➢ Example. If $X$ and $Y$ have a joint pdf

$$f(x,y)=\frac{2}{(1+x+y)^3},$$

examine whether it's a joint pdf.

for $0\le x,\ y<\infty$, then

$$f_X(x)=\int_0^\infty f(x,y)\,dy=-\frac{1}{(1+x+y)^2}\Big|_0^\infty=\frac{1}{(1+x)^2},$$

for $0\le x<\infty$. So,

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{2(1+x)^2}{(1+x+y)^3},$$

fixed; random

and, $P(Y>c|X=x)=\int_c^\infty \frac{2(1+x)^2}{(1+x+y)^3}\,dy$

$X\in x\pm\frac{\Delta x}{2}$

$$=-\frac{(1+x)^2}{(1+x+y)^2}\Big|_{y=c}^\infty=\frac{(1+x)^2}{(1+x+c)^2}.$$

a function of $x$

joint / marginal

- Mixed Joint Distribution: Definition of conditional distribution can be similarly generalized to the case in which some random variables are discrete and the others continuous (see a later example).

Recall. The 3 laws in LNp.4-11~13. $P(A_1\cap\cdots\cap A_n)=P(A_1)\cdot P(A_2|A_1)\cdots\cdot P(A_n|A_1\cap\cdots\cap A_{n-1})$

- Theorem (Multiplication Law). Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors and $(\mathbf{X},\mathbf{Y})$ have a joint pdf $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$/pmf $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$, then

$P(\mathbf{X}=\mathbf{x}\cap\mathbf{Y}=\mathbf{y})$; cf. independent

$$p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})=p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\times p_{\mathbf{X}}(\mathbf{x}),\ \text{or}$$

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})=f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\times f_{\mathbf{X}}(\mathbf{x}).$$

similar extension: $p_{X_1,\dots,X_n}(x_1,\dots,x_n)=p_{X_1}(x_1)\times p_{X_2|X_1}(x_2|x_1)\times\cdots\times p_{X_n|X_1\dots X_{n-1}}(x_n|x_1,\dots,x_{n-1})$



Proof. By the definition of conditional distribution.

pmf (LNp.7-51), pdf (LNp.7-53)

$P(B)=\sum_i P(B|A_i)P(A_i)$, partition

- Theorem (Law of Total Probability). Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors and $(\mathbf{X},\mathbf{Y})$ have a joint pdf $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$/pmf $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$, then

$\sum_{x\in\mathcal{X}} p_{X,Y}(x,y)$

$$p_{\mathbf{Y}}(\mathbf{y})=\sum_{\mathbf{x}=-\infty}^{\infty}p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})p_{\mathbf{X}}(\mathbf{x}),\ \text{or}$$

$\sum_{x\in\mathcal{X}}$

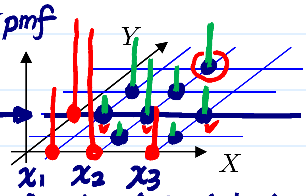
$A_x=\{X=x\},\ x\in\mathcal{X}$ form a partition of $\Omega$

$P(Y=y)$

$$f_{\mathbf{Y}}(\mathbf{y})=\int_{-\infty}^{\infty}f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x}.$$

$\int_{\mathbb{R}^n}$

joint pmf




$p_Y(y)$; $p_{Y|X}(y|x_1)$; $p_{Y|X}(y|x_2)$; $p_{Y|X}(y|x_3)$

partition: $P(A_i)\xrightarrow{\text{update}}P(A_i|B)$

Proof. By the definition of marginal distribution and the multiplication law.

- Theorem (Bayes Theorem). Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors and $(\mathbf{X},\mathbf{Y})$ have a joint pdf $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$/pmf $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$, then

update

joint / marginal

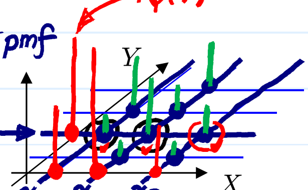
$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})=\frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})p_{\mathbf{X}}(\mathbf{x})}{\sum_{\mathbf{x}=-\infty}^{\infty}p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})p_{\mathbf{X}}(\mathbf{x})},\ \text{or}$$

meaning

$\frac{p_{X,Y}(x,y)}{p_Y(y)}$

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

$p_Y(y)$; joint pmf




$p_{X|Y}(x|y)$; $p_{Y|X}(y|x_3)$; $p_{Y|X}(y|x_2)$; $p_{Y|X}(y|x_1)$

Proof. By the definition of conditional distribution, multiplication law, and the law of total probability.

Example.

Bayesian estimation

Gold coins example in LNp.4-29

$\xrightarrow{\text{before}}$ 1st $\to$ ? $\xrightarrow{\text{after}}$ 2nd $\to$ B $\to (Y_1, \cdots, Y_n)$ (LNp.4-14)

perform Bernoulli trials (depending on $X$)

Note. $X$: continuous, $Y_1, \cdots, Y_n$: discrete

- Suppose that $X \sim$ Uniform(0, 1), and $(Y_1, \ldots, Y_n|X=x)$ are i.i.d. with Bernoulli($x$), i.e.,

conditional independent

Q: $Y_1, \cdots, Y_n$ indep.? NO

$$p_{\mathbf{Y}|X}(y_1, \ldots, y_n|x) = x^{y_1+\cdots+y_n}(1-x)^{n-(y_1+\cdots+y_n)},$$

fixed

for $y_1, \ldots, y_n \in \{0, 1\}$.

after 1st stage / before 1st stage cf.

joint dist.: a mix of pmf & pdf

- By the multiplication law, for $y_1, \ldots, y_n \in \{0, 1\}$ and $0<x<1$,

$$p_{\mathbf{Y},X}(y_1, \ldots, y_n, x) \overset{\downarrow}{=} x^{y_1+\cdots+y_n}(1-x)^{n-(y_1+\cdots+y_n)} \times \Delta x$$

$X \in x \pm (\Delta x/2)$ random

Intuition. What values of $x$ are more probable?

