7 Conditioning

7a Some examples and paradoxes

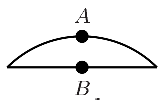
Discrete probability theory defines conditional probabilities by the elementary formula

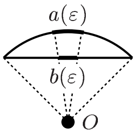
$$\mathbb{P}\left(\, A \,\middle|\, B \,\right) = \frac{\mathbb{P}\left(\, A \cap B \,\right)}{\mathbb{P}\left(\, B \,\right)} \tag{7a1}$$

whenever $A, B \in \mathcal{F}$, $\mathbb{P}\left(\, B \,\right) \ne 0$. Continuous probability theory strives to generalize $\mathbb{P}\left(\, A \,\middle|\, B \,\right)$ for the case $\mathbb{P}\left(\, B \,\right) = 0$.

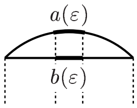
7a2 Example. A hare appears on a random distance R from a hunter. The hunter shoots, and hits the hare with a conditional probability $p_{\text{cond}}(R)$.[1] What is the unconditional probability p of hitting the hare? We'll see that $p = \mathbb{E}\, p_{\text{cond}}(R)$.[2] Discrete probability theory states that $p = \mathbb{E}\, p_{\text{cond}}(R)$ when the distribution of R is discrete; continuous probability theory states it for an arbitrary distribution of R, as we'll see soon.

7a3 Example. A point ω is chosen at random on the union of an arc and its chord, uniformly w.r.t. length.[3] Given that $\omega \in \{A, B\}$ (see the picture), what is the (conditional) probability that $\omega = A$? Naive limiting procedures give different answers.





$$\frac{\mathbb{P}\left(\, \{A\} \,\middle|\, \{A,B\} \,\right)}{\mathbb{P}\left(\, \{B\} \,\middle|\, \{A,B\} \,\right)} = \lim_{\varepsilon \to 0} \frac{a(\varepsilon)}{b(\varepsilon)} = \frac{|OA|}{|OB|} > 1\,.$$



$$\frac{\mathbb{P}\left(\, \{A\} \,\middle|\, \{A,B\} \,\right)}{\mathbb{P}\left(\, \{B\} \,\middle|\, \{A,B\} \,\right)} = \lim_{\varepsilon \to 0} \frac{a(\varepsilon)}{b(\varepsilon)} = 1\,.$$

The theory should explain, why the question is ill-formed.

7a4 Example. A random point on a circle is described by a random angle Φ distributed uniformly between 0 and 2π. Given that Φ/π is rational, what is the (conditional) probability that, say, $\Phi = \pi$?

The given distribution is invariant under all rotations. The given condition is invariant under rational rotations. The conditional distribution should be invariant under rational rotations. Thus, all rational angles should be (conditionally) equiprobable, which is impossible: their sum must be equal to 1.

The theory should explain, why the question is ill-formed.

---

[1] Say, $p_{\text{cond}}(r) = \frac{a^2}{r^2+a^2}$ where a is a given constant.

[2] Say, if R is distributed uniformly on (ba, ca), we get $p = \frac{1}{ca-ba}\int_{ba}^{ca} \frac{a^2}{r^2+a^2}\,dr = \frac{1}{c-b}\int_b^c \frac{du}{u^2+1} = \arctan c - \arctan b$.

[3] That is, pieces of equal length are of equal probability. In particular, the arc is longer than the chord, accordingly, the arc is more probable than the chord.

7b Conditional quantile functions and conditional distribution functions

A one-dimensional random variable X has its quantile function X^*, an increasing function on $(0,1)$ distributed like X. What could be its 2-dim counterpart?

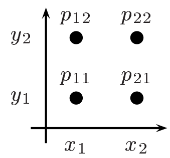
Having a 2-dim random variable (X,Y) we may consider X^* and Y^*, however, the 2-dim distribution of (X^*,Y^*) is in general quite different from the 2-dim distribution of (X,Y).

