Joint Distributions

Consider now the case of multiple random variables

$$X_1, \ldots, X_n : S \to \mathbb{R}$$

taken simultaneously

$$X_1 \times X_2 \times \cdots \times X_n : S^n \to \mathbb{R}^n$$

$$(X_1 \times \cdots \times X_n)(s_1, \ldots, s_n) \overset{\text{def}}{=} (X_1(s_1), \ldots, X_n(s_n))$$

Def. 1 They are said to be <u>jointly continuous</u> iff for $C \subseteq S^n$ there exists a joint <u>probability density function</u>

$$f : \mathbb{R}^n \to \mathbb{R}$$

such that

$$P((X_1, \ldots, X_n) \in C) = \underset{(x_1, \ldots, x_n) \in C}{\iint \cdots \int} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

Def. 2 Their joint cumulative distribution function

$$F : \mathbb{R}^n \longrightarrow \mathbb{R}$$

is given by

$$F(a_1, \dots, a_n) = P(X_1 \le a_1, \dots, X_n \le a_n)$$

$$\equiv P(\{X_1 \le a_1\} \cap \{X_2 \le a_2\} \cap \dots \cap \{X_n \le a_n\})$$

Def 3 This setup also works with discrete random variables: $X_1, \dots, X_n : S^n \to \mathbb{R}^n$; $C \subseteq S^n$;

$$P((X_1, \dots, X_n) \in C) = \sum_{(x_1, \dots, x_n) \in C} p(x_1, \dots, x_n)$$

$$= \sum_{i_1=1}^{m_1} \cdots \sum_{i_n=1}^{m_n} p(x_{i_1}, \dots, x_{i_n})$$

where

$$p(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$$

joint pmf

$$\equiv P\left(\bigcap_{i=1}^{n} \{X_i = x_i\}\right)$$

Observation 1: The separate distributions F_{X_i} of the separate random variables $X_i : S \to \mathbb{R}$ can be recovered from the joint distribution:

marginal distribution of X_i

$$
\begin{aligned}
F_{X_i}(a_i) &= P(X_1 < \infty, \ldots, X_i \le a_i, \ldots, X_n < \infty) \\
&= \lim_{\substack{a_1 \to \infty \\ \vdots \\ a_{i-1} \to \infty \\ a_{i+1} \to \infty \\ \vdots \\ a_n \to \infty}} P(X_1 \le a_1, \ldots, X_i \le a_i, \ldots, X_n \le a_n) \\
&= \lim_{\substack{a_j \to \infty \\ j \ne i}} F(a_1, \ldots, a_i, \ldots, a_n)
\end{aligned}
$$

That is, we recover F_{X_i} by letting the other $n-1$ X_j's to take any real value whatsoever.

<u>Observation</u>: Similarly, the individual pmf's p_{X_i} can be recovered from the joint pmf in the discrete case:

$$p_{X_i}(a_i) = P(X_i = a_i)$$

$$= P\Big(\{X_i = a_i\} \cap \Big(\bigcap_{\substack{j=1 \\ j \neq i}}^{n} \{X_j < \infty\}\Big)\Big)$$

$$= \sum_{\substack{j=1 \\ j\neq i}}^{n} \underbrace{\sum_{j_1=1}^{m_1} \cdots \sum_{j_n=1}^{m_n}} P(X_1 = x_{j_1}, \ldots, X_i = a_i, \ldots, X_n = x_{j_n})$$

Sum over the ranges of each X_j, $j \neq i$ / (m_j could be real or ∞)

Let us consider the computational methods in the easier case of $n = 2$ random variables X & Y:

With

$$F(a,b) \overset{\text{def}}{=} P(X \le a, Y \le b)$$

$$\overset{\text{def}}{=} P(\{X \le a\} \cap \{Y \le b\})$$

as our basis, we can compute for example

<u>Obs. 3</u>

$$P(X > a, Y > b)$$

$$= P(\{X > a\} \cap \{Y > b\})$$

$$= 1 - P\left(\left\{\{X > a\} \cap \{Y > b\}\right\}^c\right)$$

$$= 1 - P(\{X \le a\} \cup \{Y \le b\})$$

$$= 1 - \left[P(\{X \le a\}) + P(\{Y \le b\}) - P(\{X \le a\} \cap \{Y \le b\})\right]$$

$$= 1 - F_X(a) - F_Y(b) + F(a,b)$$

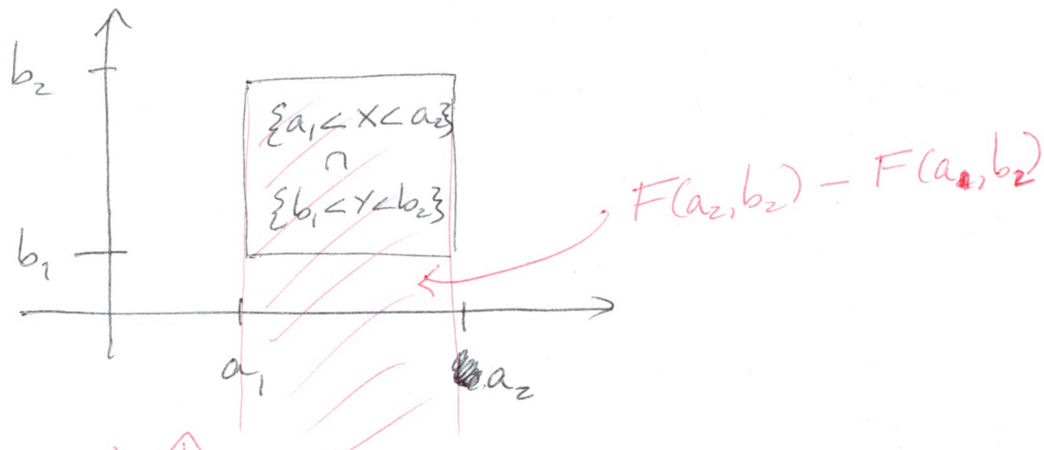
by de Morgan's law & $\{X > a\}^c = \{X \le a\}$

