PROBABILITY THEORY

2° LECTURE - 27/11/2018

We now move to dimensions >1, or equivalently to set of R.V. defined on the same probability space:

Multivariate random variable

We consider r.v.'s defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that we look globally as a random vector or a random variable on $\mathbb{R}^n$:

$$\underline{X} = (X_1, X_2, \dots, X_n) : \Omega \longrightarrow \mathcal{X} \subseteq \mathbb{R}^n$$
$$\omega \longrightarrow \underline{X}(\omega) = (X_1(\omega), \dots, X_n(\omega))$$

DISCRETE CASE: If $\mathcal{X}$ is finite or countable,

- the joint density of $\underline{X}$ is given by

$$p_{\underline{X}}(\underline{x}) = \mathbb{P}(\underline{X} = \underline{x}) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n), \qquad \underline{x} = (x_1, \dots, x_n) \in \mathcal{X}$$

which characterizes (as for r.v. in $\mathbb{R}$) the law of

$$\mathbb{P}(\underline{X} \in A) = \sum_{\underline{x} \in A} \mathbb{P}(\underline{X} = \underline{x}) \quad \forall A \subseteq \mathcal{X}$$

- the marginal density of X_i, $i = 1, \dots, n$, is given by

$$p_{X_i}(x_i) = \mathbb{P}(X_i = x_i) = \sum_{x_1, \dots x_{i-1}, x_{i+1} \dots x_n} \mathbb{P}(\underbrace{X_1 = x_1, \dots, X_n = x_n}_{\{X_1 = x_1\} \cap \dots \cap \{X_n = x_n\}})$$

Notation

and similarly, the marginal density of the subvector $(X_{i_1}, \dots, X_{i_k})$ for $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ is

$$\mathbb{P}(X_{i_1} = x_{i_1}, \dots X_{i_k} = x_{i_k}) = \sum_{\substack{x_j \\ \forall j \neq i_1, \dots, i_k}} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

<u>CONTINUOUS CASE</u>: If $\mathcal{X}$ is not countable (say $\mathbb{R}^n$ or $A \subset \mathbb{R}^n$)

- the joint distribution of $\underline{X}$ is given by

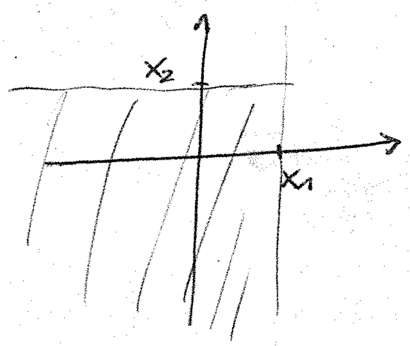
$$F_{\underline{X}}(\underline{x}) = F_{(X_1, \dots X_n)}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

for $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$

where $F_{\underline{X}} : \mathbb{R}^n \longrightarrow [0,1]$ s.t.

1. $\displaystyle\lim_{\substack{x_1 \to +\infty \\ \vdots \\ x_n \to +\infty}} F_{\underline{X}}(\underline{x}) = 1$ 2. $\displaystyle\lim_{x_j \to -\infty} F_{\underline{X}}(\underline{x}) = 0 \quad \forall j = 1, \dots n$

3. right continuity: $\displaystyle\lim_{\substack{x_1 \to y_1^+ \\ \vdots \\ x_n \to y_n^+}} F_{\underline{X}}(\underline{x}) = F_{\underline{X}}(\underline{y})$



Notice:

$$\{X_1 \leq x_1, \dots, X_n \leq x_n\} = \{\underline{X} \in \underbrace{(-\infty, x_1] \times \dots \times (-\infty, x_n]}_{\text{half-space}}\}$$

$\mathcal{B}(\mathbb{R}^n)$ = σ-algebra generated by half-spaces, as $\underline{x}$ varies.

