Rightarrow f_{Y|X}(y|x) = \begin{cases} -\frac{1}{y \ln x}, & x \le y \le 1 \\ 0, & \text{elsewhere}. \end{cases}$$

$$E[Y|X=x] = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\, dy = \int_x^1 \frac{-\cancel{y}}{\cancel{y}\ln x}\, dy = \begin{cases} \frac{1-x}{-\ln x}, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$



For Example: $Y = \underline{0.39}$, $X = 0.12$ $\Rightarrow$ our prediction of Y is $\frac{1-0.12}{-\ln 0.12} \approx \underline{0.42}$.

Continuous version of Bayes formula:

$$\boxed{f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)\, f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, f_Y(y)\, dy}}.$$

f_Y = "prior", $f_{Y|X}$ = "posterior" distributions of Y

before / after getting information about the value of X.

In our example, $f_{Y|X}(y|x) = \dfrac{\frac{1}{y}\cdot 1}{\int_x^1 \frac{1}{y}\cdot 1\, dy} = \begin{cases} -\frac{1}{y \ln x}, & x \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$ (same answer)

Ex The lifetime ($=X$) of a bulb is $Exp(\Lambda)$,

$\Lambda \sim Unif[a,b]$. Find the (marginal) distr. of the lifetime of the bulb.

$X|\Lambda \sim Exp(\Lambda)$, $\Rightarrow$ $f_{X|\Lambda}(x|\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{———} \end{cases}$, $\quad f_\Lambda(\lambda) = \frac{1}{b-a}$, $a \le \lambda \le b$.

$\Rightarrow f(x,\lambda) = f_{X|\Lambda}(x|\lambda) f_\Lambda(\lambda) = \begin{cases} \frac{\lambda}{b-a} e^{-\lambda x}, & x > 0. \\ 0, & \text{———} \end{cases}$ $\Rightarrow f_X(x) = \int_a^b \frac{\lambda}{b-a} e^{-\lambda x}\, d\lambda = \dots = \frac{e^{-ax}(1+ax) - e^{-bx}(1+bx)}{(b-a)x^2}$

Conditional vs Marginal Distributions

Simpson's paradox

From Wikipedia, the free encyclopedia

In probability and statistics, **Simpson's paradox** (or the **Yule-Simpson effect**) is a paradox in which a correlation (trend) present in different groups is reversed when the groups are combined. This result is often encountered in social-science and medical-science statistics,[1] and it occurs when frequency data are hastily given causal interpretations.

Berkeley sex bias case

One of the best known real life examples of Simpson's paradox occurred when the University of California, Berkeley was sued for bias against women who had applied for admission to graduate schools there. The admission figures for the fall of 1973 showed that men applying were more likely than women to be admitted, and the difference was so large that it was unlikely to be due to chance.[3][14]

		Applicants	Admitted
$X=1$	**Men**	8442	44%
$X=2$	**Women**	4321	35%

Marginal distribution of X is skewed toward $X=1$

However when examining the individual departments, it was found that no department was significantly biased against women. In fact, most departments had a "small but *statistically significant bias* in favor of women".[14]

Department $Y=$	Men ($X=1$)		Women ($X=2$)	
	Applicants	Admitted	Applicants	Admitted
1	825	62%	108	82%
2	560	63%	25	68%
3	325	37%	593	34%
4	417	33%	375	35%
5	191	28%	393	24%
6	272	6%	341	7%

Conditional distribution of X given $Y=y$ is fair.

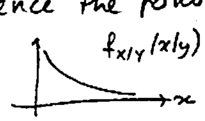
The research paper by Bickel, et al.[14] concluded that women tended to apply to competitive departments with low rates of admission even among qualified applicants (such as in the English Department), whereas men tended to apply to less-competitive departments with high rates of admission among the qualified applicants (such as in engineering and chemistry). The conditions under which the admissions' frequency data from specific departments constitute a proper defense against charges of discrimination are formulated in the book *Causality* by Pearl.[2]

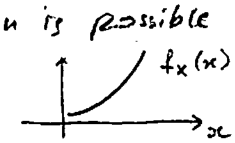
- Inferring the marginal from conditional distr should be done carefully:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\, \underbrace{f_Y(y)}\, dy$$

depends on distr. of $\underline{\underline{Y}}$!

- Hence the following situation is possible:





trend reverses!