- Suppose that we observed $Y_1=1, \ldots, Y_n=1$. (or $Y_1=y_1, \cdots, Y_n=y_n$)

$k$ "1", $n-k$ "0" $\Rightarrow k=\sum_{i=1}^n y_i$

- By the law of total probability,

marginal dist. of $Y_1, \cdots, Y_n$

$$P(Y_1=1, \ldots, Y_n=1) = p_{\mathbf{Y}}(1, \ldots, 1)$$
$$= \int_0^1 p_{\mathbf{Y}|X}(1, \ldots, 1|x) f_X(x)\, dx$$
$$= \int_0^1 x^n\, dx = \frac{1}{n+1}x^{n+1}\Big|_0^1 = \frac{1}{n+1}.$$

(exercise)

$$p_{\mathbf{Y}}(y_1, \cdots, y_n) = \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{k!(n-k)!}{(n+1)!} = \frac{(\sum_{i=1}^n y_i)!(n-\sum_{i=1}^n y_i)!}{(n+1)!}$$



- And, by Bayes' Theorem,

$\frac{\text{joint}}{\text{marginal}}$

$$f_{X|\mathbf{Y}}(x|Y_1=1, \ldots, Y_n=1)$$
$$= \frac{p_{\mathbf{Y}|X}(1, \ldots, 1|x)\, f_X(x)}{p_{\mathbf{Y}}(1, \ldots, 1)} = (n+1)x^n.$$

update

$$\frac{\Gamma(n+2)}{\Gamma(n+1)\Gamma(1)} x^{(n+1)-1}(1-x)^{1-1}$$

$p_{\mathbf{Y}}(1, \ldots, 1) = \int_0^1 p_{\mathbf{Y}|X}(1, \cdots, 1|x) f_X(x) dx$

$E(X|Y_1=1, \cdots, Y_n=1) = \frac{n+1}{n+2}$

for $0<x<1$, i.e., $(X|Y_1=1, \ldots, Y_n=1) \sim$ Beta($n+1$, 1). update

Why is $X$ still random?

(cf., marginal distribution of $X \sim$ Uniform(0, 1)=Beta(1, 1).)

- If there were an $(n+1)^{\text{st}}$ Bernoulli trial $Y_{n+1}$,

probability of getting "1" after we know $Y_1=Y_2=\cdots=Y_n=1$

$$P(Y_{n+1}=1|Y_1=1, \ldots, Y_n=1)$$
$$= \frac{P(Y_1=1, \ldots, Y_{n+1}=1)}{P(Y_1=1, \ldots, Y_n=1)} = \frac{1/(n+2)}{1/(n+1)} = \frac{n+1}{n+2}.$$

- $Y_{n+1}|Y_1=y_1, \cdots, Y_n=y_n \sim$ Bernoulli$\left(\frac{\sum_{i=1}^n y_i+1}{n+2}\right)$ (exercise)
- $Y_{n+1} \sim$ Bernoulli($\frac{1}{2}$) (exercise)
- $Y_{n+1}|X=x \sim$ Bernoulli($x$)

cf.

Why not $x$?

- (exercise) In general, it can be shown that

$$(X|Y_1=y_1, \ldots, Y_n=y_n) \sim \text{Beta}((y_1+\cdots+y_n)+1, n-(y_1+\cdots+y_n)+1).$$

$E(X|Y_1=y_1, \cdots Y_n=y_n) = (\sum_{i=1}^n y_i+1)/(n+2)$ ← a Bayesian estimator of $x$.

- Theorem (Conditional Distribution & Independent). Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors and $(\mathbf{X}, \mathbf{Y})$ have a joint pdf $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$/pmf $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$. Then, $\mathbf{X}$ and $\mathbf{Y}$ are independent, i.e.,

$$p_{Y|X}(y|x) \cdot p_X(x) = p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = p_{\mathbf{X}}(\mathbf{x}) \times p_{\mathbf{Y}}(\mathbf{y}), \quad \text{or}$$

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) \times f_{\mathbf{Y}}(\mathbf{y}),$$

if and only if

$$\frac{p_{X,Y}(x,y)}{p_X(x)} = p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p_{\mathbf{Y}}(\mathbf{y}), \quad \text{or}$$

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = f_{\mathbf{Y}}(\mathbf{y}).$$

use the concept to examine whether the $(Y_1, \cdots, Y_n)$ in LNp 7-58 are indep.?

Proof. By the definition of conditional distribution.

- Intuition.
  - The 2 graphs about the joint pmf/pdf of independent r.v.'s in LNp.7-27
  - $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ or $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ offers information about the distribution of $\mathbf{Y}$ when $\mathbf{X}=\mathbf{x}$.

    $p_{\mathbf{Y}}(\mathbf{y})$ or $f_{\mathbf{Y}}(\mathbf{y})$ offers information about the distribution of $\mathbf{Y}$ when $\mathbf{X}$ not observed.

    cf.

❖ **Reading**: textbook, Sec 6.4, 6.5