- the marginal distribution of X_i, $i = 1, \dots, n$, is

$$F_{X_i}(x_i) = \mathbb{P}(X_i \leq x_i) = \lim_{\substack{x_j \to +\infty \\ \forall j \neq i}} F_{\underline{X}}(\underline{x})$$

since $\displaystyle\lim_{\substack{x_j \to \infty \\ \forall j \neq i}} \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_i \leq x_i)$

- **Def:** $\underline{X}$ is absolutely continuous r.v., if $\exists f: \mathbb{R}^n \to \mathbb{R}^+$ s.t.

$$F_{\underline{X}}(\underline{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(x_1, \dots, x_n)\, dx_1 \dots dx_n$$

f is called joint density of $\underline{X} = (X_1, \dots, X_n)$ and satisfies:

i. $\int \cdots \int_{\mathbb{R}^n} f(x_1 \dots x_n)\, dx_1 \dots dx_n = 1$ ii. $\dfrac{\partial^n F(x_1 \dots x_n)}{\partial x_1 \dots \partial x_n} = f(x_1, \dots, x_n)$ $\forall$ all $\underline{x}$ s.t. $F(\underline{x})$ is n-differentiable

* marginal density $f_{X_i}(x_i) = \int \cdots \int_{\mathbb{R}^{n-1}} f(x_1 \dots x_n)\, d\tilde{\underline{X}}^i$ where $d\tilde{x}^i = dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$

Examples

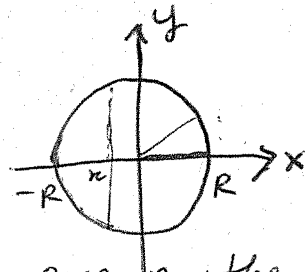
1. Let $\underline{X} = (X_1, X_2)$ a uniform r.v. on $D_0(R) = \left\{ \begin{matrix} (x,y) \in \mathbb{R}^2 \text{ s.t.} \\ x^2 + y^2 \leq R^2 \end{matrix} \right\}$

Determine the joint density of (X_1, X_2) and their marginal densities. Compute $\mathbb{P}(\underline{X} \in D_0(r))$ with $r < R$.

Sol For uniform r.v., $f(x,y) = c \cdot \mathbb{1}_{D_0(R)}(x,y)$ where $c = (\text{Leb}(D_0(R)))^{-1}$, with $\text{Leb}(D_0(R)) = \iint_{D_0(R)} dx\,dy$

Then compute

$$\iint_{D_0(R)} dx\,dy = \int_0^{2\pi} d\theta \int_0^R dr \cdot r = \pi R^2$$

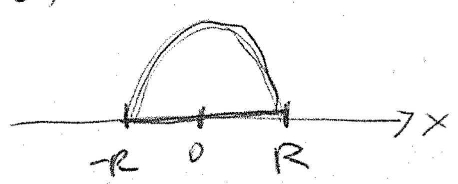


To compute the marginal densities, is useful express the range of $\underline{X}$ in simple form w.r.t to x and y:

$$D_0(R) = \left\{ (x,y) \in \mathbb{R}^2 : x \in [-R, R],\ y \in \left[-\sqrt{R^2 - x^2}, \sqrt{R^2 - x^2}\right] \right\}$$

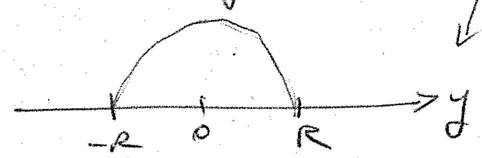
$$= \left\{ (x,y) \in \mathbb{R}^2 : y \in [-R, R],\ x \in \left[-\sqrt{R^2 - y^2}, \sqrt{R^2 - y^2}\right] \right\}$$

Then $f_{X_1}(x) = \int_{\mathbb{R}} f(x,y)\,dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy \cdot \mathbb{1}_{[-R,R]}(x)$