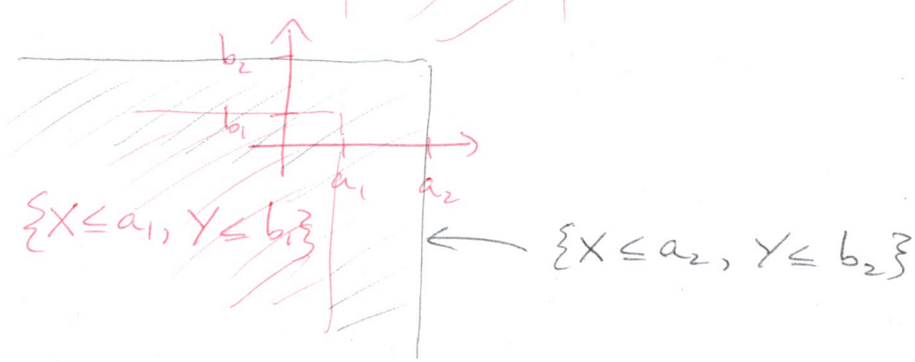
inclusion–exclusion

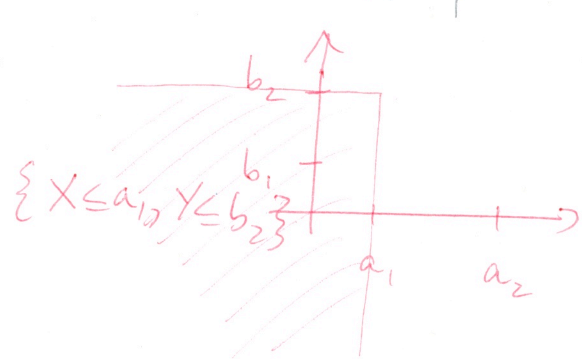
Obs. 4

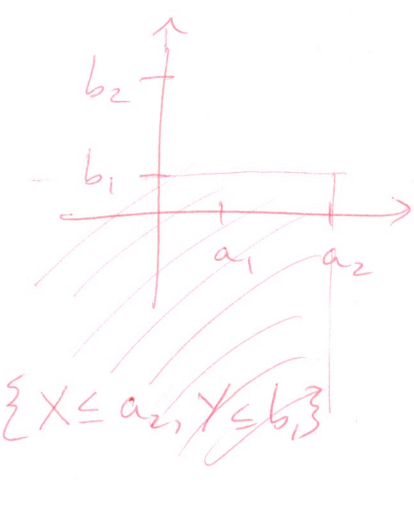
$$P(a_1 < X < a_2,\ b_1 < Y < b_2)$$

$$= F(a_1, b_1) + F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1)$$









Obs. 5 By differentiating under the integral we obtain

$$\frac{\partial^2 F}{\partial x \partial y}(a,b) = f(a,b)$$

and more generally

$$\frac{\partial^n F}{\partial x_1 \partial x_2 \cdots \partial x_n}(a_1, a_2, \ldots, a_n) = f(a_1, \ldots, a_n)$$

Def. 4 We say X & Y are <u>independent</u> if

$$P(X \in A, Y \in B) = P(X \in A) P(X \in B)$$

for all $A, B \subseteq \mathbb{R}$, i.e. if the events

$$E_A = X^{-1}(A)$$

$$E_B = Y^{-1}(B)$$

are independent for all $A, B \subseteq \mathbb{R}$.

Prop. 1 If X and Y are discrete random variables, then they are independent iff

$$p(x,y) = p_X(x)\,p_Y(y), \quad \forall x,y \in \mathbb{R}$$

joint pmf — separate pmf's

Analogously, if X & Y are continuous random variables, they are independent iff

$$f(x,y) = f_X(x)\,f_Y(y), \quad \forall x,y \in \mathbb{R}$$

joint pdf — separate pdf's ← called marginal densities

More generally, $X_1, \ldots, X_n$ are indep.

$$P(X_1 \in A_1,\ X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

$$\iff p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i) \quad \text{(discrete } X_i\text{)}$$

or

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \quad \text{(continuous } X_i\text{)}$$

pf: Case 1: X, Y discrete.

(i) Suppose X & Y are indep., so that

$$P(X\in A, Y\in B) = P(X\in A)P(Y\in B)$$

(for all $A, B \subseteq \mathbb{R}$)

then, by def.,

$$\sum_{(x,y)\in A\times B} p(x,y) = \sum_{x\in A} p_X(x) \sum_{y\in B} p_Y(y)$$

(for all $A, B \subseteq \mathbb{R}$)

so if we take $A = \{x\}$, $B = \{y\}$, we have

$$p(x,y) = \sum_{(x,y)\in A\times B} p(x,y)$$

$$= \sum_{x\in A}\sum_{y\in B} p_X(x)p_Y(y)$$

$$= p_X(x)p_Y(y) \quad \text{(for all } x, y \in \mathbb{R}\text{)}$$

(ii) Conversely, if $p(x,y) = p_X(x)p_Y(y)$ for all $x, y \in \mathbb{R}$, then

$$P(X\in A, Y\in B) = \sum_{(x,y)\in A\times B} p(x,y)$$

$$= \sum_{x\in A}\sum_{y\in B} p_X(x)p_Y(y) = P(X\in A)\cdot P(Y\in B)$$

Case 2: X, Y continuous.

(i) Suppose X & Y are indep., so that

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

that is,

$$\iint_{(x,y) \in A \times B} f(x,y)\,dx\,dy = \int_A f_X(x)\,dx \int_B f_Y(y)\,dy$$

$$= \int_A \int_B f_X(x) f_Y(y)\,dy\,dx$$

for all $A, B \subseteq \mathbb{R}$, and consequently, by Fubini's Thm. & the linearity of the integral,

$$\iint_{A \times B} [f(x,y) - f_X(x) f_Y(y)]\,dx\,dy = 0$$

for all $A, B \subseteq \mathbb{R}$. Consider $A \times B = [a, a+h] \times [b, b+h]$ for some small $h > 0$, so that the MVT