In order to reproduce the given 2-dim distribution, Y^* should depend on X^* as needed. We consider the pair

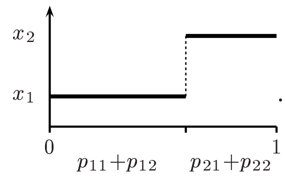
(7b1)
$$\begin{aligned} X^{**}(p_1,p_2) &= X^*(p_1)\,, \\ Y^{**}(p_1,p_2) &= Y^*(p_2|X^*(p_1))\,, \end{aligned}$$

assuming that $Y^*(\cdot|x)$ is an increasing function $(0,1)\to\mathbb{R}$ for every x (or rather, for P_X-almost every x), and $Y^*(\cdot|\cdot)$ is a Borel function on $(0,1)\times\mathbb{R}$.

7b2 Example. A discrete distribution:

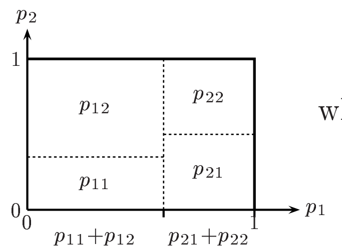


. We have X^*:

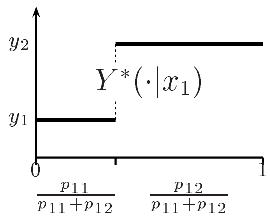


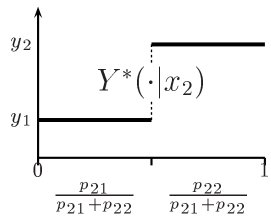
.

We want to construct two increasing functions $Y^*(\cdot|x_1)$ and $Y^*(\cdot|x_2)$ on $(0,1)$ such that (7b1) reproduces the given distribution,



which is easy to do:





The diistribution of $Y^*(\cdot|x_1)$ is just the conditional distribution of Y given $X=x_1$ as defined by discrete probability theory. The same for x_2.

7b3 Example. A random point (X,Y) is uniformly distributed on the circle $x^2+y^2=1$. We could represent it as

$$\begin{aligned} X &= \cos\Phi\,, \\ Y &= \sin\Phi\,, \end{aligned} \qquad \Phi\sim U(0,2\pi)\,,$$

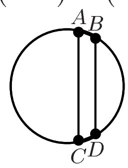
however, that is not of the form (7b1). We have $F_X(x)=\frac12+\frac1\pi\arcsin x$ for $x\in[-1,+1]$, therefore

$$X^*(p)=\sin\bigl(\pi(p-\frac12)\bigr)\,.$$

Intuition tells us that, given $X = x$, we have two equiprobable values for Y, namely, $\pm\sqrt{1-x^2}$. Accordingly, we define

$$Y^*(p_2|x) = \begin{cases} -\sqrt{1-x^2} & \text{for } p_2 < 1/2, \\ +\sqrt{1-x^2} & \text{for } p_2 > 1/2 \end{cases}$$

and ask, whether (X^{**}, Y^{**}) is distributed as needed, uniformly on the circle. The answer is positive. Indeed, the union of two arcs $(AB) \cup (CD)$ gets the right probability, and the arc (AB) gets the half of the probability.



7b4 Example. A random point (X, Y) is distributed uniformly on the disk $x^2 + y^2 < 1$. Even without calculating F_X and X^* we may guess that, given $X = x$, the distribution of Y is uniform on $(-\sqrt{1-x^2}, +\sqrt{1-x^2})$, that is,

$$Y^*(p, x) = \sqrt{1-x^2} \cdot (2p-1)\,.$$

That is true, as we'll see soon.

Representation (7b1) may be used for *defining* conditional probabilities. However, existence and uniqueness of the representation are not evident.

Assume for now that $P_{X,Y}$ admits a representation (7b1); that is, exists $Y^*(\cdot|\cdot)$ such that $F_{X^{**},Y^{**}} = F_{X,Y}$. Introduce the *conditional distribution function* $F_{Y|X}$ by

(7b5) $$F_{Y|X=x}(y) = \operatorname{mes}\{p \in (0,1) : Y^*(p|x) \le y\}\,;$$

that is, $F_{Y|X=x}(\cdot)$ is just the cumulative distribution function of $Y^*(\cdot|x)$. Still, existence and uniqueness of $F_{Y|X}$ are not evident, but first of all, what is $F_{Y|X}$ good for?

7b6 Theorem. $F_{X,Y}(x,y) = \mathbb{E}\big(\mathbf{1}_{(-\infty,x]}(X) F_{Y|X}(y)\big)$.