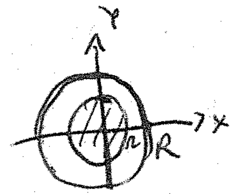
$$= 2\sqrt{R^2-x^2}\,\mathbb{1}_{[-R,R]}(x)$$

and $f_{X_2}(y) = \int_{\mathbb{R}} f(x,y)\,dx = \int_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} dx \; \mathbb{1}_{[-R,R]}(y) = 2\sqrt{R^2-y^2}\,\mathbb{1}_{[-R,R]}(y)$



Finally

$$\mathbb{P}(\underline{X} \in D_0(r)) = \iint_{D_0(r)} f(x,y)\,dx\,dy = \iint_{D_0(r)\cap D_0(R) = D_0(r)} \frac{1}{\pi R^2}\,dx\,dy = \frac{leb(D_0(r))}{\pi R^2}$$



$$= \frac{\cancel{\pi} r^2}{\cancel{\pi} R^2} = \frac{r^2}{R^2}$$

---

2. We throw 2 dices and denote by

X = number obtained from 1° dice

Y = sum of the two numbers obtained

Compute the joint density of (X,Y) and the marginal density on Y.

First identify $\mathcal{X}$ = range of $(X,Y) = \{(j,k),\ j=1,\dots,6,\ j+1 \le k \le j+6\}$

$$\mathbb{P}(X=j, Y=k) \underset{\downarrow}{=} \mathbb{P}(X=j, Z=k-j\} = \frac{1}{36}$$

where Z = number obtained from 2° dice

$$\mathbb{P}(Y=k) = \sum_{j=1}^{6} \mathbb{P}(X=j, Y=k) = \sum_{j=(k-6)\vee 1}^{(k-1)\wedge 6} \frac{1}{36}$$

$a \wedge b = \min\{a,b\}$

$a \vee b = \max\{a,b\}$

$\left(\begin{array}{l} \text{es } k=2 \quad = \frac{1}{36} \\ \quad\;\; k=3 \quad = \frac{1}{18} \end{array} \quad \text{and so on} \right)$

If $\underline{X} = (X_1, \dots, X_m)$ r.v, then

$$\mathbb{E}(\underline{X}) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_m))$$ vector of averages of the components, computed using marginal densite

Moreover, if $g: \mathbb{R}^m \longrightarrow \mathbb{R}$, and $Y = g(\underline{X})$

then
$$\mathbb{E}(Y) = \int\!\!\int_{\mathbb{R}^m} g(\underline{x})\, dF_{\underline{X}}(\underline{x}) = \begin{cases} \sum_{x_1} \cdots \sum_{x_m} g(x_1, \dots x_n)\, \mathbb{P}(X_1 = x_1, \dots X_n = x_n) & \text{discrete case} \\ \int_{\mathbb{R}} dx_1 \dots \int_{\mathbb{R}} dx_n\, g(x_1 \dots x_n)\, f(x_1, \dots, x_n) & \text{continuous ca} \end{cases}$$

Ex: Given (X,Y) with joint density $f(x,y)$, then

$$\mathbb{E}(X \cdot Y) = \int\!\!\int dx\, dy\; x \cdot y\; f(x,y)$$

Useful to compute the covariance between X and Y

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X)\mathbb{E}(Y)$$

(Also recall $\mathrm{Cov}(aX+b, cY+d) = a \cdot c\, \mathrm{Cov}(X,Y) \quad \forall a,b,c,d \in \mathbb{R}$
the bilinear property)

Independence of random variables

Def: We say that the r.v. X and Y are independent if

* $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(X \in B)$

$\forall A \subset \mathcal{X}_X$ and $B \subset \mathcal{X}_Y$ ($\mathcal{X}_X$ = range of X, $\mathcal{X}_Y$ = range of Y)

or equivalently if

* $F_{(X,Y)}(x,y) = F_X(x) \cdot F_X(y) \quad \forall (x,y) \in \mathbb{R}^2$ (joint distribution is product of marginal)

and similarly we say that $X_1, \dots, X_n$ are independent

if
$$F_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} F_{X_i}(x_i) \quad \forall \ \underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$$

Consequences

- X, Y are independent if the joint density is product of marginal densities

  * discrete case: $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ $\forall \ x, y \in \mathbb{R}^2$