for integrals gives us the existence of $(c, d) \in A \times B$ (if f_X & f_Y are continuous) such that
& $\hat{f}$

$$f(c, d) - f_X(c) f_Y(d) = \frac{1}{h_2} \int_a^{a+h} \int_b^{b+h} f(x,y) - f_X(x) f_Y(y) \, dy \, dx$$

$$= 0$$

Then,

$$f(a, b) - f_X(a) f_Y(b)$$

$$= \lim_{h \to 0} f(c,d) - f_X(c) f_Y(d) \quad \leftarrow \text{bec. } f, f_X, f_Y \text{ are contin.}$$

$$= 0$$

$$\Longrightarrow f(a,b) = f_X(a) f_Y(b)$$

<u>Remark</u>: This is still true if $f, f_X, \& f_Y$ are only Riemann integrable, because, lets say

$$g(x,y) = f(x,y) - f_X(x) f_Y(y)$$

& suppose $\iint_C g(x,y) \, dA = 0$ for all $C \subseteq \mathbb{R} \times \mathbb{R}$.

~~The only thing we require of g besides integrability is that it~~ [illegible]

g is allowed to be discontinuous, but (by a well-known theorem of Lebesgue) it can have at most countably many discontinuities. If g is nonzero only at single points (at most countably many of them), then the integral may still exist & $=0$, so we conclude that if g is only Riemann integrable (but not nec. continuous), $g \equiv 0$ except possibly at single points (of discontinuity), at most countably many. ☒

Conversely, suppose $f(x,y)$ factors as

$$f(x,y) = h(x)\,g(y) \quad \text{on } \mathbb{R}^2$$

& let us show that $h(x) = f_x(x)$ & $g(y) = f_y(y)$. Toward this end we note that since $f(x,y)$

is a joint density,

$$\underbrace{\left(\int_{-\infty}^{\infty} h(x)\,dx\right)}_{=C_1}\underbrace{\left(\int_{-\infty}^{\infty} g(y)\,dy\right)}_{=C_2}$$

$$= \iint_{\mathbb{R}^2} h(x)g(y)\,dx\,dy$$

$$= \iint_{\mathbb{R}^2} f(x,y)\,dx\,dy$$

$$= 1$$

$\Longrightarrow C_1 C_2 = 1$

$\Longrightarrow$ $$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy = \left(\int_{\mathbb{R}} g(y)\,dy\right)h(x) = C_2 h(x)$$

&

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx = \left(\int_{\mathbb{R}} h(x)\,dx\right)g(y)$$

$$= C_1 g(y)$$

$\Longrightarrow$ $$f_X(x)\,f_Y(y) = \underbrace{C_1C_2}_{=1} h(x)g(y) = f(x,y)$$

$\Longrightarrow$ $$P(X\in A, Y\in B) = \iint_{A\times B} f(x,y)\,dx\,dy = \left(\int_A f_X(x)\,dx\right)\cdot\left(\int_B f_Y(y)\,dy\right)$$

$$= P(X\in A)P(Y\in B).$$ QED

Sums of Independent Random Variables and the Convolution of Cumulative Distributions:

Case I: Discrete X and Y:

Let us enumerate $R(X)$ & $R(Y)$,

$$R(X) = \{x_i \mid i \in \mathbb{N}\}$$

$$R(Y) = \{y_j \mid j \in \mathbb{N}\}$$

Then,

$$F_{X+Y}(a) = P(X+Y \le a)$$

$$= \sum_{\substack{x+y \le a \\ x \in R(X), \\ y \in R(Y)}} p(x,y) = \sum_{x+y \le a} p_X(x) p_Y(y)$$

$$= \sum_{i=1}^{\infty} \sum_{\substack{j \text{ s.t.} \\ y_j \le a - x_i}} p_X(x_i) p_Y(y_j)$$

(inner sum $= F_Y(a - x_i)$)