Proof. Denote for convenience $\psi(x_1) = F_{Y|X=x_1}(y)$, then $\mathbb{E}\big(\mathbf{1}_{(-\infty,x]}(X) F_{Y|X}(y)\big) = \mathbb{E}\big(\mathbf{1}_{(-\infty,x]}(X)\psi(X)\big)$ and, by (7b5),

$$\psi(x_1) = \operatorname{mes}\{p \in (0,1) : Y^*(p|x_1) \le y\} = \int_0^1 \mathbf{1}_{(-\infty,y]}\big(Y^*(p|x_1)\big)\, dp\,.$$

We have

$$\begin{aligned} F_{X,Y}(x,y) &= \mathbb{P}\big(\, X \le x \,\&\, Y \le y \,\big) = \mathbb{P}\big(\, X^{**} \le x \,\&\, Y^{**} \le y \,\big) = \\ &= \operatorname{mes}_2\{(p_1,p_2) : X^*(p_1) \le x \,\&\, Y^*\big(p_2|X^*(p_1)\big) \le y\} = \\ &= \int_0^1 \mathbf{1}_{(-\infty,x]}(X^*(p_1)) \Big(\int_0^1 \mathbf{1}_{(-\infty,y]}\big(Y^*(p_2|X^*(p_1))\big)\, dp_2 \Big)\, dp_1 = \\ &= \int_0^1 \mathbf{1}_{(-\infty,x]}(X^*(p_1))\psi(X^*(p_1))\, dp_1 = \mathbb{E}\big(\mathbf{1}_{(-\infty,x]}(X^*)\psi(X^*)\big) = \\ &= \mathbb{E}\big(\mathbf{1}_{(-\infty,x]}(X)\psi(X)\big) = \mathbb{E}\big(\mathbf{1}_{(-\infty,x]}(X) F_{Y|X}(y)\big)\,. \end{aligned}$$

$\square$

Introduce conditional expectations:

$$\mathbb{E}\,\bigl(\,\varphi(X,Y)\;\big|\;X=x\,\bigr)=\int_0^1 \varphi\bigl(x,Y^*(p|x)\bigr)\,dp=\int_{-\infty}^{+\infty}\varphi(x,y)\,dF_{Y|X=x}(y)\,. \tag{7b7}$$

Note that $\mathbb{E}\,\bigl(\,\varphi(X,Y)\;\big|\;X\,\bigr)$ is a random variable (a function of X).

7b8 Theorem. (a) $\mathbb{E}\,|\varphi(X,Y)|=\mathbb{E}\bigl(\mathbb{E}\,\bigl(\,|\varphi(X,Y)|\;\big|\;X\,\bigr)\bigr)\in[0,+\infty]$ for every Borel function $\varphi:\mathbb{R}^2\to\mathbb{R}$;

(b) $\mathbb{E}\,\varphi(X,Y)=\mathbb{E}\bigl(\mathbb{E}\,\bigl(\,\varphi(X,Y)\;\big|\;X\,\bigr)\bigr)\in\mathbb{R}$ for every Borel function $\varphi:\mathbb{R}^2\to\mathbb{R}$ such that $\mathbb{E}\,|\varphi(X,Y)|<\infty$.

Proof. Assume that $\varphi:\mathbb{R}^2\to[0,+\infty)$. Denote for convenience $\psi(x)=\mathbb{E}\,\bigl(\,\varphi(X,Y)\;\big|\;X=x\,\bigr)$, then $\mathbb{E}\bigl(\mathbb{E}\,\bigl(\,\varphi(X,Y)\;\big|\;X\,\bigr)\bigr)=\mathbb{E}\,\psi(X)$, and $\psi(x)=\int_0^1\varphi\bigl(x,Y^*(p|x)\bigr)\,dp$. We have

$$\begin{aligned}\mathbb{E}\,\varphi(X,Y)=\mathbb{E}\,\varphi(X^{**},Y^{**})&=\int_0^1\int_0^1\varphi\bigl(X^*(p_1),Y^*(p_2|X^*(p_1))\bigr)\,dp_1dp_2=\\&=\int_0^1\Bigl(\int_0^1\varphi\bigl(X^*(p_1),Y^*(p_2|X^*(p_1))\bigr)dp_2\Bigr)\,dp_1=\\&=\int_0^1\psi\bigl(X^*(p_1)\bigr)\,dp_1=\mathbb{E}\,\psi(X^*)=\mathbb{E}\,\psi(X)=\mathbb{E}\bigl(\mathbb{E}\,\bigl(\,\varphi(X,Y)\;\big|\;X\,\bigr)\bigr)\,.\end{aligned}$$

Applying it to $\varphi_+=\max(0,\varphi)$ and $\varphi_-=\max(0,-\varphi)$ we get both (a) and (b). $\square$

7b9 Exercise. Show that 7b6 is a special case of 7b8.