  * continuous case: $f_{XY}(x, y) = f_X(x) f_Y(y)$

- If $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$, and X, Y indep.

$$\Longrightarrow \mathbb{E}(g(X) \cdot h(Y)) = \mathbb{E}(g(X))\,\mathbb{E}(h(Y))$$

Es: If X and Y are independent, then

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

$h(y) = y$
$g(x) = x$ $\longrightarrow$

$$\Longrightarrow \mathrm{Cov}(X, Y) = 0$$

that is, if X and Y are indep. $\Longrightarrow$ X and Y are uncorrelated

Note: The viceversa is not true in general!

(More generally, if X, Y are independent, then
$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\text{)}$$

CONDITIONAL DISTRIBUTION

A) We already notice that given $B \in \mathcal{F}$ st $\mathbb{P}(B) > 0$, then $\mathbb{P}(\cdot \mid B)$ is a probability on $(\Omega, \mathcal{F})$.

<u>Def</u>: Given $X : \Omega \to \mathcal{X}$ a discrete R.V., we can define

- the <u>density of X conditionated to B</u> as

$$\mathbb{P}(X = x \mid B), \quad x \in \mathcal{X}$$

- the <u>average of X conditionated to B</u> as

$$\mathbb{E}(X \mid B) = \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}(X = x \mid B)$$

In particular if $(A_n)_{n \in \mathbb{N}}$ are disjoint sets st $\Omega = \bigsqcup_{n=1}^{\infty} A_n$

$$\Longrightarrow \underline{\mathbb{E}(X)} = \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}(X = x) = \sum_{n=1}^{\infty} \underbrace{\sum_{x \in \mathcal{X}} x \cdot \mathbb{P}(X = x \mid A_n)}_{= \mathbb{E}(X \mid A_n)} \mathbb{P}(A_n)$$

$$= \underline{\sum_{n=1}^{\infty} \mathbb{E}(X \mid A_n) \mathbb{P}(A_n)}$$

#

<u>Example</u>: We have two boxes:
- Box A = balls numbered from 1 to 6 (6 total
- Box B = balls numbered from 3 to 10 (8 total

From a box chosen uniformly at random we extract a ball and denote by X its number.

If $E = \{\text{we have chosen box } A\}$, compute $\mathbb{E}(X \mid E)$ and $\mathbb{E}(X)$.

<u>Sol</u>: We have $\mathbb{P}(X = k \mid E) = \frac{1}{6}$ $\forall k = 1, \dots, 6$, and $= 0$ otherwise

$\mathbb{P}(X = k \mid E^c) = \frac{1}{8}$ $\forall k = 3, \dots, 10$, and $= 0$ otherwise

Then $\mathbb{E}(X \mid E) = \sum_{k=1}^{6} k \cdot \frac{1}{6} = \frac{7}{2}$, $\mathbb{E}(X \mid E^c) = \sum_{k=3}^{10} k \cdot \frac{1}{8} = \frac{52}{8}$

and $\mathbb{E}(X) = \mathbb{E}(X \mid E) \underbrace{\mathbb{P}(E)}_{= 1/2} + \mathbb{E}(X \mid E^c) \underbrace{\mathbb{P}(E^c)}_{= 1/2} = \frac{7}{4} + \frac{52}{16}$ #

B) Let $X, Y : \Omega \longrightarrow \mathcal{X}$ discrete r.v.

with joint density $P_{X,Y}(k,j) = \mathbb{P}(X=k, Y=j)$, $(k,j) \in \mathcal{X}$

Def: the density of X given that $\{Y=j\}$, for $\mathbb{P}(Y=j)>0$

$$(*) \quad P_{X|Y=j}(k) := \mathbb{P}(X=k \mid Y=j) = \frac{\mathbb{P}(X=k, Y=j)}{\mathbb{P}(Y=j)} = \frac{P_{X,Y}(k,j)}{P_Y(j)}$$