Def. 5 Convolution of F_X & F_Y:

$$= \sum_{i=1}^{\infty} F_Y(a - x_i) p_X(x_i)$$

$$\stackrel{\text{def}}{=} (F_X * F_Y)(a)$$

This is the same expression found in the convolution, or Cauchy product, of ~~two~~ infinite series

$$\left(\sum_{n=0}^{\infty} a_n\right) * \left(\sum_{m=0}^{\infty} b_m\right)$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{n} a_m b_{n-m}$$

To see why this formula emerges from the finite case, let us compute

$$\left(\sum_{n=0}^{N} a_n\right)\left(\sum_{m=0}^{M} b_m\right)$$

$$= \underbrace{a_0 b_0}_{} + \underbrace{(a_0 b_1 + a_1 b_0)}_{} + \underbrace{(a_0 b_2 + a_1 b_1 + a_2 b_0)}_{}$$

$$\underbrace{= \sum_{m=0}^{0} a_m b_{0-m}}_{n=0} \qquad \underbrace{= \sum_{m=0}^{1} a_m b_{1-m}}_{n=1} \qquad \underbrace{= \sum_{m=0}^{2} a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0}_{n=2}$$

<u>ex.</u> $(a_1 + a_2 + a_3 + a_4)(b_1 + b_2 + b_3)$

$$= (a_1 + a_2 + a_3 + a_4) b_1 + (a_1 + \cdots + a_4) b_2 + (a_1 + \cdots + a_4) b_3$$

$$
\begin{aligned}
= a_1 b_1 &+ a_2 b_1 + a_3 b_1 + a_4 b_1 \\
&+ a_1 b_2 + a_2 b_2 + a_3 b_2 + a_4 b_2 \\
&\quad + a_1 b_3 + a_2 b_3 + a_3 b_3 + a_4 b_3
\end{aligned}
$$

$$
\begin{aligned}
= a_1 b_1 &+ (a_1 b_2 + a_2 b_1) + (a_1 b_3 + a_2 b_2 + a_3 b_1) \\
&+ (a_2 b_3 + a_3 b_2 + a_4 b_1) + (a_3 b_3 + a_4 b_2) \\
&+ a_4 b_3
\end{aligned}
$$

$$
= \sum_{n=1}^{4} \sum_{m=1}^{n} a_m b_{n-m+1}
$$

This is important, for we recall that

$$
\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{m=0}^{\infty} b_m\right) \overset{\text{def}}{=} \lim_{\substack{N\to\infty \\ M\to\infty}} \left(\sum_{n=0}^{N} a_n\right)\left(\sum_{m=0}^{M} b_m\right)
$$

$$
= \lim_{N\to\infty} \sum_{n=0}^{N}\left(\sum_{m=0}^{n} a_m b_{n-m}\right)
$$

Case II: X and Y continuous:

In this case, if X & Y are indep.,

$$F_{X+Y}(a) = P(X+Y \le a)$$

$$= \iint_{\substack{x+y \le a \\ (x,y) \in R(X) \times R(Y)}} f_X(x) f_Y(y)\, dx\, dy$$

$$= \lim_{N, M \to \infty} \left(\sum_{i=1}^{N} \sum_{j=1}^{M} f_X(x_i^*) f_Y(y_j^*) \Delta x \Delta y \right)$$

still requiring $x_i^* + y_j^* \le a$

$$= \lim_{N \to \infty} \sum_{i=1}^{N} \sum_{j=1}^{i} f_X(x_j^*) f_Y(y_{i-j}) \Delta x \Delta y$$

$$= \int_{\mathbb{R}} \int_{-\infty}^{a-y} f_X(x) f_Y(y)\, dx\, dy$$

$= F_X(a-y)$

$$= \int_{\mathbb{R}} F_X(a-y) f_Y(y)\, dy$$

$$\overset{\text{def}}{=} (F_X * F_Y)(a)$$

* this is, of course, nothing but an explanation of Fubini's Thm.:

$$\iint_{x+y \le a} f_X(x) f_Y(y)\, dA = \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y)\, dx\, dy$$

$\int_{-\infty}^{a-y} \ldots dx = F_X(a-y)$

$$= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\, dy$$

Convolution of F_X & F_Y

Conditional Distributions

Def. 6 The conditional probability mass function of a discrete random variable X given that another discrete random variable $Y=y$ is

$$P_{X|Y}(x|y) \overset{\text{def}}{=} P(X=x \mid Y=y)$$

$$\overset{\text{def}}{=} \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{P(x,y)}{P_Y(y)}$$

whenever $P(Y=y) = P_Y(y) > 0$.