Hint: $\varphi(x_1,y_1)=\mathbf{1}_{(-\infty,x]}(x_1)\cdot\mathbf{1}_{(-\infty,y]}(y_1)$.

7b10 Exercise. $\mathbb{E}\,\bigl(\,\varphi_1(X)\varphi_2(X,Y)\;\big|\;X\,\bigr)=\varphi_1(X)\,\mathbb{E}\,\bigl(\,\varphi_2(X,Y)\;\big|\;X\,\bigr)$. Prove it.

7b11 Exercise. Show that $\mathbb{E}\,\bigl(\,\varphi_1(X)\varphi_2(Y)\;\big|\;X\,\bigr)=\varphi_1(X)\,\mathbb{E}\,\bigl(\,\varphi_2(Y)\;\big|\;X\,\bigr)$ which in general differs from $\varphi_1(X)\mathbb{E}\,\varphi_2(Y)$ (give an example).

7b12 Exercise. Show that the following conditions are equivalent.

(a) X and Y are independent.

(b) $Y^*(p\pm|x)=Y^*(p\pm)$ for all p.

(c) $F_{Y|X=x}(y)=F_Y(y)$ for all x,y.

(d) $\mathbb{E}\,\bigl(\,\varphi_1(X)\varphi_2(Y)\;\big|\;X\,\bigr)=\varphi_1(X)\cdot\mathbb{E}\,\varphi_2(Y)$ for all φ_1,φ_2.

7b13 Exercise. Show that

$$F_{X+Y}(u)=\mathbb{E}\,F_{Y|X}(u-X)$$

in general, and

$$F_{X+Y}(u)=\mathbb{E}\,F_Y(u-X)=\mathbb{E}\,F_X(u-Y)$$

when X,Y are independent.

Hint: $\varphi(x,y)=1_{(-\infty,u]}(x+y)$.

7b14 Exercise. Show that $\mathbb{P}\bigl(X^2+Y^2\le 1\bigr)=\mathbb{E}\bigl(\mathbf{1}_{[-1,+1]}(X)\bigl(F_Y(\sqrt{1-X^2})-F_Y(-\sqrt{1-X^2}-)\bigr)$ for independent X,Y.

Now we are in position to prove uniqueness of conditional quantile functions (as well as conditional distributions).

7b15 Theorem. *(Uniqueness)* Let $X^* : (0,1) \to \mathbb{R}$ be an increasing function, and $Y_1^*, Y_2^* : (0,1) \times \mathbb{R} \to \mathbb{R}$ be two Borel functions increasing in the first argument and such that the corresponding pairs (X^{**}, Y_1^{**}), (X^{**}, Y_2^{**}) defined by 7b1 are identically distributed. Then $Y_1^{**} = Y_2^{**}$ almost everywhere on $(0,1) \times (0,1)$.

Proof. We have $\mathbb{E}\,\varphi(X^{**}, Y_1^{**}) = \mathbb{E}\,\varphi(X^{**}, Y_2^{**})$ for every bounded Borel function $\varphi : \mathbb{R}^2 \to \mathbb{R}$. In particular, $\mathbb{E}\big(\mathbf{1}_A(X^{**})\mathbf{1}_B(Y_1^{**})\big) = \mathbb{E}\big(\mathbf{1}_A(X^{**})\mathbf{1}_B(Y_2^{**})\big)$ for every Borel sets $A, B \subset \mathbb{R}$. That is,

$$0 = \mathbb{E}\big(\mathbf{1}_A(X^{**})(\mathbf{1}_B(Y_1^{**}) - \mathbf{1}_B(Y_2^{**})\big) = \mathbb{E}\big(\mathbf{1}_A(X^{**})\,\mathbb{E}\,\big(\,\mathbf{1}_B(Y_1^{**}) - \mathbf{1}_B(Y_2^{**}) \bigm| X^{**}\,\big)\big)\,.$$