Remark:

i. $\mathbb{P}(X=k \mid Y=j) \in [0,1] \quad \forall k \in \mathcal{X}_X \implies \mathbb{P}(X=k \mid Y=j)$, $k \in \mathcal{X}_X$

ii. $\sum_{k \in \mathcal{X}_X} \mathbb{P}(X=k \mid Y=j) = 1$ is a density on $\mathcal{X}_X$

Similarly, if $\{X=k\}$ is s.t. $\mathbb{P}(X=k)>0$, one can define

$$P_{Y|X=k}(j) := \mathbb{P}(Y=j \mid X=k) = \frac{\mathbb{P}(X=k, Y=j)}{\mathbb{P}(X=k)}.$$

Remark

1) $(*)$ is equivalent to $P_{X,Y}(k,j) = P_{X|Y=j}(k) P_Y(j)$

2) If X and Y are independent, then

- $P_{X|Y=j}(k) = \frac{P_X(k) P_Y(j)}{P_Y(j)} = P_X(k)$
- $P_{Y|X=k}(j) = \dots = P_Y(j)$

3) $$F_{X|Y=j}(x) = \sum_{k \leq x} P_{X|Y=j}(k) = \sum_{k \leq x} \frac{P_{XY}(j,k)}{P_Y(j)}$$

Ex) Show that if X_1, X_2 are indep. r.v.,

a. If $X_1 \sim Bi(m_1,p)$, $X_2 \sim Bi(m_2,p) \Rightarrow X_1 | X_1+X_2 = n \sim Iper(m_1, m_1+m_2, n)$

b. If $X_1, X_2 \sim Geo(p) \Rightarrow X_1 | X_1+X_2 = n \sim U\{1,\dots,n-1\}$

c. If $X_1 \sim Po(\lambda)$, $X_2 \sim Po(\mu) \Rightarrow X_1 | X_1+X_2 = n \sim Bi\left(n, \frac{\lambda}{\lambda+\mu}\right)$

Sol:

a. We want to compute

$$P(X_1=k \mid X_1+X_2=n) = \frac{P(X_1=k,\ X_1+X_2=n)}{P(X_1+X_2=n)} \overset{*}{=}$$

Denominator: Since $X_1+X_2 \sim Bi(m_1+m_2, p)$,

$$P(X_1+X_2=n) = \binom{m_1+m_2}{n} p^n (1-p)^{m_1+m_2-n}$$

Numerator: $P(X_1=k,\ X_1+X_2=n) = P(X_1=k,\ X_2=n-k) =$

$$= P(X_1=k)P(X_2=n-k) = \binom{m_1}{k} p^k (1-p)^{m_1-k} \cdot \binom{m_2}{n-k} p^{n-k} (1-p)^{m_2-n+k} =$$

from indep.

if k s.t. $\begin{cases} 0 \le k \le m_1 \wedge n \\ 0 \le n-k \le m_2 \end{cases} \Rightarrow \begin{cases} 0 \le k \le m_1 \wedge n \\ n-m_2 \le k \le n \end{cases} \Rightarrow \begin{matrix} k \le m_1 \wedge n \\ k \ge 0 \wedge n-m_2 \end{matrix}$

$$= \binom{m_1}{k}\binom{m_2}{n-k} p^n (1-p)^{m_1+m_2-n} \quad \text{, and then}$$

$$\overset{*}{=} \frac{\binom{m_1}{k}\binom{m_2}{n-k} \cancel{p^n (1-p)^{m_1+m_2-n}}}{\binom{m_1+m_2}{n} \cancel{p^n (1-p)^{m_1+m_2-n}}} = \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}} \quad \forall \begin{matrix} k \le m_1 \wedge n \\ k \ge 0 \wedge n-m_2 \end{matrix}$$

which corresponds to the density of $Iper(m_1, m_1+m_2, n)$ #

Solve point b. and c. for exercise.