Def. 7 The conditional probability distribution function of discrete X given that a discrete $Y=y$ is

$$F_{X|Y}(x|y) \overset{\text{def}}{=} P(X \le x \mid Y \le y)$$

$$\overset{\text{def}}{=} \sum_{a \le x} P_{X|Y}(a|y) = \sum_{a \le x} \frac{P(a,y)}{P_Y(y)}$$

Remark: If X & Y are independent, so that $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for all $A, B \subseteq \mathbb{R}$, then

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} = \frac{p_X(x)\,p_Y(y)}{p_Y(y)}$$

$$= p_X(x) \qquad \boxtimes$$

Def. 8 The conditional probability density of a continuous random variable X given that another contin. random variable $Y = y$ is

(X & Y must be jointly distributed!)

$$f_{X|Y}(x|y) \overset{\text{def}}{=} \frac{f(x,y)}{f_Y(y)}$$

(idea: $f_{X|Y}(x|y)\,dx = \dfrac{f(x,y)\,dx\,dy}{f_Y(y)\,dy}$

$\approx \dfrac{P(x \leq X \leq x+dx,\ y \leq Y \leq y+dy)}{P(y \leq Y \leq y+dy)}$)

Def. 9 Of course, all of the above can be generalized a little: for any $A \subseteq \mathbb{R}$, we may define

$$P(X \in A \mid Y = y) \overset{\text{def}}{=} \sum_{x \in A} p_{X|Y}(x|y) = \sum_{x \in A} \frac{p(x,y)}{p_Y(y)}$$

in the discrete case, and

$$P(X \in A \mid Y = y) \overset{\text{def}}{=} \int_A f_{X|Y}(x|y)\, dx = \frac{1}{f_Y(y)} \int_A f(x,y)\, dx$$

in the continuous case.

Def. 10 Then, the <u>conditional probability distribution</u> of $X \leq a$ (i.e. $X \in A \overset{\text{def}}{=} (-\infty, a]$) may be stated as

$$F_{X|Y}(a,y) \overset{\text{def}}{=} P(X \leq a \mid Y = y) \overset{\text{def}}{=} \int_{-\infty}^{a} f_{X|Y}(x|y)\, dx$$

Joint Probability Distributions of Functions of Two or More Random Variables

Consider two jointly distributed random variables X and Y, and consider the transformation

$$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

$$T(X, Y) = (U, V)$$

where

$$U = g_1(X, Y)$$

$$V = g_2(X, Y)$$

and consider the probability

$$P(U \in A, V \in B)$$

[signature]

Thm. 2 In the discrete case, if T is invertible,

$$T : \mathbb{R}^2 \to \mathbb{R}^2, \quad T = (g_1, g_2)$$
$$T^{-1} : \mathbb{R}^2 \to \mathbb{R}^2, \quad T^{-1} = (h_1, h_2)$$

then, with $(U,V) = T(X,Y)$, so that $(X,Y) = T^{-1}(U,V)$, we have

$$P(U \in A, V \in B) = \sum_{(x,y) \in T^{-1}(A \times B)} p(x,y)$$

In that case, we have

$$p_{UV}(u,v) = (p_{UV} \circ T)(x,y)$$
$$= p_{XY}(x,y)$$

In the continuous case, we require T to be invertible and the Jacobian

$$|J(x,y)| = |\det DT(x,y)|$$
$$= \left|\det \begin{pmatrix} \partial_x g_1 & \partial_y g_1 \\ \partial_x g_2 & \partial_y g_2 \end{pmatrix}\right| > 0$$

In that case

$$P(U\in A, V\in B) = \iint_{(x,y)\in T^{-1}(A\times B)} f_{XY}(x,y)\, |J(x,y)|^{-1}\, dA$$