Taking $A_\pm = \{x : \int_0^1 \big(\mathbf{1}_B(Y_1^*(p|x)) - \mathbf{1}_B(Y_2^*(p|x))\big)\,dp \gtrless 0\}$ we conclude that $\mathbb{P}\,\big(\,X \in A_\pm\,\big) = 0$. Taking $B = (-\infty, y]$ we get $F^{(1)}_{Y|X}(y) = F^{(2)}_{Y|X}(y)$ almost surely; here $F^{(1)}, F^{(2)}$ are conditional distribution functions corresponding to Y_1^*, Y_2^* respectively. The exceptional set (of values of X) may depend on y. However, a single exceptional set serves all rational y, therefore (by right continuity of the distribution functions), all y. So, for P_X-almost every x, the two increasing functions $Y_1^*(\cdot|x)$, $Y_2^*(\cdot|x)$ are identically distributed, which means that they coincide up to a countable set (that depend on x). $\square$

Due to one-one correspondence between quantile functions and distribution functions, uniqueness of conditional quantile function implies uniqueness of conditional distribution function, as follows.

7b16 Corollary. Let $X, Y : \Omega \to \mathbb{R}$ be random variables, and $F : \mathbb{R}^2 \to [0,1]$ a Borel function such that $F(x,\cdot)$ is a cumulative distribution function for P_X-almost all x, and

$$F_{X,Y}(x,y) = \mathbb{E}\,\big(\mathbf{1}_{(-\infty,x]}(X)F(X,y)\big)$$

for all x, y. Then F is a conditional distribution function in the sense that

$$\mathbb{P}\,\big(\,\forall y \;\; F(X,y) = F_{Y|X}(y)\,\big) = 1\,.$$

Existence is more complicates than uniqueness, and I formulate it without proof.

7b17 Theorem. *(Existence.)* For every two-dimensional distribution there exist a conditional quantile function, and (therefore), a conditional distribution function.

7c Special cases

The following result shows that the general definitions conform with the elementary definition (7a1) when both are applicable.

7c1 Exercise. If x is an atom of X then

$$F_{Y|X=x}(y) = \frac{\mathbb{P}\,\big(\,Y \le y \,\&\, X = x\,\big)}{\mathbb{P}\,\big(\,X = x\,\big)}\,.$$

Prove it.

Hint: $\varphi(x_1, y_1) = \mathbf{1}_{\{x\}}(x_1) \cdot \mathbf{1}_{(-\infty,y]}(y_1)$.

7c2 Corollary. If X is discrete then

$$F_{Y|X=x}(y) = \mathbb{P}\left(Y \le y \,\middle|\, X = x \right) = \frac{\mathbb{P}\left(Y \le y \,\&\, X = x \right)}{\mathbb{P}\left(X = x \right)}$$

for all $x, y \in \mathbb{R}$ such that $\mathbb{P}\left(X = x \right) > 0$.

Turn to an absolutely continuous X (but arbitrary Y). By Radon-Nikodym theorem, X has a density f_X. Take any y such that $F_Y(y) > 0$ and consider[4]

$$F_{X|Y\le y}(x) = \mathbb{P}\left(X \le x \,\middle|\, Y \le y \right) = \frac{F_{X,Y}(x,y)}{F_Y(y)}\,.$$

That is also an absolutely continuous distribution, since

$$\text{(7c3)} \qquad \operatorname{mes} S = 0 \quad \Longrightarrow \quad \mathbb{P}\left(X \in S \right) = 0 \quad \Longrightarrow \quad \mathbb{P}\left(X \in S \,\middle|\, Y \le y \right) = 0\,.$$

Thus, it has a density,

$$F_{X|Y\le y}(x) = \int_{-\infty}^{x} f_{X|Y\le y}(x)\,dx\,.$$

We have

$$\begin{aligned} F_{X,Y}(x,y) &= \mathbb{P}\left(X \le x \,\&\, Y \le y \right) = \mathbb{P}\left(X \le x \,\middle|\, Y \le y \right)\, \mathbb{P}\left(Y \le y \right) = \\ &= F_{X|Y\le y}(x) F_Y(y) = \int_{-\infty}^{x} f_{X|Y\le y}(x)\,dx \cdot F_Y(y) = \int_{-\infty}^{x} \frac{f_{X|Y\le y}(x) F_Y(y)}{f_X(x)} f_X(x)\,dx\,, \end{aligned}$$

which is $\mathbb{E}\left(\mathbf{1}_{(-\infty,x]}(X) F_{Y|X}(y) \right)$, if we put

$$\text{(7c4)} \qquad F_{Y|X=x}(y) = \frac{f_{X|Y\le y}(x) F_Y(y)}{f_X(x)}\,.$$

Due to uniqueness (recall 7b16), (7c4) is thus proven for absolutely continuous X and arbitrary Y.