Accordingly with the definition of average condition to an event, we define

$$* \quad \mathbb{E}(X|Y=j) = \sum_{k \in \mathcal{X}} k \cdot P_{X|Y=j}(k) \qquad \left(\begin{array}{c}\text{average of } X \\ \text{given that } Y=j\end{array}\right)$$

Remark:

i) Notice that $\mathbb{E}(X|Y=j) = g(j)$ $\left(\begin{array}{c}\text{deterministic quantity} \\ \text{depending on } j\end{array}\right)$

ii) If X and Y are independent, then

$$\underline{\mathbb{E}(X|Y=j) = \mathbb{E}(X)} \qquad \left(\begin{array}{c}\text{as } P_{X|Y=j}(k) = P_X(k) \\ \text{under independence}\end{array}\right)$$

From i), we can think

$$\underbrace{\mathbb{E}(X|Y)}_{g(Y)} : \mathcal{X}_Y \longrightarrow \mathbb{R}$$
$$j \longrightarrow \mathbb{E}(X|Y=j)$$

Then, on one hand:

$$\underline{\mathbb{E}(X)} = \sum_{j \in \mathcal{X}_Y} \underbrace{\mathbb{E}(X|Y=j)}_{g(j)} \mathbb{P}(Y=j) =$$

$$= \mathbb{E}(g(Y)) = \underline{\mathbb{E}(\mathbb{E}(X|Y))}$$

Ex Let $T_1, \dots, T_n$ independent and s.t. $T_k \sim Bi(k, p)$

Let $Y \sim \mathcal{U}\{1, 2, \dots n\}$ indep. of T_k's and define

$$X := T_Y = \begin{cases} T_1 & \text{if } Y=1 \\ \vdots & \\ T_n & \text{if } Y=n \end{cases}$$

Compute $\mathbb{E}(Y)$ and $\mathbb{E}(X)$.

Sol: First $\mathbb{E}(Y) = \sum_{k=1}^{n} k \cdot \frac{1}{n} = \frac{1}{n} \frac{n \cdot (n+1)}{2} = \frac{n+1}{2}$

Also recall that $\mathbb{E}(T_k) = k \cdot p$, $\forall k = 1, \dots, n$

Then
$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$$
$$= \sum_{k=1}^{n} \mathbb{E}(X|Y=k)\mathbb{P}(Y=k)$$
$$= \sum_{k=1}^{n} \frac{1}{n} \mathbb{E}(T_k) = \sum_{k=1}^{n} \frac{p \cdot k}{n} = p\frac{n+1}{2}$$

\#

Note: We may think that $X = T_Y \sim \text{Bin}(Y, p)$ and thus: $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(T_Y)) = \mathbb{E}(Y \cdot p)$
$$= p \cdot \mathbb{E}(Y) = p\frac{(n+1)}{2}$$

\#

---

c) let $X, Y : \Omega \to \mathcal{X} \subseteq \mathbb{R}^2$ absolutely cont. r.v's with joint density $f_{X,Y}(x,y)$, $(x,y) \in \mathbb{R}^2$ and let y s.t. $f_Y(y) \neq 0$.

The density of X conditioned to $\{Y=y\}$ is

$$(*) \quad f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad \forall x \in \mathbb{R}$$

Remark:

i. $f_{X|Y}(x|y) \geq 0$ ii. $\int_{\mathbb{R}} f_{X|Y}(x|y)\,dx = 1 \quad \forall y \in \mathcal{X}_y$

Similarly, for x s.t. $f_X(x) > 0$:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad \forall y \in \mathbb{R}$$

Remark

1. ⊛ is equivalent to $f_{X,Y}(x,y) = f_{X|Y}(x|y) \cdot f_Y(y)$

2. If X and Y are independent, then

$$f_{X|Y}(x|y) = f_X(x) \quad \text{and} \quad f_{Y|X}(y|x) = f_Y(y)$$

3. $$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(t|y)\,dt = \int_{-\infty}^{x} \frac{f_{X,Y}(t,y)\,dt}{f_Y(y)}$$

Similarly to the discrete case, we define the average of X conditioned on $\{Y=y\}$ as

$$\underbrace{\mathbb{E}(X|Y=y)}_{g(y)} = \int_{\mathbb{R}} x\, f_{X|Y}(x|y)\,dx$$

then
$$\begin{aligned} \mathbb{E}(X|Y) : \mathcal{X}_Y &\to \mathbb{R} \\ y &\to \mathbb{E}(X|Y=y) \end{aligned}$$

and it holds

$$\boxed{\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))} = \int_{\mathbb{R}} f_Y(y)\,\mathbb{E}(X|Y=y)\,dy \overset{*}{=}$$

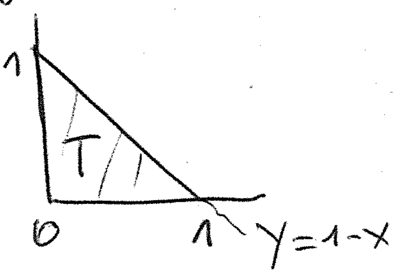
Indeed:

$$\overset{*}{=} \int_{\mathbb{R}} dy\, f_Y(y) \int_{\mathbb{R}} dx\, f_{X|Y}(x|y)\,x$$

$$= \iint_{\mathbb{R}^2} x\, \underbrace{f_Y(y) f_{X|Y}}_{f_{X,Y}(x,y)}(x|y)\,dx\,dy$$

Ex) Let X, Y r.v. with joint distribution $\mathcal{U}(T)$
where $T = \{(x,y) \in \mathbb{R}^2 : x \in [0,1],\ y \in [0, 1-x]\}$
Determine the joint density $f_{X,Y}$, the marginal densities f_X and f_Y (of X and Y), the conditional density of X given $\{Y = y\}$, for $y \in (0,1)$, and the conditioned average $\mathbb{E}(X \mid Y = y)$.

Sol



$$T = \{(x,y) \in \mathbb{R}^2 : x \in [0,1],\ y \in [0,1-x]\}$$
$$= \{(x,y) \in \mathbb{R}^2 : y \in [0,1],\ x \in [0,1-y]\}$$

By def. of uniform law,

$$f_{X,Y}(x,y) = c \cdot \mathbb{1}\{(x,y) \in T\}$$

where $c = (\mathrm{Leb}(T))^{-1} = \left(\frac{1}{2}\right)^{-1} = 2$

- $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy = 2\int_0^{1-x} dy \cdot \mathbb{1}(x \in (0,1)) = 2(1-x)\,\mathbb{1}(x \in (0,1))$
- $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx = 2\int_0^{1-y} dx\ \mathbb{1}(y \in (0,1)) = 2(1-y)\,\mathbb{1}(y \in (0,1))$

If $f_Y(y) \neq 0$, then

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\cancel{2}\,\mathbb{1}(y \in (0,1))\,\mathbb{1}(x \in (0,1-y))}{\cancel{2}(1-y)\,\mathbb{1}(y \in (0,1))} = \frac{1}{1-y}\,\mathbb{1}(x \in [0,1-y$$

Notice that this density correspond to $\mathcal{U}[0,1-y]$

that is $X|Y=y \sim \mathcal{U}[0,1-y]$.

Finally $\mathbb{E}(X|Y=y) = \int_{\mathbb{R}} x \cdot f_{X|Y}(x|y)\,dx = \int_0^{1-y} \frac{x}{1-y}\,dx$

$= \frac{1}{1-y}\frac{(1-y)^2}{2} = \frac{1-y}{2}$ # $\left(\begin{array}{l}\text{Indeed, if } Z \sim \mathcal{U}[a,b] \\ \Rightarrow \mathbb{E}(Z) = \frac{b-a}{2}\end{array}\right)$

As the conditional average is an average (w.r.t. to a conditioned probability), it has the same properties of usual average, that is:

a. Linearity : $\mathbb{E}(aX_1 + bX_2|Y=y) = a\mathbb{E}(X_1|Y=y) + b\mathbb{E}(X_2|Y=y)$

b. Positivity : If $X \geq 0 \Rightarrow \mathbb{E}(X|Y=y) \geq 0$

Moreover it holds:

c. $\mathbb{E}(g(X,Y)|Y=y) = \mathbb{E}[g(X,y)|Y=y]$

d. $\mathbb{E}(g(Y)|Y=y) = g(y)$

e. $\mathbb{E}(g(X)|Y=y) = \mathbb{E}(g(X))$ if X and Y are indep.