The case of absolutely continuous (X,Y) is especially important; having $f_{X,Y}$ we get

$$\begin{aligned} F_{X|Y\le y}(x) &= \frac{F_{X,Y}(x,y)}{F_Y(y)} = \int_{-\infty}^{x} \left(\frac{1}{F_Y(y)} \int_{-\infty}^{y} f_{X,Y}(x_1,y_1)\,dy_1 \right) dx_1\,, \\ f_{X|Y\le y}(x) &= \frac{1}{F_Y(y)} \int_{-\infty}^{y} f_{X,Y}(x,y_1)\,dy_1\,, \end{aligned}$$

and (7c4) gives

$$F_{Y|X=x}(y) = \frac{1}{f_X(x)} \int_{-\infty}^{y} f_{X,Y}(x,y_1)\,dy_1\,,$$

which proves the following result.

[4] $F_{X|Y\le y}$ was not defined in this section; rather, we use the elementary definition (7a1).

7c5 Theorem. If the joint distribution of X, Y is absolutely continuous, then the conditional distribution of Y given $X = x$ is absolutely continuous (for P_X-almost all x), and

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} .$$

Assume now that X is absolutely continuous but Y is discrete. We have $f_X(\cdot)$ and $p_Y(\cdot)$ but neither $f_{X,Y}(\cdot,\cdot)$ nor $p_{X,Y}(\cdot,\cdot)$. Rather,

$$F_{X,Y}(x,y) = \sum_{y_1 \le y} \int_{-\infty}^{x} g_{X,Y}(x_1, y_1)\, dx_1 = \int_{-\infty}^{x} \Big(\sum_{y_1 \le y} g_{X,Y}(x_1, y_1) \Big)\, dx_1 \tag{7c6}$$

where $g_{X,Y}$ is defined by

$$\mathbb{P}\big(X \le x \,\&\, Y = y \big) = \int_{-\infty}^{x} g_{X,Y}(x_1, y)\, dx_1 . \tag{7c7}$$

Its existence follows from Radon-Nikodym theorem; indeed, $F_{X|Y=y}(\cdot)$ is absolutely continuous (check it similarly to (7c3)), therefore it has a density $f_{X|Y=y}(\cdot)$, and $g_{X,Y}(x,y) = f_{X|Y=y}(x)\mathbb{P}\big(Y = y \big)$ fits, since $\mathbb{P}\big(X \le x \,\big|\, Y = y \big)\; \mathbb{P}\big(Y = y \big) = \mathbb{P}\big(X \le x \,\&\, Y = y \big)$.

7c8 Exercise. Prove that

$$f_X(x) = \sum_y g_{X,Y}(x,y) ,$$
$$p_Y(y) = \int_{-\infty}^{+\infty} g_{X,Y}(x,y)\, dx .$$

We have from (7c6), assuming $F_Y(y) \ne 0$,

$$F_{X|Y\le y}(x) = \frac{F_{X,Y}(x,y)}{F_Y(y)} = \int_{-\infty}^{x} \Big(\frac{1}{F_Y(y)} \sum_{y_1 \le y} g_{X,Y}(x_1, y_1) \Big)\, dx_1 ,$$