Def: The variance of X given $\{Y=y\}$ is defined:

$$\mathrm{Var}(X|Y=y) = \mathbb{E}\left[(X-\mathbb{E}(X|Y=y))^2 \,|\, Y=y\right]$$

$$\mathrm{Var}^y(X) = \mathbb{E}^y\left((X-\mathbb{E}^y(X))^2\right)$$

Ex: Prove that

$$\mathrm{Var}(X) = \mathbb{E}\big(\underbrace{\mathrm{Var}(X|Y)}_{h(Y)}\big) + \mathrm{Var}\big(\underbrace{\mathbb{E}(X|Y)}_{g(Y)}\big)$$

Sol:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mathbb{E}(X))^2\right] = \mathbb{E}\left[\mathbb{E}\left[(X-\mathbb{E}(X))^2 | Y\right]\right]$$

$$= \mathbb{E}\left[\mathbb{E}(X^2|Y) - 2\,\mathbb{E}(X|Y)\mathbb{E}(X) + \mathbb{E}(X)^2\right]$$

$$= \mathbb{E}\Big[\underbrace{\mathbb{E}(X^2|Y) - \mathbb{E}(X|Y)^2}_{\mathrm{Var}(X|Y)}\Big] + \mathbb{E}\Big[\big(\underbrace{\mathbb{E}(X|Y)}_{g(Y)} - \mathbb{E}(X)\big)^2\Big]$$

$$= \mathbb{E}\big(\mathrm{Var}(X|Y)\big) + \mathrm{Var}\big(\mathbb{E}(X|Y)\big)$$

#

Applications: Distribution with random parameters

1) Already seen $X|Y=k \sim Bi(k,p)$ for $Y \sim \mathcal{U}\{1,\dots,n\}$
In that case $X \sim Bi(Y,p)$ → random parameter

2) Assume that $X|Y=y \sim Po(y)$, with $Y \sim Exp(1)$
Compute density and average of X.

Sol: $X \sim Po(Y)$ → random parameter

$$\mathbb{E}(X) = \mathbb{E}\big(\underbrace{\mathbb{E}(X|Y)}_{Po(Y) \Rightarrow \mathbb{E}(Po(\mu))=\mu}\big) = \mathbb{E}(Y) = 1$$

or $$= \int_0^\infty e^{-y}\cdot \mathbb{E}(X|Y=y)\,dy = \int_0^\infty y e^{-y}\,dy = 1$$

To compute the density, $\forall k = 0, \ldots, n, \ldots$

$$\mathbb{P}(X=k) = \int_{\mathbb{R}} P(X=k \mid Y=y) \cdot f_Y(y)\,dy =$$

$$= \int_0^\infty e^{-y} \cdot \frac{y^k}{k!} e^{-y}\,dy = \frac{1}{k!} \int_0^\infty y^k e^{-2y}\,dy$$

$2y = z$
$2dy = dz$

$$= \frac{1}{k!}\frac{1}{2^{k+1}} \underbrace{\int_0^{+\infty} z^k e^{-z}\,dz}_{\Gamma(k+1) = k!} = \frac{\cancel{k!}}{\cancel{k!}}\frac{1}{2^{k+1}}$$

that is $(X+1) \sim Geo(\frac{1}{2})$.

Ex: Compute the density of X, knowing that

a) $X \mid Y=n \sim Bi(n,p)$ and $Y \sim Po(\lambda)$ $(X \sim Bi(Y,p))$

b) $X \mid Y=\lambda \sim N(0, \frac{1}{\lambda})$ and $Y \sim \Gamma(\frac{n}{2}, \frac{2}{n})$