which means that

$$f_{X|Y\le y}(x) = \frac{1}{F_Y(y)} \sum_{y_1 \le y} g_{X,Y}(x, y_1) ,$$

and (7c4) gives

$$F_{Y|X=x}(y) = \frac{1}{f_X(x)} \sum_{y_1 \le y} g_{X,Y}(x, y_1) ,$$

which means that the conditional distribution is discrete, and

$$p_{Y|X=x}(y) = \frac{g_{X,Y}(x,y)}{f_X(x)} . \tag{7c9}$$

Especially, let Y be an indicator, $Y = \mathbf{1}_A$. Then

$$\begin{aligned} g_{X,Y}(x,1) &= f_{X|A}(x)\mathbb{P}\left(A\right), \\ g_{X,Y}(x,0) &= f_{X|\overline{A}}(x)(1-\mathbb{P}\left(A\right)), \\ f_X(x) &= f_{X|A}(x)\mathbb{P}\left(A\right) + f_{X|\overline{A}}(x)(1-\mathbb{P}\left(A\right)), \\ \mathbb{P}\left(A \mid X=x\right) &= \frac{f_{X|A}(x)\mathbb{P}\left(A\right)}{f_X(x)}, \end{aligned}$$

and

(7c10)
$$\mathbb{P}\left(A\right) = \int_{-\infty}^{+\infty} \mathbb{P}\left(A \mid X=x\right) f_X(x)\,dx = \mathbb{E}\mathbb{P}\left(A \mid X\right)$$

(recall 7a2).

Another singular case appears when X is absolutely continuous and $Y = \varphi(X)$ where φ is a smooth (but usually non-invertible) function (say, $\varphi(x) = x^2$); thus Y is absolutely continuous, however, the pair (X,Y) is singular. Given $X = x$, we have only one possible value $y = \varphi(x)$ for Y. However, given $Y = y$, we have a discrete set $\varphi^{-1}(y)$ (that is, $\varphi^{-1}(\{y\})$, of course) of possible values for X. Each $x \in \varphi^{-1}(y)$ contributes $f_X(x)/\varphi'(x)$ into f_Y;

$$f_Y(y) = \sum_{x\in\varphi^{-1}(y)} \frac{f_X(x)}{\varphi'(x)}.$$

Also,

$$f_{Y|X\le x}(y) = \frac{1}{F_X(x)} \sum_{x_1\in\varphi^{-1}(y)\cap(-\infty,x]} \frac{f_X(x_1)}{\varphi'(x_1)},$$

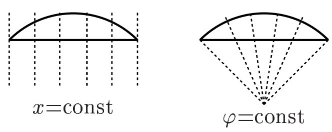
and we apply (7c4) to (Y,X) instead of (X,Y):

$$F_{X|Y=y}(x) = \frac{f_{Y|X\le x}(y)F_X(x)}{f_Y(y)} = \frac{1}{f_Y(y)} \sum_{x_1\in\varphi^{-1}(y)\cap(-\infty,x]} \frac{f_X(x_1)}{\varphi'(x_1)},$$

which means that the conditional distribution is discrete, and

(7c11)
$$p_{X|Y=y}(x) = \frac{f_X(x)}{f_Y(y)\varphi'(x)} \qquad \text{for } x \in \varphi^{-1}(y).$$

Return to paradoxes of 7a. In 7a3, one solution corresponds to the conditional probability of the event $A =$ "arc, not chord" given X; the other solution corresponds to the conditional probability of the same event A given Φ (see the picture).



True, $(x=0) \iff (\varphi=0)$. Nevertheless,[5] $\mathbb{P}\left(A \mid X=0\right) \ne \mathbb{P}\left(A \mid \Phi=0\right)$. Another

[5]You see, an event of probability 0 is not autonomous, its conditional distribution involves its infinitesimal neighborhood. In contrast, an event of positive probability is autonomous, it determines its conditional probability with no help of other events.

example of the same kind:

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{+\infty} f_{X,Y}(x_1,y)\,dx_1}\,;$$

$$f_{X|\frac{Y}{X}=u}(x) = \frac{f_{Y/X,X}(u,x)}{f_{Y/X}(u)} = \frac{|x| f_{X,Y}(x,ux)}{\int_{-\infty}^{+\infty} |x_1| f_{X,Y}(x_1,ux_1)\,dx_1}\,;$$

$$f_{X|Y=0}(\cdot) \ne f_{X|\frac{Y}{X}=0}(\cdot)\,.$$

Another difficulty appears in 7a4. The event "Φ/π is rational" of probability 0 is not included into a continuum of events. Though, using the axiom of choice, one can prove that there is a function $X : (0,2\pi) \to \mathbb{R}$ such that $X(\varphi_1) = X(\varphi_2) \iff (\varphi_1 - \varphi_2)/\pi \in \mathbb{Q}$. However, there is no such Borel function!